

Putting Assessment *for* Learning into practice in a higher education EFL context

Edmund White

B. A. B. Ed. MSc. TESOL

Submitted to: Department of Linguistics

Faculty of Human Sciences

Macquarie University

Sydney, Australia

This thesis is presented for the degree of Ph.D. in Applied Linguistics

January 2010

Contents

Introduction -----	p. 6
Part A: AfL fundamentals and theoretical background -----	p. 8
Part B: Literature overview - AfL in classroom practice -----	p. 17
Part C: Research focus -----	p. 23
Chapter 1. Feedback through marking ----- p. 45	
Part A: Teacher self-assessment of written feedback -----	p. 57
Part B: Does the feedback feed forward?	
Student response to and views of teacher feedback -----	p. 103
Appendices (Chapter 1) -----	p. 129
Chapter 2. Student as Assessor ----- p. 135	
Part A: Self-assessment for learning in a Communicative English course --	p. 137
Part B: Peer-assessment for learning and student presentations -----	p. 164
Appendices (Chapter 2) -----	p. 194
Chapter 3. Student as Questioner, and Assessment Synergy - p. 199	
Part A: Student-generated questioning for learning -----	p. 201
Part B: Synergizing formative and summative assessment	
of presentation slide shows -----	p. 249
Appendices (Chapter 3) -----	p. 272
Conclusion ----- p. 280	
Part A: Assessing the AfL research studies -----	p. 282
Part B: Distinctions and considerations - applying AfL to a HE setting -----	p. 290
References -----	p. 298

Synopsis

In the past twenty years or so, assessment in education has become increasingly viewed as a means of guiding and improving student learning, rather than simply measuring it. The assessment for learning (AfL) movement, arising primarily from mainstream primary and secondary education in the UK, strongly advocates this formative, learning-centered view of assessment. At the level of classroom practice, AfL is comprised of a number of key procedures including: the use of effective feedback, self-and peer assessment, questioning and classroom dialogue, and the formative use of summative tests. Based on a series of classroom research projects, this thesis examines these AfL procedures in a higher education EFL context with adult students at *Tokyo Woman's Christian University*. Providing the impetus, the theoretical framework, and serving as a crucial resource for this research, has been a book entitled, *Assessment for Learning: Putting it into Practice* (Black, Harrison, Lee, Marshall & Wiliam, 2003).

After an introductory chapter in which AfL theory and fundamentals are presented, the three body chapters of this thesis are comprised of a series of five reports of AfL procedures being used in various classroom and teaching contexts. Chapter 1 describes a research project centered on the issue of feedback on the first draft of student essays, and begins with a teacher self-assessment of written feedback. This is followed by a report of the impact of the feedback on student essays, and how students felt about the process.

Chapter 2 focuses on the student in the assessor role, and begins in Part A with a report focused on self-assessment of class participation in a freshman English class. This is followed in Part B by examining the issue of peer assessment in a public speaking course.

Chapter 3 first investigates the issue of questioning in classroom dialogue, in particular, student-generated questioning. The final research project report then examines the issue of using summative assessment for formative purposes in promoting student learning. Following Chapter 3, a conclusion draws the five AfL reports together for a more holistic view of the process of putting these AfL procedures into practice in a higher education EFL context with adult learners.

Statement of Candidate

I certify that the work in this thesis entitled "**Putting Assessment for Learning into practice in a higher education context**" has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

The research presented in this thesis was approved by Macquarie University Ethics Review Committee. The reference numbers and approval dates are as follows:

Ethics approval number	Date of approval
HE23FEB2007-D04992	January 15, 2007
HE28MAR2008-D05724	April 9, 2008

Edmund Michael White
Student number: 40376656

January 10, 2010

The goal of assessment has to be, above all, to support the improvement of learning and teaching.

(Frederiksen & Collins, 1989, p. 32)

Any attempt to turn ideas into practice will be a learning experience . . .

(Black, Harrison, Lee, Marshall & Wiliam, 2003, p. 118)

Introduction

Learning, assessment, and higher education

How people learn is undoubtedly a complex business. However, it seems that a core component of learning involves the notion of *change* on the part of the learner – whether in developing knowledge and understanding, or skills and abilities in a particular area or activity. Nichols (2004, p. 54) writes that learning is defined by “changes in knowledge, understanding, skills and attitudes brought about by experience and reflection on that experience, whether that experience is structured or not”. Explaining that it is more than just a process of ingesting information, Maki (2004) defines learning as “a multidimensional model of making meaning” (p. 1). In a classroom context then, the educators’ primary role may be described as ‘a facilitator of student meaning-making’. This simple description belies the complexity and challenge of effective teaching, the kind which promotes and maximizes student learning. Assessment has come to be seen as a crucial bridge in aiding this teaching and learning process.

For professionals at all educational levels, the primary goal of promoting and deepening student learning would be an idea that few would take issue with, despite the wide-ranging, often conflicting, views of the most effective ways of achieving this goal. In the past twenty years or so, with the publication of such works as Sadler’s (1989) influential “Formative assessment and the design of instructional systems”, there has been a continuous and developing recognition of the vital role assessment can play in promoting, rather than simply measuring, student learning. While research and publications have focused largely on primary and secondary schooling in mainstream education, the potential beneficial role of assessment in higher education (HE) has also become more recognized in recent years. For example, in writing about tertiary assessment, Bryan and Clegg (2007) assert that assessment and learning should be seen as working in tandem, with each contributing to the other.

According to Atkins (1995, p. 25) four overlapping purposes for higher education may be distinguished:

1. To provide a general educational experience of intrinsic worth in its own right.
2. To prepare students for knowledge creation, application, and dissemination.
3. To prepare students for a specific profession or occupation.
4. To prepare students for general employment.

Supplementing these purposes, the role of tertiary education in establishing a framework for

continued and lifelong learning has also been recognized (for example, Boud, 2000). According to Boud and Falchikov (2007), the provision of a foundation for lifelong learning beyond the academy in work and other social settings is the very “raison d’être of a higher education” (p. 399).

The impact of assessment in working toward these higher education purposes has become increasingly appreciated. For university instructors, classroom assessment plays a key role in and significant influence on both their teaching and student learning (Cheng, Rodgers & Wang, 2008). The role, influence and beneficial potential of assessment in higher education are well captured by McInnis and Devlin (2002) below in Box 1.

Box 1. Influence of HE assessment

Assessment is a central element in the overall quality of teaching and learning in higher education. Well-designed assessment sets clear expectations, establishes a reasonable workload (one that does not push students into rote reproductive approaches to study), and provides opportunities for students to self-monitor, rehearse, practice and receive feedback. . . Carefully designed assessment contributes directly to the way students approach their study and therefore contributes indirectly, but powerfully, to the quality of their learning.

(McInnis and Devlin, 2002, p. 7)

As will be seen, a number of the issues raised here by McInnis and Devlin (2002), and their general argument that effective assessment promotes more effective student learning and teaching practice, will be evident throughout this thesis. The impact of classroom-based assessment on student learning is the central focus under consideration here. A higher education institution in Japan is the setting for this assessment and learning research report, and while describing a variety of classroom practices and processes, “assessment is our focus but learning is the goal” (Gardner, 2006, p. 2).

In writing about the issue of originality in HE doctoral research, Dunleavy (2003) points out that it rarely entails coming up with an entirely new way of looking at things. Instead, originality in modern social sciences and humanities often involves,

. . . encountering an established idea or viewpoint or method in one part of your discipline (or neighboring discipline) and then taking that idea for a walk and putting it down somewhere else, applying it in a different context or for a different purpose. (Dunleavy, 2003, p. 40)

This thesis involves taking assessment for learning (AfL) theory and practice from mainstream school education for a walk and putting them down in Japan at a tertiary setting – a woman’s university in Tokyo - with students for whom English is a foreign language. Yet, while the context does indeed differ, the purpose remains the same – using assessment to promote and maximize student learning.

This introduction will delineate in more detail the nature of AfL in practice, particularly with reference to tertiary education, and provide an overview of the research conducted at the Japanese university. But first, it is appropriate to clarify the fundamental ideas, practices and theory associated with AfL, or formative assessment. This will provide an underlying framework for the reports and discussions of classroom assessment procedures to follow. It should be noted that this introductory section will simply present some of the essential ideas associated with AfL, and that subsequent chapters, comprised of five research reports, will each offer further details and connections to the now extensive AfL-related literature.

Part A: AfL fundamentals and theoretical background

Fundamental principles and practices

More than ten years after the publication of Black and Wiliam’s (1998a) seminal review of formative assessment research, AfL may be considered as ‘an established idea or viewpoint’ in classroom-based assessment literature for primary and secondary education, particularly in the UK. AfL theory and practice have become widely disseminated and appreciated, if not implemented. A grouping of educators and researchers advising on assessment policy and practice in the UK, the Assessment Reform Group (ARG), is perhaps the key information source for AfL research and literature. The ARG is comprised of some of the most important AfL writers and researchers, and its Internet website provides a rich fount of fundamental summaries of what AfL entails.

In 1999, the ARG published a follow-up to the critical research review by Black and Wiliam (1998a), and noted that while assessment is one of the most powerful educational tools for promoting effective learning, it must be properly used to have this effect. They explain that the research shows that assessment can be used to improve learning depending on five key factors, shown in Table 1 below.

Table 1. *Improving learning through assessment* (ARG, 1999, p. 4-5)

Five key factors for using assessment to improve learning
1. Providing effective feedback to students
2. The active involvement of students in their own learning
3. Adjusting teaching to take account of assessment results
4. Recognizing the profound influence assessment has on student motivation and self-esteem (both crucial influences on learning)
5. The need for students to assess themselves and understand how to improve

According to the ARG (1999), these features are essential for effective day-to-day learning in the classroom, as well as being key components for successful lifelong learning. However, this report also noted that in reality, classroom practice often fails to meet these key factors for both students and their teachers.

In a 2002 publication, the ARG provides the most commonly referred to definition of AfL, and ten founding principles which have become the qualities collectively attributed to this assessment concept (Gardner, 2006). The definition, and founding principles are provided below in Table 2.

Table 2. *AfL definition and ten founding principles* (ARG, 2002)

Assessment for learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there.	Assessment for Learning
	- is part of effective planning
	- focuses on how children learn
	- is central to classroom practice
	- is a key professional skill
	- is sensitive and constructive
	- fosters motivation
	- promotes understanding of goals and criteria
	- helps learners know how to improve
	- develops the capacity for self-assessment
	- recognizes all educational achievement

A noticeable theme in both the definition and the ten principles is the primary focus on *the learner* and *learning*, rather than on the teaching or subject matter, as part of the assessment process. Another key idea is determining where students are in their learning, and ways to move forward. One further relevant point here, in the context of this thesis, is the second principle listed above and use of the term ‘children’. This is a common reference label for students in the AfL literature, and the ideas and practices are commonly associated with younger learners, rather than adults.

Terminological usage (assessment of, for, and as learning, formative assessment)

When considering or constructing assessment tools and procedures, a critical factor is the purpose for their use. A key distinction, and common reference point in the literature, is between assessment *for* learning, or formative assessment and assessment *of* learning, or summative assessment. James (2006) explains the difference as follows:

A distinction between formative and summative (summing-up) purposes has been familiar since the 1960s, although the meaning of these two terms has not been well understood. A more transparent distinction, meaning roughly the same thing, is between assessment *of* learning, for grading and reporting, and assessment *for* learning, where the explicit purpose is to use assessment as part of teaching to promote pupils’ learning. (p. 8)

Stiggins (2002) similarly observes that the crucial distinction between assessment *of* and *for* learning is “between assessment to determine the status of learning and assessment to promote greater learning” (p. 761). This distinction in assessment purposes is an important one, but in reality, and as will be seen in this thesis, summative and formative purposes are often intertwined.

Some educators, particularly Earl (2003), differentiate between teacher and student roles within AfL, and have introduced the further distinction of ‘assessment *as* learning’. This term essentially refers to the role played by the student, while AfL focuses more on the teachers’ role in the classroom. Following Glassen (2009), this thesis uses the broader term ‘assessment *for* learning’ as most of the research literature uses the term in this manner, it is less confusing, and “it confirms that the role of assessment is an integrated one, involving both teachers and students in a mutually responsible, symbiotic and potentially productive relationship” (p. 5).

A point should be made here also about the distinction between the terms ‘assessment for learning’ and ‘formative assessment’. Various writers have linked these two terms (Marsh, 2007), and they are commonly assumed to carry the same meaning. The potential problem, as noted by Black, Lee, Marshall, and Wiliam (2003), is that ‘formative assessment’ may be interpreted in different ways by teachers and can often mean simply that assessment is frequently carried out and planned at the same time as teaching. The assessment in question may be formative for the teacher, but not for the students (Black et al., 2003). Formative assessment *for the learner* should be the point of emphasis here.

While noting that ‘assessment for learning’ and ‘formative assessment’ may convey meanings which are slightly different (Hodgen & Marshall, 2005), they are used interchangeably in this thesis because, as Gardner (2006) notes, “in the final analysis, there is little of substance to distinguish the two terms” (p. 2). Indeed, in the body of AfL literature the two terms are often interchangeably used (for example, Wiliam, Lee, Harrison & Black, 2004; Hodgen & Marshall, 2005; James & Pedder, 2006a; and Harlen, 2007). Wiliam (2008) captures the similar essence of the two terms when writing, “Just as we use the term *formative* to describe the experiences that shape us as we grow up, a formative assessment is one that shapes learning” (p. 9). Shaping, promoting and deepening student learning are also fundamental elements encapsulated in the term ‘assessment *for* learning’.

AfL in practice

The above terminological clarification leads us to a final point to be made in this section about how AfL looks, or should look, in practice. Tables 1 and 2 above from the Assessment Reform Group (1999, 2002) present the key principles associated with AfL, but how these translate into classroom practice are not explicated. Because formative assessment may be misunderstood, or weakly implemented by teachers (Black & Wiliam, 1998a) the ARG (1999) insists that it is important to distinguish AfL from other inadequate interpretations of formative classroom assessment. Such practices may include what becomes essentially summative assessment in nature; activities like simply adding procedures or tests to existing work, or on-going assessment which only feeds marks or grades back to students. In such cases, the problem is that *student learning* and its promotion are not fore-fronted. Black et al. (2003) are also explicit about avoiding any misunderstanding when they assert,

... It is important to emphasize the critical criterion – formative assessment is a process, one in which information about learning is *evoked* and then *used* to modify the teaching and

learning activities in which teachers and students are engaged. (p. 122, original emphasis)

In order to make good AfL practice explicit and show teachers how it is related to effective teaching and learning, the ARG (1999) identify seven “key characteristics of AfL practice” to correspond with the principles delineated in the tables above. The practical classroom applications of AfL for teachers and students are laid out here in Table 3.

Table 3. *Fundamental characteristics of AfL in practice (ARG, 1999)*

Seven key characteristics of AfL
1. It is embedded in a view of teaching and learning of which it is an essential part.
2. It involves sharing learning goals with pupils.
3. It aims to help pupils to know and to recognize the standards they are aiming for.
4. It involves pupils in self-assessment.
5. It provides feedback which leads to pupils recognizing their next steps and how to take them.
6. It is underpinned by confidence that every student can improve.
7. It involves both teacher and pupils reviewing and reflecting on assessment data.

These fundamental characteristics of putting AfL into practice are important reminders for educators of the essentials that need to be focused on. This ARG (1999) characterization of the practices AfL principles should exemplify is an important point of reference, especially considering the “plethora of products and services that describe themselves as ‘formative assessment’ ” (Wiliam, 2007, p. 1), but in reality do not promote student learning. The seven characteristics also provide excellent criterion for evaluating the AfL practices and procedures implemented in the five research projects reported on in this thesis. They will be used in this manner in the conclusion of this thesis as part of an overview of the entire research project, and to assess the individual five studies according to recommended AfL practice.

Recent descriptions of AfL

While there has been some confusion and misunderstanding of the meaning of ‘assessment for learning’ or ‘formative assessment’, there has been some clarification of what these terms encompass in recent years (James, Black, McCormick & Pedder, 2007). We end this section and its presentation of fundamental ideas with three relatively recent descriptions of AfL. These explanations insightfully and descriptively gloss the central assessment/education concept of this thesis.

Recent descriptions of AfL:

- ◆ Assessment *for* learning: the use of the formative assessment process and its results as an instructional intervention designed to increase - not merely to monitor and grade - pupil learning. Research evidence gathered in hundreds of studies conducted around the world over the last decade . . . shows that the consistent application of principles of assessment for learning can give rise to unprecedented gains in pupil achievement, especially for perennial low achievers. (Stiggins, 2007, p. 17)
- ◆ The ultimate goal of AfL is therefore to involve pupils in their own assessment so that they can reflect on where they are in their own learning, understand where they need to go next and work out what steps to take to get there. The research literature sometimes refers to this as the processes of self-monitoring and self-regulation. In other words, pupils need to understand both the desired *outcomes* of their learning and the *processes* of learning by which these outcomes are achieved, and they need to act on this understanding. (James et al., 2007, p. 7, original emphasis)
- ◆ Assessment for Learning (AfL) aims to inform and improve student learning within the regular flow of teaching and learning through students becoming active meaning makers and thoughtful judges of their own learning. (Willis, 2007, p. 52)

As we shall see, putting AfL into practice can also result in *teachers* becoming ‘active meaning makers’ and more ‘thoughtful judges’ in the classroom.

In short, AfL is *for* students, *for* learning, and also *for* more effective teaching.

Theoretical background

As the Stiggins (2007) description of AfL above acknowledges, empirical research evidence

for the effectiveness of formative assessment practices is both widespread and extensive (documented in research reviews by Natriello, 1987; Crooks, 1988; Kluger & DeNisi, 1996; Black & Wiliam, 1998a; Nyquist, 2003; and Brookhart, 2005). Theoretical support for AfL principles and practices has also been established, for example from contemporary psychology and the contention that understanding is created by the learner (James, 2006). According to Biggs (2003), this learning theory termed *constructivism* enjoys a long history in the field of cognitive psychology, having the central idea that “what the learner has to do to create knowledge is the important thing” (p. 12, original emphasis). As we have noted, implicit in the concept of learning are notions of change and development of understanding or ability. Regarding learning in an educational context, Harlen (2007b) explains:

Although learning processes in detail vary among subjects, underpinning learning in all subjects is a view of learning as the progressive development of understanding, in which new experience is linked to existing knowledge. (p. 117)

In this constructivist view of learning, the development of understanding involves a process of construction and reconstruction of knowledge by the student (Harlen & James, 1997). It is a view of learning that provides a basis for the active participation of the learner in classrooms in which formative assessment is practiced.

Teaching is based on certain assumptions of how people learn, and the three main perspectives on learning identified by learning theory writers are *behaviorist*, *constructivist* and *socio-cultural* (James et al., 2007). Table 4 below presents a summary of the main ideas connected with each of these learning theories. Two variations of constructivist theory are identified: cognitive constructivism and social constructivism.

Table 4. *Summary of three main learning theories (based on James, et al., 2007, pp. 16-20)*

Behaviorist	Constructivist	Socio-cultural
Assumptions about learning originate in behaviorist psychology which sees learning as a conditioned response to a stimulus. Behaviorism is most concerned with behavior, not the mental life going on in a persons' head.	Cognitive constructivist approaches focus attention on the mental models that a learner uses when responding to new information or problems. Learning always involves analyzing and transforming any new information. The reception of new knowledge depends on existing knowledge and understanding.	People learn through participating in 'communities of practice', like apprentices. Through membership and activity they come to understand what to pay attention to and what counts as quality in a particular group.
For teaching and assessment: 1) rewards (or withholding them) are powerful ways to establish desired behaviors, 2) a complex skill can be taught by breaking it down and teaching and testing the pieces separately, 3) it is best to learn facts and basic skills first, understanding will come later.	Self-awareness and self-regulation (meta-cognition) are viewed as key elements for learning. Students need to understand what it means to learn and they need to monitor how they go about planning, monitoring and revising. They need to reflect upon their learning and to learn to determine for themselves whether they understand or can do something.	Learning through participation in communities of practice is important. Collective expertise can be more productive than individuals working alone. Group work is essential for learning.
Behaviorist theory has lost favor in recent years, but many practices associated with it are still widespread.	Social constructivist approaches extend these cognitive ideas. Contends that learning proceeds by interaction between the teacher and learners in a social context, mediated by the social norms that value the search for understanding.	The role of experts guiding novices is an important concept.

Constructivist learning theory is highlighted and centered in Table 4 to reflect the fact that it provides a central theoretical base for AfL. James et al. (2007) affirm that,

Most approaches to assessment for learning have been developed within a cognitive constructivist framework for understanding learning, although, Black and Wiliam (2006) have begun to develop a theory of formative assessment drawing on socio-cultural perspectives. (p. 18)

Indeed, in the past ten years the theoretical basis of formative assessment has undergone a number of formulations (Black & Wiliam, 2009) and shifting points of emphasis, including

more attention being paid to the social constructivist and socio-cultural theories noted in Table 4 above. Yet, the cognitive constructivist paradigm remains a central one for AfL, and for this thesis, as students are seen to be actively involved in constructing understanding and developing skills during the learning process.

As noted in Table 4, while still influential in practice, behaviorist learning theory has fallen out of favor in recent years. Ellery (2008) observes in Box 2 below that, at least in terms of effective classroom-based assessment, the sands have been shifting away from assessment practices and frameworks associated with behaviorist educational theories and towards more constructivist approaches.

Box 2. Shifting perspectives on assessment: from behaviorism to constructivism

The academic value of such an approach [traditional behaviorist assessment methods] has been increasingly questioned and in recent years there has been a shift towards a constructivist, student-centered approach that integrates assessment with learning. This approach assumes knowledge is not a fixed, identifiable entity to be absorbed by the learner but instead is constructed by students based on their own understanding, which is influenced by their background, perspectives and experiences. As a result this type of assessment tends to be more flexible, integrative, contextualized, process- oriented, criteria-referenced and formative. This 'assessment for learning' approach encourages student independence and self-evaluation and can lead to active and deeper learning. (Ellery, 2008, p. 421)

This constructivist paradigm underlies this investigation of putting AfL ideas into practice in a higher educational setting. A desideratum for the university classes discussed in this thesis was for students to participate in knowledge and meaning construction as they actively engaged in the learning process; which is the critical emphasis of constructivism (Ellery, 2008). In the classroom, Murphy (1994) describes how learning, teaching and assessment may be intertwined within a constructivist approach:

... with a constructivist framework, students are seen as active collaborators in the building of knowledge. Learning takes place through interaction, existing in the transaction between student and student, student and text, student and teacher. Viewed from a constructivist perspective then, assessment procedures are inevitably a part of the dialectic of teaching and learning, part of the process which defines what knowledge is, what is learned, and how students learn. Assessments which reflect this perspective provide a means for engaging students in self-reflection and for acknowledging their role as collaborators in the learning process. In sum, a constructivist perspective acknowledges the reciprocity and interdependency of assessment and curriculum. (p. 190)

The emphasis here is on effective assessment promoting student engagement with peers and teacher in the collaborative construction of understanding and meaning, and promoting student reflection on this process.

This connection with constructivist learning principles and formative assessment is commonly made and referred to in the assessment literature (for example, Sheppard, 2000; Black, 2001; Marshall & Drummond, 2006; Carless, 2007; MacMillian, 2007).

To briefly summarize Part A, AfL principles and practices are now well established; as are arguments supporting this formative assessment model, both empirically and theoretically.

Part B: Literature overview – AfL in classroom practice

As noted earlier, each of the studies reported in this thesis have reviews of the pertinent AfL literature related to the particular chapter focus in question (for example, regarding feedback or self-assessment). Prior to delineating and previewing the research conducted and presented in those chapters, this brief literature overview takes a broader perspective of the spread of AfL ideas and practice, and considers this model of formative assessment as related to subject differences, tertiary settings and to an EFL context. A rationale for the research engaged in and reported on here is also provided.

Dissemination of AfL ideas and practices

For AfL, the essential orientation point and key text is the formative assessment research review commissioned by the UK's Assessment Reform Group (ARG) and published by Paul Black and Dylan Wiliam in 1998. Black and Wiliam clearly showed the existence of strong research evidence that formative assessment can promote learning and raise levels of student achievement. In the decade following this seminal and highly influential work, AfL has become an established element in official education policy throughout the UK (Daugherty & Ecclestone, 2006), as well as receiving much attention internationally. In Europe, the Organization for Economic Co-operation and Development (OECD) has published formative assessment research using case studies from eight countries: Canada, Denmark, England, Finland, Italy, New Zealand, Australia and Italy (OECD, 2005). This OECD report recognized that formative assessment is a powerful teaching and learning tool in the new

century: “Teachers using formative assessment approaches guide students toward development of their own learning to learn skills that are increasingly necessary as knowledge is quickly outdated in the information society” (OECD, 2005, p. 22). It should be noted that this extensive case study publication focused on lower secondary education contexts.

Lee (2007) also observes that the classroom potential for AfL has become increasingly recognized, and impacted classroom assessment practices in places such as Hong Kong and Australia, in addition to the UK. Leung & Lewkovicz (2006) report that ARG’s ten principles of AfL (listed earlier in Table 2) have received strong professional endorsement, and been incorporated in a number of policy statements in a variety of diverse international jurisdictions, including those countries previously mentioned. At the mid-point of this decade, and reflecting on the years since the research review by Black and Wiliam (1998), Black summarized the situation as follows:

Formative assessment has been a growth industry in the last seven years, fuelled by evidence of two kinds – from research showing that formative practices improve pupils’ achievement, and from practice showing that teachers can transform ideas from the research into productive practices. While its impact has been most extensive in the UK, there have been significant innovations in other countries. (Black, 2005, p. 133)

Later, in 2009, Black noted that a good deal of research attention had recently been focused on formative assessment, and this included second language education. He also commented on the lively discussion among language educators and applied linguistics regarding the relationship between assessment and learning (Black, 2009).

This decade has seen a considerable amount of work in terms of research and professional conferences in the Asia-Pacific region where interest in AfL has expanded in recent years (Klenowski, 2009). To take just one example, since 2006 an annual International Conference on Assessment for Learning has been held, most recently in New Zealand in 2009. AfL has become part of education policy in such places as Singapore, Malaysia and Hong Kong, and the subject of published research in Korea (e.g. Butler, 2009). Primary and secondary education has been the main contextual setting for this development of AfL theory and practice.

However there has been some limited development of formative assessment practice in higher education contexts in Asia. For example, in 2005, tertiary educators from the region met for *ELTA 2005: The First International Conference on Enhancing Teaching and Learning through Assessment*. A range of papers were presented here, and this was later

followed up by a related book describing the alternative use of assessment with illustrations of selected cases from different universities (Frankland, 2007). Content is primarily related to courses such as business, computing, graphic design and engineering rather than the teaching and learning of English as a foreign language.

Despite such developments, and while AfL ideas and practice have indeed become widely disseminated and incorporated in policy statements and classroom practice, its implementation, despite being significant, still remains limited. In particular, high-stakes state-mandated testing which exists in such places as most American states, and in England, makes it difficult to effectively implement formative assessment practices (Wiliam, Lee, Harrison & Black, 2004). Cumming (2009) points out that while AfL principles have been well established, there are scarce examples of their programmatic applications.

Carless (2007) makes this cautionary note also, disclosing that while there are indeed ‘pockets of successful implementation in schools’ in such places as England, Australia and New Zealand, large-scale implementation is considerably challenging and has not occurred (p. 173). Heavy workloads and large class sizes present barriers for teachers wishing to implement formative assessment practices in their classrooms, and may lead teachers to believe that while AfL is theoretically sound, it is rather impractical, time-consuming and incompatible with schooling demands (Carless, 2007). Making a further observation related to non-western cultural contexts, Carless (2007) writes: “Prospects for the implementation of formative assessment are even more daunting in various international contexts where transmissive teaching and summative assessment have characteristically dominated” (p. 173). As will be noted, transmissive teaching and the domination of summative assessment are characteristics of the Japanese educational context.

Yet, it must be acknowledged that AfL has had an impressive impact in the relatively few years since the work of ARG educators and researchers, and others, have promoted and disseminated its ideas and practices. James et al. (2006) write, “the huge interest in assessment for learning in recent years is because it shifts assessment practices to serve formative purposes; to improve and not just measure learning” (p. 1). Yet despite the extensive interest and body of literature already created, there is much to learn and explore, as Black (2005) affirmed: “There is a new field here [formative assessment], which will continue to be rich in possibilities for fruitful interactions between research and practice, to the benefit of both” (p. 135). One possible avenue for fruitful research and interaction with practice, one little explored thus far, is implementing AfL in a higher education context and

with EFL-based subjects. As such, the research described in this thesis is focused on the ‘programmatic applications’ of AfL principles (Cumming, 2009) and their impacts on student learning, rather than engaging in the various and ongoing theoretical debates that exist around this assessment model.

Subject differences and AfL practice

While AfL in classroom practice, and related research and writing, mostly occurs in primary and secondary schooling, it is also concentrated on certain subject areas. According to Harlen and Winter (2004),

There is now a considerable body of literature on the type of classroom assessment that benefits learning, much of it carried out in relation to mathematics, science and English (*as a first as opposed to a foreign or additional language*) . . . (p. 391, emphasis added)

Marshall (2007) posits a rationale for the reasons why much of AfL literature arises from work in math and science subjects.

Much of the literature on formative assessment, or assessment for learning (AfL), has arisen out of the work of math and science. This is possibly because the constructivist view of learning in these disciplines lends itself to very clear paths of progression onto which formative assessment can be readily mapped. Work of a similar nature does not exist in English and the humanities or in the social sciences, where progression is a much messier business. It is hard, for example, in English to be precise about the developmental trajectory of the imagination. (p. 136)

The ARG (2002) recognizes that the way AfL principles manifest themselves in different subjects may differ. However, Black et al. (2003) explain that while different formative assessment techniques may have more or less use in different subjects, all of the broad strategies (e.g. self-assessment, feedback, sharing criteria with learners) are applicable across the subject spectrum. It should be noted here that this claim is a subject of ongoing debate in the related literature, with some writers questioning the application of these formative techniques across all subjects (e.g. Rea-Dickens, 2001; Davison and Leung, 2009). Based on a consideration of the research conducted with teachers in all subject areas in the UK, Black and Jones (2006) observe that “formative assessment has *generic* features, which will apply to learning across all stages and school subjects, and features which are *specific* – to primary teachers and to individual schools subjects” (p. 4, original emphasis). It is these generic

features (such as self-assessment or effective feedback) and their practical application to an EFL context in a Japanese tertiary setting that are the focus of the research projects reported on in this thesis.

One difference noted by Black et al. (2003) is concerned not with a particular school subject itself, but rather how it is interpreted by the teacher in the classroom. Black et al. (2003) contend that when learning goals are very specific, teachers will tightly regulate students' work. On the other hand, regulation of students work will be looser when the goals are less well defined. As we shall see, this notion of learning goal specificity and teacher regulation of student work has implications for the EFL courses described in this thesis, where "subject areas may not have a clear linear or hierarchical structure" (James et al., 2007, p.). Wiliam and Leahy (2007) are more specific about the teachers' classroom regulation of learning in different subjects:

In most teaching of mathematics and sciences, the regulation of learning will be relatively tight. . . . In contrast, in the teaching of language arts and social studies, the regulation will be much looser. Rather than a single goal, there is likely to be a broad *horizon* of appropriate goals, all of which are acceptable (p. 35, original emphasis)

Thus, putting AfL into practice may be affected by the learning goals being focused on, and the classroom subject being taught. However, Hodgen and Marshall (2005) report that, rather than subject specific approaches, AfL research has focused largely on generic pedagogic approaches and strategies that are applicable to all teaching and learning in all subject areas.

There is some limited amount of literature available regarding language teaching and AfL (for example, Ofsted, 2003; Black & Jones, 2006), particularly coming from UK schools and the teaching of what is termed Modern Foreign Languages (MFL) (e.g., French, German, and Spanish). In a report about good assessment practices in MFL, the Office for Standards in Education (Ofsted) in the UK remind teachers that "Assessment in modern foreign languages (MFL) is most effective when it is seen as an integral part of teaching and learning" (Ofsted, 2003, p. 1). This is a common underlying theme of AfL practice: using assessment to inform teaching and learning.

The Black and Jones (2006) article on the learning and teaching of foreign languages and formative assessment is one of the few available resources for modelling how to put AfL into practice in such courses. Aside from the generic/specific formative assessment features distinction mentioned earlier, two further points in particular are worthy of note from Black

and Jones (2006, p. 5). They emphasize:

- [implementing] a formative assessment framework where language learning is demystified through the sharing of learning intentions and success criteria, and that
- the ultimate aim of language learning is for learners to make progress in their foreign language competence and to monitor their own progress as independent language learners.

While Black and Jones (2006) are referring to primary/secondary language learners in a UK curriculum with government set standards (and students and teachers typically sharing the same first language of English), this ‘ultimate aim’ for both learner progression of language competence and progress monitoring, is universally applicable; including the context of the research reported here. These crucial language learning aims (progress in language competence, and monitoring of this progress) are also seen in this thesis as being embedded in the broader aims of language teaching and assessing, with the “ultimate goal” being “improved learning through effective assessment practices” (ARG, 2008, p.18).

Research rationale: AfL and higher education

As with the OECD (2005) research report referred to above, the bulk of AfL research carried out has been in primary and secondary schooling (Yorke, 2003; Murphy, 2006). Indeed, while formative assessment principles may be generic and universally applicable, the AfL literature is heavily based on, and geared towards, the education of younger learners, largely from developed Anglophone cultures (Stobart, 2006). Murphy (2006) observes that the body of knowledge is growing regarding the application of formative assessment in school classrooms, “*but in higher education we are still at an early stage of understanding how effectively this approach to assessment can be developed within different higher education institutions and in different areas of the curriculum*” (p. 42, emphasis added). Reflecting the point made earlier about behaviorist learning theory still retaining influence, and despite the shift towards constructivist theories of assessment noted earlier (Ellery, 2008), Biggs (2003) writes that in HE contexts “the view of university teaching as transmitting information is so widely accepted that delivery and assessment systems the world over are based on it” (p. 12). Pryor and Crossouard (2005) concur, and claim: “Within universities, especially the more prestigious ones, traditional forms of assessment have been largely taken for granted and

developments in formative assessment ignored” (p. 1).

Willis (2006) observes that a lot is known about “the complexity of AfL as a classroom pedagogy” (p. 57), but she also acknowledges that there is still much more to understand and calls for more studies of AfL practice in varying contexts. Little has been written about the application of AfL practice in language learning settings, particularly in higher education. Cumming (2004) writes of surprise at the small number of systematic analyses of assessment practices in language classrooms. Brindley (2007) reports, “To date, relatively few studies have been undertaken in language learning contexts that have investigated classroom assessment” (p.1), and further observes that “relatively little information is available” with regard to the assessment practices of tertiary English language teachers (p. 4).

In one of the few journal discussions of formative assessment in HE, Yorke (2003) lists some of the higher educational pressures which threaten formative assessment usage. Among others, these include: an increasing concern with standards of attainment and resulting emphasis on summative assessment; as well as the increase in staff/student ratios which consequently decrease attention given to individual students. While acknowledging these pressures, Yorke (2003) proposes that, “A major challenge for higher education is to respond to the main ‘message’ of Black and Wiliams’ (1998) review [that] formative assessment is, after all, a key tenet of good teaching.” (p. 482). The research conducted and reported on here is one response to Yorke’s (2003) HE formative assessment challenge.

Finally, in one of their more recent discussions of AfL in the classroom, Black and Wiliam (2006) propose areas of further exploration and research. These include, “the need to extend work of this nature to other groups, notably pupils in infant and junior school and *students in post-16, tertiary, and non-statutory assessment settings*” (p. 24, emphasis added). This thesis extends AfL research to other groups and contexts, and is the only work of this kind implementing the full range of key formative assessment procedures into classroom practice in a tertiary setting.

Part C: Research focus

As noted at the beginning of this introductory chapter, and reflecting a point about research foci made by Dunleavy (2003), this thesis involves taking established ideas (AfL) from one part of a discipline (mainstream education in primary/secondary schools) for a walk and putting them down somewhere else, applying them to a different context (a higher education

EFL setting, with adult L2 students). In the past 30 years, the English Language Teaching (ELT) profession has developed a strong knowledge base for effective professional practice, much of it derived from research and thinking from a wide range of contributing disciplines, including applied linguistics, psychology and education (Hedge, 2000). This thesis continues that borrowing tradition of deriving knowledge from a neighboring discipline; in this case, from mainstream education's formative assessment principles and practices intended to promote student learning. While each of the five studies included in this thesis has a corresponding research focus related to it, the overall thrust of the thesis concerns this central research question:

◆ **What does AfL practice look like in a HE context with adult EFL students?**

Set in a Japanese context at a woman's university in Tokyo, this thesis will "put classroom flesh on the conceptual bones of the idea of assessment for learning" (Black & Wiliam, 2006, p. 25) by describing and analyzing "a variety of living examples of implementation" (Black & Wiliam, 1998b, p. 16). The research reports comprising this thesis will show AfL principles and practices brought to life in a series of five different classroom contexts, each focusing on a particular key assessment procedure or strategy.

In the earlier section of this introduction concerning AfL fundamentals, the key ARG (2002) definition of AfL explained it as a process for using assessment evidence to determine where learners are in their learning, where they need to go and the best way to get there. With regard to the practicalities of implementation, Gardner (2006) observes:

Unpacking this deceptively simple definition, in terms of classroom practice, reveals *a complex weave of activities involving pedagogic style, student-teacher interaction, self-reflection (teacher and student), motivation and a variety of assessment processes*. (p. 2, emphasis added)

This description of AfL in practice as a 'complex weave of activities' provides a succinct summary of the various assessment procedures carried out and reported on in this thesis.

Background

This thesis is comprised of a series of research projects, which "set out to investigate the effects of pedagogical interventions in real classrooms" (Tomlinson, 2005, p. 150). The five project reports that make up the thesis examine various key components of the AfL framework.

The impetus for this research lies in the 1998 Black and Wiliam research review, and in particular with a later, related research project in schools in England described in the book entitled *Assessment for Learning: Putting it into Practice* (Black, Harrison, Lee, Marshall & Wiliam, 2003).

Black and Wiliam's (1998a) formative assessment research review identified the following four groups of practices that were shown to be effective in promoting student learning:

- ◆ questioning
- ◆ feedback
- ◆ sharing criteria
- ◆ self-assessment

In a follow-up research project, these four practices were developed and re-shaped by teachers as they put AfL into classroom practice. Working with colleagues from King's College London, Black and Wiliam and a group of fellow researchers collaborated in a two-year development and research project with a group of 36 science, math and English teachers from six secondary schools in England. The findings of the King's Medway, Oxfordshire Formative Assessment Project (KMOFAP) were summarized for teachers in a short booklet (Black et al., 2002), and reported in more detail in the book mentioned above (Black et al., 2003).

In the course of implementing the four dominant strategies or procedures identified by Black and Wiliam (1998) and listed above, they underwent some changes and alterations when faced with classroom and school realities; 'where the rubber meets the road', as it were. For example it became obvious to the KMOFAP researchers and teachers that an important part needed to be played in making formative use of summative testing. In addition to self-assessment, the development of peer-assessment practices was also recognized. Also, the initial separate focus on sharing criteria was changed. Because sharing criteria with learners served a number of other areas, rather than standing alone it was subsumed into both the self-assessment and feedback categories (Black et al., 2003). As a result of these alterations and modifications, the KMOFAP project focused on teachers experimenting with formative practice in these four updated areas:

- questioning
- feedback through marking
- peer-and self-assessment by students

- the formative use of summative tests

The secondary school teachers in the project focused on these specific areas in order to implement and experiment with AfL in their classes. As Marshall (2007) points out, even though each can be considered discretely the four strategies overlap; “How can students assess themselves or their peers without understanding the criteria? How can feedback be meaningful unless the task is appropriate?” (p. 138).

Through the process of researching and writing a report about self-assessment, I learned about the KMOFAP project, and subsequently the formative assessment framework identified as AfL. I decided that putting these fundamental ideas about formative assessment and classroom learning into practice in my teaching context at a Tokyo university would be a worthy doctoral research project to undertake. In addition, it would make some original contribution to the body of formative assessment research by dealing with adult students in a tertiary setting, and also for learners experiencing English as a foreign language. Essentially, the research described in this thesis is a solo (i.e. single teacher) version of the KMOFAP research described in Black et al. (2003) involving dozens of secondary school teachers.

In considering the design of my classroom research, I decided to separate the student-as-assessor procedures of self and peer assessment into two separate research projects. In the KMOFAP research both self and peer assessment are developed in tandem, with peer assessment seen as an important complement to improving a students’ ability to self-assess their own work. My rationale for separating them is that they are significant subject areas in their own right; each with their own associated body of research literature. Separating them into two distinct research foci would make them more manageable to implement and examine in classroom practice. However, as will be seen, self-assessment is an element of all five of the project reports included here, and both peer and self-assessment are important and complementary procedures in the chapter focused on feedback through marking.

As for ‘sharing criteria with learners’, following the KMOFAP practice, this crucial element is subsumed within each of the research projects, particularly through the use of rubrics as both learning and teaching tools. Implicit in making the sharing of assessment criterion an element of each project is recognition of the fact that research shows the importance of students understanding what counts as success; both in different curriculum areas and in their own stages of development as learners (James et al., 2007).

The Afl wheel

Thus, the four clusters of classroom practice developed and put into practice in the KMOFAP research were taken and applied to university courses I taught in Japan, with the added modification of individually examining the issues of self-assessment and peer-assessment. This slight modification resulted in a total of five Afl research projects. These five studies, and the key assessment procedures they are based on, are visually represented in the Afl wheel seen below in Figure 1. The figure represents the strategies identified by Black & Wiliam (1998a) and updated by Black et al. (2003) as being effective in enhancing learning in the classroom. The figure is accompanied by an important reminder of the primacy in formative assessment design and practice of promoting student learning.

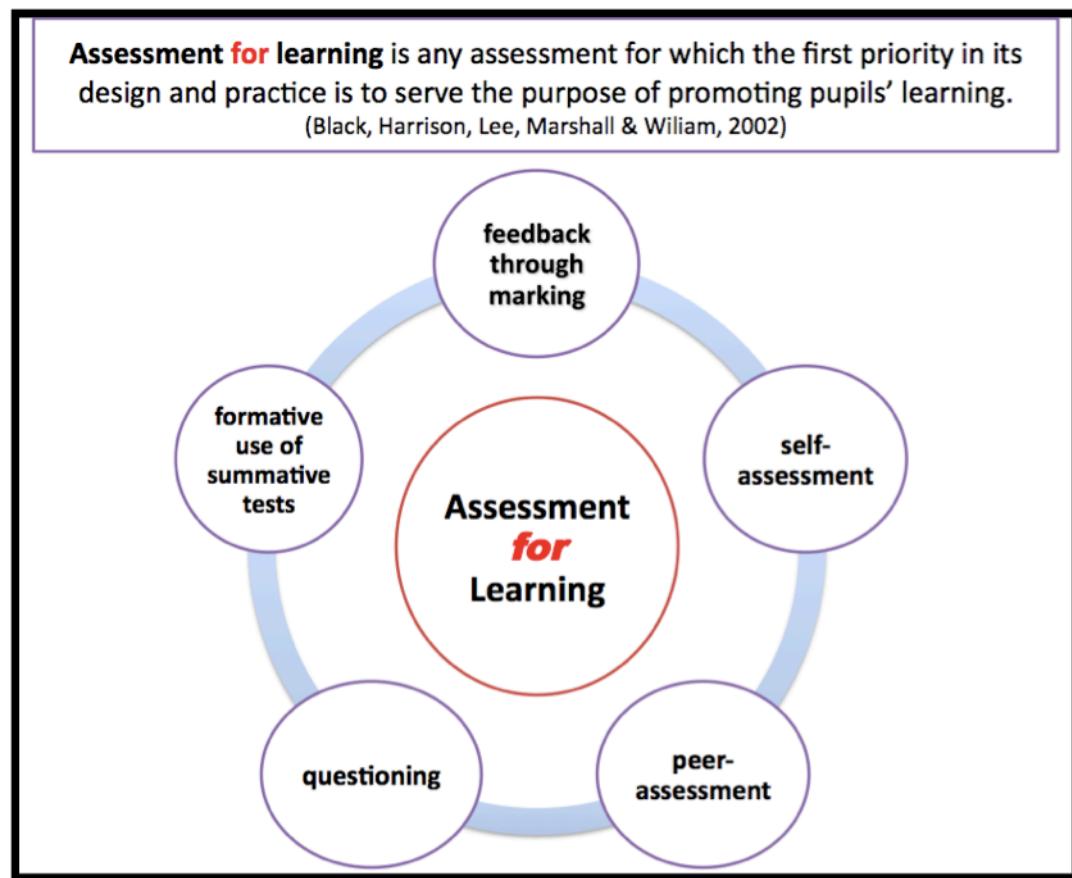


Figure 1. The Afl wheel with five key components

The wheel format reflects the overlapping, interconnected nature of the procedures, as noted by Marshall (2007). This circular representation also shows that it is a moving, dynamic system that affects, and is affected by, the classes, individuals and contexts surrounding Afl in practice. In addition, the Afl wheel serves as an organizational mechanism for the

individual project reports presented in this thesis. It will be used periodically as a signposting device to mark the formative assessment procedures being focused on in each chapter, and as a visual reminder of the larger framework of AfL principles and practices.

Collaboration with colleagues?

A key distinction between the research projects in this thesis and the KMOFAP research is the fact that rather than a group of teachers collaborating together, here we have a description of one teacher working alone to put AfL ideas into practice. Black et al. (2002, p. 23) contend that working with a collaborative group that is trying out similar AfL innovations is ‘almost essential’. Yet the same Black et al. (2002) booklet also provides some advice for what someone can do as an individual teacher, presumably recognizing the possibility of a solo teacher working in isolation to effectively implement these practices. According to James et al. (2007),

Evidence for the effectiveness of AfL is derived mainly from carefully controlled but small-scale experiments which have involved intensive support to teachers. If these innovations are to be scaled-up and sustained across the system, they will have to grow with much less support. (p. 7)

The research described in this report involved no intensive support or collaboration with colleagues, showing that a teacher working alone can implement these ideas into practice.

Investigator triangulation (Denzin, 1978), involving multiple researchers, has it’s advantages in establishing research validity, and lack of such triangulation may be viewed as a limitation of this research project. However, it is often neither practical or possible to assemble a team of different teacher-investigators given such factors as time constraints, individual schedules, teaching styles and, indeed, willingness to be involved. Such was the case in with this research agenda at TWCU in Japan, and a reason for this being the report of one individual instructor rather than a collaborative project.

I had considered getting colleagues involved in this research project but this posed a number of problems. Most of the classes taught in the Department of English at TWCU are done by part-time teachers, who have little time or incentive in engaging in classroom-based AfL research. Of the small number of full-time EFL colleagues, in my department, they were either just beginning employment with a heavy course load to teach, or coming to the end of

their term-limit contracts and naturally preoccupied with other things. Attempting to recruit other teachers into a KMOFAP-style project, and providing intensive support to them, would have proved very difficult with trying to complete my own heavy course load of teaching, and writing up my doctoral research. Also, as already noted, formative assessment can be misunderstood or weakly implemented by teachers, and without an extensive support and training component (e.g. KMOFAP), such misunderstanding or misapplication might have been a distinct possibility in collaborating with a group of colleagues in researching AfL.

After mulling over these considerations, working alone and engaging in my own construction and application of AfL knowledge, was consciously deemed the best way to proceed. As Black and Wiliam (1998b) made clear, changes in classroom practice are not easily implemented and the rewards of formative assessment will only be realized, “if each teacher finds his or her own ways of incorporating the lessons and ideas . . . into his or her own patterns of classroom work”(p. 15). This thesis is a report of a university instructor finding his own way of embedding these AfL ideas into ‘patterns of classroom work’ with the guiding aim of promoting student learning.

Five research studies

In the KMOFAP research, the researchers encouraged the 36 teachers involved to select and experiment with some of the AfL strategies and techniques. Teachers were then asked to formulate an action plan of the practice they wished to work on (for example, questioning or self-assessment) and “to identify a single focal class” in which the strategies would be introduced at the beginning of the school year (Black et al., 2003, p. 14). In a similar manner, regarding the research presented in this thesis, a focal course or class was used to describe how the key components of AfL were put into practice (deliberately, or reflectively as noted below) to promote student learning.

As outlined in the synopsis, the three body chapters of this thesis are comprised of a series of five reports of AfL procedures being used with a central class or course. Chapter 1 focuses on the issue of feedback, and includes reports of teacher self-assessment of written feedback, and the impact of this feedback on subsequent essay drafts. The second chapter focuses on the student in the role of assessor, and begins with a report of self-assessment of student class participation in a freshman English class. A report on peer-assessment of presentations in a public speaking course completes this chapter. It should be noted here that versions of both research reports included in Chapter 2 have already been published (White, 2009a; White,

2009b). The two research reports in the final chapter focus on the issues of student-generated questioning and classroom dialogue, and using summative assessment for formative purposes.

Table 5 below presents an overview of the five research reports, in order of appearance in this thesis. It includes: the focal course or class in which the procedures were implemented, the number (and year) of university students involved in each project, as well as the research focus and the data collected for each individual study.

Table 5. Overview of five AfL research studies

AfL procedure	Focal course/class	Students	Research focus	Data sources
1. Feedback through marking	Junior Composition	23 juniors	1. Nature and form of first-draft teacher feedback on essays 2. Impact of feedback provided, and student views of feedback	1. Examining and coding teacher feedback provided on 21 essays (first draft) 2. Comparing feedback on draft 1 with changes in final draft 2 3. Student survey
2. Self-assessment	Communication Skills (3 classes)	70 freshmen	Assessing the self-assessment of class participation framework used according to AfL principles, and issues of validity, reliability, etc.	1. Self-assessment tool data (completed 3 times per student in one semester) 2. Student survey
3. Peer-assessment	Effective Public Speaking (2 classes)	55 sophomores	Student views about peer assessment: its usefulness in promoting public speaking skills, and use for determining final grades	1. Student survey 2. Peer rating score sheets for presentations
4. Questioning (student-generated)	Communicative Writing	22 sophomores, juniors	Whether student-generated questioning promoted learning about course content	1. Question samples from students 2. Two mini-surveys and final student survey
5. Formative use of summative assessment	Public Presentations	20 juniors	Synergizing formative and summative assessment of students' final presentation slide shows	1. Assessment 1 - self-assessment 2. Assessment 2 - teacher formative assessment 3. Assessment 3 - final summative assessment

The individual research project reports included in the body chapters of this thesis will clarify in more detail the nature of each of these ‘living examples of implementation’. A total of 190 students were included in this research regime conducted at a university in Tokyo. None of the students were part of more than one research project, so there was no overlap between

participants and studies. These were all Japanese adult females; first, second or third year university students in the 18-21 years old age range. The research projects previewed in Table 5 above took place over a three-year period, from 2005 to 2008. All research was approved by, and conformed to, ethical requirements for conducting human research at Macquarie University.

Cycles of formative assessment

The timing of AfL strategy usage is a crucial factor, considering the importance of such factors as feedback and self-monitoring of understanding of performance while a task or activity is in progress. For clarification purposes, Wiliam and Thompson (2007) categorizes formative assessment types into three groupings: long-cycle, medium-cycle and short-cycle. Table 6 clarifies these terms with regard to focus and length.

Table 6. *Formative assessment types (from Wiliam & Thompson, 2007)*

Type	Focus	Length
Long-cycle	Across marking periods, quarters, semesters, years	4 weeks to 1 year
Medium-cycle	Within and between instructional units	1 to 4 weeks
Short-cycle	Within and between lessons	5 seconds to 2 days

According to Wiliam (2007) the research literature shows improvements in student achievement are evident only from short- and medium-cycle formative assessments.

The length of formative assessment cycles described in the five reports in this thesis generally would fall in the medium-cycle category, primarily because the course context described below includes weekly classes. Thus, teacher formative feedback, for example in first draft of student essays (Chapter 1) or presentation slide show first drafts (Chapter 3, Part B) entail feedback response one week following submission. However, as will be seen, short-cycle formative assessment is also evident throughout this thesis, in the form of self-assessment and peer-assessment activities, as well as ongoing, informal teacher assessment within lessons. Chapter 2 describes a research project regarding student self-assessment of class participation completed across a semester - for three times approximately every three

weeks – and, as such, may be considered an example of long-cycle formative assessment.

Research methodology overview

In applied linguistics, second language research has been defined as “any systematic and principled inquiry in language learning and teaching” (Brown & Rodgers, 2002, p. 12). As such, this thesis is a systematic and principled inquiry into the application of AfL theory and practice in university classroom settings in Tokyo where English is studied as a foreign language. While each of the individual reports provides its own account of the research methodology used, an overview of procedures used across the group of studies provides a broader preview of the research as a whole. First of all, an important point about the timing of data collection should be noted.

Timing of data collection

According to Allright (1999), the central methodological question for any research investigation is the issue of collecting relevant data. Such data collection is indeed a central issue in this doctoral research. Allright (1999) goes on to assert that decisions about data collection techniques can be determined for a particular research project only after considering the type of data needed, and where they are to come from.

As mentioned, this research project covered a three-year period, and during this time I taught and conducted research in the five courses included in this thesis. It should be noted here that the first three of these reports dealing with feedback, self-assessment, and peer assessment made use of student and course data collected *prior* to my decision to focus on examining AfL practices for a doctoral thesis. After settling on investigating these five AfL procedures for my research, I also decided to make use of data previously collected in the natural course of teaching those particular classes. These three particular assessment frameworks were put into place without a particular knowledge of AfL principles or practices. Rather they were simply part of the assessment make-up of the course. Thus, the reports in Chapters 1 and 2 are in a sense reflective exercises, looking back at assessment procedures previously used and examining them in light of AfL recommended practice. The use of such previously collected data does not, I believe, detract in any way from an examination of these key procedures or a discussion of the reality of putting them into practice. While the data was previously collected with the thought that it may potentially prove useful, the particular research focus for each study was later based on the AfL literature. For example, previously

collected student essay drafts were reflectively examined with a focus on teacher feedback effectiveness after reading about the key AfL concept of ‘closing gaps’ between present understanding or skills and desired objectives (reported in Chapter 2, Part B). In this manner, previously collected data was used to exemplify and analyze AfL procedures in practice for the first three of the five research reports.

For the final two reports, regarding questioning and the formative use of summative assessment, focal classes and procedures were specifically chosen to experiment with these strategies, following recommendations set out in the related AfL literature. In that sense, these final two research projects adhere more closely to the KMOFAP format used by teachers in the UK of specifically choosing a focal class and putting an AfL procedure into practice.

In any case, despite this distinction in the timing of data collection, all five research studies involved “the practical application of formative classroom assessment, with the goal of enhanced meaningful student learning” (McMillan, 2007, p. 5).

Characteristics of research methodology

While there are a number of levels at which research can be categorized and analyzed, it is the responsibility of the researcher to demonstrate that methods and findings are credible and important (Duff, 2001). One useful model for categorizing research methodology is the model proposed by Grotjahn (1987). In his model, three different aspects of the research process can be distinguished:

1. the method of data collection (whether through experiment, or non-experimentally)
2. the type of data (qualitative or quantitative)
3. the type of analysis (statistical or interpretative)

Following this model, this thesis may be described as non-experimental, primarily qualitative, and utilizing interpretative data analysis. Grotjahn (1987) labels such research as falling within an ‘exploratory-interpretative’ paradigm.

While having some quantitative elements, the research conducted in these university classes in Tokyo may be broadly categorized as a “qualitatively oriented inquiry” (Cumming, 2004, p. 9). In analyzing different research methodologies, Burns (2000) explains that, “The qualitative researcher attempts to gather evidence that will reveal qualities of life, reflecting the ‘multiple realities’ of specific educational settings from participants perspectives”; and

further notes that this is done by way of qualitative methods which “attempt to capture and understand individual definitions, descriptions and meanings of events” (p. 388). Reflecting the spirit of assessment for learning, this thesis may be considered as ‘research for learning’; an attempt to reveal, understand, and learn from the ‘multiple realities’ of putting AfL into practice. According to Cumming (2004), “Serious consideration of the uses of language assessment requires adopting research methods that investigate people’s attitudes, beliefs, cultural values and ways of interacting” (p. 9). Research methods adopted here reflect this view. While four of the five reports, except the final one, involve the quantifying of data (particularly for analyzing survey results), use of such descriptive quantitative techniques plays an important but supportive role in the overall qualitative research design of the set of reports contained in this thesis. Croker (2009) explains that unlike quantitative researchers who focus on measuring outcomes, those engaged in qualitative research focus more on the process of what is going on in a setting.

Some further research characteristics to be found throughout the research presented in this thesis include the following three points regarding the use of surveys, self-assessments, and rubrics:

- 1.** Extensive use of student views and commentary are a strong presence throughout this thesis, primarily gleaned from survey responses. Four of the five reports (except the final one on the formative use of summative assessment) include a student survey of the AfL procedures used. In second language research surveys and questionnaires are used primarily to collect data on such things as attitudes, motivation, self-concepts and other phenomenon which are not easily observed (Selinger & Shohamy, 1989). In some sense, such survey research is to be expected in a report about an assessment framework which puts students at the center. As Harlen (2007a) observes, “The students are at the center of the process, since it is they who do the learning” (p. 120). Just as AfL is student-centered, the contents of this thesis, with its extensive use of learner views and voices, also reflect a student-centered philosophy. Making extensive use of student thinking and feeling about the assessment tools and frameworks used is important also because their perceptions of classroom activities and assessment processes have an impact on their learning. According to Prosser and Trigwell (1999), “the way students perceive their learning and teaching situations is central to the quality of their learning” (p. 81). In addition, the inclusion of this element in the research methodology addresses the criticism made by Broadfoot and Black (2004) that attention to student perspectives has played too small a part in assessment research.

It is perhaps useful at this point to note the potential advantages and disadvantages of using surveys.

Potential benefits and drawbacks of survey research

A survey, usually in the form of a questionnaire, gathers data at a specific point in time and is the most commonly used descriptive method used by educational researchers (Burns, 2000). It is also the most commonly used data collection method used in these research studies. In Table 7 below are presented some of the potential advantages and disadvantages that may be present when engaging in survey methodology.

Table 7: *Advantages and disadvantages of survey research* (from Burns, 2000, p. 581)

Advantages
It is one of the few techniques available to provide information on beliefs, attitudes and motives.
It can be an efficient way of collecting data in large amounts in a short period of time.
It can be used on all normal human populations, except young children (i.e. flexibility of use).
If confidentiality can be guaranteed, more truthful responses may be elicited (as compared to an interview or focus group, for example).
Structured surveys are amenable to statistical analysis.
Information from a number of survey respondents can go beyond description to provide patterns in data.
Disadvantages
Poor responses from respondents will be caused by instruments that are complex, ambiguous or vague.
Ambiguous, incomplete or inaccurate information cannot be followed up.
Respondents may be limited from providing free expression of opinions due to the design of the survey instrument.
Open-ended parts of a questionnaire may produce data that is difficult to analyze systematically.
Possibility of misinterpretation of the questions by the respondents.

This final possible disadvantage above is particularly important when the survey language is not the respondents first language, as in the case of this research. All care was taken to ensure

that questionnaire items were presented in clearly written English, and the self-assessment survey used with freshman students described in Chapter 2 included a Japanese translation. The surveys used were designed for a number of very specific purposes and courses; in some cases (e.g. the peer assessment and student-generated questioning research) the only section of the course on campus. As such, piloting these surveys would have posed some problems and, indeed, at times there was simply no time to go through a careful piloting process. However, the surveys used in these reports were typically shown to colleagues, both native English speakers and Japanese, prior to administering them, and adjustments in language or content were sometimes made.

A related point should be made here about another key feature of some survey research - interviews, whether of an individual nature or with a group. Focus groups, or group interviews, are another are a form of qualitative research. Krueger and Casey (2000) define a focus group as "a carefully planned series of discussions designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment" (p. 5). Interviews with students, or groups of them, can lead to a deepening of understanding of the phenomenon being researched and students' impressions of it (e.g. feedback, engaging in self-assessment) that a written survey may not be able to tap into. As such, focus group interviews could have been potentially a valuable source of data in this series of research studies. However, they were not used in these research studies. There are a number of reasons for not utilizing this qualitative research method, including: simply neglecting to consider the focus group option; my lack of interviewing expertise; the logistical challenges of administration (arranging times, rooms, recording equipment); effects of lack of anonymity for frank student responses to questions; and the consequential challenges of data transcription and analysis. In terms of the research conducted at this university in Japan, an additional factor should be considered. Interviews, including focus groups, pose heightened challenges when conducted in a foreign language with students of varying levels of proficiency, including beginning levels of communicative competence of in English. This was the main reason why interviewing was not part of the research methodology employed in the research reported in this thesis.

2. As noted previously, a self-assessment element is included in all five research projects. This fact is a reflection of the importance of learner meta-cognition in determining learning destination, current location and routes forward. James et al. (2007) refer to self-assessment as a three-stage process of: stepping back, reflecting, and then stepping back in. They describe

self assessment as involving, “a cognitive ‘stepping back’ from the learning process during engagement with it, in order to reflect on it critically and strategically, often in dialogue with others, and then to ‘step back in’ to restructure or transform the learning process” (James et al., 2007, p. 28). Also included in this thesis are examples of *the teacher* self-assessing, and engaging in this three-move process of stepping back, considering, and stepping back in.

3. The use of rubrics as learning and teaching tools is a common feature of the research presented here. Four of five reports, except the student-generated questioning study, include assessment guidelines for the purposes of self, peer, or teacher assessment. The importance of sharing criteria with learners is fundamental to AfL practice, and the effective use of rubrics enables this criterion sharing to occur. Reddy (2007) notes the connection between rubrics, assessment and learning as follows: “Rubrics support constructivist theories of learning, which emphasize upon students and teachers to look to assessment as a source of continuous feedback for improvement of learning processes rather than as an evaluative process” (p. 8). As such, the effective use of rubrics accords well with AfL theory and practice.

The reader will note that the range of investigations, data sets and analytical perspectives used in this thesis is wide-ranging, and contains elements of a variety of particularly qualitative approaches to research. Croker (2009) provides a snapshot of a number of qualitative approaches used in applied linguistics research. These include case study, ethnography, phenomenology and mixed methods. They are included below in Table 8, with a brief description. Elements of all four of these approaches may be found throughout the range of classroom research projects described in this thesis.

Table 8: *Qualitative research approaches evident in this thesis*

Research Approach	Description
Case Study	Creates an in-depth description and analysis of a ‘bounded system’ – one individual institution, or educational context. By concentrating on a single (or few) case(s) this approach can describe particular learning or teaching processes or research setting in great detail. Case study uses multiple sources of data and data collection methods, and is often combined with other qualitative and quantitative research approaches.
Ethnography	Refers to both a research process and also the product of that research. It describes and interprets the common patterns of a culture-sharing group. Ethnography is not defined by how data is collected, but rather by the lens through which data is interpreted; the goal is to recreate for the reader the shared beliefs, practices, knowledge, and behaviors of a group of people.
Phenomenology	Describes the meanings that several individuals make from experiencing a single phenomenon. The purpose of a phenomenological study is to reduce individual experiences of such phenomenon to a description of the basic ‘essence’ of that experience, by creating a composite description of that experience for all of the participants. In a broader sense, phenomenology as a school of philosophical thought underpins all qualitative research, because of its interest in understanding and representing the subjective experience of participants.
Mixed Methods	Combines both qualitative and quantitative research methods in a single study. A mixed methods study could emphasize qualitative and quantitative data equally, or give one type greater emphasis.

(adapted from Croker, 2009, pp. 14-15)

While coming more at the qualitative end of the research continuum where understanding is sought by the observation of phenomenon in natural (e.g. classroom) settings (Nunan, 2004), this thesis also employs a mixed methodology with quantitative elements included, particularly in the first report about feedback on student essays. Case study research elements are also evident in that the research takes place at an individual institution, and in five individual courses where great detail is provided about each. This doctoral research is ethnographical in that common patterns of behavior and attitudes are described for particular groupings of students, for example, for seventy students in the self-assessment report in the second chapter. The phenomenological aspect of qualitative research is also evident

throughout this thesis as individual student voices reporting their subjective experiences of different AfL techniques are incorporated.

In addition to providing the useful snapshot of approaches outlined above in Table 7, Croker (2009) is also perceptive in reminding us that in these qualitative research approaches, it is *the researcher* who acts as the primary instrument of data collection and analysis.

Researcher as learner

As already indicated, and to be subsequently followed up, the teacher/researcher is also considered a *learner*, both in the AfL framework itself, and throughout this research process. In a book entitled *Assessment for Learning in Higher Education* (one of the first published usages of the AfL phrase, later popularized by the ARG), Boud (1995) writes:

Good assessment is that which both closely reflects desired learning outcomes and in which the process of assessment has a directly beneficial influence on the learning process. This is a major challenge for all staff . . . They will need to become researchers of students' perceptions, designers of multi-faceted assessment strategies, managers of assessment processes and consultants assisting students in the interpretation of rich information about their learning. (p. 42)

As well as the description of what effective assessment entails, this challenge provides a good description of my role in this AfL research investigation in a HE context. In this research process, among other things, I become:

- a student perceptions researcher,
- an assessment strategies designer,
- an assessment processes manager, and
- a student learning consultant.

The teacher/researcher is a learner in this process, and, reflecting a constructivist perspective, one actively engaged in “a progressive development of understanding” (Harlen, 2007b, p. 116) of the practicalities and challenges of implementing a range of formative assessment procedures in a series of university classes in Japan.

Research context

In the final section of this thesis introduction, we take a brief look at the broader assessment context of Japan, and briefly introduce the university where this research was conducted.

Japan

In Japan, English is considered as a foreign language, that is: “a language studied in an environment where it is not the primary vehicle for daily interaction and where input in that language is restricted” (Oxford, 2001, p. 359). Tomlinson (2005) notes that the surrounding social context will have an affect on EFL learners’ behavior and expectations. To use an example from this research, Chapter 2 deals with a spoken English university course which is mandatory for all freshman students, despite the fact that opportunities to communicate in the target language are very limited (and not needed) outside the language classroom. These ‘outside’ contextual realities may impact the community of learners (including the teacher) ‘inside’ the classroom.

In implementing formative assessment, the learning context - which includes the social and political environment occurring outside the classroom - is viewed as a critical consideration (Stobart, 2006). In referring to the USA, Gardner (2006) makes the following observation, which could equally apply to assessment in the broader context of Japan: “[with its] long-established variations of summative assessment, much of it geared to high-stakes selection, assessment for learning is barely on the horizon” (p. 202). Student experiences of assessment in Japan essentially means those of the summative variety, often of a high-stakes nature; particularly the competitive entrance exams that provide access to upper secondary schools and tertiary educational institutions. Timothy McVeigh, a social anthropologist who lived, taught and researched in Japan for 15 years, summarized the local assessment context as follows:

Examinations, of course, are not unique to Japan. But while in other places testing is used to enhance and facilitate learning (examinations for learning), the argument can be made that in Japan the relationship between testing and learning is often reversed (learning for examinations). Official rhetoric has it that schooling is for learning, self-cultivation, and personal development. But the reality is far different: schooling is a type of training for how to sit and pass examinations. This reality has profound ramifications for how students are socialized to view their role as students, for school culture, and the ultimate purpose of learning. (McVeigh, 2002, p. 35)

Such a wider ‘assessment reality’ indeed puts AfL ‘barely on the horizon’, affording very little room or consideration for formative assessment principles and practices in secondary education in Japan. Stubbs (2000) warns that “language teachers must pay attention to local conditions rather than taking a set of ideas around the world with them” (p. 16). This thesis does involve ‘taking a set of ideas around the world’, and the local Japanese context and students’ assessment experience background are important consideration in the reports of AfL in practice to follow.

According to Black and Wiliam (1998a), students bring to their work models of what learning entails, which may in fact interfere with their own learning. In the same manner, based on past experience, tertiary students bring to their work models of what assessment entails (or should entail) which also may cause learning interference. This is an important consideration in this thesis, as it makes significant use of student opinions and views of assessment frameworks. Boud (1995) reminds us that, “Students are not simply responding to the subject – they carry with them the totality of experiences of learning and being assessed and this certainly extends far beyond concurrent and immediately preceding subjects” (p. 37). As we shall see, one reason why many students responded positively to the formative assessment procedures used may have been the refreshing change from the typical assessment experiences they have had.

In ending this brief, but necessary, discussion of broader contextual considerations of the research described here, we turn to Tomlinson (2005) and his writing about the universal application of principled procedures on EFL teaching. He concludes as follows:

It seems that no particular pedagogical procedure can be used effectively without some modification from context to context, but also that a procedure that proves effective in one context of learning has the potential to be effective in other contexts of learning too. (Tomlinson, 2005, p. 150)

The effectiveness of AfL practices have been shown in primary and secondary schools with younger learners in classes where, for the most part, they share English as a first language with the teacher. Here, we consider the potential effectiveness of formative assessment practices with adult EFL students who do not share the same L1 as the teacher and have various degrees of English proficiency in both productive skills (speaking, writing) and receptive (listening, reading) skills.

It should be noted here that I, as the sole teacher involved in planning and teaching the

courses and classes reported in this thesis, have what may be generously described as ‘limited’ proficiency in Japanese. Regarding the classes described in this thesis, in addition to course books and supplementary materials, all classroom instruction was conducted in English, as was all teacher-student communication (whether in groups or individually).

Tokyo Woman's Christian University (TWCU)

While students’ background experience with assessment, and the wider cultural context are important considerations, Stobart (2006) also points out that “the opportunities for formative assessment in a centralized curriculum with high-stakes national testing, will be different for those teachers who enjoy more autonomy over what they have to cover and how to assess it” (p. 17). Teaching in a tertiary setting did provide me with such autonomy over course curriculum and assessment frameworks.

Tokyo Woman's Christian University (TWCU) is a four-year liberal arts college in Japan, with approximately 3000 female students. Almost all students are Japanese, with a minimal number of international attendees. The comparatively small university is comprised of ten departments, such as History, Philosophy, Psychology, Japanese Literature and the Department of English, where I teach and which includes both literature and linguistics majors. Set in the heart of Tokyo, TWCU is a popular HE institute for Japanese women and attracts students from all over the country. While TWCU may be considered a ‘religious’ institution and there are a few mandatory freshman courses in Christianity, in reality it is also very secular in nature. Reflecting Japanese society in general, the number of practicing Christians among the student body and faculty is, in fact, minimal. Figure 2 below shows a campus view from the main entrance gates.



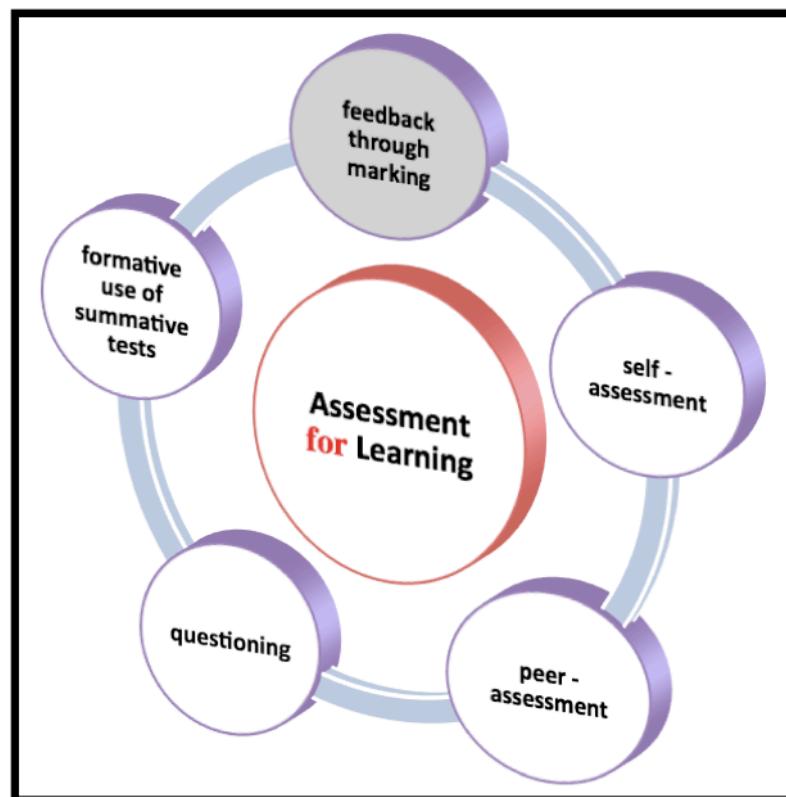
Figure 2. TWCU's main building and square

Engraved on the façade of the main building is the university's Latin motto "Quaecunque Sunt Vera". It is a biblical phrase (from Philippians 4:8), and is translated in TWCU brochures as "Whatsoever Things Are True". With regard to this thesis, the research conducted at this tertiary institute may be viewed as an attempt to determine whether 'truths' established in the AfL literature about the types of formative assessment practices that are effective in promoting student learning, are also 'true' when put into practice in a completely different educational context with completely different groups of learners.

We now move on to the main body of this thesis, containing the five research project reports investigating AfL procedures in this context. We begin with a focus on the issue of feedback, an optimum starting point considering its primary importance in formative assessment. The reader will note that three other key AfL components (self and peer assessment, as well as using summative assessment for formative purposes) are also elements of the feedback for learning chapter which follows.

Chapter 1: Feedback through marking

- **Part A:** Teacher self-assessment of written feedback
- **Part B:** Does the feedback feed forward?
Student response to and views of teacher feedback



Preview

With classroom-based assessment, in all its variety and complexity, the teacher engages in a process of gathering information about what a student understands, knows and can do. Giving feedback to students on how we view that information, and to help improve their understanding, knowledge and skills, is at the heart of formative assessment. As already noted in the introduction (Table 1), the provision of effective feedback is first on the list of ARG's (1999) five factors for using assessment to improve learning.

This thesis chapter focuses on this crucial role that feedback plays in the teaching-learning process, and within AfL theory and practice itself. In the context of a third-year academic writing course, and one set of 21 student essays, an exploration will be made of: 1) teacher feedback practices (What will the results of a self-assessment of feedback practices reveal?), and 2) student response to the feedback provided (Do the teacher's practices become 'feedback for learning'?). This thesis chapter is comprised of two parts, each a separate, yet related, study of feedback in one particular academic writing course.

Part A. Teacher self-assessment of written feedback

In order to analyze the effectiveness of the feedback provided to students, it is first necessary to closely examine the nature and form of the feedback itself. In this section I provide an account of myself as a teacher, shifting the observer's focus to closely examine my own feedback practices; the what, how and why of the feedback responses I have provided to one particular group of students on one particular essay writing task. I will take a detailed, systematic look at my feedback practices (for the first time) and aim to see what can be learned, and assess these practices against recommended practice from the related literature on the provision of feedback. Considering AfL theory about providing feedback, and in particular two recent publications with recommendations and suggestions for optimum feedback practice, I conduct a teacher self-assessment of my feedback practices.

Part B. Does the feedback feed forward? Student response to and views of teacher feedback

Following up on my self-assessment of feedback, this complementary study will examine the results of my feedback practices. Does the feedback provided help enhance students' future understanding and level of achievement- is it feedback that feeds forward? Are student essays

improved through my feedback and student responses to it? Also, through analysis of student survey responses, this section will examine how students feel about the feedback process implemented in the academic writing course being discussed.

To sum up, this chapter looks at the feedback issue from two different perspectives; first from the teacher's view, followed in Part B by a switch to how things look from the student side—the receiving end of the process. From exploring the issue from these two different angles, we can come to a more holistic view of the role of feedback for learning in this EFL context in Japan.

Before proceeding to these two main parts of this chapter, it is useful to frame the ensuing discussion by briefly; a) introducing and clarifying the issue of feedback, b) discussing its place in the AfL framework, and c) providing an overview of the academic writing course at Tokyo Woman's Christian University in which these two reports are situated.

A. Feedback overview

In their recent review of the feedback literature, Hattie and Timperly (2007) noted:

Feedback is one of the most powerful influences on learning and achievement, but this impact can be either positive or negative. Its power is frequently mentioned in articles about learning and teaching, but surprisingly few recent studies have systematically investigated its meaning. (p. 81)

The potential powerful influence of feedback has indeed been well noted. In his detailed review of 87 meta-analyses of the things that make a difference in student achievement, (Hattie, 1987) reported that feedback was the single most powerful influence. Feedback needs to be carefully considered and provided by teachers however, because of the potential negative effects it can have, something also documented in the research literature. In their review of 131 feedback studies, Kluger and DeNisi (1996) found that in 40 per cent of studies, the feedback provided led to a deterioration in student performance. More recently, Price and O'Donovan (2006) wrote: "A review of the literature indicates that feedback is the most important part of the assessment process in its potential to affect future learning and student achievement". Feedback research has, among other things, emphasized its central role in being a catalyst for productive interaction between students and instructional materials, its importance in student development and retention, but also the fact that for many academic staff, feedback is one of the most time-consuming elements of professional practice (Ross et

al., 2006).

The past two decades have seen an increasing number of studies examining the nature of feedback provided to students and how it can be best used to promote learning (Brown & Glover, 2006). Yet, no single definition of feedback has been agreed upon in the literature, and there are a variety of examples offered as to what the term entails. Starting with Ramaprasad's widely cited definition, Table 1 below presents a few descriptions of the term, similar yet with shades of difference, that are useful frames of reference for the focus of this chapter.

Table 1. *Some relevant definitions of feedback*

Feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way. (Ramaprasad 1983. p. 4)
Feedback is information that provides the performer with direct, usable insights into current performance, based on tangible differences between current performance and hoped for performance. (Wiggins, 1993, p. 182)
Feedback is information about how a student has performed in relation to some standard or goal. (Nicol & Milligan, 2006, p. 64)
Feedback is information provided by an agent (e.g., teacher, peer, book, parent, experience) regarding aspects of one's performance or understanding. It occurs typically after instruction that seeks to provide knowledge and skills or to develop particular attitudes. (Hattie & Timperley, 2007, p. 102)

The reader will note that in all the above definitions feedback is described as ‘information’, which the American Heritage Dictionary defines as “knowledge derived from study, experience, or instruction.” The quality and quantity of this information, and, just as importantly, how it is used by students, are key points of consideration when examining the feedback issue.

It is useful to point out some distinctions in types of feedback, particularly between internal and external feedback, as well as between the formative and summative versions of this information provided by the teacher.

Nicol and Milligan (2006) make an important point in reminding us about the distinction between *internal* and *external* feedback. They note that as students monitor their engagement

with learning activities and tasks, they continuously generate their own internal feedback which evaluates their performance in the context of what they are trying to do. In contrast, feedback can also be provided externally by teachers and peers. While this investigation deals with externally provided teacher feedback, it should be remembered that students are also generating their own internal feedback throughout this process, making judgments about their own ongoing performance.

An important distinction between formative and summative feedback should also be made clear. Here is how the two terms may be conceptualized:

- **Formative feedback** - response to student work, while it is in progress, which helps identify strong and weak aspects of performance and provides suggestions for improvement. Formative feedback plays a part in ‘forming’ or shaping student response to the task being worked on in order to draw out the best possible performance from students. The feedback is intended to *in-form* change and improve student learning. Formative feedback is ongoing feedback which does not include a final assessment of the work under consideration; it is feedback *for* learning.
- **Summative feedback**- a summary of students final output or performance, which will include a grade or score that is recorded and aggregated into final course grades. It may help shape the next performance or process, but is too late to play a part on the task being evaluated. The feedback is intended to *sum-up*, and is a judgment or measurement of student learning. Summative feedback is end-point feedback which includes a final assessment of the work (typically a letter grade or numerical score); it is feedback *on* learning that has (or has not) taken place.

This summative understanding of feedback is the most common application of the term for many teachers, losing sight of the possibilities and importance of formative feedback. Stefani (1998) makes the following insightful observation:

It is still the case that too many academics believe that a grade, and a short series of comments, usually of a simple praise or blame nature constitute feedback, when what students actually want, is user-friendly information, relating to how they are doing and how, specifically, they might be able to improve on what they are doing (p. 348).

This chapter deals primarily with the issue of providing students with such ‘user friendly’ formative feedback, although summative feedback also has a role to play. In her extensive literature review of feedback, Shute (2008) defines formative feedback as “information

communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning” (p. 154). Such ‘feedback for learning’ is the focus for this chapter. Wiggins (1997) provides a forceful summary of the importance of feedback and formative assessment:

You can’t learn without feedback . . . It’s not teaching that causes learning. It’s the attempts made by the learner to perform that cause learning, dependent upon the quality of the feedback and the opportunities to use it . . . (p. 33).

B. Summary of feedback in AfL

In their comprehensive review of formative assessment, Black and Wiliam (1998a) also emphasized the large and consistent positive effects feedback has on student learning, compared with other aspects of the teaching process. This view of feedback has been a consistent and central theme in AfL literature and practice.

Black and Wiliam (1998a) noted that there are two main functions of feedback: *directive* and *facilitative*. Directive feedback is explicit in telling the student what needs to be fixed or revised. Facilitative feedback is less specific and provides comments and suggestions to help guide students in their own revision and conceptualization. As will be seen, both directive and facilitative elements are part of the feedback I provide to my writing students.

According to Black et al. (2003), “An essential part of formative assessment is feedback to the learner, both to assess their current achievement and to indicate what the next steps in their learning trajectory should be” (p. 42). In this work, the authors provide a warning call to teachers against providing summative feedback in the form of marks or grades, and promote the need for more formative-style ‘feedback for learning’. The following excerpt from Black et al. (2003) is quoted at length due to the number of important points it makes about the concept of feedback in AfL:

In general, feedback given as rewards or grades enhances ego rather than task involvement—that is, it leads students to compare themselves with others and focus on their image and status rather than encourages them to think about the work itself and how they can improve it. Feedback by grades focuses students’ attention on their ‘ability’ rather than on the importance of effort, damaging the self-esteem of low attainers. *Feedback which focuses on what needs to be done can encourage all to believe that they can improve.* Such feedback can enhance

learning, both directly through the effort that can ensue and indirectly by supporting the motivation to invest such effort. A culture of success should be promoted where every student can make achievements by building on their previous performance, rather than by being compared with others. Such a culture is promoted by informing students about the strengths and weaknesses demonstrated in their work and by giving feedback about what their next steps should be. (p. 46, emphasis added)

Feedback provided with the expectation that all students can improve on their initial performance was a key element in the teacher response to student essays in this chapter. Black et al. (2003) also note in this book that a key feature in any formative assessment procedure is the quality of the feedback that is provided. Commenting on the nature of such feedback, (Black & Wiliam 1998b) observe “Feedback to any pupil should be about the particular qualities of his or her work, with advice on what he or she can do to improve, and should avoid comparisons with other pupils” (p.9).

In England, the Qualifications and Curriculum Authority (QCA) is an organization which develops the national curriculum and associated assessments. In their website, the QCA discuss the key components of AfL, including the characteristics of effective feedback. These are as follows:

- Feedback is more effective if it focuses on the learning intention of the task and is given regularly while still relevant.
- Feedback is most effective when it confirms that pupils are on the right track and when it stimulates correction or improvement of a piece of work.
- Suggestions for improvement should act as 'scaffolding', i.e. pupils should be given as much help as they need to use their knowledge. They should not be given the complete solutions as soon as they get stuck and should learn to think things through for themselves.
- Pupils should be helped to find alternative solutions if simply repeating an explanation continues to lead to failure.
- Feedback on progress over a number of attempts is more effective than feedback on one attempt treated in isolation.
- The quality of dialogue in feedback is important and most research indicates that oral feedback is more effective than written feedback.
- Pupils need to have the skills to ask for help and the ethos of the school should encourage them to do so. (QCA, 2005)

It should be remembered that these characteristics are about feedback in general, and are provided in the context of primary and secondary school education in the UK, where students and teachers usually share the same first language. Yet, they provide a general picture of how feedback is viewed in AfL, and most of these characteristics will appear throughout this chapter on my written feedback practices in an EFL context at a Japanese university.

There are two final points to be made here in this brief overview of feedback in AfL. One is the important point made by Black et al. (2003, p. 58) that good feedback starts with good planning, making it essential that tasks are specifically designed and framed to support feedback. And finally, getting to the heart of the issue, “The central point here is that, to be effective, feedback should cause thinking to take place” (Black, 2004, p.7). As noted above, AfL theory directs that such feedback-induced thinking should be focused on the task at hand, rather than on the students themselves.

C. Feedback study setting – overview, context and assessment

Overview of the Junior Composition course

As in any description of teaching and learning, contextual factors often play a key role and must be considered in any evaluation of a course or the assessment framework used in it. With respect to a Writing course, Hyland and Hyland (2006) explain that “context is a combination of factors related to the institution and the writing program and to the factors that teachers and students bring to the interaction” (p. 212).

While this chapter describes the teachers written feedback in a third-year academic writing course in the Department of English at TWCU, what happens to students in their fourth year is of significant import to this discussion. In order to graduate, the Dept. of English requires all students to write a graduation thesis, 20-25 pages in length, in English. This graduation thesis must meet the standards of a traditional academic research report showing skills in paraphrasing, summarizing and quoting source material, as well as showing writing competence in both essay content and organization. This fourth year graduation essay casts a long shadow over the third-year writing course described here, and how both teachers and students view it.

The third-year writing course is called Junior Composition (JC), taught by five teachers in classes of 20-25 students. This yearlong program is divided into two parts: JC 1 (April – July)

and JC 2 in the second semester (Sept. - Jan.). There are approximately 15 weekly classes during the semester, each for 90 minutes. Teachers are required to submit course grades after each semester.

JC serves as a writing workshop, one intended to help students develop the required writing skills necessary to produce good essays and be able to move on to writing a competent graduation thesis in the fourth year. The departmental guidelines for teachers describes this course as follows:

Junior Composition (JC): This is a required course for all English majors. The aim of this course is to develop their academic writing skills to the point where students are able to write good 3-5 page essays on literary or language topics using MLA documentation style. All English majors must, in their senior year, write a 20-25-page graduation essay in English on either a literature or linguistic topic. By the end of the Junior Composition course, students should be capable of writing short, smoothly developed, technically correct, interesting papers in their selected area. They will thus be able to enter the senior year well equipped to undertake their graduation essay in English. Junior Composition works best when it is taught as a writing workshop. Depending on your students' area of interest, please choose texts which you think appropriate as the basis for assignments.

From the third year, English Department students must focus on either Literature or Linguistics as an area of specialization for their senior year courses and graduation essay. The 23 students in the JC class discussed in this report all chose Linguistics and therefore would write essays on linguistics-related topics. JC is a high-stakes, mandatory course for English majors and students must pass it to move on to writing the graduation thesis in their senior year.

JC students are made fully aware that the course is strongly linked to the work they will do in the fourth year in producing their graduation essay. As such, because of this extrinsic motivation, students are serious about developing their writing skills and class attendance and motivation to produce quality essays are quite high. Students know the writing and research skills and competencies they develop (or do not) in the JC course will have a direct bearing on the their fourth year work and, indeed, their prospects for graduating from the university on time (with their cohort) and in good standing.

Despite the high stakes nature of the course, the curriculum is loosely structured and it is up to individual teachers to choose course content (of either a literature, or linguistics nature) and the writing activities that will develop student skills. Teachers are also left to their own devices with regard to the assessment of students work, and the provision of feedback to

them. It is generally understood that students will produce at least two essays per semester.

While there is no class textbook included in the course syllabus, there is a writing reference book which all students must purchase; a writing guide entitled *The Pocket Wadsworth Handbook, 3rd Edition* (Kirschner & Mandell, 2006). This is a quick reference guide including sections on the writing process, sentence grammar and style, the research process and MLA documentation style. Sections about thesis-statement writing are particularly useful, as is the discussion of the process of writing essays. Students read that “Writing is a constant process of decision making- of selecting, deleting, and rearranging material as you plan, shape, draft, revise, edit, and proofread your paper” (Kirschner & Mandell, 2006, p. 6). Interestingly, providing written feedback on student essays may also be considered ‘a constant process of decision making’.

In general, the overall goal for this academic writing course is to develop competent, self-sufficient writers of academic English, who are ready to proceed to writing a graduation report in their senior year.

Course structure and assessment

In my particular JC class, students were required to write three essays in the first semester and two in the second, all with linguistics-related subject matter (e.g. comparing speaking and writing, gender and communication). The class follows a standard ‘process approach’ for teaching writing, defined as “A teaching approach to writing which emphasizes the development of good practices by stressing that writing is done in stages of planning, drafting, revising and editing, which are recursive, interactive and potentially simultaneous” (Hyland, 2002, p. 230). For all essays students were required to write both a first draft and final draft. The first draft of the essays was submitted and formative feedback provided on them. These drafts were returned, and students had one week to edit and revise before submitting the final version. This version was then summatively assessed and given a final grade. In the weekly 90-minute classes students were provided with instruction, and carried out tasks related to both the particular linguistics topic being focused on in the essay task, and to academic essay writing itself (e.g. writing effective thesis statements, paraphrasing sources). Figure 1 below shows the seven steps in a typical cycle of essay writing and feedback in the JC course.

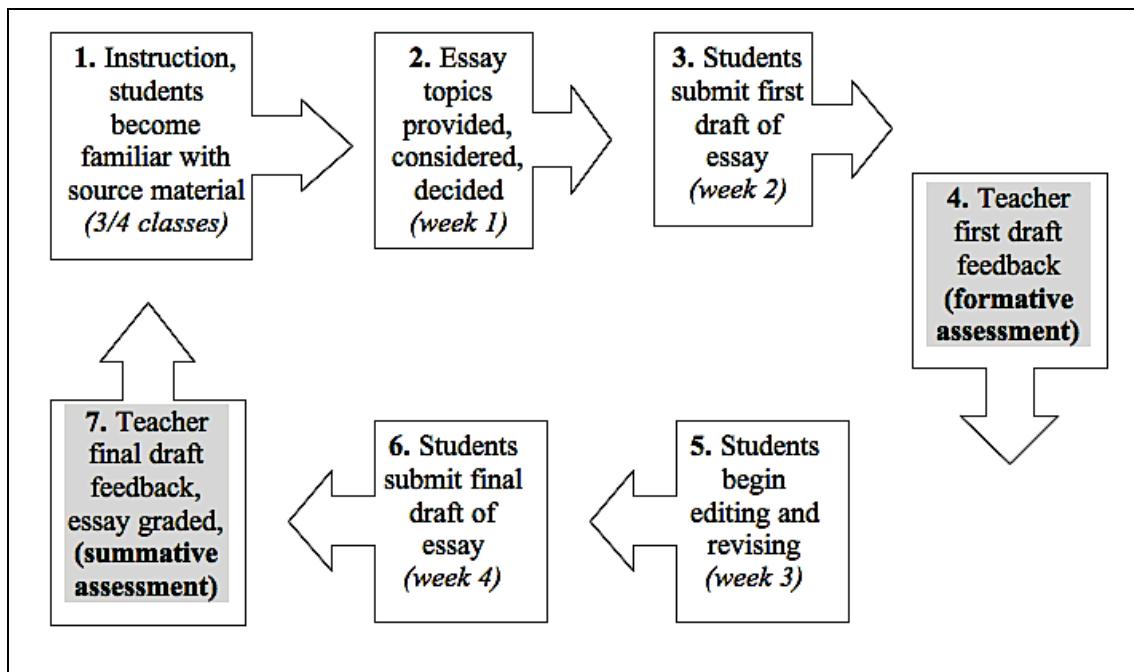


Figure 1. JC essay writing and teacher feedback cycle

The one-week response time for students receiving first draft feedback would be labeled as medium-cycle formative assessment in the Wiliam and Thompson (2007) categorization, shown in Table 6 of the introductory section (p.31).

All JC essays followed the same seven-step pattern, before proceeding to the next essay topic. Brown and Glover (2006, p. 43) refer to such a writing-feedback cycle as ‘the performance-feedback-reflection-performance-feedback loop’. Step 4 in this cycle, related to teachers written feedback on student first drafts, is the particular area of interest for Part A of this chapter. Part B, dealing with student reaction to feedback will be more concerned with Step 5 - the editing and revising process (i.e. closing the gaps between current and optimum performance). It should be noted, as will be later explained, that opportunities for self and peer assessment were included during this feedback cycle.

This chapter deals with the second half of the writing course, JC 2, and one particular essay students wrote (on the topic of ‘slang’) and my feedback on it. It is perhaps important to consider what students have done in the first semester of this course, as it influences the course content and the feedback and assessment discussion described here in JC 2.

In the first semester students wrote essays on the linguistics-based topics of 1) comparing

speaking and writing, 2) cell phone communication and 3) gender differences in communication. For these essays in JC 1, the focus was on clear, concise essay structure and expression of ideas. Particular attention was paid to writing thesis statements and paraphrasing, summarizing sources. For the three essays, source material was provided to students. These three or four sources per essay typically consisted of a short 2-4-page essay or textbook excerpt about the essay topic focus. Students would familiarize themselves with the content of these readings and use this material to construct an academic essay. Students were assessed using a three-criteria scoring rubric: content, organization and language use. In the second semester, the same assessment rubric was used with the addition of a new criteria; ‘use of source material’ (see **Appendices A1** and **A2** for the summative assessment rubric used for final drafts of essays). New for JC 2, particular focus was placed on students’ use of MLA style documentation of sources, especially the use of parenthetical citations and a Works Cited list at the end of the essay.

After completing the first semester of the course, at the start of JC 2: students were familiar with each other and the teacher; had already written three linguistics-related essays (but without MLA documentation); were familiar with the assessment rubric used; knew the writing cycle followed for each essay, knew the types of feedback provided on first drafts; and were familiar with responding to this teacher feedback to make their essays better, before submitting it for summative assessment.

Now that this necessary background information (a feedback overview and how it is viewed in Afl, as well as the necessary information about the JC course) has been provided, we can now move on to the two main parts of this examination of teacher written feedback in an academic writing course.

Part A. Teacher self-assessment of written feedback

For many years I taught in universities . . . I marked thousands of scripts without examining what the scripts could teach me about my capacity as a teacher and examiner.
(Ashby, 1984, p. v)

The teacher . . . is continually exerting influence on the students and the learning situation. By studying his own behavior in some systematic, objective manner, the teacher may gain further insights into his own pattern of influence. (Amidon & Flanders, 1967, p. 72)

Introduction

This report of teacher self-assessment (TSA) is an examination of feedback provided on student essays to gain insight into my roles and performance as teacher and ‘feedbacker’ (i.e. feedback provider) in a Japanese college writing class. It is an exercise in critical appraisal and analysis of one facet of pedagogic practice. According to Black and Wiliam (1998b),

We use the general term *assessment* to refer to all those activities undertaken by teachers, and by their students in assessing themselves, that provide information to be used as feedback to modify the teaching and learning activities. Such activities become *formative assessment* when the evidence is actually used to adapt the teaching to meet student needs. (p.139, original emphasis)

The TSA process engaged in here may also be considered as a formative assessment of a very important aspect of teaching practice - the provision of written feedback on student performance. Teachers’ written feedback can have a significant impact on improving second language (L2) students’ writing, however “this role is complex and requires careful reflection to be used effectively” (Hyland, 2003, p. 192). A careful, systematic assessment of the written feedback I produced and provided to students in one academic writing course, and what can be learned from this self-assessment, is the focal point here. This investigation in the context of the Junior Composition course is guided by two key questions:

1. What is the nature and form of the written feedback I am giving to students on their essay first drafts?
2. How do my feedback practices compare with recommended practice in the recent feedback literature?

The results of this investigation on feedback will actually *serve as feedback* into my role as a

learning facilitator in the academic writing classroom.

In this section, I describe the process of assessing my written feedback practices in order to learn from the process, and perhaps modify the practice as a result. In the context of a third-year academic writing course, teacher written feedback on 21 first drafts of student essays will be analyzed and critiqued in a systematic manner. The teacher feedback provided with this particular set of student scripts will be assessed according to principles and conditions proposed in two key articles regarding teacher feedback practices (Gibbs & Simpson, 2004-05; Nicol & Macfarlane-Dick, 2006).

Teacher self-assessment is also referred to as ‘teacher self-evaluation’, and Airasian and Gullickson (1997) offer a good description of what that term entails:

Teacher self-evaluation is a process in which teachers make judgments about the adequacy and effectiveness of their own knowledge, performance, beliefs, or effects for the purpose of self-improvement. . . . In self-evaluation, the teacher becomes responsible for examining and improving his or her own practice. It is the teacher who collects, interprets, and judges information bearing on personal practice. It is the teacher who frames criteria and standards to judge the adequacy of his or her beliefs, knowledge, skills and effectiveness. Teacher self-evaluation is evaluation of the teacher by the teacher and for the teacher (p.3).

This explanation is also a good description of the process engaged in for this report into the nature of my feedback practices. It may be added that TSA can also be considered ‘for the students’ as well as for the teacher, as they may also benefit from such reflective actions and any resulting refinement of pedagogic knowledge or practice.

As will be seen, the TSA described here is unusual in the literature on writing feedback, as most L2 research has centered on student self-assessment of their writing, rather than teachers’ self-assessment of their own written feedback. Little is known about teacher self-assessment of the written feedback they provide, especially for second language writing instructors (Montgomery & Baker, 2007). However, the literature in general education that encourages teachers to become ‘reflective practitioners’ is well established. The role of *teacher as learner* within the Afl framework has also been recognized and promoted. We now briefly turn to these two areas of teacher self-examination of their own ideas and practices.

Being a reflective practitioner

It is widely recognized that a central tenet of the teaching-learning process is reflective practice (Harford & MacRuairc, 2008), and there are numerous references to this reflective teaching concept in the language teaching education literature (Akbari, 2007). Perhaps less recognized, but no less true, is the insightful comment by Hounsell (2005) that “reflective teaching and quality of learning go hand in hand”(p. 257).

Lucas (1991) defines the term ‘reflection’ as systematic inquiry into one’s own practice in order to improve that practice and form a deeper understanding of it. In the context of language teaching, according to Murphy (2001, p. 499), reflective teaching has three purposes:

- To expand one’s understanding of the teaching/learning process
- To expand one’s repertoire of strategic options as a language teacher
- To enhance the quality of learning opportunities one is able to provide in language classrooms

Similarly, Richards and Lockhart (1994) write that reflective teaching is an approach in which teachers “collect data about teaching, examine their attitudes, beliefs, assumptions, and teaching practices, and use the information obtained as a basis for critical reflection” (p. 1). This description is a close approximation of the processes described in Part A of this chapter. Teacher feedback data, on a set of essay first drafts, was collected, examined and used as a basis of critical reflection on written feedback practice.

Much of the discourse related to reflective practice stems from the writing of Schon (1983, 1987) who differentiated between ‘reflection-in-action’ and ‘reflection-on-action’. While the former refers to teachers being conscious of their actions as they work, reflection-on-action focuses on the importance of reflecting back on and critiquing one’s practice. Reflection-on-action, a hind-sighted, rearview look is the type of practice engaged in here as the critical analysis of student scripts and written feedback on them began approximately six months after this particular JC class had finished.

It should be noted here that the data set for this investigation was collected months prior to deciding to use it to analyze my feedback practices. The first and final drafts collected here were done so with the idea that they may provide useful research data in the future. The feedback provided on student drafts discussed in this chapter was in no way influenced by

considering recommended AfL practices, or by the research process itself. As such, the data set is an unbiased representation of my usual feedback practice for student essays in the JC course.

At the time of writing I continue to teach this academic writing course (and others), therefore the reflection-on-action engaged in here is eminently relevant to my ongoing teaching and feedback practices. According to Pedder, James and MacBeath (2005, p. 237), “Classrooms need to become crucibles of learning for teachers as much as for their students”. This idea of *the teacher being a learner* is an overarching concept in this discussion, and is also noted and promoted in AfL theory and practice.

Teacher as learner in AfL

The first and foremost purpose of assessment in education is to support learning (Black & Wiliam, 2006). The primary learner in the classroom is clearly the student. Clearly, but not only the student. The teacher also should be an active learner in the classroom, as all practitioners have room for improvement, either in subject knowledge or pedagogical skills in maximizing student learning. Black et al. (2003) phrase this additional teacher role as “teachers casting themselves as learners and working . . . to learn more” (p. 98). James and Pedder (2006b) go even further by stating that a teachers’ professional learning is an essential condition for AfL and note “parallels between processes of assessment for learning for students and inquiry-based learning by teachers” (p. 27). The professional development of teachers is explicitly stated as one of the ten principles of AfL laid out by the Assessment Reform Group:

Assessment for learning should be regarded as a key professional skill for teachers. Teachers require the professional knowledge and skills to: plan for assessment; observe learning; analyze and interpret evidence of learning; give feedback to learners and support learners in self-assessment. Teachers should be supported in developing these skills through initial and continuing professional development (ARG, 2002).

As noted here, such professional skills need to be developed and this includes the provision of effective feedback to learners on their work. Part of this professional development requires teachers to become reflective practitioners because “implementing assessment for learning/formative assessment may require a teacher to rethink what effective learning is, and

his or her role in bringing it about" (James, 2006, p. 49).

While AfL is considered to be assessment for *student* learning, in this case we shift the emphasis to assessment of written feedback practices for *teacher* learning. James and Pedder (2006b) make explicit reference to such a shift in focus:

If teachers are prepared and committed to engage in the risky business of problematizing their own practice, seeking evidence to evaluate in order to judge where change is needed, and then to act on their decisions, *they are thus engaging in assessment for learning with respect to their own professional practice*. Helping students to do the same with respect to their learning becomes less challenging because teachers are familiar with the principles and processes through inquiry into their own practices (p. 40, emphasis added).

By taking part in this critique of my pedagogical practices with written feedback and assessing how they stand up against recommended feedback practice, I become the learner in the AfL framework. The goal here is to have a better understanding of what I do when giving written feedback and perhaps how it can be improved. "For students to be actively engaged in creating their own understanding, they must learn to be critical assessors who make sense of information, relate it to prior knowledge, and use it for new learning" (Rethinking Classroom Assessment, 2006, p. 41). In this report I play the role of 'critical assessor' of my written feedback practices. In order to do this, effective assessment instruments are needed to help examine my practices. These tools and related ideas will be described in the next literature review section, prior to them being used in this TSA of feedback process.

Literature review

The review of the literature pertinent to this report will focus on three main areas: teacher response to student writing, TSA of written feedback research, and finally a description of the assessment framework to be used in this TSA process. We begin with a general summary of some of the key ideas and findings related to teacher response to student writing and have some relevance to the following discussion.

1. Teacher feedback on student writing.

This has been an area of extensive research, exploring the following issues in particular: the effectiveness of grammar correction; different points of feedback focus-error, content,

organization; the use of coding schemes; student correction behaviors; teacher correction behaviors; negative versus positive feedback; students views on types of feedback and the clarity of teacher feedback (O'Brien, 2004).

According to Hyland & Hyland (2006), one of the ESL writing teachers most important tasks is the provision of feedback to students, but they point out that feedback practices have been transformed in the past 20 years: “Summative feedback, designed to evaluate writing as a product, has generally been replaced by formative feedback that points forward to the students’ future writing and the development of his or her writing processes” (p. 1). While feedback practices may have undergone changes in recent years, the day-to-day questions teachers have about feedback remain the same:

- What should I give feedback on?
- How should I express it?
- What mode should I use?
- How will feedback affect my relationship to the student?
- Will my feedback make a difference to students’ writing?

(Hyland & Hyland, 2006)

There are a considerable number of studies examining teacher written feedback. Table 1 below summarizes some key findings from this body of research.

Table 1. *Summary of teacher written feedback research (from Goldstein, 2006; Ferris, 2006)*

- Some scholars have called for all student errors to be corrected in order to prevent fossilization.
- Selective correction has also been called for from some quarters, focusing on patterns of error that can be productively addressed.
- Some researchers recommend that error correction be eliminated as it is either unnecessary, ineffective or even counterproductive.
- For academic and professional audiences, accuracy is important and L2 errors may stigmatize writers in some contexts.
- L2 student writers report that they need and value error feedback from teachers.
- Students report being sometimes confused by teacher feedback commentary.
- Students report using feedback despite not understanding the reasons why it was provided.
- Students sometimes think they have understood feedback when they do not.
- Students may not know how to use teacher feedback to revise their writing.

While noting the conflicting ideas and stances in the literature regarding teacher response to student writing, and the complexity of the issues related to such feedback, Goldstein (2006) points to some common ground for writing teachers:

We would all agree that the quality of feedback matters and that students will most benefit from feedback that is text specific, relevant, and clear; does not appropriate the text or the writers' responsibilities; and allows students to determine a way to revise in response to the feedback (p. 203).

Few thoughtful writing teachers would disagree with this concise summary of what quality feedback entails.

In summarizing some of the previous research regarding teacher feedback on writing, both L1 and L2, Montgomery and Baker (2007, pp. 83-84) add to our understanding of the issue by noting the following:

- both teachers and students feel that teacher-written feedback is an important part of

- the writing process, particularly for L2 writing
- while some scholars have argued that feedback is not helpful, students still believe it to be so and use it to improve both L2 writing and L2 grammar
- teacher feedback is often not text specific
- feedback can be incorrect,
- feedback may not address the issues it intends to
- there may be a mismatch between feedback desired by students and that given by teachers
- some teachers focus more on local issues (e.g. grammar, mechanics) rather than global issues (e.g. content, organization)

Montgomery and Baker (2007) also point out that complex relationships in the classroom may affect how feedback is provided and received.

In an article entitled “Effective faculty feedback: the road less traveled” Stern and Solomon (2006) remind us that “feedback provides students with a way to determine if they (students) have communicated their ideas clearly and effectively” (p. 24). After an extensive review, Stern and Solomon (2006) identify three key principles that appear to be consistent across the writing and response literature. These are:

1. Provide positive comments in addition to corrections
2. Provide feedback only on a select few areas that are considered important for that particular writing assignment- those which are tied to the student learning goals for the paper assignment
3. Provide comments that identify patterns of weaknesses, errors, and strengths rather than overwhelm students with identification of individual errors.

They add two additional suggestions related to these three principles: tell students how errors can be fixed because they may not know how to do so; and teachers should set writing assignments that require essay revision by students (Stern & Solomon, 2006).

With regard to second language students and teacher feedback on their writing, Hylands’ (2006) state-of-the art review article addresses many of the issues mentioned above. He also makes some additional points of interest which provide useful background commentary to this investigation. These include the following:

1. In classrooms that are process-based and learner-centered, feedback is viewed as an important developmental tool helping learners move through multiple drafts towards the capability for effective self-expression (p.83).
2. Despite earlier conflicting views, more recent empirical research seems to suggest that feedback does lead to improvements in writing (p. 84).
3. Teachers approach texts with a number of purposes in mind, and these purposes may change with different students, assignments and different drafts (p. 86).
4. Teachers adopt a variety of strategies for commenting on students' work which change according to the type of essay, the point in the semester that feedback is provided, and the abilities of the student (p. 86).
5. While L2 students want to have teacher written feedback, how such feedback contributes to writing development remains unclear, both with regard to immediate impact on revised drafts, and longer term development of writing skills (p. 87).

After thoroughly reviewing the issue of feedback for L2 writers in this article, Hyland (2006) captures the central role of teachers responding to student writing: “The most important role of response is to help students develop into independent writers who are able to critique and improve their own writing” (p. 96).

The wide-ranging summary of ideas presented above provide a useful backdrop to the teacher written feedback context and many of the points mentioned will be evident in the discussions and analyses which follow.

Despite the plethora of research in teacher feedback on student writing, and the sometimes conflicting results and points of view, the importance of such feedback has become widely recognized in both mainstream education and second language writing. Across the education spectrum of courses and contexts, effective feedback is viewed as critical for both encouraging and consolidating student learning (Hyland & Hyland, 2006).

2. Teacher self-assessment of feedback practices

The scarcity of research and publications on teacher self-assessment of the feedback they provide to students lies in stark contrast to the extensive body of teacher feedback research. Despite the increasing number of studies looking at teacher feedback to students in the past two decades, very few of them focus on teacher self-assessment of their own feedback

practices. Almost all of the writing on teacher feedback discusses what other teachers or research subjects do when giving feedback, and very few writers examine their own practices (in print, at least). Brown and Glover (2006) sum up the state-of-play in this area as follows:

Given the high value that students place on individualized written feedback, the role that good-quality feedback may play in aiding student learning (Black & Wiliam, 1998) and the significant time costs to teachers in its delivery, it is surprising that few attempts have been made to classify systematically the different types of teacher comments that constitute feedback so that the quality of feedback can be analyzed (p. 82).

Surprising indeed, and such a systematic self-examination to analyze the quality of teacher written feedback is the focus of the investigation reported here. Only a few reports in the related literature pay attention to TSA of their written feedback, and these are primarily from L1 writing contexts. Most L2 writing research has focused on student perceptions rather than the teachers' perspectives of their practice. Self-assessment studies have focused on student self-assessment to improve their writing performance, rather than on teacher self-assessment to improve their feedback performance (Montgomery & Baker, 2007).

In a study at an American university, Montgomery and Baker (2007) obtained data from 15 teachers and 98 students in an intensive ESL program. Their investigation had three foci: how much local and global written feedback teachers give; how TSA and student perceptions coordinated; and how well teachers' self-assessments matched their performances. Among other things, they found that the coordination between TSA and actual performance was not strong, showing that teachers may not be fully aware of the amount and type of feedback they were providing. Montgomery and Baker (2007) also found that teachers did not provide the same amount of feedback to each student and that in general the 15 teachers gave substantial amounts of local feedback, but little of the global variety (e.g. content, organization). These results were not reflected in the TSA's the teachers provided, pointing out the discrepancy between perceived and actual feedback practice. Montgomery and Baker (2007) call for further TSA research in L2 writing to encourage teachers to become more aware of their feedback practices, and help improve its effectiveness as a result.

One of the few accounts of a teacher evaluating and critiquing their own written feedback practices is by Straub (2000), which is a case study into teacher response. In an American context of a first-year college writing class, Straub investigated his written comments on one student's work. This occurred over a period of time, and he provided a detailed contextual examination of his feedback response practices. Taking a teacher-as-researcher stance, Straub

found that his teacher responses to student writing were shaped by such factors as: the sequence of assignments and classroom instruction, his teaching style, and the work and needs of individual students. Straub frames his TSA in light of ten practical strategies for responding to student writing that he identified in the writing response literature. He compares his feedback practices to those recommended practices, listed here in Table 2 below.

Table 2. *Practical strategies from the writing response literature (Straub, 2000)*

1. Turn your comments into a conversation.
2. Create a dialogue with students on the page.
3. Do not take control over the text: instead of projecting your agenda on student writing and being directive, be facilitative and help students realize their own purposes.
4. Limit the scope of your comments.
5. Limit the number of comments you present.
6. Give priority to global concerns of content, context, and organization before getting overly involved with style and correctness.
7. Focus your comments according to the stage of drafting and the relative maturity of the text.
8. Gear your comments to the individual student behind the text.
9. Make frequent use of praise.
10. Tie your responses to the larger classroom conversation.

Straub (2000) was convinced that for students to develop as writers they need not only a lot of writing practice, but also “a lot of response from readers about how that writing is working for them” (p. 27).

His case study is one of the few examples of TSA of written feedback in the literature.

Additionally, the article is noteworthy because of the following commentary:

This case study shows how we can use theory to become engaged with, talk back to, and learn something about classroom practice; it shows how we can study our classrooms to become engaged with, talk back to, contribute to, and increase the practical value of contemporary theory (p. 51).

While using a different set of TSA tools, my report is also an attempt to use contemporary theory to engage with and learn something from my written feedback practices, and from the self-assessment process itself.

3. Seven/Seven framework for feedback assessment

In order to self-assess my feedback practices, as well as having a set of student scripts that include my feedback to analyze and scrutinize, a criterion-based framework of recommended practices is required to compare them with. Two articles in particular, Gibbs & Simpson (2004-05) and Nicol & MacFarlane-Dick (2006) both contain a number of recommended feedback conditions and principles, providing a set of criteria for an assessment instrument. The seven feedback criteria from each article have been combined here into a system I call the ‘Seven/Seven framework for feedback assessment’, or simply ‘the Seven/Seven framework’. A brief review of each article is necessary before later putting these feedback conditions and principles to use to self-assess my own practices.

1. Gibbs and Simpson (2004-05): Conditions under which assessment supports students learning

In this paper, Gibbs and Simpson argue that student learning is best supported by assessment when a series of conditions are met. After examining a wide range of case studies, these authors identified 11 conditions for assessment to support learning. Seven of these assessment conditions are specifically related to teacher feedback. The other four important conditions relate to assessments’ influence on the volume, focus and quality of student studying. These four assessment conditions which support learning are as follows:

1. Assessment tasks should capture sufficient study time and effort
2. Tasks should be evenly spread over topics and weeks

3. Tasks should lead to productive activity (not surface learning, but deep learning)
4. Tasks should communicate clear and high expectations.

(Gibbs & Simpson, 2004-05, pp. 12-15)

While these conditions are important and worthy in themselves of discussion and analysis, the focus of this paper is on the remaining conditions, those dealing with feedback. Seven of these assessment conditions which support learning are related to the quantity, timing, and quality of teacher feedback and how students respond to it. Gibbs and Simpson write that these conditions are concerned with “how the provision of feedback affects students learning behavior-with how feedback results in students taking action that involves, or does not involve, further learning” (p. 17). Table 3 lists these seven feedback conditions intended to promote student learning.

Table 3: *Seven feedback conditions influencing learning* (Gibbs & Simpson, 2004-05)

1. Sufficient feedback is provided, both often enough and in enough detail
2. The feedback focuses on students' performance, on their learning and on actions under the students' control, rather than on the students themselves and on their characteristics
3. The feedback is timely in that it is received by students while it still matters to them and in time for them to pay attention to further learning or receive further assistance
4. Feedback is appropriate to the purpose of the assignment and to its criteria for success
5. Feedback is appropriate, in relation to students' understanding of what they are supposed to be doing
6. Feedback is received and attended to
7. Feedback is acted upon by the student

These seven conditions will form one strand of the Seven/Seven framework used here to self-assess my written feedback practices. Gibbs and Simpson (2004-05) make reference to theory, empirical evidence and their own practical experience to justify the list of conditions

they have set. In describing how these ideas may be used, they write, “These conditions are offered as a framework for teachers to review the effectiveness of their own assessment practice” (p.3). The investigation reported here takes up this offer.

An additional perspective on feedback practices, helping to widen and deepen the feedback assessment criteria, is provided in the second key article supporting the TSA discussed here.

2. Nicol and Macfarlane-Dick (2006): *Formative assessment and self-regulated learning: a model and seven principles of good feedback practice*

In this article, these two educators and researchers show that formative assessment and feedback processes can help students become self-regulated learners - taking control over their own learning. Based on a review of the research literature, the article lays out seven good feedback principles which support such self-regulation. As with Gibbs and Simpson (2004-05), the research underpinning each principle is presented and discussed. Nicol and MacFarlane-Dick define ‘good feedback practice’ as “anything that might strengthen the students’ capacity to self-regulate their own performance”(p. 205). Table 4 below presents these seven principles of good feedback practice, developed from a synthesis of the research literature, which facilitate student self-regulation.

Table 4. *Seven principles of good feedback practice* (Nicol & MacFarlane-Dick, 2006)

1. Feedback helps clarify what good performance is
2. Feedback facilitates the development of self-assessment (reflection) in learning
3. Feedback delivers high quality information to students about their learning
4. Feedback encourages teacher and peer dialogue around learning
5. Feedback encourages positive motivational beliefs and self-esteem
6. Feedback provides opportunities to close the gap between current and desired performance
7. Feedback provides information to teachers that can be used to help shape the teaching

These seven principles form the second strand of the Seven/Seven framework used in this TSA of written feedback practices.

These seven principles deal with the learning process itself, students understanding of what good performance entails and the effects feedback has on students' motivation and self-esteem. As with Gibbs and Simpson, the principles are intended for use by teachers for self-assessment. Nicol and MacFarlane-Dick (2006) propose, "that teachers examine current assessment practices . . . An audit of this kind might help identify where assessment practice might be strengthened" (p. 215).

When combined, the seven conditions (Gibbs and Simpson, 2004-05) and the seven principles (Nicol & MacFarlane-Dick, 2006) provide a comprehensive, yet practical, assessment instrument for teachers to use in a self assessment of written feedback practices – the Seven/Seven framework. Using my formative feedback on one set of 21 student essays (first drafts), I will assess my practices against the two systems in this framework, 14 criterion items in total. As far as can be determined, this report is the first research paper to conduct such a systematic TSA of feedback practices using these two conceptual models provided in the recent feedback literature.

Both of these articles used for the Seven/Seven framework refer to the formative assessment and feedback writings of Black and Wiliam (1998) and other writers associated with AfL. Nicol and Macfarlane-Dick also refer to the key text in this thesis, *Assessment for Learning: Putting it into Practice* (Black et al., 2003). The recommended feedback conditions and principles in Gibbs and Simpson (2004-05) and Nicol and Macfarlane-Dick (2006) also reflect those effective feedback practices described and promoted in the AfL model to maximize student learning.

Methods

In describing that the TSA process is comprised of a series of steps or stages, Airasian and Gullickson (1997) write,

Self-evaluation should (a) have a clear focus, (b) collect information that will provide teachers with an objective awareness of their practice, (c) provide opportunity for teacher reflection, (d) result in a decision about practice, and (e) lead to strategies to improve teaching, if necessary (p. 4).

This section will briefly describe how the collected feedback information was collected, coded and analyzed to help me form such an ‘objective awareness of my practice’. It will describe the writing task students engaged in, the feedback that was provided on their writing, and the data coding and analysis procedures used.

Essay task

This TSA report discusses the feedback provided for the first essay of JC 2, on the linguistic topic of ‘slang’. Three short source readings (2-4 pages in length) were provided to students related to this topic. Students were also required to find two of their own additional sources, from the Internet or campus library, for possible use in the essay. A minimum of three sources were required in writing the paper. After three or four classes focused on the topic of slang, familiarizing students with the sources provided and essay construction (particularly MLA documentation of sources), students were required to bring the sources they had found to class. These were then evaluated together for reliability and usefulness, and students were presented with the slang essay questions. The essay topics were the following:

1. *What are the main characteristics of slang?*
2. *For some people, slang has a negative image and the use of such words or expressions are disapproved of and discouraged. Why does such a negative impression of slang exist?*
3. *What are the main reasons why the use of slang is so popular and commonplace?*

On the essay information sheet provided with these questions, students were told that these were topics only and they had to construct an effective essay - one with a thesis statement based on the topic question selected and supported by the source material. Students were required to submit a five-paragraph essay, approximately 700-800 words in length (2 ½-3 pages). These basic requirements were similar to those used in the three essays for JC 1 in the previous semester, so students were well acquainted with the essay format. For the first time, as well as finding some of their own source material, students were required to include parenthetical citations of sources, and a Works Cited page at the end of the essay.

Documentation was required in Modern Language Association (MLA) format, the same style as would be required in the fourth-year graduation essay. Students were reminded to use the course writing reference book for help in constructing their essay. As usual in the JC writing

cycle, they were given one week to produce the first draft of the essay.

First draft feedback

Feedback on the slang essay first drafts came in three forms: use of indirect correction code symbols; direct marginal comments; and, an overall feedback end sheet attached to the back of the essay.

In the JC course, I decided to follow the now commonly accepted practice of using a correction code indicating the location and type of errors or other problems with the writing. Research suggests that the use of such a correction code is effective in stimulating student response and developing self-editing strategies (Hyland, 2003). Student first drafts were marked using the following codes shown here in Table 5.

Table 5. *Correction symbols used on essay first drafts*

? = meaning not understood	0 = missing word
R = rewrite (awkward, unclear English)	pl = plural
g = grammar problem	cap = capitalization
ww = wrong word	sp = spelling
wf = wrong word form	= - join together (sentences)

It should be remembered that by the time of this slang essay, students had already written three essays in which this coding scheme was used on first drafts. In the first semester, students became quite familiar with the correction symbols (and were provided with a hard copy of the coding scheme), and needed simply to be quickly reminded about the coding system for deciphering feedback on this slang essay. Students were also reminded why it would be more effective if they could fix their own errors rather than having the teacher do so. These particular symbols arose from the most commonly used in the course reference

book, shortened and simplified so as not to overwhelm students with excess correction coding on their first drafts.

The second form of feedback was handwritten commentary on the essay, perhaps the most common type of teacher written feedback. These comments are related to content or presentation of ideas in the text (for example, ‘citation needed’) and are instances of the teacher as reader responding to texts ‘on the fly’ (Hyland, 2003).

The final type of written feedback was the completion of a feedback end sheet, an overall indicator of essay strengths and weaknesses, attached to the back of the first draft. This feedback end sheet used the same criteria as would be used in the final (summative) assessment. The feedback end sheet consisted of a set of four criteria, and brief handwritten commentary was provided informing students how they had performed in relation to each one. Figure 2 on the following page shows the feedback end sheet attached to the back of each first draft. This formative assessment rubric (and final, summative version in Appendices A1 and A2) is adapted from a written project evaluation rubric included in the teachers’ manual for a University of Reading textbook called *Extended Writing and Research Skills*, by McCormack and Slaght (2005). For formatting purposes, the version of the rubric shown here has been reduced in terms of size and spacing.

Student: _____

1. **Content:** Clearly focused content, relevant to the essay topic. Length, scope and level of detail are appropriate/relevant. Ideas are well presented and developed, with supporting evidence from a variety of sources. It is evident that the writer knows the topic well.

(strong <----- * -----> weak)

2. **Organization:** Overall structure and main ideas are clearly organized and easy to follow. Introduction has general comments about the topic, followed by a well-written thesis statement. In the body of the essay, supporting ideas are effectively linked together and ‘flow’ coherently, making it easy for the reader to follow. Conclusion summarizes main points and effectively brings the essay to a finish.

(strong <----- * -----> weak)

3. **Language use:** Ideas are clearly expressed, with wide, accurate usage of *vocabulary* and *grammar*; any errors do not interfere with communication. Formal academic style is used (e.g. formal expressions, longer sentences, impersonal tone, etc.). Use of a range of linking words and phrases to join ideas at paragraph and sentence level.

(strong <----- * -----> weak)

4. **Use of source material:** Effective use of a range of sources. These are appropriately incorporated into the body of the essay through *paraphrase*, *summary*, and *quotation*. Shows ability to synthesize well from several sources to support ideas. Works Cited page and in-text referencing follow MLA format. No obvious or conscious plagiarism.

(strong <----- * -----> weak)

Overall:

Figure 2. First draft feedback end sheet (adapted from McCormack and Slaght, 2005)

Feedback provided on this sheet involved making a circle on the dotted line for the strong-weak continuum for each criterion, some brief comments in the space after each of the four criteria, and some final overall comments at the bottom. The overall comments were intended

to particularly remind students about areas to focus on to improve the essay. As already noted, the same four criteria (content, organization, language use, use of source material) were also used and graded individually for the final assessment. Students were fully aware of this connection between the rubric criterion for both the formative (first draft) and summative (final draft) assessments.

As noted earlier, providing written feedback is a time-consuming business and the amount of time involved in checking these slang essays should be briefly noted here. On average, responding to each first draft took anywhere from 15-to 25 minutes. This varied depending on the content of the essay and the writing proficiency exhibited by the student. First drafts were checked over a series of five sessions, providing feedback for approximately five essays per feedback session.

With regard to writing assessment criteria, obviously different focus areas are more or less important, depending on institutional or departmental guidelines and the course being taught. Yet, there are some commonalities across the range of possible assessment criteria of academic writing. Harrington et al. (2006) write:

The criteria that are employed in the assessment of essays vary between institutions and disciplines, just as individual tutors vary in what they see as the most important qualities in students' written work. However, some criteria are commonly employed across different disciplines and institutions, and appear to have a central role in the shared perception of what constitutes a good student essay. These include the following criteria: addressing the question, demonstrating understanding, developing argument, using evidence, structuring, critically evaluating, and using language well. (pp. 110-111)

In this JC course, the assessment criterion used for evaluating student essays, both formatively and summatively, contain the 'commonly employed' criteria mentioned above in determining what makes a good student essay.

3. Data Collection and Analysis

While considering possible avenues of research in my doctoral studies, it occurred to me that the issue of teacher feedback on writing would be one possibility. By the start of the second semester, I had decided that I would administer a survey to students at the end of the JC course to gain their perspectives on being on the receiving end of this feedback process. At the end of the slang essay process in the second semester, I realized that having access to the

student essays (both first and final drafts) would possibly be a valuable data source for a future feedback research project. So, at the end of the writing cycle for the slang essay, after summative assessments of final drafts were completed and as we began the final essay topic for JC 2, I asked students if I could use their slang essays for my doctoral research. Specifically I asked them to return to me the final draft of their essay (with my attached summative assessment) to be copied, as well as the first draft of the essay containing the feedback I had provided. Students readily agreed with my requests. In the following two weeks I proceeded to photocopy all 23 final drafts and 21 of 23 first drafts were returned to me. Two students did not return these first drafts, for undetermined reasons. Students agreed that I could keep the original first drafts, as they were no longer useful to them. After copying the final drafts, the originals of this version were returned to students. In this manner, I was able to collect 21 essay sets, including a first draft containing my feedback, and the edited, revised final version of the same essay.

For the purposes of the TSA of feedback part of this chapter, the 21 first drafts of student essays are the primary data materials. These scripts were closely examined and the types and amount of feedback were systematically counted, coded, and analyzed.

Each of the first drafts was closely examined and the number and type of feedback interventions were coded. A ‘feedback intervention’, or FI, according to Kluger and DeNisi (1996) is defined as “actions taken by an external agent to provide information regarding some aspects of one’s task performance” (p. 255). In this context, such interventions will refer to the corrective symbols and comments I provided on each essay. The correction code FI’s were categorized into the symbol types listed above in Table 6. My handwritten commentary was also counted and coded using a system for categorizing types of written feedback established by Haines (2004). Finally a record of ‘other FI’s’ was also counted. These included such actions as: crossing out things (words, sentences); fixing errors; indicating that sentence part or words should be relocated (by using arrows, circles, etc.) and a number (1, 2 or 3) indicating the strength of the thesis statement. The type and style of my FI responses to student writing can be seen in Figure 3, which shows the first page of a student first draft.

Slang -Out of Favor with Some People

"Yo! Whassup, G?" It is an example of slang and it means "Hi! What's up with my friend?" From young to adults, both men and women have ever used ^R slang at least once in a life. Slang is very common and exists in every languages. Some people use slang to show solidarity with a member of certain group such as ^a gang or a community (Gardiner 100). Other people, especially among young people, prefer using it just because slang ^{is seemed to be cool.} However, slang is not ^{9 seems} always accepted by all people and in all situations. Some people has minus image to slang and does not like to use it. ^{for some reasons} There are three reasons – slang itself, who uses ⁹ it and where is used ^{doG} ^{slang}. ^{(2) people/place/slangu}

The first reason why slang has a negative image is that, as Gardiner points out, traditionally the source of slang is ^{negative R} some bad image, for example money, alcohol or sex etc. There are a lot of slang words for these image in dictionary such as "buck"(dollar), "tipsy"(slightly drunk) and "nicky"(kiss mark) (TAK slang dictionary). The original meaning of slang is ^R also bad. Looking at one example of slang words, 'shit' is often used to express one's feeling like ^{has negative} ^{WF regret} regrettable or anger and the original meaning is "feces". ^{That is R why} As knowing from this example, slang words are often dirty and coarse. Moreover, there are many words to show swearing or taboo words in slang dictionary, but slang to praise or show pleasure is a few. ^{WW As?} From these reasons, slang words or expression is ^{WW} coarse and vulgar so people do not like or hesitate to use it. ^{are} The thought that slang has bad image is rooted in many people's mind.

^{9 other} The another reason that some people ^{9 do} does not like to use slang words or expression comes from the fact ^{? who what?} (that people) who ^{use slang} In general, slang is used by more young people than adults, more working or middle class people than

Figure 3. Student essay extract (blue = teacher feedback, red = student revision notes)

The approximately 60 pages of the total set of student first drafts look similar to the feedback provided in this sample page from one student essay.

The counting and coding of symbols and written commentary, as well as a consideration of responses provided on the feedback end sheets, will enable a self-assessment of feedback practices using the Seven/Seven framework. This data provided a clear, yet complex, picture of what exactly I do when providing first draft feedback on student compositions in this Junior Composition class.

Results

Approximately six months after the completion of this particular JC 2 class, analysis of the 21 essay sets was begun. The start point was the counting and coding of each feedback intervention (FI) on essay first drafts. These FI's were separated into the following three categories:

1. **Correction code symbols** - the range of ten different symbols, shown earlier in Table 6, were used to direct students in editing and revising first drafts;
2. **Comments** - these included marginal comments (single words, phrases, sentences), and those comments written directly above or below sentences;
3. **Other FI's** – these included: (1) crossing things out (words, sentences), (2) making corrections (fixing, adding, re-writing), (3) indicating that words or phrases be moved, through using arrows, lines, or circles, etc. (4) a number providing an evaluation of the essays' thesis statement as follow: 1= 'good, fine'; 2= 'OK, but rewrite to make clearer, smoother'; or 3 = 'major changes needed, completely rewrite'. These three number codes were written next to the thesis statement for each essay.

For all of the 21 first drafts, a total of **828** FI's were identified and recorded. Table 6 is a record of the FI breakdown per student for each of the three categories mentioned above. At the bottom, a percentile breakdown of the total number of FI's for all essays is also provided for each of the three categories.

Table 6. *Feedback Intervention (FI) record on essay first drafts (N=21 essays)*

Essay writer	Correction code symbols	Other FI's	Comments	FI Total per essay
1. Natsumi	15	8	7	30
2. Chisako	12	10	9	31
3. Chihiro	24	6	10	40
4. Yukako	26	2	7	35
5. Kano	17	9	6	32
6. Mariko	33	10	13	56
7. Kurumi	30	6	14	50
8. Sachiko	12	5	11	28
9. Asami	22	4	25	51
10. Mio	19	3	17	39
11. Aki	24	7	13	44
12. Akiko	21	9	10	40
13. Shoko	28	3	17	48
14. Aya	34	4	20	58
15. Yumiko	28	6	19	53
16. Keiko	5	5	18	28
17. Misaki	19	6	15	40
18. Akiko	15	5	20	40
19. Yumiko	19	6	8	33
20. Asako	11	3	6	20
21. Saki	8	6	18	32
Totals	422	123	283	<u>828</u>
% of FI total	51%	15%	34%	

As can be seen from the table, the number of FI's per essay ranged from a low of 20

(Asako) to a high of 58 (Aya). The mean average for all 828 responses was approximately **39** FI's per essay. More than half of the FI total (51%) was comprised of correction code symbols. Teacher written commentary on the essay itself, rather than on the feedback end sheet, comprised 34% of the FI's provided on this group of essays.

The commentary was coded using a system for categorizing types of written feedback established by Haines (2004). Haines used seven categories for coding written feedback, and these are shown below in Table 7 with examples from my JC students slang essays.

Table 7. *Types of written comments (Haines, 2004), with slang essay examples*

Written comment types	Examples
1. Regulatory instructions	<i>No first names for in-text citations</i>
2. Advisory comments	<i>Use an example from English, not Japanese</i>
3. Descriptive observation	<i>Almost the entire paragraph is a Wikipedia quote</i>
4. Rhetorical questions	<i>People use slang for this reason?</i>
5. Direct criticism	<i>Difficult to read and understand</i>
6. Praise	<i>Very good paragraph</i>
7. Correctness	<i>Not all slang disappears</i>

The category of ‘correctness’ here does not refer to language use errors, but to any content (ideas or information) in the essay which may be mistaken or erroneous. The first type, ‘regulatory instructions’, in this categorization system refers to feedback on adhering to prescribed rules or instructions for the writing task. In the case of this JC essay, using proper MLA format and documentation of sources was particularly important.

From the 21 first drafts, each of my FI comments, in the form of a word, phrase, or sentence were counted, and these 283 written comments were catalogued according to Haines’s (2004) categorization. Table 8 below shows the numbers for the written feedback commentary I provided to students on their first drafts. A percentage of the total is also provided for each category.

Table 8. *Record of first-draft feedback commentary (N=21 essays)*

Type of Written feedback	Number	%
1. Regulatory instructions	28	10 %
2. Advisory comments	34	12 %
3. Descriptive observation	11	4 %
4. Rhetorical questions	156	55%
5. Direct criticism	25	9 %
6. Praise	5	2 %
7. Correctness	24	8 %
	<u>283</u>	

This table shows that the overwhelming type of commentary was rhetorical questions, comprising more than half the total. The ‘praise’ category contained the fewest comments, just 2% of the total number.

As noted earlier in Table 6 above, 15% of all interventions were categorized as ‘Other FI’s’ (crossing things out, making corrections, instructions to move words, and evaluating thesis statements). These FI’s were not individually numbered and categorized due to the difficulty of doing so.

However, the correction code symbols used, 51% of all FI’s, were categorized after being counted. A breakdown of the 422 correction code symbols provided on the 21 essays is provided in Table 9 below, including a percentile breakdown.

Table 9. Correction code symbol record for individual essays ($N = 422$ symbols)

	R = rewrite (awkward, unclear English)					0 = missing word					
	? = meaning not understood					pl = plural					
	g = grammar problem					cap = capitalization					
	ww = wrong word					sp = spelling					
	wf = wrong word form					= - join together (sentences)					
Essay writer	R	?	g	ww	wf	0	pl	cap	sp	=	Total
1. Natsumi	6		1	5	3						15
2. Chisako	5	3	1	1		2					12
3. Chihiro	16	3	3				1			1	24
4. Yukako	16	3	5	1			1				26
5. Kano	10	2	3	1		1					17
6. Mariko	10	16	4		1		1		1		33
7. Kurumi	15	13	1						1		30
8. Sachiko	9	3									12
9. Asami	7	8		1	2	1		2		1	22
10. Mio	10	4	1	1	1		1		1		19
11. Aki	11	3	5	3	1	1					24
12. Akiko	5	8	7	1							21
13. Shoko	13	11	3		1						28
14. Aya	16	14	1	2				1			34
15. Yumiko	12	10	3	1	1			1			28
16. Keiko	3	2									5
17. Misaki	12	6	1								19
18. Akiko	1	10	1	2				1			15
19. Yumiko	11	4	1	2	1						19
20. Asako	7	3	1								11
21. Saki	4	3		1							8
Symbol Totals	199	129	42	22	11	5	4	3	5	2	422
Percentages	47%	31%	10%	5%	2.6%	1.1%	.9%	.7%	1.1%	0.4%	

The most common symbols used ('R' and '?') comprised a total of 78% of the correction code FI's. The next highest number of symbols used were the grammar symbol (10%), followed by the vocabulary symbols ('ww', 'wf') making up approximately 8% of FI's used. All together these five most common symbols made up approximately 95% of correction code used on these essay first drafts. The number of symbols per essay ranged from a low of 5 (Keiko), to a high of 34 (Aya). The mean average of correction code symbols for the entire set of 422 symbols was approximately 20 symbols per essay.

Table 10 below shows a simple, descriptive statistical analysis of numeric data. High and low scores show the largest and smallest number of feedback interventions on an essay for each grouping.

Table 10. *Statistical overview of FI data (N = 21 essays)*

FI data group	Total	Mean (per essay)	SD	High	Low
Correction code symbols	422	20.1	8.0	34	5
Other FI's	123	5.8	2.3	10	2
Comments	283	13.4	5.4	25	6
All FI's	828	39.4	10.3	58	20

While the wide range in number of responses per individual essay, and in standard deviation scores, may have been affected by teacher idiosyncrasies or other factors at the time of feedback provision, the key reasons for this wide range is considered to be related to both the English language proficiency and writing competency of individual students. Some students are simply better English writers than others, and the amount and type of feedback provided to individuals on their essays is a reflection of this fact.

As mentioned, an important part of teacher written feedback was the end sheet attached to the back of each first draft. Feedback was given on the four key areas of content, organization, language use and use of source material. Table 11 below will give the reader a feel for the type of commentary provided on this feedback end sheet.

Table 11. *End sheet feedback for three students first-drafts*

Feedback criterion	Natsumi	Chisako	Mariko
1. Content	Focused and relevant content. Essay shows you know the topic well. The BBC section doesn't fit well with the essay so you need to think about that part.	Content is focused on one main point, but the essay is too short. Ideas need to be developed into longer, stronger paragraphs. Some information is unnecessary.	The content is focused, but some ideas need to be better presented and developed. Some parts are difficult to read and understand.
2. Organization	Good organization. Structure is easy to follow. Intro and conclusion work well also. Again, the BBC section interrupts the flow of the essay.	Structure of the essay is organized but this also needs to be improved. A more clearly written thesis statement and concluding paragraph are essential.	Structure and main ideas need to be better organized. The thesis statement should be clearer. Summarizing main points in the conclusion is also important.
3. Language Use	Some problems here, but the essay is mostly easy to read and understand.	Expression of ideas is sometimes unclear. Better grammar/vocabulary will make the essay stronger.	Some serious problems with language use, communication often breaks down. You need to express your ideas more clearly.
4. Use of source material	Sources are effectively used and cited. Page numbers are needed with some in-text citations.	Sources are used and cited, too much perhaps at times (ex. final paragraph). More of your own analysis, commentary is needed.	While you do use source material effectively at times, your Works Cited page and in-text citations need to follow MLA format.
<u>Overall</u>	A good first draft Natsumi. Editing and revising should make it even stronger.	This is a rough draft, Chisako. It is focused and organized, but serious editing and revising are needed to make the essay longer and stronger.	Lots of work needed to turn this into an effective essay, Mariko. Do a good job on editing and revising for a better final draft.

One noteworthy point here is that on all of the end sheets, in the ‘overall’ section, each student was addressed by name, helping to personalize the feedback. The reader is reminded again here that the feedback criterion on the end sheet for the first draft is the same as the assessment criterion on the summative grading of students final drafts.

This section provides a summary of the quantity and types of teacher feedback provided on these 21 first drafts. Now that the style, types and amount of feedback has been determined and catalogued, the Seven/Seven framework instrument can be used for the self-assessment of my feedback practices.

Discussion

The previous Results section provides a response to the first key question posed in the introduction regarding the nature and form of the written feedback I have been providing to students. I now have a much clearer picture of what I have been doing as in my role as ‘feedbacker’ in the JC course. After examining the results of my feedback practice for these 21 essays, I am also now in a position to respond to the second key question in the introduction: How do my feedback practices compare with those recommended in the feedback literature? I can now engage in a self-assessment of these practices, using the Seven/Seven framework for feedback assessment.

This discussion will be divided into three sections, one for Gibbs and Simpson’s (2004-05) seven feedback conditions which support learning, followed by Nicol and Macfarlane-Dick’s (2006) seven principles of good feedback practices. A summation of the results of the TSA will then be provided at the end of this section. Both the seven conditions and seven principles making up the two strands of the Seven/Seven framework are formatted as questions in this discussion. The TSA is comprised of responses to these 14 questions.

Part 1. Seven feedback conditions which support learning (Gibbs & Simpson, 2004-05)

1. Is sufficient feedback provided, both often enough and in enough detail?

Yes, students are provided with sufficient feedback, in terms of frequency and specificity. With respect to timing, one week after the first draft is submitted students receive feedback on it in the following class. Timing is crucial issue for feedback, the faster the better. Ideally, students would be able to receive teacher feedback within a day or two of producing the first

draft. This is not practically possible in such an academic writing setting as this, and dealing with more than 20, five-paragraph essays. Students receiving essay feedback one week later, is considered to be timely in this case. According to Gibbs and Simpson (2004), in order for feedback to be useful, it has to be quite specific. The writers of these 21 essays received a lot of detailed feedback through correction code symbols, teacher on-essay commentary, and the attached end sheet providing an overall formative evaluation of each script. The feedback is also specifically connected to the assessment criterion for the essay.

2. Does the feedback focus on students performance, on their learning and actions under their control (rather than on students themselves and their characteristics)?

The feedback provided does focus on the essay produced. The feedback, guided by the assessment criteria, is focused on the skills and knowledge that students need to develop (e.g. MLA documentation). The essay editing and revising process is under the students' control, and the final version produced is directly related to how much effort students put into this process of making the final draft of the essay stronger. Through the feedback provided, students are informed "where they have gone wrong and what they can do about it" (Gibbs & Simpson, 2004-05, p. 18). As previously mentioned with regard to the view of feedback in AfL theory, Gibbs and Simpson also note the fact that when personal characteristics are tied with critical feedback it can have a negative effect on students' sense of competence, or self-efficacy. For this JC slang essay, data analysis shows that feedback is not directed toward students themselves; it is task-focused, rather than ego-focused. One minor exception is the use of students name in the overall comment at the bottom of the end sheet, which may serve to create the impression of 'a conversation' with each student about their work. In any case, the feedback provided is performance-focused, not student-focused

3. Is the feedback timely, received by the students while it still matters to them? Yes, students receive a host of feedback about their performance, while that performance is still going on. The feedback is directly of import to the final draft of the essay they must produce one week after the formative feedback has been received. The teacher feedback provided on the essay first draft can be used to improve the 'work-in-progress'. In his seminal article about formative assessment, Sadler (1989) wrote that an essential condition for improvement is that the student "is able to monitor continuously the quality of what is being produced

during the act of production itself"(p. 121, emphasis in original). Gibbs and Simpson also note that feedback must be relevant to ongoing work for it to be most meaningful.

4. Is the feedback appropriate to the purpose of the assignment and to its criteria for success?

In this particular case, the focus of the essay was to produce a quality five-paragraph research essay based on at least three sources. Particular focus was paid to use of source material and MLA documentation. The feedback provided was appropriate to these assignment purposes, and was directly related to them. The feedback end sheet, in particular, emphasized the key criterion (content, organization, language use, use of source material) and teacher comments included here informed students clearly of essay expectations (e.g. 'Works Cited page and in-text referencing do not follow MLA format.'). In order to orient themselves effectively to the assignment task, student understanding of success criteria is important (Gibbs and Simpson, 2005-05). Sadler (1989) also deemed it essential "that the student comes to hold a concept of quality roughly similar to that held by the teacher" (p.21). First draft feedback provided on the slang essay was directly related to assignment purposes and success criterion.

5. Is feedback appropriate, in relation, to students' understanding of what they are supposed to be doing?

Gibbs and Simpson (2004-05) tell us that "Feedback needs to be sensitive to what kind of writing is expected and what students are likely to understand about it" (p23). The task presented to students for this slang essay, including the choice of essay questions and teacher expectation of student knowledge and usage of the related slang source material, is clearly laid out for students. The feedback provided is also appropriate to their understanding of the slang topic and ability to write a clear research essay focused on the question they had selected. It should be remembered that this was the fourth essay students had written (and received feedback on) in this JC course; students were quite clear as to 'what they were supposed to be doing' and the feedback provided was appropriately focused.

6. Is feedback received and attended to?

A number of studies have described students paying little or no attention to (summative) feedback provided by teachers, a point of much frustration and irritation for the feedback provider. Gibbs and Simpson (2004-05) identify a number of steps that can be taken to engage students with feedback, including “using two-stage assignments with feedback on the first stage, intended to enable the student to improve that quality of work for a second stage submission, which is only graded” (p. 24). According to Cooper (2000), such an assessment system, synergizing formative and summative elements, can improve the performance of almost all students, especially the weaker ones. This slang essay (as with all JC papers) was a two-stage assignment, and as such feedback was received by students and attended to in the editing and revising process.

7. Is feedback acted upon by students?

Yes. As noted, students are given feedback in sufficient time and quantity to edit and revise their first draft. The class in which first drafts are returned is always designated as an ‘editing/revising workshop’ in the essay production cycle. This is the beginning of the re-writing process, and over the coming days students work on improving their essay prior to submission of the final draft in the following weeks’ class. James (2006) writes: “Only learners can do the learning, so they need to act upon information and feedback if their learning is to improve” (p. 8). JC students did engage with the feedback provided, often making extensive revisions (and improvements) between first and final drafts.

According to Gibbs and Simpson (2004-05), feedback works best to support learning when these seven conditions are adhered to. My written feedback practice in this JC course meets all seven of these standards, and, as such, is supportive of student learning and their efforts to be better academic essay writers.

We now move on to the second strand in the Seven/Seven framework in this teacher self-assessment of feedback practices.

Part 2. Seven principles of good feedback practice (Nicol & MacFarlane-Dick, 2006)

The seven effective feedback principles laid out by Nicol and MacFarlane-Dick (2006) place special emphasis on students becoming self-regulated learners, in control of their own learning. These feedback principles incorporate Gibbs and Simpsons' (2004-05) seven conditions discussed above, but are broader in scope. This wider perspective provides a complimentary and more comprehensive picture of feedback practices, enabling a more effective self-assessment of these practices to be completed. Following the earlier format, these principles are here posed as questions.

1. Does the feedback clarify what good performance is?

As Sadler (1989) and Black and Wiliam (1998a) stress, it is only if students understand learning goals that they will be able to self-assess progress toward them, and be able to achieve them. Such an understanding is also crucial in making sense of teacher feedback, which should also make abundantly clear what good performance entails. As previously stated, the feedback provided on this slang essay does indeed clarify criteria, standards and goals for students as they work to improve the first draft of their essay. This is particularly true for the end sheet attached to each essay. Here are two examples of overall commentary provided at the bottom of two feedback sheets:

- *It is a well-organized essay Akiko, and I can see what you are trying to do. But explanation and support for main ideas in your body paragraphs need to be stronger and clearer in the final draft.*
- *Following MLA format when using sources, and organization are both good Asami. Content and language use are the two weak areas to focus on for making a better final draft.*

These two examples show that letting students know what good performance entails was incorporated in the feedback provided.

2. Did the feedback facilitate the development of self-assessment (reflection) in learning?

In the class in which students are required to submit the first draft, students also complete tasks in which they engage in both self-and peer assessment. Such activities help generate

both the internal and external feedback that promotes reflective learning.

Prior to submitting their first drafts, students are given a self-assessment essay checklist to evaluate their ‘work in progress’ and identify strong and weak areas. Following this activity, by using the reverse side of the checklist, students spend some time assessing a partner’s first draft and giving oral feedback on it. At this peer-assessment stage, students are told they may provide feedback in Japanese and generally everybody chooses to discuss their drafts in their first language. These self-and peer assessment elements included in the JC course may be considered examples of ‘short cycle’ formative assessment in Wiliam and Thompson’s (2007) categorization, as feedback is provided within a lesson period. **Appendix A3** shows the self/peer assessment instrument used.

Before submitting the final draft of their essay students also complete a self-assessment essay checklist and attach it to the final draft of their essay. Gibbs (2006) points out the key reason for the importance of such student assessments of their own and others work:

. . . the value of self- and peer assessment is that students internalize academic standards and are subsequently able to supervise themselves as they study and write and solve problems, in relation to these standards. It is the act of students making judgment against standards that brings educational benefits, not the act of receiving a grade from a peer (p. 27).

In addition to these formally structured self-and peer assessments (i.e. use of checklists and related tasks), the teacher commentary provided on first drafts was intended to promote student reflection about the slang topic and what they have written about it. As indicated in the results section, more than half (55%) of the written feedback provided on the body of student essays were rhetorical questions. Here are a few examples:

- Do you ‘think carefully’ about when to use slang?
- If this negative image is the main feature of slang, why is it so popular and widely used?
- Is it illegal to use slang?
- In what way? The connection is not clear.

According to Black and Jones (2006), questions are useful ways of framing feedback comments because they initiate thinking, and their questioning nature encourages students to start improving on the work in question. As this investigation was a reflective exercise, I was not aware of this Black and James (2006) rationale for framing feedback in question format

until much later than when the essay feedback was provided. Indeed, there was no plan to make use of a questioning format in the feedback, and I was quite surprised to see how extensively it was used across the entire set of 21 first drafts. It was hoped that such commentary would serve as prompts for students to reflect on what they had written, how the reader might interpret it, and whether intended meaning was effectively communicated or not.

I am satisfied that the feedback generated by the assessment framework itself, as well as my written feedback on essay first drafts, promoted the type of self-regulation which Nicol and Macfarlane-Dick (and, indeed, AfL) rightly place so much emphasis on; the idea that “feedback should be used to empower students as self-regulated learners” (p. 199).

3. Does the feedback deliver high quality information to students about their learning?

Nicol and Macfarlane-Dick point out the research finding that much external feedback given to students is not of good quality because it may be delayed, irrelevant, uninformative or confusing for the student. The authors provide the following definition of what ‘good quality’ feedback entails:

Good quality external feedback is information that helps students troubleshoot their own performance and self-correct: that is, it helps students take action to reduce the discrepancy between their intentions and the resulting effects. (Nicol & MacFarlane-Dick, 2006, p. 208)

Considering this explanation, and the nature of the feedback provided with its focus on student editing and revising their own work ('trouble shooting their own performance'), the feedback provided to the JC students may be considered 'of good quality'. Each student receives feedback on strengths and weaknesses of their efforts and how any deficiencies may be repaired to produce an improved final draft. The feedback enables students to see if the meanings they intended to convey in their writing resulted in successful communication with the reader, or fell short of that goal. Nicol and Macfarlane-Dick write that the quality of student feedback may be improved by such things as having predefined assessment criteria and be received in a timely manner; practices already noted as being part of the feedback process used in the JC class.

4. Does the feedback encourage teacher and peer dialogue around learning?

Nicol and MacFarlane-Dick (2006) remind us that before external feedback can be used to make improvements it must be understood by students; yet the research literature suggests that often students do not understand the feedback given by their teachers and tutors. In order to maximize student understanding of the feedback provided, the authors suggest that feedback be conceptualized “as *dialogue* rather than as information transfer. Feedback as dialogue means that the student not only receives initial feedback information, but also has the opportunity to engage the teacher in discussion about the feedback” (p. 210, emphasis in original).

As mentioned, the class in which student first drafts are returned is designated as an editing/revising workshop. For most of the class time students work on improving their essays. This becomes a busy time for me as I move around the class responding to student questions and queries. Sometimes they ask for advice, or perhaps they do not understand what point my commentary is trying to make. While student numbers can make it difficult to have a dialogue with everyone, during this workshop, I inevitably talk briefly to most of the students in class. At times, prior to class I make a note of particular students I want to talk to about their weak first drafts. Students are also directed to e-mail me or come and visit my office if they needed further assistance, before the final draft is due. Some students take advantage of this opportunity, and make queries or ask me to check things like a revised thesis statement.

Nicol and MacFarlane-Dick note that, “Discussions with the teacher help students to develop their understanding of expectations and standards, to check out and correct misunderstandings and to get an immediate response to difficulties” (p. 210). It should be remembered, however, that these students are L2 learners with differing levels of both writing and speaking proficiency in English. This factor may impede an effective dialogue taking place, as students may not be able to clearly articulate in face-to-face dialogue what they have trouble communicating in their writing. However, such dialogue was indeed encouraged in the JC class.

As for peer dialogue, Nicol and Macfarlane-Dick note the difficulty that class sizes pose in enabling teacher-student dialogue about feedback, and propose a possible solution in small groups of students discussing feedback together. In this JC class the feedback framework in place did not particularly promote such dialogue about teacher feedback among peers in the class. Students generally worked on their own to make improvements for the final version of their essays.

5. Does feedback encourage positive motivational beliefs and self-esteem?

In both learning and assessment, student motivation and self-esteem have very important roles to play (Nicol and Macfarlane-Dick, 2006). Assessment, in particular, can have a strong impact on students and, as Bryan and Clegg (2006) note, having someone make a judgment about the quality of your work can certainly be an emotional, and potentially humiliating, business. As noted earlier, the AfL literature also makes a point that feedback should be ‘task-centered’ and not ‘ego-centered’, and that feedback can have a negative effect on student attitudes and performance if attention is drawn away from the task itself and towards self-esteem (Black & Wiliam, 1998a). Nicol and MacFarlane (2006, p. 212) write, “it is important that students understand that feedback is an evaluation, not of the person, but of the performance in context”.

After reviewing studies on self-esteem and motivation, Nicol and Macfarlane Dick summarize the teaching practices which will be more likely to enhance student motivation and self-esteem:

- Low-stakes assessment tasks with feedback directed toward the provision of information about progress and achievement, rather than just high-stakes summative assessment tasks providing information only about success or failure, or how students compare with peers
- The provision of marks on written work only after students have responded to feedback commentary
- Giving students time to rewrite initial efforts at the task
- Inclusion of drafts and re-submissions in the assessment process.

My JC course incorporated all of these elements in its assessment regime, and, as such an affirmative response may also be offered to this feedback question. The initial low-stakes formative assessment (first draft) was followed later by the high-stakes summative assessment (final draft). Students had time, and feedback directing their efforts, in the interval between these two assessments to make the essay better, stronger, and clearer.

A related, noteworthy point on the issue of self-esteem and motivation is the fact that analysis of my feedback commentary shows very few praise comments used, much less than comments which make direct criticisms. As noted in the results section, of my 283 examples of written commentary on student first drafts only 5 (2% of the total) offered praise on students work, while the number of comments giving direct criticism amounted to 25, (9% of

the total). Table 12 below shows all five praise comments and a representative sampling of ten comments providing direct criticisms.

Table 12. *First-draft praise comments and some criticism commentary*

Praise comments	Critical comments	
Very good paragraph	Confusing	Repetitious
Good point	These arguments are weak	Essay title is confusing
Good summary	This paragraph is confusing, difficult to catch your point	Too short. Develop this paragraph more
Interesting beginning	No, not true	Too simplistic
Good ending	Very difficult to follow. I don't see the connection to slang	This paragraph has a weak argument

It could be argued that, I could provide more self-esteem building praise in my commentary, something recommended in the literature. However, such praise is included in the overall comments section on the end sheet attached to each first draft. Additionally, overall commentary on the final summative assessment always starts with an inclusion of a few words of praise about essay strong points.

6. Does the feedback provide opportunities to close the gap between current and desired performance.

Yes, in fact the assessment framework is set up to encourage this gap-closing process. The JC essays students write are set up as ‘two-stage assignments’ with the feedback provided at the first stage designed to result in an improved performance at the second stage. In an oft-quoted part of his seminal formative assessment paper, Sadler (1989) identifies three key factors in establishing good-quality formative assessment: (1) the ability to understand the goals being aimed for, (2) some way of comparing the actual performance with the goals, and (3) the

skills to engage in activities which close the gap between the two. In the context of this JC course, the desired performance is made clear to student as it is embedded in pedagogical practice and the assessment framework at both key formative and summative points. The feedback provided on the first draft is a key factor in enabling students to close the gap between that goal and their current performance. Boud (2000) explains the situation as follows:

The only way to tell if learning results from feedback is for students to make some kind of response to complete the feedback loop (Sadler, 1989). This is one of the most often forgotten aspects of formative assessment. Unless students are able to use the feedback to produce improved work, through for example, re-doing the same assignment, neither they nor those giving the feedback will know that it has been effective. (p. 158)

In the JC class, students are provided with extensive feedback to help them produce improved work, and this inevitably results in better final drafts of their essays. While students may have some difficulties in closing gaps (due to English language ability or academic writing proficiency, for example) the key point here is that *opportunities are provided* for them to do so. Nicol and Macfarlane-Dick use the phrase ‘providing feedback on a work-in-progress’ (p. 213), and this is an important part of the feedback cycle in the JC course.

This issue of closing gaps will be addressed in more detail in Part B of this chapter, which deals with student response to feedback.

7. Does the feedback provide information to teachers that can be used to help shape the teaching?

While students are the primary beneficiaries of feedback, ideally providing them with information that is accessible and usable, it should also provide good information to teachers as well (Nicol & Macfarlane-Dick, 2006). As Yorke (2003) explained;

The act of assessing has an effect on the assessor as well as the student. Assessors learn about the extent to which they [students] have developed expertise and can tailor their teaching accordingly. (p. 482)

Indeed, and in the JC course being an assessor of first drafts had a significant impact on subsequent teaching, and especially the oral feedback given at the start of the editing/revising

class. For example, it became clear in the formative assessment of these slang essay first drafts that particular sources of difficulties for students were (1) the composing of clear, concentrated thesis-supporting arguments within body paragraphs, as well as (2) uncertainty as to the rules for when and how parenthetical citations must be included in the text. These two points were centers of attention in the subsequent lesson, and continued to be recurring teaching points as we moved on to the next essay topic and task. My making note of the most common type and location of student difficulties in the first drafts, easily identifiable from the quantity of my related feedback, regularly turned into teaching points that were recycled as students developed their expertise in academic essay writing. Thus, yes, the feedback did indeed provide information to me which helped shape the instruction provided in class.

This brings us to the end of Nicol and Macfarlane-Dick's (2006) group of seven principles provided for teachers to examine their current assessment practices. The authors note that their seven principles of effective feedback practice "address a wide spectrum – the cognitive, behavioral and motivational aspects of self-regulation" (p. 215). This wide spectrum of points to assess feedback practice against, combined with the seven feedback conditions that support learning laid out by Gibbs and Simpson (2004-05), together provide a comprehensive framework for teachers to engage in a self-assessment of their feedback practices.

Summary of Seven/Seven framework findings

Overall, how do my feedback practices on this slang essay, representative of my standard feedback modus operandi for all JC essays, measure up against the recommended practices in the Seven/Seven framework? Here I would like to provide a tabulated summary of this TSA process. In comparing my actual practice with the recommended practice in the Seven/Seven framework, I use the following simple three-category format:

- **yes** = feedback does meet this condition or principle
- **no** = feedback does not meet this condition or principle
- **somewhat** = there is some evidence for this condition or principle being met.

The results of my TSA of feedback practice are recorded below in Table 13.

Before examining the results shown, the reader is reminded of two points here. First, the

self-assessment results refer to first draft feedback provided in the JC course *as a whole*, based on the feedback practices shown in examining the representative sample of 21 essays. Strengths and weaknesses of feedback provided on individual slang essays showed some degree of variability (to be expected considering an average of 39 FI's per essay). Secondly, the 'yes' indicates that it is evident that this feedback condition or principle is part of the feedback process. It does not mean that there is no room for improvement in making the practice work better.

Table 13. *Results of TSA of JC feedback practices with the Seven/Seven framework*

Seven/Seven framework of feedback assessment in JC course	TSA evaluation		
Seven feedback conditions (Gibbs & Simpson, 2004-05)	yes	somewhat	no
1. Sufficient feedback is provided, both often enough and in enough detail	✓		
2. The feedback focuses on students' performance, on their learning and on actions under the students' control, rather than on the students themselves and on their characteristics	✓		
3. The feedback is timely in that it is received by students while it still matters to them and in time for them to pay attention to further learning or receive further assistance	✓		
4. Feedback is appropriate to the purpose of the assignment and to its criteria for success	✓		
5. Feedback is appropriate, in relation to students' understanding of what they are supposed to be doing	✓		
6. Feedback is received and attended to	✓		
7. Feedback is acted upon by the student	✓		
Seven feedback principles (Nicol & Macfarlane-Dick, 2006)	yes	somewhat	no
1. Feedback helps clarify what good performance is	✓		
2. Feedback facilitates the development of self-assessment (reflection) in learning	✓		
3. Feedback delivers high quality information to students about their learning	✓		
4. Feedback encourages teacher and peer dialogue around learning		✓	
5. Feedback encourages positive motivational beliefs and self-esteem	✓		
6. Feedback provides opportunities to close the gap between current and desired performance	✓		
7. Feedback provides information to teachers that can be used to help shape the teaching	✓		

This rating table shows I have determined that my feedback practices stand up well against all 14 criteria in the Seven/Seven framework. All categories received a positive evaluation, except the Nicol and MacFarlane-Dick (2006) principle related to teacher/peer dialogue for reasons already noted. I have tried to be as objective as possible in my self-assessment, fully aware of the fact that, by definition, any such ‘self’ evaluation is a subjective exercise, and that the main criticism of student self-assessment is inaccurate judgment and unreliable scoring. Just as both Gibbs and Simpson (2004-05) and Nicol and Macfarlane-Dick (2006) based their feedback recommendations on extensive reviews of the research literature, the JC research data and subsequent analysis provide a supporting rationale for my assessment for each of the 14 items in the framework.

It should be noted that this self-assessment applies only to my feedback practices in this particular Junior Composition course. Had I applied the Seven/Seven framework to other writing courses I teach, for example a first year course in which no formative feedback is provided (primarily due to large class size), weak areas of my practice would predominate. One reason why my feedback practices are more careful and considered for this class, and score well in this framework, is that I realize the high stakes nature of the JC course in helping students develop the necessary essay writing skills to be successful in completing the senior year graduation thesis.

With regard to judging the effectiveness of formative assessment, Yorke (2003, p. 483-484) writes that two questions can be asked:

1. Is what the assessor has done regarding feedback the best that could have been done (or-more weakly-reasonable in the circumstances)?
2. Did the formative assessment influence student behavior?

While I do think that the formative assessment (i.e. first draft feedback) did positively influence student behavior, the first of these questions is of most relevance here in Part A of this chapter. While perhaps there is room for improvement with regard to my feedback practices (e. g. an improved correction code, more praise of effective parts of essays, using the Internet to receive drafts and provide faster feedback), considering the context of the JC academic writing class, the numbers of students and amount of writing to respond to, I am satisfied that the feedback I have provided was at least ‘reasonable in the circumstances’.

The self-assessment process engaged in here with The Seven/Seven framework shows that my feedback practices in this JC course rate positively against recommended feedback

practice in the assessment literature. More importantly, these feedback practices are effective in helping students produce better academic essays, and hopefully (though less clear), become better English writers.

Conclusion

Thus, the critical educator will see him/herself as a co-worker with students in pursuit of education rather than a provider of knowledge for passive recipients. (Grundy, 1989, p. 96)

This report is a documentation of my attempts to become a more ‘critical educator’ through the self-assessment of my written feedback practices in an academic writing course. Among other things, it has resulted in, and reinforced, my feeling of also being a learner, ‘a co-worker with students’, as Grundy (1989) phrases it, as we all try improve our knowledge and skills in that ‘crucible of learning’ - the classroom.

This teacher self-assessment of written feedback provided in an academic writing course focused on two aims: gaining a clearer picture of the form and nature of the feedback being provided, and, subsequently, assessing these feedback practices against recent recommended practice in the feedback literature. A close look at 21 essay first drafts resulted in a systematic documentation of the feedback interventions provided, an average of 39 such FI’s per essay. These were coded, counted and analyzed. The feedback end sheet attached to each essay was also reviewed and considered as part of this TSA exercise. Two important publications from the feedback literature, Gibbs and Simpson (2004-05) and Nicol and Macfarlane-Dick (2006), provided the key criterion of conditions and principles for the Seven/Seven framework, the instrument used to assess my feedback practices in the JC course. This Seven/Seven framework provides a comprehensive summary of what effective feedback should entail, and is a functional self-assessment instrument teachers can use to put their feedback practices under the microscope. Overall, the TSA results of my feedback in the Junior Composition class must be seen as satisfactory (i.e. supportive of student learning), and I seem to be following recommend feedback practice with this particular writing course.

Assessment for *teacher* learning has been a central theme in this feedback report. But such a teacher self-assessment as described here is a not a summative activity, but a formative one (McCloskey & Egelson, 1993); improvement in pedagogic understanding and practice is seen as an ongoing, continuous process. Just as AfL is firmly grounded in a constructivist theory of

learning, so my learning about feedback is knowledge that is being constructed and continuously refined. According to Lambert and Coombs (1998, p. 10):

Learning is a constructive process that occurs best when what is being learned is relevant and meaningful to the learner and when the learner is actively engaged in creating his or her own knowledge and understanding by connecting what is being learned with prior knowledge and experience.

At the time of this writing I continue to teach the Junior Composition course, and a number of other writing classes, so my ongoing learning about feedback is indeed ‘relevant and meaningful’. This process will require my continued active engagement to ensure that I am providing the best possible ‘feedback for learning’ in the courses I teach.

The ultimate aim of AfL is student involvement in their own assessment so that they can reflect on where they are in their learning, understand where to go next and the steps required to help them get there (James, 2006). While students and their learning remain the primary focus in AfL, involving teachers in assessing their own practice has also become a fundamental consideration. This TSA exercise has enabled me to critically reflect on my feedback practice and understanding. An obvious next step seems for me to apply the Seven/Seven feedback framework to the other courses, writing and otherwise, that I teach and use it as a guide for improving practice. Through the continued application of the Seven/Seven framework to the feedback I provide in JC, and extending this to the other courses I teach, I have developed an understanding of where I need to go next in improving my feedback practices and the steps required to get there.

Nicol and Macfarlane-Dick (2006) point out that any model of feedback must take account of the way students understand and make use of feedback information. Part B of this chapter will focus on this student side of the issue and consider how feedback is used by students to close gaps between current and desired essay writing performance. Results of a student survey on feedback will help complete this examination of feedback for learning in the AfL framework.

Part B. Does the feedback feed forward? Student response to and views of teacher feedback

Teachers need to view feedback from the perspective of the individuals engaged in the learning . . . (Hattie and Timperley, 2007, p. 101).

Introduction

After a close examination of the feedback issue from the viewpoint of teacher as ‘feedbacker’ in Part A of this chapter, it is necessary to complete the picture by considering the student perspective. This student angle forms a crucial part of this feedback for learning equation, as noted by Lee (2008):

Without understanding how students feel about and respond to teacher feedback, teachers may run the risk of continually using strategies that are counter-productive. As teachers give feedback on student writing, it is crucial that student responses to the feedback are fed back to teachers as a heuristic to help them develop reflective and effective feedback practices (p. 144).

Part B of this chapter takes a closer look at student responses to and feelings about the feedback provided in order to develop a more comprehensive understanding of feedback in the JC class, and in AfL practice. As Murphy (2000) warns, because of the fact that knowledge is interactively constructed by teachers and students in the classroom, looking only at how teachers respond to student writing without examining how students react to such feedback would be an incomplete view of the revision process and limit the conclusions that could be drawn about it.

Effective, timely feedback is a fundamental element of formative assessment; “the two concepts of formative assessment and feedback overlap strongly” (Black & Wiliam, 1998a, p. 47). Considered one of the central components of AfL theory and practice, feedback is “defined in terms of ‘closing the gap’ between actual and desired performance” (Stobart, 2006, 141). Influenced by the seminal writings of Sadler (1989), the core of formative assessment was defined by Black and Wiliam (1998 b) as comprising two actions: the student must recognize that a gap exists between the current performance and the desired one; and the student must engage in effective action to close the gap. Student knowledge and action in

being aware of and attempting to close the performance gap are seen as pivotal in the feedback process. In their 2003 book, Black and his colleagues stress this point as follows:

The learner first has to understand the evidence about this gap and then take action on the basis of that evidence. Although the teacher can stimulate and guide this process, the learning has to be done by the student. It would be a mistake to regard the student as a passive recipient of any call to action: there are complex links between the way in which a message is understood, the way in which that perception motivates a selection among different courses of action, and the learning activity that might follow. (Black et al., 2003, p. 14).

While the teacher acts as stimulator and guide to action in closing the gaps in performance, the work (and the learning) must be done by the student.

Based on the work of Black and Wiliam and their formative assessment colleagues, and in the context of a writing class, Lee (2007) described four essential conditions required in order to use feedback to promote assessment for learning (i.e. formative feedback):

1. Students are told about their strengths and what needs to be done in their writing – e.g., areas for improvement, in terms of content, organization, language, etc.; the assessment is prospective;
2. Information is communicated clearly and made intelligible to students in terms of what they have learnt, hence a close link between teaching, learning and assessment;
3. Students act on the teacher feedback and are provided with opportunities to improve their learning based on the teacher feedback;
4. Students play an active role in managing their own learning. (p. 182)

While Part A of this feedback for learning chapter discussed the first two of these essential elements, Part B below will pay attention to conditions 3 and 4; student actions in response to the feedback provided, and their active role in this process.

Part A of this chapter focused on providing information to students about the existing gaps in the first draft of the slang essay (current performance) and the requirements for a good quality final draft, academically researched and referenced in MLA format (the desired performance). In Part B we will examine the gap-closing actions taken by students to produce a better final draft. Part B takes a look at the final drafts of the slang essay and compares them

to the first drafts, to examine whether gaps are being closed. While Part A focused on draft 1 of the slang essay, Part B compares the first and final drafts in response to the feedback interventions (FI's) provided. Part B of this chapter will address the following two questions related to both the impact of the feedback provided and student perceptions of it:

1. Are students producing better writing based on the teacher feedback provided? (Are gaps being closed?)
2. How do they feel about the feedback provided in the JC course?

We begin, in Section 1 by examining student response to first draft feedback and the changes that resulted in the final draft. Some forty examples of student response to the teacher FI's are presented and provide, a 'before (feedback) and after (feedback)' picture. As will be seen, many successful revisions took place, while others were less successful. Section 2 below will describe the results of a student survey administered in the final class of JC about the feedback provided to them throughout the course. After a brief overview of the literature related to student response to teacher feedback, we complete this introduction with a brief review of the meanings associated with the critical concept of 'feedforward'.

Student response to feedback: brief literature review

Most feedback research focuses on 'the input side of the equation', with much less focus on how students interpret and deal with the feedback (Polos & Mahoney, 2008). Ellery (2008) summarizes the sometimes conflicting evidence in the research literature in regard to how students respond to feedback: feedback is often misunderstood; it is often not read; it may be read but not acted upon; and it sometimes has no effect on student learning. According to Lee and Schallert, (2008) studies of second language writing show that ESL students were willing to follow closely the feedback provided by teachers, but such commentary "had the potential of miscommunicating and of being misunderstood" (p. 165). Weaver (2006) contends that students' intellectual maturity and previous experience both play a large part in their learning and as a result the extent to which they are able to engage with teacher feedback. (The reader is reminded here that JC students are third-year university students in the 20-21 age range, and have had previous experience with essay writing and dealing with teacher feedback in the first semester of the JC course.)

With EFL/ESL and L2 writing, feedback is more positively viewed. It is recognized that L2 writers want and benefit from effective teacher feedback and produce better final products in a multiple-draft writing class. Wojtas (1998) reported that many students improved their work after understanding the purpose of feedback and the assessment criteria being used. In looking at writing research involving multiple-draft classrooms, Lee (2008) concluded that the research evidence shows teachers' feedback comments are attended to, and students believe such feedback helps their writing improve. In an extensive review of the formative feedback literature, Shute (2008) reported "formative feedback has been shown in numerous studies to improve students' learning and enhance teachers' teaching to the extent that the learners are receptive and the feedback is on target (valid), objective, focused, and clear" (p. 182). Shute (2008) concluded that the research evidence indicates that if feedback is delivered correctly, learning processes and outcomes can be improved significantly.

Feedback as feedforward

Building on the work of previous writers (for example, Ramaprasad, 1983; Sadler, 1989; Black & Wiliam, 1998a), Hattie and Timperley (2007) provide an effective framework for considering feedback. After explaining that the purpose of feedback is "to reduce discrepancies between current understandings/performance and a desired goal" (Hattie and Timperley, 2007, p. 86), the authors conceive of three major feedback questions as follows:

1. Where am I going? (the goals) = **feed up**
2. How am I going? = **feed back**
3. Where to next? = **feed forward**

According to Hattie and Timperley (2007), when both students and teachers strive to answer these questions an ideal learning environment or experience will occur. This simple, yet effective formulation of the crucial feedback questions bears a strong similarity to the definition of AfL provided by the Assessment Reform Group in the UK; "[AfL is] 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" (ARG, 2002). Again, the role of effective feedback in the AfL framework is sine qua non, a focal and indispensable component of its theory and practice.

Feedback becomes feedforward when it is “forward-looking so that it can improve students’ learning and enhance their future performance on assessed tasks” (Carless et al. 2006, p. 12). Thus, feedforward means providing useful information to the student, *while it still matters*, that will help them recognize where gaps in performance (and learning) are, and to use that information to close the gaps and move forward. Feedback that feeds forward can enhance students understanding and achievement level. Students learn faster and much more effectively when they have a clear sense of how well they are doing and what they might need to do in order to improve (Carless, 2006). In the context of the Junior Composition course, this feedforward idea would entail students successfully responding to teacher feedback to produce a better final draft of the slang essay , and hopefully improve their learning regarding academic writing skills.

We now move on to the main sections of Part B of this feedback chapter, looking at how students responded to teacher feedback (Section 1), followed by their perceptions of the first draft feedback provided in the JC course (Section 2).

Section 1: Closing gaps - student response to teacher feedback

While feedback is widely regarded as central to writing development (Hyland, 2002), how students respond to the feedback provided is another matter. After noting that feedback is required for learning to take place, Gibbs (2006) writes: “The crucial variable appears not to be the quality of the feedback (which is what teachers focus on) but the quality of engagement with that feedback” (p. 26). Students must engage with the feedback for it to feed forward and become ‘feedback for learning’. Polos and Mahoney (2008) echo this view, explaining that “how the student interprets and deals with feedback is critical to the success of formative assessment” (p. 144).

Hyland and Hyland (2006) remind us that feedback is ‘an act of communication’ and like all such acts it occurs in particular cultural, institutional and interpersonal contexts (p.10). The cycle of reading student texts and providing feedback, and students responding to this feedback involves a process of meaning negotiation between reader (teacher) and writer (student). Knoblauch and Brannon (1984) describe the process as follows:

The comments of the facilitative reader are designed to preserve the writers control of the discourse, while also registering uncertainty about what the writer wishes to communicate. The questions posed suggest the possibility of negotiation between writer and reader, leading

to richer insights and more meaningful communication. Negotiation assumes that the writer knows better than the reader the purposes involved, while the reader knows better than the writer the actual effects of authorial choices. (p. 128)

This section takes a closer look at this complex negotiated interaction between teacher and students; their attempts at collaborative engagement in ‘meaningful communication’ to help produce the best possible final draft of the slang essay.

Comparing first and final drafts of the slang essay

How did students respond to the feedback interventions (FI’s) that I provided? Were they able to close the gaps between current and desired final performance? Did the teacher feedback feed forward to better student writing in the final essay draft? Examples of student writing provided in the tables below help provide answers to these questions.

The reader will remember from Part A that an average of 39 FI’s per script was provided to the 21 first drafts of the slang essay. An overall total of 828 such interventions were documented, including correction code symbols, commentary and other FI’s. In this section, 40 responses to teacher feedback are provided (including 15 examples in **Appendix B1**) as a representative sample of students’ attempts to close the writing gaps between initial (first draft) and desired performance (final draft). These examples were randomly selected from all 21 essay sets. Table 1 below shows one such example of a revised thesis statement.

Table 1. Closed gaps: thesis statement revision from one student

First draft
WW Teenagers use slang to get solidarity with friends, to express their identity, and to insist the independence from adult. (1)
Final draft
Teenagers use slang to show solidarity with friends, to express their identity, and to assert the independence from adult.

As explained in Part 1, the numerical correction code symbol (1) indicates that I thought this

to be a good quality thesis statement. Two other correction code symbols are provided here (ww, R) on the first draft. The better, more clearly written final draft version is considered a successful response to the FI's provided. The above example shows one instance of a student 'closing the gaps' to produce an improved thesis statement in the final draft. (As a brief aside, I wonder why I did not interject any revision request for the phrase 'from adult' in the above example. This is a good case in point of the sometimes inconsistent, idiosyncratic nature of teacher feedback).

We will examine some other gap-closing attempts by students, but first a brief introductory note is required about the distinction between 'global' and 'local' errors or concerns in student writing.

Ferris (2002) makes the distinction between "global errors—those that interfere with the overall message of the text—and local errors, which do not inhibit a readers' comprehension" (p. 57). In broad terms, global-level concerns deal with content and overall argument and paragraph structures, while local-level concerns deal with sentence structure and grammatical problems. In looking at student response to teacher feedback, we make note of this global/local distinction.

Table 2 below shows some examples of local concerns being successfully revised in the final draft in response to the teacher feedback, the FI, on the first draft.

Table 2. Closed gaps: local concerns successfully revised

	First draft	Feedback intervention	Final draft revision (response to feedback)
1	“It is generally . . .” (<u>Alan</u> 100)	<i>R, use family name</i>	“It is generally . . .” (Gardiner 100).
2	Slang is sometimes considered to be rebellious and subversive (Gardiner).	<i>page # needed</i>	Slang is . . . subversive (Gardiner 102).
3	Frompkin et el. (1998) <u>definite</u> that	<i>sp, ?</i>	Frompkin et al. (1998) say that one reason . . .
4	. . . feelings like <u>regrettable</u> or . . .	<i>wf [word form]</i>	. . . feelings like regret or . . .
5	Slang has strong power to make a great <u>expression</u> on readers and . . .	<i>ww [wrong word]</i>	Slang has strong power to make a great impression on readers and . . .
6	If people do not use slang, <u>they sometimes</u> isolated.	<i>g [grammar problem]</i>	If people . . . they are sometimes isolated.
7	Slang is useful and familiar to <u>us</u> .	<i>R, no personal pronouns like this</i>	Slang is useful and familiar to people.
8	. . . slang itself, who uses it and where_is used.	<i>O [missing word/phrase]</i>	. . . slang itself, who uses it and where it is used.

Examples 1 and 2 in the table above are related to the mechanics of proper MLA documentation formatting, and considered a local rather than global issue. The local writing concerns presented in Table 2 are relatively minor items, and of a more grammatical- and lexical-level type “that do not impede understanding” of the text (Ferris, 2002, p. 22). However, one of the four criteria which students are evaluated on in the assessment rubrics (both formative and summative versions) is ‘language usage’. Through oral feedback to the class, students were reminded that a preponderance of such local errors in their final drafts would have a negative impact on the reader (me), indicating a less than thorough editing/revising of the final draft and would negatively impact their essay grade.

Upon being given the topics for writing the first drafts of their essays, students were told that in the first draft they should focus on content and organization of ideas, while grammatical and other language use problems could be focused on for the final draft. The written feedback provided to students covered both global and local errors/concerns, but particular emphasis was provided to the more serious writing issues that led to communication problems with the reader. In her examination of error treatment in L2 student writing, Ferris (2002) stressed the need for teachers to prioritize students' most frequent errors, and global errors: "Errors that should receive the greatest attention should include serious ('global') errors that interfere with the comprehensibility of the text" (p. 22). Below we pay particular attention to student attempts to close gaps dealing with these more serious global concerns related to content, organization and which lead to communication breakdown with the reader. In this JC essay, due to the emphasis placed on proper documentation of sources used, and the care needed to avoid suspicions of plagiarism, use of such documentation (or neglecting to use it) was considered a more serious global concern.

Global writing concerns, dealing with conceptual-and structural-level issues, as well as use of MLA documented source material, were particular focus points on the end sheet feedback. This formative assessment rubric addressed such global concerns and aimed to provide a more holistic picture for students of the strengths and weaknesses of their first draft.

In examining the issue of formative feedback and draft revision, McGarrell and Verbeem (2007) write:

... formative feedback (also sometimes referred to as facilitative or intermediate feedback) typically consists of feedback that takes *an inquiring stance towards the text*. Addressing the particular needs of individual writers, it often consists of questions intended to raise awareness of the reader's understanding of the meaning of the text as a means to encourage substantial revision on the next draft. (p. 229, emphasis added)

Such an inquiring stance can be detected in many of the feedback interventions I provided on student first drafts. As reported in Part A, more than half (55%) of my feedback commentary (of 283 comments documented) on student first drafts were coded as 'rhetorical questions'. Hounsell (1995) suggested that feedback commentary phrased as questions rather than directions, would be better received and thus acted upon by students. Responding to my questions was a key challenge to students in revising their first drafts. McGarrell and Verbeem (2007) encourage teachers to provide formative feedback "that seeks to discover and clarify intended meanings" (p. 15). This is also a useful conceptual reference for the

correction code symbols I used, to discover (the ‘?’ symbol) and clarify (the ‘R’ symbol) what students were intending to communicate, and to show that there was a gap between actual and desired communication of ideas. In the feedback process described in this JC class, students were pushed “to refine their intended messages and to consider alternative ways of expressing their ideas” (McGarrell and Verbeem, 2007, p. 229).

In this discussion, ‘a successful response’ to a feedback intervention is considered as one which clarified the writers’ intended meaning, or otherwise improved the content and structure of the text. An unsuccessful response would be considered final draft student writing that did not achieve these results of clarity, or message-refinement, with in the final version of the essay.

In Table 3 below we can see some examples of students ‘closing the gaps’ and successfully revising global problems in their first drafts, to produce clearer and more comprehensible writing (expression of meaning) in the final draft. It should be noted that the ten examples of global concerns shown in Table 3 are taken from ten different student writers.

Table 3. Closed gaps: global concerns successfully revised

	First draft	Feedback intervention	Final draft revision (feedback response)
1	<u>Reliability</u> of slang [essay title]	? <i>Which essay question are you answering?</i>	Negative image caused by slang
2	What are the main reasons why the use of slang is so popular and commonplace? [thesis statement]	<i>A thesis statement should not be a question.</i>	There are three features of slang which affect its popularity.
3	On the other hand, <u>there are many slang disappear</u> . For example, Japanese word ‘nau’ means modern.	<i>R + ‘Use an example from English, not Japanese.</i>	On the other hand, many slang words disappear. For example in England, ‘bob’ was a slang expression for shilling.
4	From these reasons, it is concluded... [concluding paragraph]	<i>Repeat the reasons again here for reader</i>	By examining its origin, character, words and expressions, it is concluded . . .
5	Since it is <u>absolutely free</u> to make and use slang, . . .	<i>What do you mean?</i>	It is true that everyone has a chance to produce and use slang, . . .
6	A possible reason why slang . . . expressions disappear rapidly.	<i>Source? Citation needed</i>	A possible reason why slang . . . expressions disappear rapidly (Ross 105).
7	. . . so people who do not know the background cannot understand.	<i>What background?</i>	. . . so people who do not know the background of the group or subculture or generation cannot understand.
8	Even though there is <u>such risk</u> , people use it.	<i>What risk?</i>	Even though there are such risks as some people not understanding or having a negative impression, people use it.
9	The reason some people have a negative reaction to slang is because slang <u>does not gain its social position</u> .	?	The reason some people have a negative reaction to slang is because slang is not suitable to serious settings.
10	In addition, young people . . . [This section continues for seven lines of text]	<i>This is a somewhat different issue from your main point. You should revise this part. Or delete it.</i>	In brief, slang is informal and can be mean as compared to standard language. [Edited to become one sentence, and off-topic information is removed.]

Table 3 above reflects a point about teacher feedback commentary made by Black et al. (2003) in the context of teacher response to students work in UK schools: “The comments provided the vehicle for personal dialogue with each learner about his or her work to which the learner could respond” (p. 48). The examples provided in this section may be viewed as ‘snippets of dialogue’ in which the teacher is responding to students’ original writing, and then the student responds to the feedback intervention. Through the FI’s provided here, students were told that there were some points of communication breakdown or lack of meaning clarity as a result of what they had written, and therefore signaling the need for a revision of the text. These examples from tables 2 and 3 show students successfully responding to my FI’s with improved final draft writing.

An additional set of 15 examples of successful revision of global issues can be found in **Appendix B1**. As noted, this total collection of 25 examples of successful revisions of global concerns uses all 21 essay sets. These examples are representative of students’ successful improvements to final drafts of the slang essay.

Once again, McGarrell and Verbeem (2007) provide an insightful commentary which describes the feedback process engaged in here, and is worth quoting at length:

Formative feedback . . . is an instructional approach that encourages revision by raising potential questions readers might have about ideas presented in the text, leaving it to the developing writer to make the final decisions about the content. The approach is based on the rationale that writers will be motivated to revise if they are confident that the intended audience will treat their work seriously, will want to know what writers have to say, and will respect their authority as writers to make decisions. Teachers' probing questions about key areas of a text show writers where and what kind of additional information readers might need. Meeting this need prompts writers to go beyond surface level revisions and to rethink their intended meanings at a deeper level of engagement with their texts. (p. 231)

Encouraging such deeper level of text engagement was a hoped for goal for my JC course, promoting student learning and competence as academic writers of English, and better positioned to cope with the senior-year graduation essay. Final decisions about content, as the quote above notes, were left to the developing JC student writer to make.

However, not all final draft revisions were successful and the feedback intervention did not feed forward to students producing writing of better clarity or other improvements. Gaps remained unclosed and therefore problematic. Table 4 below provides some examples of unsuccessful revisions which did not clarify the writers intended meaning or otherwise improve the content or structure of the message. These are taken from seven different student

essay sets.

Table 4. Open gaps: unsuccessful revisions

	First draft	Feedback intervention	Final draft revision (response to feedback)
1	Slang is Considered Corrupt Words [essay title]	<i>R</i>	Slang is Considered Corrupt Words except for specific situation.
2	<u>In the following sentences, some main characteristics are shown.</u> [thesis statement]	<i>R</i>	Therefore, . . . it has some main features and the following sentences show some of them.
3	However, if they use slang that involves rebellion against it, they will not endanger themselves, because the slang is more rebellion of language.	<i>confusing</i>	However, if she uses slang, she can express the same meaning as the word by it because it replaces the word. [Still confusing].
4	... he may feel that <u>he will be extreme</u> and use slang	?	... he may feel that he will take the fashion in advance and use slang . . . [Revision also marked with a '?'.]
5	For these reasons, the use of slang . . . [concluding paragraph]	<i>Repeat the reasons here for the reader.</i>	For these reasons, the use of slang . . . [No change]
6	(Introductory paragraph is four lines long.)	<i>Intro is too short</i>	[No change, virtually the same contents and length.]
7	However, it is notable... (Wikipedia)	<i>Direct quote?</i>	[No change in final version.]

How can such unclosed gaps be accounted for? As noted in the introduction to Part B, teacher commentary has the potential of being “miscommunicated or misunderstood” (Lee and Schallert, 2008, p. 165). Example 7 above is perhaps a good example of me ‘miscommunicating’ the intended message. A clearer, more explicit FI from me would be something like this: *Direct quote? If so, be sure to include quotation marks.* Lea & Stierer (2000) noted that a possible reason for unsuccessful revisions may be the potential discrepancy between the intended meaning of feedback and the student interpretation of it. Student difficulty in making sense of teacher comments has also been noted by Norton and

Norton (2001). It should also be remembered that these JC students are L2 writers with various levels of proficiency, and may not have the writing competence to successfully revise and clarify intended meaning. This factor may help explain the ‘unclosed gaps’ remaining in examples 1, 3 and 4 above. For other reasons, for example, time pressure or consciously choosing to ignore the FI, teacher feedback is not addressed and the message remains unchanged. Examples 5 and 6 in Table 4 above, in which the final and first drafts are the same despite the FI provided, may be a reflection of this fact.

Despite the challenges and difficulties students sometimes had in improving the intended meaning of their messages, more often than not they were able to successfully revise their writing and produce better work for the final draft of the slang essay. While difficult to quantify, considering the more than 800 feedback interventions provided on student first drafts, on the whole, revisions linked to the feedback provided did lead to text improvement in the final draft. This supports empirical research that feedback in multiple-draft classrooms does lead to improved student writing (for example in Ferris, 1997; Ferris, 2006, Hyland and Hyland, 2006, Lee, 2008). As already mentioned, JC students were serious about improving academic writing skills and becoming better writers. They worked hard to produce the best possible final drafts for the graded summative assessment. While there are examples of ‘unclosed gaps’ in the student essay data set, without a doubt the 21 essays improved through the drafting process, sometimes significantly so, through students effective editing and revising practices in responding to the feedback provided.

The feedback interventions I provided were signals and signposts for first draft editing and revising, but it was the student engagement with the feedback, ‘the crucial variable’ as Gibbs (2006) wrote, that produced better final drafts. Active student engagement was essential in closing the gaps indicated by the FI’s provided. Taras (2002) captures well the key AfL idea that the active involvement of students is required for true learning from feedback to occur:

[Feedback] does not count as formative feedback unless the student has understood what the purpose of the assessment was, how it was assessed or judged, and how they can use their shortfalls in the future. Even this is not sufficient. Formative feedback is not complete until the students have produced an equivalent piece of work where the issues have been addressed and remedied. (p. 506)

In terms of the JC class discussed in this chapter, students: (1) understood the assessment purpose; (2) knew how their work would be judged; (3) were provided with feedback indicating shortfalls and how to strengthen their first drafts; and (4) they produced a related

piece of work (the final draft) where issues were usually addressed and remedied for a better final product. The first draft feedback provided to students was usually successfully acted upon, feeding forward to future work- an improved final essay draft.

Despite some weaknesses in both my feedback responses and students attempts at revision, the data set collected and examined shows the feedback did feed forward and helped students successfully improve their first drafts. As such, this leads to an affirmative answer to the research question posed in the introduction of Part B as to whether gaps were closed in the quality of student drafts in response to the feedback provided.

While the formative assessment of student first drafts were not graded (following recommended AfL practice), and therefore cannot provide a direct point of comparison, it may be useful here to briefly record the students' summative grades for the slang essay. These grades are provided below in Table 5. The reader is reminded that the criterion-referenced summative assessment is determined based on the four core criteria of content, organization, language use, and use of source material (see Appendices A1, A2).

Table 5: Summative grading breakdown for slang essay final draft (N=21)

A+	A	A-	B+	B	B-	C+	C	C-	D	F
0	6	8	2	3	1	1	0	0	0	0

As a class, the slang essays received quite good grades. The formative feedback provided to students, and (just as importantly) their active engagement with it to close the gaps between the quality of first and final drafts, were key factors in the generally high final grade results achieved.

Section 2: Feedback survey - student views of teacher feedback

In the feedback literature, there is a body of research related to student views, primarily from one-off surveys concerned with student preferences and expectations. But as Lee (2008) notes much of this research is decontextualized, and “there have rarely been any attempts to link student reactions to actual teacher feedback in specific contexts”(p.145). This report of

student perceptions does deal with actual feedback in a specific context, and these views come at the end of a year-long JC course in which students received formative feedback (and summative grades) on five different essays.

In the final class of the Junior Composition course, I administered a student survey about the first draft feedback program in place. During this class, students also submitted the final draft of their last essay for summative assessment. This was to be graded and picked up by them at the departmental office one week later. Following the typical writing/feedback cycle for the course, the first draft of this final essay would have been returned to students in the penultimate class. So, in the week before completing the feedback survey, students had engaged in their last cycle of responding to teacher feedback in the JC course.

In completing the survey, which only deals with first draft feedback, students are reflecting back throughout the academic year and are provided with the opportunity for expressing their views not on individual first draft feedback, but on the feedback process as a whole and their views of it.

With these contextual factors in play, the student survey was administered at the end of the last class, and took students about 10 minutes to complete. The survey can be found in **Appendix B2**.

Questionnaires, such as the one used here, can be useful for tapping into students' preferences, attitudes and judgments. But, as Hyland (2002) notes, it is important to remember that surveys only give reports of what people say they think or do but are not direct evidence of it. Nevertheless, the survey did reveal some broad trends and perceptions, and was effective in gauging the first draft feedback perspectives of 23 student writers in the JC class.

The 12-item feedback survey asks students about such things as how they viewed and responded to the teacher feedback provided. The feedback survey was comprised of two main parts, followed by a third open-ended part for any additional comments about the feedback process. The survey items were devised after doing some preliminary background reading about the feedback issue, and prominent related ideas in the literature (for example, understanding feedback, use of correction codes, emotional reactions to feedback). In the survey results presented here, I simply use descriptive measures of response numbers and percentages to identify general features of student attitudes toward the feedback process.

In both parts of the survey, two response formats are used. One format presents a

statement and asks for students' level of agreement, with a Likert scale of four options available along a response continuum (*agree, tend to agree, tend to disagree, disagree*). A neutral option is purposely not included in this scale. Including a mid-point 'undecided' or 'unsure' choice in a Likert scale can cause some difficulties. As Frary (2003) points out, we cannot be sure that a respondent choosing a middle scale position has a neutral opinion. They may choose this option for other reasons, including uncooperativeness (not wanting to go through the trouble of forming an opinion) or reluctance to answer (not wishing to display his/her true opinion). The four option format used here enables students to take a position, while allowing them to express some degree of reservation (*tend to agree/disagree*) if they wish to do so. The other format used in the survey, a five-point Likert scale, simply asks students to indicate a frequency response on an *always - never* continuum. The one survey item which differs from these two formats is the final item (# 12). This asks the student to rate the first draft feedback system on a four point continuum from *poor* to *excellent*.

Survey: Part 1

Part 1 of the survey sought to ascertain student views of the correction code used and the teacher commentary provided. Table 6 below presents the survey items and student responses. The first three items in part one deal with the correction code symbols used, while the remaining items and responses documented in this table focus on the issue of teacher written commentary. Responses in this table (as well as in the following Table 2) are simply presented numerically, followed by a percentile breakdown.

Table 6. Survey Part 1: correction code, teacher comments (N = 23)

1. I understand why the teacher used a correction code rather than just fixing my essay problems/mistakes.	agree	tend to agree	tend to disagree	disagree	
	16 (70%)	6 (26 %)	1 (4 %)	0	
	always	usually	sometimes	seldom	never
2. From the correction code used, I could understand the type of errors/problems in the essay.	9 (39%)	13 (57%)	1 (4%)	0	0
3. I had trouble knowing how to fix problems/mistakes marked with a correction code.	1 (4%)	4 (17%)	14 (61%)	4 (17%)	0
4. Teacher comments on the first draft were clearly written and easy to read .	9 (39%)	11 (48%)	3 (13%)	0	0
5. The meaning of comments was clear and easy to understand.	6 (26%)	14 (61%)	3 (13%)	0	0
6. Any negative comments about my writing made me feel a bit upset or hurt.	0	1 (4%)	4 (17%)	7 (30%)	11 (48%)
7. The feedback end sheet helped me understand the strong and weak parts of my first draft.	agree	tend to agree	tend to disagree	disagree	
	13 (57%)	10 (43%)	0	0	
8. Generally, the number of teacher comments written on essay first drafts were: too many = 1 (4%) appropriate = 21 (92%) too few = 1 (4%)					

In item 1 in the survey, the majority of students (16 of 23) revealed a clear understanding of why a correction code was used, rather than me directly correcting their errors or other problematic aspects in the first draft texts. I usually reminded students of the rationale for

such a feedback methodology when returning first drafts, presenting, in student-friendly terms, Ferris's (2002) contention that such indirect feedback "is more helpful to student writers in most cases because it leads to a greater cognitive engagement, reflection, and guided learning and problems solving" (p. 19).

As noted in Part A, Gibbs and Simpson (2004) highlight the importance of feedback being understandable to the student. Survey item 2 reported that 22 of 23 students could either *always* (39%) or *usually* (57%) understand the type of error or problem indicated from the correction code used. As noted earlier, students became quite familiar with the correction code feedback symbols, and types of concerns they represented, throughout the JC course.

However, recognizing that a problem exists and fixing it ('closing the gap') are not the same. In terms of closing the gaps between the actual performance (first draft) and the desired performance (final draft), survey item 3 is an important one. A combined 78% of students (18 of 23) indicated that they either *usually* (17%) or *sometimes* (61%) had trouble knowing how to fix problems with their writing. One student reported *always* having trouble knowing how to fix problems indicated by correction code symbols and 'close the gaps'.

Again, the context of an L2 writing class must be kept in mind here, and one of the reasons why teaching writing is so challenging is that most classes contain a mixture of students (Kroll, 2001). This mixture of writing skills and English language proficiency was evident in this JC class with some students showing weaknesses in both areas as they "struggle to make meaning in a foreign language" (Hyland and Hyland, 2006, p. 207). In a finely crafted phrase, Pintrich and DeGroot (1990) tell us that; "Pupils need to have both the 'will' and the 'skill' to be successful in classrooms" (p. 53). While all JC students had the motivation (will) to become better essay writers (as they consider the looming senior year graduation thesis), their writing ability and English language proficiency (skill) did vary. Some students were more skillful than others in revising and reshaping their texts, negotiating and producing clearer meaning messages for the reader. Yet, as stated in the previous section, on the whole, students were able to make successful revisions on the final draft of their essays, a fact reflected in the generally good grade breakdown depicted at the end of the previous section. It should also be remembered from Part A that an average of 20 correction code symbols was provided per essay, and some degree of difficulty in successfully responding to these could be expected in even the more proficient writers in the JC class.

Items 4-8 on the survey deal with the teacher written commentary provided in the body of student first drafts. In item 4, a total of twenty students reported that teacher comments were

always (9 students) or *usually* (11 students) clearly written and easy to read. This reported degree of my commentary being understandable is good news, and somewhat surprising given the sheer number of comments provided on first drafts, an average of 13 per essay. Three students reported that clear commentary was only *sometimes* provided, indicating that perhaps comment clarity is something I need to pay more attention to in responding to student writing.

The reader is reminded again here that feedback was provided on these students first drafts prior to me determining that this particular data set would be used for any feedback research. Thus my feedback was not at all influenced by the research process itself.

Related survey item 5 deals not with the clarity of presentation of my feedback commentary, but the meaning the comments intended to convey about closing the gaps and fixing the indicated problem. A similar total percentage in the always/usually categories, 87%, was reported as for item 4, with 20 students saying that they could *always* (6 students) or *usually* (14 students) easily understand the meaning of the comments that I wrote on their first drafts. Again, three students reported *sometimes* having problems understanding feedback comment meaning. As mentioned, feedback is ‘an act of communication’ and item 5 responses signal to me that there is room for improvement in providing comments which are “consistent, clear, helpful, and constructive” (Hyland and Hyland, 2006, p. 223).

Survey item 6 attempted to tap into some of the affective elements at play in the feedback process by inquiring whether students felt ‘a bit upset or hurt’ by any negative comments I provided on their first drafts. Less than half (48%) responded *never* here, indicating that 12 students had experienced some degree of such upset or hurt feelings. A combined 21% of students reported *usually* (1 student) or *sometimes* (4 students) having such reactions. Responses to this item are a reflection of the fact that assessment is a process rife with emotion and because students put their time and themselves into the assessment tasks we set, the feedback we provide engages them on an emotional level (Lee, 2007). Hyland and Hyland (2006) write, “Our comments can transform students’ attitudes to writing and lead to improvements, but our words can also confuse and dishearten them” (p. 223). While I need to remain aware of the potential damaging effects of feedback commentary, the number of students reporting *never* or *seldom* feeling upset or hurt was 78%.

A key component of effective feedback is letting students know the main strengths and weaknesses of the draft they have written. Among the feedback processes described here, the feedback end sheet in particular was intended to show these characteristics of the first draft. In survey item 7, thirteen students (57%) reported that the end sheet helped them understand the

strengths and weaknesses of their first draft. Ten students ticked the *tend to agree* response here indicating some degree of reservation and perhaps showing that the end sheet, and my commentary on it, could have done a better job of doing this. However, no students disagreed with item 7 and this fact reveals a degree of effectiveness for the end sheet in the feedback framework for the course.

In calculating the amount of feedback commentary provided to students, I noted with surprise that some essays contained from 15 to 20, or more, of my written comments (as noted, the mean average was 13). I began to wonder whether I was providing excessive feedback, something that can be as much of a problem as too little response. However in item 8 on the survey sheet, only one student reported the number of comments as being *too many*, with the same singular response to the *too few* comments option. Responses showed that 92% of students indicated that the number of teacher feedback comments provided was *appropriate*. Hyland and Hyland (2006) reported that “ESL students, particularly those from cultures where teachers are highly directive, generally welcome and expect teachers to notice and comment on their errors and may feel resentful if their teachers do not do so” (p. 3). This cultural element may be at play here in this Japanese university classroom context, but, in any case, the responses to this item show that for almost all JC students the quantity of feedback written commentary was neither excessive nor deficient.

Survey: Part 2

In this section students were asked to express their views on the first draft feedback system as a whole. The survey sheet reminded students that the ‘feedback system’ includes all three elements; correction code, teacher comments, and feedback end sheet. Four items (numbers 9-12) are included in this section and student responses are presented below in Table 7.

Table 7. Survey Part 2: feedback system overall (N = 23)

	agree	tend to agree	tend to disagree	disagree	
9. The feedback system made it clear what I needed to do to improve my first draft.	15 (65%)	8 (35%)	0	0	
	always	usually	sometimes	seldom	never
10. The feedback system gave me enough advice in how to write a better final draft.	8 (35%)	13 (57%)	2 (8%)	0	
11. I ignored (did not use) first draft feedback when editing/revising for the final draft.	0	0	2 (8%)	7 (30%)	14 (61%)
12. In terms of giving clear, useful feedback for helping to write a better final draft, the first draft feedback system for this Junior Composition course was:	poor	average	good	excellent	
	0	0	9 (39%)	14 (61%)	

With regard to closing the gap between the first draft performance and the desired final draft, item 9 is important in asking students if the feedback system made it clear what they needed to do to improve the first draft (i.e. close the gap). While 65% agreed, 8 students seemed to express some reservations here, opting for the *tend to agree* response. This seems to indicate that there is room for improvement here in terms of my feedback clarity and specificity with regard to making improvements. However, none of the 23 students expressed disagreement with item 9, indicating that all students, at least to some degree, knew what they needed to do to ‘close the gaps’.

Item 10 on the survey is a continuation of this theme, asking students whether the feedback system provided enough advice for closing the gaps and writing a better final draft. Only 35%

of students (8 of 23) agreed that they received enough advice about how to write a better final draft. More than half (13 students) chose the *tend to agree* response, while two students disagreed with the statement. Perhaps the message here is that, while I am providing plenty of advice about what the gaps are, students seem to be reporting that the teacher feedback could do more to tell them how to close the gaps. One student commented as follows on the survey: “*I think first draft feed back is very useful for me to improve my essay and to make it better, but I want you to give me more detailed advice for improving*”. Perhaps there is room for more clarity to individual students about the specifics of improving their first draft, but there is an obvious conflict here between student numbers, the amount of feedback provided (three different types in this system) and the specificity of diagnostic advice. Prioritizing specific advice for making improvements to the first draft may be a possible solution here.

Item 11 asks student if they ever ignored my feedback when editing and revising the first draft. Almost all students (91%) responded *never* or *seldom* here, providing an indication of student earnestness in closing gaps and producing better final drafts. Two students responded that they sometimes ignored feedback, presumably because it involved miscommunication, they did not know how to make a revision, or even being pressed for time in completing the final draft. Hyland (2003) provides some useful commentary about students’ desires and uptake of teacher feedback:

It is also important to note that what individual students want from feedback-and the use they make of it-varies considerably. Some students want praise, others see it as condescending; some want a response to their ideas, others demand to have all their errors marked; some use teacher commentary effectively, others ignore it altogether. It can be difficult for teachers to cater to all these different perceptions and expectations. (p. 180)

The difficulties in catering to all students’ needs and wishes to everybody’s satisfaction should be noted when considering student responses to feedback; and indeed when considering their responses to this feedback survey.

The final survey item asks students to offer an overall assessment of the JC first draft feedback system in terms of giving clear, useful feedback to write a better final draft. Of the 23 students survey, 14 (61%) considered it to be *excellent*, while the remaining 9 students (39%) evaluated it as a *good* system. This response is encouraging, and offers some supporting evidence from the students’ perspective for my strong evaluative rating in the Seven/seven framework for teacher self-assessment reported in Part A. The key argument in Sadlers’ (1989) view of feedback, and fundamental to AfL, is that the power of feedback

comes from closing the gap between where the students are and where they are aiming to be. Responses to survey item 12 shows that, while there may be room for improvement, students feel that the feedback system is a very good one; a powerful feedback response framework in enabling them to close the gaps in their writing and produce their best work for the final drafts of the JC essays.

Part 3 of the survey asked if students had any additional comments. Only 5 of the 23 students made a brief comment, including the one recorded above in the discussion of item 10. There was nothing particular noteworthy in the commentary, all included something to the effect that the feedback system was ‘good’, ‘useful’ and ‘effective’.

According to Lantolf and Pavlenko (2001) students are active agents in the feedback process and “construct the terms and conditions of their own learning” (p. 145). The 23 students in this Junior Composition class worked hard in constructing their own learning and endeavoring to become better academic writers in English. Hyland (2002) reminds us that, fundamentally, writing is learned rather than taught. While the feedback system in place provided signals of gap identification and guidance in closing them, it was primarily the student’s efforts that resulted in gaps being closed and better-written final essay drafts being produced.

Conclusion: Feedback for learning

In *Rethinking Foreign Language Writing*, Scott (1996) observes that “Writing is clearly not a simple act, but rather an intricate set of steps and choices” (p.31). To this perceptive statement we can also add that neither teacher provision of feedback or student response to it are simple acts. As this chapter and the research it describes reiterates, a formative feedback process also involves ‘an intricate set of steps and choices’ by both the provider and receiver of these informational responses to the work produced. Cohen, Boud, and Sampson (2001) capture some of the intricacies of this process as follows:

The giver of feedback has to identify what constitutes good work in a given subject area and express these ideas in a coherent form, while also communicating effectively with the prospective receivers about their feedback needs. The recipient of feedback benefits from identifying and articulating these needs, . . . and from responding appropriately to the feedback. (p. 249)

The communicative dialogue that is often produced in this teacher-student ‘feedback waltz’ is

a fundamental element of the feedback for learning that is generated when the process works well.

This two-part thesis chapter examined the feedback component of the AfL wheel, through the close examination of a set of 21 student essays (first and final drafts) in a third year academic writing course called Junior Composition (JC). These 21 essay sets are representative of the feedback provided for the series of five essays written and responded to throughout this year-long course.

In Part A, teacher self-assessment of written feedback, I engaged in a systematic evaluation of my feedback practices in light of AfL feedback theory, and in particular against two recent publications regarding recommended feedback provision practices. Creating and using an evaluative tool named ‘the Seven/Seven framework for feedback assessment’, based on Gibbs and Simpson (2004-05) and Nicol and MacFarlane-Dick (2006) my teacher feedback provided on 21 first drafts of student essays was closely inspected. After documenting and coding a total of 828 feedback interventions, and considering the additional feedback end sheet attached to each first draft, this TSA of written feedback in the JC course concluded that my written feedback practices were self-assessed positively overall, and considered effective in generating ‘feedback for learning’ with this particular group of students.

Part B of this chapter examined the feedback issue from the student side of the coin. This part focused on two issues: whether the feedback provided feeds forward to the ‘closing of gaps’ in student writing and the production of an improved final draft; and student perceptions of the first draft feedback framework implemented in JC. In examining whether student achievement levels were being enhanced by the feedback, the first and final drafts of student essays were compared to examine whether gaps were being closed. While examining student responses to the more than 800 feedback interventions documented showed examples of remaining problems in the final drafts (‘unclosed gaps’), generally speaking, student final drafts were stronger, better and more clearly written due to student response to the feedback provided. The first draft feedback survey, completed by 23 students, also showed an overall high level of satisfaction with the JC feedback framework, despite some indications that the feedback provided has room for improvement (for example, in more detailed advice in how to make improvements). This encouraging student survey response to the feedback system in place provides an alternative and supporting view of the positive TSA results from Part A of this chapter.

In their seminal review of the formative feedback literature, a publication which served as a catalyst for the ensuing AfL movement, Black and Wiliam (1998a) wrote,

All [classroom] work involves some degree of feedback between those taught and the teacher, and this is entailed in the quality of their interaction which is at the heart of pedagogy. (p. 16)

This notion of quality feedback and teacher-student interaction being at the heart of effective pedagogy helps explain the importance placed on feedback within AfL (and its central place in this thesis). At the beginning of Part B of this chapter, Lee's (2007) four essential conditions needed for feedback to promote assessment for learning in a writing class were presented. They include: providing students with information about strengths and areas for improvement in their writing; communicating information to students in a clear, intelligible manner; providing students with opportunities to act on feedback; and students being active managers of their own learning. The feedback framework and processes described in this chapter meet these four essential conditions, revealing an AfL-centered course in which all participants collaborate and learn together by "using feedback as a pedagogical tool for improving the teaching and learning of writing" (Lee, 2007, p. 180).

While this chapter begins with an exercise in the teacher engaging in self-assessment, we continue this important AfL theme of reflection and self-monitoring from the students point of view as we move into Chapter 2.

Appendix A1 Essay final draft assessment rubric (front page). Adapted from McCormack & Slaght, (2006)

Class: Junior Composition

Instructor: Eddy White

Essay evaluation - final draft

Student name: _____

<u>Assessment criteria</u>	<u>Grade</u>
1. Contents	
2. Organization	
3. Language use	
4. Use of source material	

Overall grade	
----------------------	--

Comments:

Appendix A2 Essay final draft assessment rubric (back page). Adapted from McCormack & Slaght, (2006)

1. Contents

Clearly focused content, relevant to the essay topic. Length, scope and level of detail are appropriate/relevant. Arguments are well presented and developed with supporting evidence from a variety of sources. It is evident that the writer knows the topic well.	A
Generally well-focused content. May be lacking in level of detail or development of ideas and/or limited in scope (which may affect length). Much of the content describes rather than critically analyzes. Arguments/main ideas may be inconsistent or insufficiently developed.	B
At times, essay focus may be lost; some content may be irrelevant. Clearly limited in level of detail, superficial treatment of subject with no development of ideas. Shows lack of knowledge of the topic. May be very short. Little or no evidence of evaluation of ideas, mostly at level of describing. No clear argument/thesis evident.	C
No obvious focus; clearly content inadequately researched; unable to deal with topic (probably very short) or widespread plagiarism has made it impossible to assess true level of the essay. Too much personal/anecdotal material	D

2. Organization

Overall structure and main ideas are clearly organized and easy to follow. Introduction has general topical sentences, followed by a well-written thesis statement. In the body of the essay, supporting ideas are effectively linked together and “flow” coherently making it easy for the reader to follow. Conclusion summarizes main points and effectively brings the essay to a finish.	A
Overall structure and main ideas are generally easy to see. Introduction and conclusion are appropriately linked to the main body. At times there may be a tendency to move from one idea to another with no attempt to link them.	B
Difficult for reader to determine overall structure/identify main ideas. May be due to poor language control (i.e. grammar, vocabulary), which also affects cohesion. Introduction/conclusion may be inadequate. Frequent move from one idea to another with no attempt to link them.	C
Ineffective attempt to organize the essay. Very difficult for the reader to follow the text. The introduction fails to give the reader an overview/clear idea of what will follow or widespread plagiarism has made it impossible to assess true level of the essay	D

3. Language Use

Ideas are clearly expressed, with wide, accurate usage of vocabulary and grammar. Any errors do not interfere with communication. Formal academic style used(e.g. formal expressions, longer sentences, impersonal tone, etc.) Use of a range of linking words and phrases to join ideas at paragraph and sentence level.	A
Ideas are usually clearly expressed. Linking of ideas within paragraphs generally appropriate, but at times may be lacking between sections. Some vocabulary and/or grammar problems, but generally do not interfere with communication. Spelling and or punctuation errors do not interfere with comprehension.	B
Some ideas are simply expressed, but others are not clearly expressed. Linking between and within sentences may be inconsistent. Fairly serious vocabulary and/or grammar problems; can interfere with communication. Spelling and/or punctuation may be seriously flawed.	C
The level of vocabulary and grammar is so consistently weak that the end product fails to achieve its purpose due to ineffective communication (or widespread plagiarism has made it impossible to assess the true level of the essay).	D

4. Use of source material

Effective use of a range of sources. These are appropriately incorporated in the body of the essay through paraphrase/quotation/summary. Shows ability to synthesize well from several sources to support ideas. Works Consulted page and in-text referencing follow MLA conventions and a range of sources are used. No obvious/conscious plagiarism.	A
Effective use of sources, mostly when summarizing/paraphrasing ideas clearly. Shows some evidence of synthesis of information. Works Consulted and use of sources show an understanding of the concept of referencing, though this is not always followed (e.g., not in alphabetical order, name of publisher missing, in-text references include first name, etc.). No obvious/ conscious plagiarism.	B
Limited sources used, and summary/paraphrase of ideas not always clear. Some attempt at synthesis of ideas. Clearly has problems writing a Works Consulted page and incorporating in-text sources in an appropriate way, although there is some attempt to do this. Poor language control (grammar, vocabulary) may be a factor. Suspicion of plagiarism in some sections.	C
Inadequate attempt to use source material, e.g., may only use one source or none. Content based mainly on student's views with little or no evidence to support it. Shows little understanding of the importance of referencing and academic writing conventions. No Works Consulted page, or, where this exists, does not follow appropriate MLA format.	D

Appendix A3 Self-and peer evaluation checklist for essay first draft

Junior Composition

Self / Peer Evaluation essay checklist

A. Take a close look at the first draft and write **Y (yes)** or **N (no)** after each one of these sentences. If you are not sure, write a question mark (?) .

Introduction	
1. The introduction has some general sentences which provide background to the topic	
2. There is a clear thesis statement.	
Body	
3. Each paragraph has a clear topic sentence, which is connected to the thesis statement.	
4. Each paragraph has supporting details or information to back up the main idea.	
5. There is no unnecessary, or irrelevant information included.	
Conclusion	
7. The conclusion restates the ideas of the thesis statement (in different words).	
8. The main ideas in the body of the essay are summarized.	
9. The final sentence is a good one, and effectively brings the essay to a finish.	
Use of sources	
10. Each body paragraph has in-text citations showing sources which have been paraphrased, summarized or quoted.	
12. A Works Cited (WC) page has been included at the end of the essay.	
13. The WC page has at least 3 references, which tell the reader the source for all in-text citations	
14. MLA format has been closely followed for the in-text citations and the WC page.	

B. Strong points of essay first draft:

1. _____
2. _____

Weak points of essay first draft:

1. _____
2. _____

Appendix B1 Closed gaps: additional global concerns successfully revised (N = 15)

	First draft	Feedback intervention	Final draft (response to feedback)
1	To show the authority of the word is to show how the word is prosperous.	?	Deleted from final draft. [Oral class feedback included my teaching point that deleting something can sometimes be a better writer decision than attempting revision.]
2	Therefore slang is a good means to express their thinking, feeling, and <u>insistence</u> freely.	?	Therefore . . . thinking, feeling and emotion freely.
3	In other words, the attraction of slang and the <u>inseparable connection</u> between people . . .	<i>What is this connection?</i>	As stated in this essay, the attraction of slang and its capacity to strengthen the solidarity among group members . . .
4	Because of <u>slang's fertility</u> , people . . .	<i>What does this mean?</i>	Slang has a large amount of words, so it is possible to express one subject in several ways. People . . .
5	Slang has mainly four characteristics which cannot see in the Standard one. [thesis statement]	R	There are mainly four characteristics for slang that shows its essence.
6	Finally, <u>slang is made up of many purposes of usages</u> .	R	Finally, slang has some important characteristics.
7	Slang is over flowed in the world.	R	Generally, slang is widely used by many people.
8	The second one is that it is informal, <u>alive</u> , and rebellious. [Paragraph topic sentence]	<i>Second what?</i> <i>+ Meaning? (alive)</i>	The second characteristic is that it is informal and rebellious. ['Characteristic' has been inserted. 'Alive' deleted.]
9	Therefore, the use of slang <u>may give interest to our communication</u> .	R	That is to say, it makes communication more interesting.
10	In this way, slang is a kind of <u>trending</u> . . .	? <i>Is this a word?</i>	In this way, slang is a kind of popularity that people use . . .
11	In conclusion, slang is <u>flourished by people</u> . People use it for many reasons.	?	In conclusion, people use slang for many reasons.
12	As has been mentioned, it was marijuana smoker who use the slang word 'weed'.	<i>Why mention this again here?</i>	[Deleted]
13	The most important reason is that the use of slang can give people the opportunity to <u>Maintain their belongings</u> .	<i>Meaning what?</i>	The most . . . slang can give people the opportunity to enter a group and make a bond between their group members stronger.
14	There are <u>some reasons</u> why the use of slang . . . [thesis statement]	<i>How many will you discuss? Tell the reader</i>	There are three main reasons why the . . .
15	By understanding the meanings of slang members can <u>make their community easy</u> and deepen familiarity.	?	By understanding the meanings of slang, members can make their communication easy and...

Appendix B2. End of course feedback survey

Junior Composition questionnaire: First draft feedback

As part of my research, I am interested in ways to make teacher feedback on the first drafts of student essays more effective. Student views on this process will be very helpful in giving better feedback on essay drafts in the future.

During this Junior Composition course you have written five essays and were given feedback on the first drafts of each one. The first draft feedback you received was of three kinds:

1. The use of a correction code to point out problems, errors (ex. R=rewrite, WW= wrong word, etc.)
2. Teachers comments/questions in the margins, etc. of your essay (ex. “Confusing, what do you mean?”, “You need to give a source here” , etc.)
3. First draft feedback sheet attached to your essay. This sheet included information about how strong/weak the main elements of the essay were (content, organization, language use, use of source material). The sheet also usually contained some end comments about how good the draft was overall and main areas to work on.

Please respond to the following statements about first draft feedback. Do not write your name on this paper and be as honest as possible in your responses.

Circle the answer that best matches with your views or feelings.

Part 1: Correction code, teacher comments

1. I understand why the teacher used a correction code rather than just fixing my essay problems/mistakes.

agree tend to agree tend to disagree disagree

2. From the correction code used, I could understand the type of errors/problems in the essay.

always usually sometimes seldom never

3. I had trouble knowing how to fix problems/mistakes marked with a correction code.

always usually sometimes seldom never

4. Teacher comments on the first draft were clearly written and easy to read.

always usually sometimes seldom never

5. The meaning of comments was clear and easy to understand.

always usually sometimes seldom never

6. Any negative comments about my writing made me feel a bit upset or hurt.

always **usually** **sometimes** **seldom** **never**

7. The feedback end sheet helped me understand the strong and weak parts of my first draft.

always **usually** **sometimes** **seldom** **never**

8. Generally, the amount of teacher comments written on essay first drafts were:

too many **appropriate** **too few**

Part 2: Feedback system overall

[Feedback system includes all three elements; correction code, teacher comments, feedback end sheet]

9. The feedback system made it clear what I needed to do to improve my first draft.

agree **tend to agree** **tend to disagree** **disagree**

10. The feedback system gave me enough advice in how to write a better final draft.

agree **tend to agree** **tend to disagree** **disagree**

11. I ignored (did not use) first draft feedback when editing/revising for the final draft.

always **usually** **sometimes** **seldom** **never**

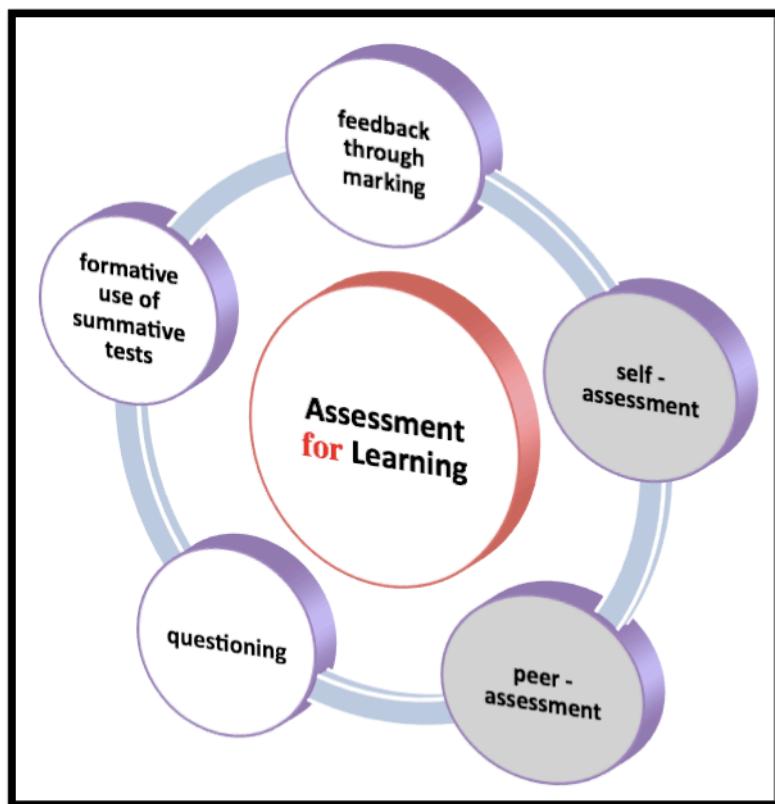
12. In terms of giving clear, useful feedback for helping to write a better final draft, the first draft feedback system for this Junior Composition course was:

poor **average** **good** **excellent**

Part 3. Final comments (or suggestions to improve the first draft feedback system)

Chapter 2: Student as Assessor

- **Part A:** Self-assessment for learning in Communicative English course
- **Part B:** Peer-assessment for learning and student presentations



Preview

In this chapter, the focus is on the student in the assessor role. First, Part A describes a communicative English course where students self-assess their active participation as learners in the classroom community. This study is followed in Part B by an investigation which focuses on students assessing the performance of their peers. One key difference between the two research projects reported on here is that the self-assessment framework does not incorporate the self-scoring into students' summative course grades. On the other hand the peer-assessment model used does do this, and, as we shall see, the issue of formative and summative tensions that are often part of the reality of classroom-based assessment come into play.

The self-assessment (SA) part of this chapter is a pilot study which reports on the use of a self-assessment of class participation procedure used in three freshman classes. In judging its effectiveness and potential use for future courses, the self-assessment process is itself assessed according to the five principles of practicality, reliability, validity, authenticity, and washback. After implementation with approximately 70 students in three communicative English classes, usage and student feedback point to the effectiveness of the self-assessment instrument, particularly as a consciousness-raising tool, in promoting more class participation. However, the main potential problem with self-assessment, reliability of student scoring, is also apparent in the pilot study.

The peer assessment (PA) research reported on in Part B was conducted to determine student feelings about this type of student-centered assessment procedure, and whether it was useful in promoting more effective learning. Set in a Public Speaking course, this investigation reports on a PA framework in which 30% of students' final course grades was comprised of peer assessment scores of oral presentations. Based on survey responses, student perspectives on using peer assessment were positive, on the whole, and the process did indeed promote student learning.

Part A. Assessing the assessment: an evaluation of a self-assessment of class participation procedure

What successful learner has not developed the ability to monitor his or her own performance and to use the data gathered for adjustments and corrections? (Brown, 2004, p. 270)

Introduction

Like the effective use of feedback, student self-assessment (SA) is also viewed as fundamental to AfL. According to Black and Wiliam (1998b), “self-assessment by pupils, far from being a luxury, is in fact an essential component of formative assessment” (p.10). As we have already seen from the ARG (1999, 2002) sources used in our discussion of AfL fundamentals in the introductory section of this thesis, much emphasis is put on developing student capacity to assess themselves and their work so they can understand how to improve. The reader is reminded of the earlier-noted James et al. (2007) three-stage description of SA as a cognitive stepping back from the learning process, reflecting, and then stepping back in. With the pilot study described in this chapter, we see a living example of students engaging in this meta-cognitive process of becoming more self-regulated learners (Nicol & Macfarlane-Dick, 2006)

In the past 25 years or so SA has become a more advocated and widespread assessment option, both in mainstream education and in English language teaching. It has generated quite an extensive body of research and been a prominent area of inquiry and discussion, particularly in the areas of learner autonomy and language testing (Benson, 2001). In language learning contexts, SA has primarily focused on issues of proficiency, ability and task-performance (see, for example, Alderson & Banerjee, 2001; Dochy, Segers, & Sluijsmans 1999; Oscarson, 1997; Ross, 2006). For the purposes of this thesis, SA may be defined as “any assessments that require students to judge their own language abilities or language performance” (Brown, 1998, p. 53), and “provides an approach in which learners typically rate themselves according to a number of criteria or dimensions” (Bachman, 2000, p. xi). Cassidy (2007) comments that, for students, self-assessment is defined by the acceptance of responsibility for their own learning and performance.

This research report describes a pilot study regarding self-assessment of class participation set

in three oral communication classes at TWCU. It examines the effectiveness of the SA procedure used with regard to five fundamental principles of assessment: practicality, reliability, validity, authenticity, and washback. In this investigation, students in communicative English classes used a performance-based assessment instrument to self-assess their degree of active class participation.

AfL, reflecting constructivist learning theory, emphasizes active student participation in the learning process. According to Glassen (2009):

Self-assessment – including self-evaluation, self-regulation and self-monitoring – is very *active* learning. It can not occur in a classroom where the teacher maintains control of learning, and it cannot occur in a classroom where students are willing to be passive recipients of what the teacher offers them. It is hard work, but it is worth it. (p. 93, original emphasis)

A primary responsibility for teachers is to “engineer an effective learning environment” (Black, Harrison, Lee, Marshall, & William, 2002, p. 20). This is premised on the idea that for learning to be effective, this active involvement of students is essential. Encouraging such active participation can sometimes be problematic for language teachers, especially in mandatory courses which may include students with little need or desire to improve their spoken proficiency in the target language. The use of a student self-assessment procedure is one possible way to encourage active class participation and maximize learning.

Class participation and language learning

Evaluating students’ class participation may be regarded as “assessing the quality of a students’ non-academic performance within a subjective criteria” (Shindler, 2003, p. 20). An examination of motivation in the classroom by Skinner and Belmont (1993) discusses *student engagement* and their operationalization of this term offers a close approximation of how class participation is considered in this study:

Engagement versus disaffection in school refers to the intensity and emotional quality of children’s involvement in initiating and carrying out learning activities . . . Children who are engaged show sustained behavioral involvement in learning activities accompanied by a positive emotional tone. They . . . initiate action when given the opportunity, and... they show generally positive emotions during ongoing action, including enthusiasm, optimism, curiosity and interest. The opposite of engagement is disaffection. Disaffected children are passive, do not try hard, and give up easily in the face of challenges . . . [they can] be bored, depressed, anxious or even angry about their presence in the classroom; they can be withdrawn from learning opportunities or even rebellious towards teachers and classmates. (Skinner &

Belmont, 1993, p. 572).

For learning to take place, students need to exert effort and be engaged and involved as active participants in the learning process. This is especially true in the communicative language learning environment where the target language is both the object of study and the medium of student interaction.

The English language course discussed in this study uses a communicative teaching methodology focused on maximizing student speaking time through pair/group work. The rationale for such student interaction is rooted in the *interaction hypothesis* (Long, 1996), which posits that acquisition of language will occur as learners actively engage in attempting to communicate in the target language. According to Allwright (2000): “interaction is language learning . . . It is not merely the process whereby learned linguistic knowledge is practiced, but rather the process whereby linguistic knowledge, and also linguistic ability, are themselves developed”(p. 6). Without students’ active engagement in the communicative language learning environment very little fluency development can take place. Such classes can be an unproductive waste of time for disaffected, passive students.

SA pilot study

This report describes the implementation of a self-assessment of class participation framework in English oral communication classes for first-year students. The SA procedure was conducted as a pilot study in a course called Communication Skills (CS). Pilot studies enable researchers to determine whether the research instrument may be inappropriate, too complicated or otherwise ineffective (van Teijlingen & Hundley, 2001). This investigation was conducted over a one-semester period (Sept. - Dec.) in order to determine whether the SA procedure should become a more permanent feature of the CS course.

A self-assessment of class participation score sheet (**Appendix A1**) was used with three different CS classes, involving approximately 70 students. It was hoped that the formative use of the SA checklist would prove a teaching and learning tool by promoting students’ English language development, encouraging active engagement with the classroom community, and discouraging disaffection and passivity.

Evaluating an assessment procedure

The thesis introduction has already noted that in the AfL framework the first priority in assessment design and practice should be for the purposes of promoting student learning (Black et al., 2002). While keeping in mind this primary priority, there are other considerations that come into play when putting an assessment scheme into practice in the classroom and evaluating its effectiveness. Incorporating ideas from more traditional assessment theory, there are a number of common principles which are generally applicable across a range of assessment frameworks. According to Brown (2004), in his book *Language Assessment: Principles and Classroom Practices*, when designing and evaluating assessment procedures the “five cardinal criteria” to be considered are *practicality, reliability, validity, authenticity* and *washback* (p.19). Brown notes that while these criteria are most often applied to formal tests, ultimately, the principles apply to all kinds of assessment procedures and provide guidelines useful for both evaluating existing procedures and designing new ones. In the same way that sources from the assessment literature (Gibbs & Simpson, 2004-05; Nicol & Macfarlane-Dick, 2006) were used to scrutinize and evaluate feedback in the previous chapter, a similar path is followed here with respect to using Brown’s (2004) ‘five lenses’ to take a critical look at an SA framework implemented in a freshman course focused on communicating in English.

Based on these five fundamental criteria, Brown (2004) poses six essential questions for ‘testing a test’. Slightly modified here (for example, replacing the word ‘test’ with ‘assessment’), the questions used to assess the SA procedure used in this pilot study are as follows:

1. *Is the assessment procedure practical?*
2. *Is the assessment reliable?*
3. *Does the procedure demonstrate content validity?*
4. *Does the procedure demonstrate face validity?*
5. *Is the assessment authentic?*
6. *Does the assessment offer beneficial washback to the learner?*

By answering these six questions, and thereby forming an overall assessment of the SA procedure used, the main objective of this pilot study can be achieved; determining the effectiveness of the procedure and deciding whether it should continue to be used in future CS

classes.

Review of the Literature

As mentioned, SA has been a prominent area of research on issues of proficiency, ability and task-performance for language learners. This extensive literature identifies a number of potential benefits and drawbacks associated with using SA. Potential problems with students using self-assessment include: lack of accuracy in student judgments, SA being prone to evaluative biases, students having limited experience of assessing themselves, and student perception of assessment as being the teachers' responsibility. However, the potential benefits of self-assessment have also been widely recognized. They include the fact that SA can be directly integrated into the teaching/learning process, encourages learner autonomy and may increase student motivation, it can reduce the teachers' assessment burden, and can result in increased student involvement in monitoring and assessing their language performance (Brindley, 1989; Brown 1998).

Language teaching practice has been usefully informed by SA research from a variety of educational contexts. Chappelle & Brindley (2002) summarize the major insights that have been provided:

1. The importance of providing students with training in the use of SA techniques; ability to self-assess should not be taken for granted.
2. The transparency of the assessment instrument impacts accurate self-assessment.
3. SA scales are most effective when statements they include are situation specific and have a close relation to students' personal experiences.
4. A student's willingness to self-assess and also the accuracy of that assessment may be affected by cultural factors.

Reviewing the research evidence, Ross (2006) reported finding, across a variety of grades and subjects, persuasive evidence that SA made contributions to improved behavior and student learning, as well as higher achievement by students. Harris (1997) explained that the potential power of SA procedures in affecting students is based on the simple fact that the assessment focus is on student-controlled behavior.

In Black and William's (1998a) formative assessment literature review, they contend that self-assessment lies at the heart of AfL for two reasons: it is a way of informing and involving the students themselves in the assessment process, and it is a means by which they take

responsibility for their own learning. Black and Jones (2006) get to the heart of the importance of student self-assessment (and peer-assessment) by explaining:

The overall aim here is to achieve meta-cognition, which is the power to oversee and steer one's own learning so that one can become a more committed, responsible and effective learner. (p. 8)

Indeed, this metacognitive 'overall aim' of enabling students to oversee and direct their own learning could also be seen as an ultimate goal of AfL itself.

While an extensive body of SA literature exists, the issue of self-assessment of class participation in an EFL/ESL context has been little researched or reported. The number of such SA empirical studies or reports available in the literature is minimal.

In an unpublished paper, Philips (2000) created a self-assessment rubric in which students rated their class participation in a pre-university ESL class in Hawaii. This SA instrument was completed by students in the middle of the semester and followed up by a teacher-student conference in which students set future goals. The results of Phillips investigation are unreported, but a copy of his self-assessment of class participation instrument is reproduced in Brown (2004), making it one of the few examples of such a tool available in the literature. It includes such criteria as attendance, asking/answering questions, participation in pair/group work, active listening and completion of peer reviews

Another such report comes from a Japanese context. Harrison, Head, Haugh, & Sanderson (2005), focused on self-assessment and its uses to motivate active class participation. The authors describe a number of SA approaches used in their classes at a Japanese university, and student reactions to them. These included scoring of class participation in note books, action logs and class journals related to class learning and progress made, self-evaluation handouts, and learning journals. A questionnaire to gauge student reactions to SA indicated that self-assessment may lead to: positive learner outcomes such as increases in active participation and L2 communication, student thinking about progress, student confidence, and increased awareness of the connection between active participation and English language skill improvement. Harrison et al. (2005) also include a number of principles to guide instructors who wish to implement a SA framework. These principles include the importance of the fit between the teacher's style, classroom approaches and the kind of SA carried out, the need for repeated opportunities for SA, and the fact that students will find SA easier if the procedures used have detailed, specific criteria. They concluded that students can make a connection

between self-evaluation and active class participation and that this mode of assessment can provide a tool to assist students in more fully realizing their learning potential.

The paper by Harrison et al. (2005) does not include any self-assessment rubrics actually used by students. Also, in the otherwise extensive body of SA literature, not enough is known about what students actually do, think, and feel when they are asked to engage in self-assessment (Andrade & Du, 2007). As such, the pilot study reported on in this chapter is one of the few examples of self-assessment of class participation in an EFL or ESL context which includes both the assessment tool used, a detailed description of the process and outcomes, and student views on the assessment procedure. We now turn to the methodology of the SA research that was conducted.

Methods

Course and Context

For all students entering TWCU, English language classes are mandatory. One of two required oral communication courses organized by the Dept. of English is called *Communication Skills* (CS), and is geared towards fostering students' spoken fluency and listening comprehension skills. The year-long course is in two parts; CS-A is taken in the spring semester (April-July) and CS-B is taken in the fall (Sept-Dec.). As usual for TWCU courses, classes are once weekly for 90 minutes. Beyond a broad goal of developing students' communicative fluency in English, and the use of a textbook reflecting a communicative language teaching methodology, there is no explicit syllabus containing course objectives and assessment procedures in place. Departmental guidelines for the CS course point out that 50% of the final grade is to be comprised of language lab work, attendance, and a final speaking test. The remaining 50% is to be determined by students' weekly classroom participation/performance. It is up to individual CS teachers to interpret and to put these general guidelines into practice.

Participants

Approximately 70 female students from three different CS classes were involved in this pilot study. Each class was comprised of 23 or 24 first-year students, all Japanese. These first year students were grouped together according to their declared major (English, Psychology,

History). As recent high school graduates, these CS classes are typically the first time for students to have an entire course taught in English by a native speaker of the target language.

The 18 to 19 year-old students in these classes exhibited various degrees of spoken proficiency in English. Using the generic descriptions for speaking in the American Council on the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines (1999), the English abilities of the students in these groups would range from *Novice Mid* ('communicate minimally and with difficulty by using a number of isolated words and memorized phrases') to *Intermediate Mid* ('able to handle successfully a variety of uncomplicated communicative tasks'). Many students, in particular the English majors, were eager to communicate in English and improve their fluency. However, because this is a mandatory course, classes invariably contained students who had less desire to improve their speaking skills, were passive in terms of participation and use of the target language, or may have had a negative attitude about learning English.

The pilot study was conducted during the second semester (Sept.- Dec.) of the school year. By September, students had already had a spring semester of 15 classes together and were familiar with each other, the teacher, class materials and lesson routines. It would be fair to say that for most, if not all, of these students this SA procedure would have been their first experience of assessing themselves in this way.

Materials

This investigation made use of two sources of data: a SA score sheet and a student survey. When used in class, a complete translation into the students' native Japanese language was included for both documents.

The SA score sheet is a criterion-referenced analytical scoring instrument. It was compiled and revised after a period of considering previous CS classes and the types of attitudes and behaviors exhibited by students in being active class participants. Out of this process, six criteria were established. These were as follows:

- 1. Punctuality and preparation** - coming to class on time and prepared (including doing any review or homework).
- 2. Being attentive and completing tasks** - staying focused on English and not wasting time chatting, checking cell phones, sleeping, etc. Actively completing textbook

exercises or other activities.

3. **Speaking English** - trying to communicate as much as possible in English by giving lots of information, asking questions and volunteering thoughts and ideas.
4. **Active listening** - listening actively to classmates and to the teacher.
5. **Speaking Japanese** - trying hard not to use any Japanese during English speaking activities and discussions.
6. **Overall effort and attitude** - being an active member of class, not a passive one. Making strong efforts to communicate in English with other students and improve speaking and listening skills.

I determined that these six criteria would cover the essential elements of active class participation. They would enable students to assess their degree of active engagement with the target language, and with their peers during class time. Students were told that while their SA scores would not be used for determining final course grades, I would use the same criterion to make decisions about the class participation component of their grades at the end of the semester.

The SA score sheet (shown below on p. 143) has two sections. In Section 1, students were first asked to write the date of the self-assessment, and then think back on what they had been doing and saying in class during the previous three or four lessons. They then used the following Likert scale to give themselves a score from 1 to 4 for each of the six categories:

1= seldom true for me **2** = sometimes true for me

3= generally true for me **4** = almost always true for me

On Section 2 of the score sheet, students were asked to take a few minutes to write some comments about their class participation and/or goals for future classes. Examples of student commentary from six students are provided in **Appendix A1**.

On the last day of class a student survey (**Appendix A2**) was administered to get some insight into student attitudes and understanding of the purpose, criteria and perceived benefits of both the SA score sheet and the self-assessment process. Students used a Likert scale from 1 to 6, to show their level of agreement with the ten items on the survey. At the end of the survey, space was provided for any additional comments about the SA procedure.

Approximately 70 SA score sheets, and 65 student surveys from three different class groupings of students were collected and analyzed. A few students were absent during the final class when surveys were completed.

Procedures

On the first day of class students were told, in English, about the importance of active class participation and how it would be the primary consideration in determining their course grades. This class then included a ten-minute introduction to the SA score sheet (with Japanese translation included), the six criteria to be assessed, and explanation of the SA procedure. Students were told that their SA scores would not be used for determining final grades. They were then given a few minutes to discuss with each other, in Japanese, whether they understood what SA was about and how the score sheet would be used. They were also given the opportunity to ask questions about the SA process during this first class, although no questions were asked. This was the extent of student training and preparation for the SA process to come.

At three different points throughout the semester, students in the three CS classes were asked to complete the self-assessment rubric. The first of the approximately 15 classes began in late September and students self-assessed their class participation in late October, November and December. They gave themselves a 1 to 4 score for each of the six categories. A total score for that assessment period (usually 3 or 4 classes) was then added up, out of a possible maximum score of 24. The final self-assessment was completed during the last regular class meeting in late December. It should be reiterated that this scoring system was intended only to provide a numerical means of identifying strong or weak areas of class participation for self-assessment purposes, not for summative grading. Table 1 below shows the main section of the SA instrument used by students, without Japanese translation.

Table 1. SA score sheet front, showing Section 1

Self- assessment (SA) Score sheet: Communication and Participation					
Class: _____	Student name / #: _____ / _____				
<p>Think about what you have been saying and doing during class for the past 3 or 4 lessons. Consider your efforts to communicate in English and your behavior. Use the score sheet below to evaluate your performance <i>honestly</i>. Write one of the following numbers to describe your personal contribution as an <u>active</u> member of the class:</p>					
<p>1= seldom true for me 2 = sometimes true for me 3= generally true for me 4 = almost always true for me</p>					
<p>Section 1</p>					
<p>CLASS PARTICIPATION CRITERIA</p>		<u>Date:</u> →	SA#1	SA#2	SA#3
<p>1. Punctuality and preparation</p> <p>I came to class on time and was prepared (including doing any review or homework).</p>					
<p>2. Being attentive and completing tasks</p> <p>I stayed focused on English and did not waste time chatting, checking my cell phone, sleeping, etc. I actively completed textbook exercises or other activities</p>					
<p>3. Speaking English</p> <p>I tried to communicate as much as possible in English by giving lots of information, asking questions and volunteering my thoughts and ideas.</p>					
<p>4. Active listening</p> <p>I tried to listen actively to my classmates and to the teacher.</p>					
<p>5. Speaking Japanese</p> <p>I tried hard not to use any Japanese during English speaking activities and discussions.</p>					
<p>6. Overall effort and attitude</p> <p>I have been an <u>active</u> member of this class, not a passive one. I made strong efforts to communicate in English with other students and improve my speaking skills.</p>					
		SA # 1	SA # 2	S # 3	
Total scores: →					

The reader will note that the statements for the six criteria are framed from the first person perspective (e.g. ‘I tried’). The vertical ‘SA # 1, 2, and 3’ columns on the sheet were correspondingly completed monthly by students for October, November and December. On each of the three SA times for these months, the last 10 minutes of a lesson was allotted for students to complete the score sheet. Score sheets were then taken up by the teacher and held until the next assessment, in the following month. No written commentary or feedback was given by the teacher on the score sheets.

On the back of this score sheet is included Section 2. It contains three boxes and students are instructed as follows: “After completing the score sheet, take a few moments to write (in English) comments about your performance or future goals”. Students did this for October and November. At the end of the course, in the third box of Section 2, students were asked to write some final comments about their class participation overall during the entire period (Sept.-Dec.). They were required to complete the comments section of the SA checklist in English, with the aid of a dictionary if they wished.

The SA survey was distributed to students on the final day of class. Students completed it shortly after finishing the last of the three self-assessments.

Results

This section will report findings from both the SA score sheet and the student survey.

SA score sheet

With approximately 70 students scoring themselves in six different categories on three occasions during the semester, the SA rubric generated a large amount of numerical data. As mentioned, the scoring options were from 1 (*seldom true for me*) to 4 (*almost always true for me*). The total added score for the six criteria is a maximum of 24.

While a detailed analysis and comparison of individual or class scores was not undertaken in this pilot study, an examination of the score sheets revealed some overall patterns:

- The large majority of scores for the three groups were 3’s or 4’s (*generally* or *almost always true for me*).
- Most students’ scores fell in the 20-24 range for each total score.

- Some scores of 2 (*sometimes true for me*) were evident throughout the score sheets. Scores of 1(*seldom true for me*) were very rare.
- The category which received the highest number of 2 scores was ‘Speaking Japanese’ (trying hard not to use Japanese during English speaking activities).
- Total scores generally showed a pattern of increase over the three assessments (for example, Oct. total = 18, Nov. = 21, Dec. = 22). However some total scores remained the same, or even decreased from one assessment month to another (possibly due to lateness or absences).
- Individual and total scores for the English majors group were slightly higher overall than the other groups (History, Psychology).

In Section 2 of the SA score sheet the students were asked to write some comments about their active participation in class or future goals. The following ideas/attitudes were recurring themes among the student commentary:

- A desire to improve vocabulary skills (in order to help students better express themselves in pair/group discussions).
- Enjoyment in being able to communicate with fellow students.
- A desire to use less Japanese.
- Frustration with inability to express themselves in English (“So many times I can’t say things I want to!”). Some students note that this is reason for switching to Japanese.
- Shyness, anxiousness, lack of confidence in ability to communicate.
- Expressing the feeling of having “poor English skills”.
- Worrying about making mistakes when speaking.
- Wanting to be more fluent English speakers.
- Feeling that ability to communicate in English was getting better.

Student comments in Section 2 give a deeper insight into their perspective on the struggle to communicate in English and be active members of the classroom community. The reader is again directed to **Appendix A1** for examples of such student commentary.

In consideration of whether to make self-assessment of class participation a more permanent component of the CS course, gaining student perspectives on the process was essential. Consequently, a student survey (**Appendix A2**) was administered and completed by 65 students during the final class. For this pilot study, the most important points were the final three issues on the survey: if the SA process encouraged students to actively participate in class (item 8), whether they spoke more English in class due to the SA process (item 9), and if they recommended using SA for future classes CS classes (item 10). Student responses to these three items are represented in Figures 1, 2, and 3 below. In addition to student attitudes toward these final three items on the survey, some of the most interesting survey results relate to item 7; whether students' assessments on the checklist were an honest reflection of class efforts. Figure 4 below represents student responses to this reliability-related item.

As mentioned, the survey used a six-point Likert scale on an agree-disagree continuum. The four charts presented here show frequency of responses and percentages from the 65 students who responded to the survey.

Figure 1 below shows student responses to the survey item checking whether they thought the SA process encouraged them to actively participate in class.

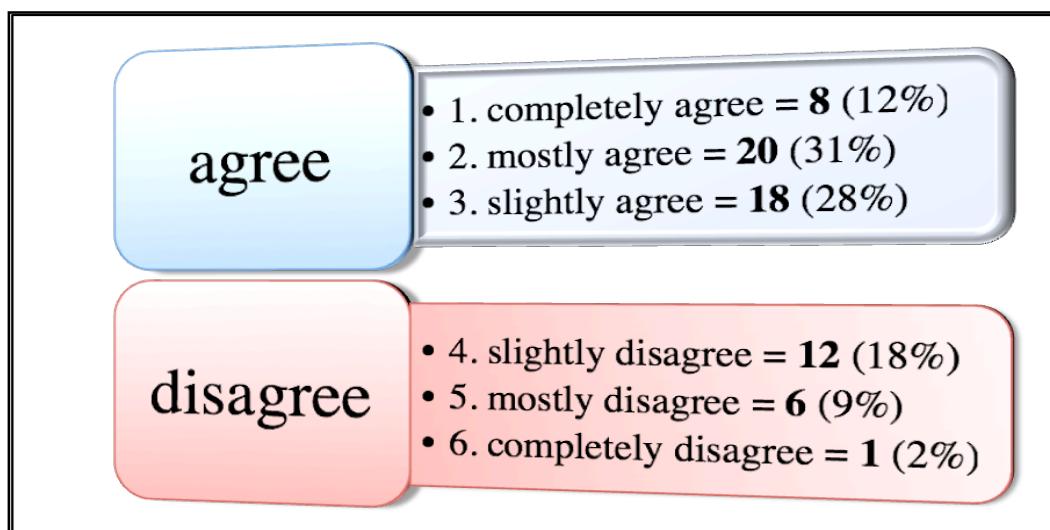


Figure 1. SA encouraged active class participation (N=65)

Combining the completely agree (1) and mostly agree (2) responses, 43% of students showed strong agreement that SA promoted active class participation. Another large grouping, 28%,

indicated slight agreement with this proposition. A combined total of 29% disagreed that more active class participation resulted from the SA procedure.

Figure 2 shows student responses as to whether they spoke more English in class because of the SA process that was embedded in the course.

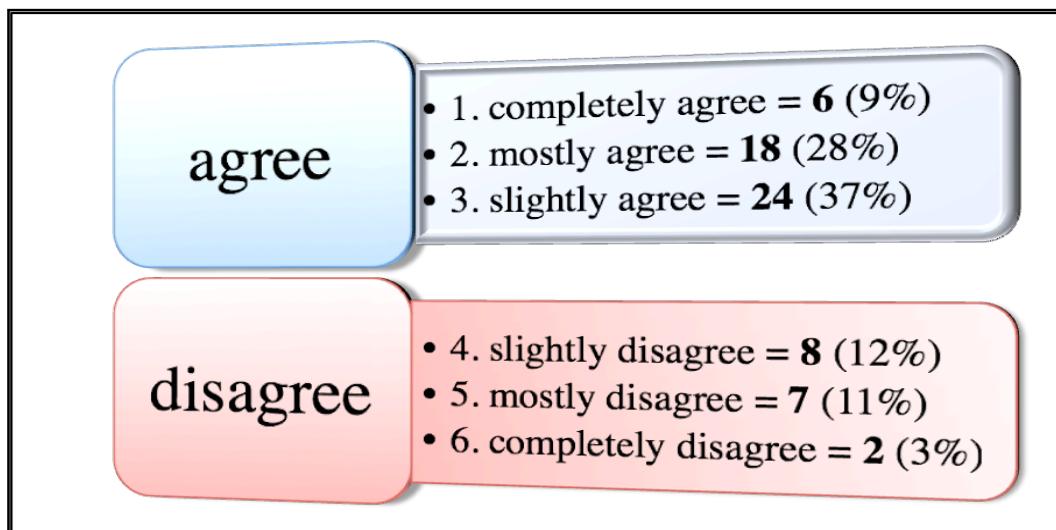


Figure 2. Students spoke more English in class due to SA (N=65)

A primary reason for implementing the SA framework was to promote students' fluency by encouraging more spoken English during class time. According to responses in item 9 on the checklist, a combined total of 74% agreed that SA encouraged them to speak more English. Combining responses 1 and 2 show that 37 % of students expressed strong agreement. However, the same percentage of respondents indicated only slight agreement with this proposition. A combined 14% of students (choosing 5 and 6) expressed strong disagreement that SA resulted in them speaking more English.

Student responses as to whether they would recommend the use of self-assessment in future CS classes are shown in Figure 3.

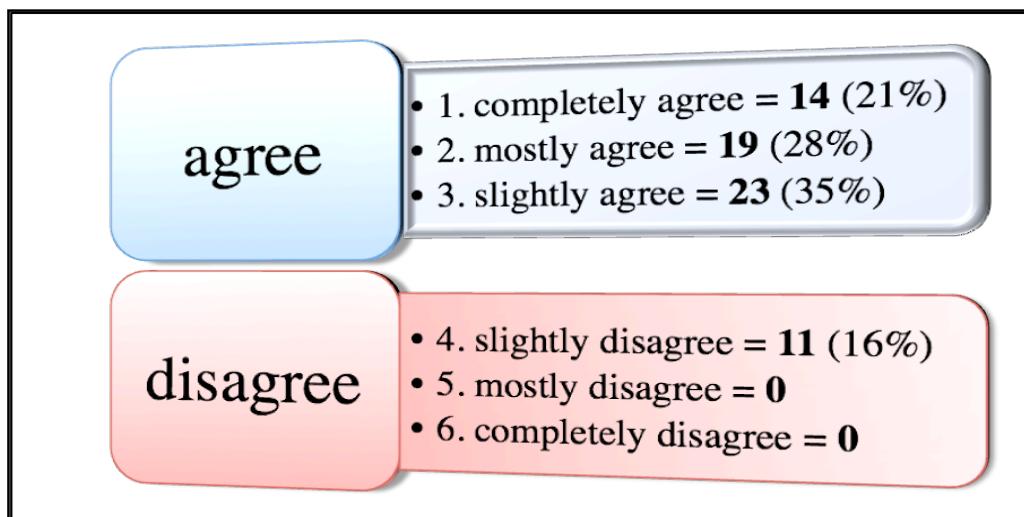


Figure 3. Students recommend SA in future CS classes (N=65)

From the 65 students in three classes, responses to the final item on the questionnaire show widespread agreement that the SA procedure should be used with future CS classes. Out of the 84% of students agreeing, 49% (responses 1 and 2) expressed strong agreement. However, 35% expressed lukewarm (slight) agreement and 16% of students thought the SA framework should not be used with future classes.

The last figure in this section, Figure 4, displays responses to the survey item asking whether the scores students had written during the three month SA process were an honest reflection of their efforts to be active class participants. The reader is reminded here that on the version of the survey used in class, a Japanese translation was provided for each item.

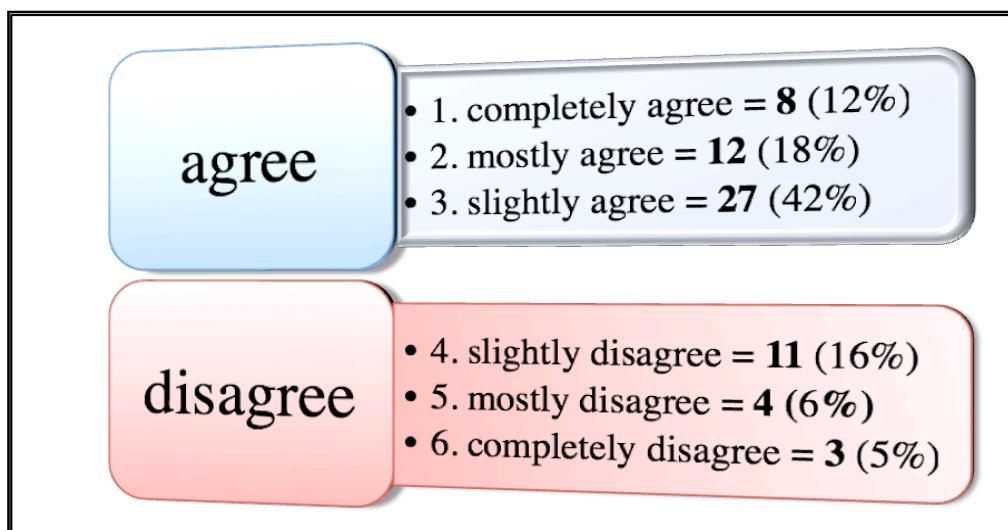


Figure 4. SA scores were honest reflection of class effort (N=65).

The number of students who expressed strong agreement (mostly or completely agree) totaled 30% of students. But the largest grouping, 42%, expressed only slight agreement that SA scores honestly reflected the class participation efforts they made in class. The total number of students disagreeing here is 27%, with strong disagreement (mostly or completely disagree) totaling 11%. These results bring the issue of reliability to the fore, pointing to discrepancies between reported scores for the categories on the score sheet and ‘true’ scores (ones which would honestly reflect efforts made in class).

Responses to the remaining six survey items are presented below in Table 2. Once again, response frequencies and percentages are given. Bold numbers indicate largest response

grouping for each survey item.

Table 2. SA survey responses for items 1-6 (N= 65)

1 = completely agree 6 = completely disagree	1	2	3	4	5	6	Total Percentages
1. I understood why SA is used in this class	15 (23%)	24 (37%)	20 (31%)	4 (6%)	2 (3%)	0	Agreement = 91% Disagreement = 9%
2. I understood SA criteria	27 (41%)	20 (31%)	15 (23%)	3 (5%)	0	0	Agreement = 95% Disagreement = 5%
3. SA criteria are helpful	18 (28%)	25 (38%)	14 (21%)	7 (11%)	1 (2%)	0	Agreement = 87% Disagreement = 13%
4. SA of class participation is fair.	25 (39%)	10 (15%)	19 (29%)	9 (14%)	2 (3%)	0	Agreement = 83% Disagreement = 17%
5. SA is a waste of time	0	2 (3%)	7 (11%)	18 (28%)	15 (23%)	23 (35%)	Agreement = 14% Disagreement = 86%
6. SA is easy to do	15 (23%)	19 (29%)	20 (31%)	7 (11%)	2 (3%)	2 (3%)	Agreement = 83% Disagreement = 17%

While a total of 91% of students understood why self-assessment was used in the CS classes, a large number (31%) expressed only slight agreement. This indicates some confusion on the part of students for the purpose of the whole exercise. Understanding of the checklist criteria (item 2) was strong (72%, combining 1 and 2). Students also indicated that the checklist criteria were helpful (item 3), with a combined total of 66% (choosing 1 and 2) showing strong agreement.

A total of 83% of students agreed that SA was fair, with the number showing complete agreement (39%) making up the largest response grouping. Disagreement with the fairness of the procedure was expressed by 17% of students. In considering whether or not the whole process was a waste of time (item 5), 14% seemed to think so. A total of 58% (responses 5 and 6) showed strong disagreement with this idea.

A combined total of 52% (responses 1 and 2) expressed strong agreement that SA was easy to do. However, the largest grouping (31%) expressed only slight agreement. A total of 17%, 11 of 65 students, disagreed. What exactly was not easy about the SA process remains unclear (for example, choosing numerical scores, consciously assessing their class

participation in an objective way, remembering behavior from previous classes).

At the end of the survey, students were given the opportunity to write some additional comments about self-assessment. Only 7 of the 65 students decided to write something, and these will be referred to in the discussion section.

Discussion

We now return to six key questions posed in the introduction, from Brown (2004), that will help determine the overall effectiveness of the SA procedure and whether it should be retained for future CS classes.

1. Is the assessment procedure practical?

In terms of time-efficiency and ease of administration, the SA procedure was quickly and easily implemented. About ten minutes at the end of class was sufficient for completion of each of the three assessments in the cycle. Student scoring of the six criteria on the rubric could be completed quickly and the 1- 4 scoring scale was specific, providing a range of choice for students. Administration was also unproblematic in requiring minimal time for the checklists to be collected and returned after and prior to each monthly assessment.

With regard to feedback on the SA checklist, no scoring, comment-writing or other such action was required by the teacher. Written feedback to student comments in Section 2 of the score sheet was considered when planning and organizing the SA process. However, with approximately 70 students doing a series of three assessments each, this idea was rejected as impractical and too time-consuming. In any case, the SA score sheet criterion let the student know the kinds of attitudes and behaviors they needed to work on to improve levels of class participation. Each of the three CS classes was periodically reminded about the SA criterion and the need to be actively involved for their communicative ability in English to improve, and that their degree of active participation would be a main part of their final course grade.

The importance of teacher feedback in AfL was thoroughly examined in Chapter 1 of this thesis, in the context of an academic writing class with clearly defined learning goals for students (for example, being able to properly use MLA documentation format in essays). The nature of the CS course discussed in this chapter involves a very different context with a

much looser set of course objectives. Here active class participation in English is a means to an end, the development of communicative English abilities. As noted in the previous chapter, “Feedback which focuses on what needs to be done can encourage all they can improve” (Black et al., 2003). The SA tool used, with its criteria clearly defined for students, and class feedback in the form of reminders and encouragement about the need for and purpose of active class participation was deemed sufficient considering the context of this particular course. Individual student feedback on progress, as used in Chapter 1 with essay-writing students, was neither practical nor possible in these CS classes.

As a whole, the process of administering the SA procedure three times, including introducing the system in the first class could all be completed within a total one hour of class time. For both students and the teacher, this SA procedure was easy to administer and may be rated as high in terms of practicality.

2. Is the assessment reliable?

This self-assessment procedure is complicated by the fact that the student is both the rater and the subject being rated. Rater-reliability may influence scoring due to human error, subjectivity and bias (Brown, 2004). In terms of assessment reliability, the consistency and dependability of the assessment tool itself may be impacted by sections or criteria which are poorly written, or ambiguous. These two aspects of reliability may increase the likelihood that a reported score deviates from the students’ ‘true’ score.

On the score sheet, student scores may have been affected by a number of different considerations. Class participation criterion (speaking Japanese, active listening) would have been an unconscious experience at times, making it more difficult to score accurately. Additionally, this assessment framework required students to engage in reflective assessment by looking back on participation and communication in a previous set of classes over a number of weeks. This time factor, and student recall, may also have impacted the reliability of SA scores.

Oscarson (1997) noted the possible cultural dimensions of engaging in such types of alternative assessment as in this pilot study: “In some cultures, the notion of learner autonomy may not be highly esteemed. In others, social etiquette requires modesty, which may affect the degree of accuracy in the assessments” (p.183). These factors may impact reliability, and are relevant in dealing with Japanese students who have minimal experience of this type of

self-assessment and are culturally conditioned to express modesty.

The scoring scale used on the rubric may have caused difficulties for students. It was hoped that the four scoring options provided a good range of choice ('seldom' to 'almost always') without being too large and unwieldy or too small and not comprehensive. However, the boundaries between scores may have been unclear or been interpreted differently by students, and thus impacted scoring choices.

Responses to item 7 on the survey ('scores were an honest reflection of efforts made') indicate that some students may have inflated their scores. A large number of students (42%) only slightly agreed that scores honestly reflected classroom behavior. A further 28% disagreed that this was so. One possible reason for inflated scoring may have been due to the fact that students thought that their scores would be aggregated into final course grades (despite being told this would not be the case). Saito (2005) noted that SA can be severely influenced when there is perceived advantage to a higher rating. Responses to this survey item reflect the two most common concerns about using SA: inflated student perceptions of their performance, and motivation by self-interest (Ross, 2006). However, some elements of the SA instrument should have improved reliability. These include the fact that the score sheet has a limited number of six carefully specified criteria for students to score, and the provision of an L1 translation.

In terms of whether this SA instrument generated scoring that was dependable (fairly reflecting students' in-class performance and participation), reliability may receive a lower evaluation than the other assessment principles being considered here. However, it should also be noted that by its very nature self-assessment is a subjective process. As such, it would be inappropriate to apply the same reliability standards of more formal tests to a self-assessment procedure.

3. Does the procedure demonstrate content validity?

Content validity, requiring students to perform the behavior being measured, is the major source of validity in a classroom assessment and it may be evaluated by considering two key factors: 1) whether classroom objectives are identified and appropriately framed, and 2) whether lesson objectives are represented in the form of assessment specifications (Brown, 2004).

In the SA rubric, objectives for classroom participation are identified for students. The six criteria were repeatedly referred to throughout the semester, through the cycle of self-assessments and periodic reminders by the teacher of what active participation entails. Survey responses showed that 95% of students agreed that they understood the SA criteria. Content validity also considers whether objectives are framed in a form that lends itself to assessment. The six criterion included in the score sheet (for example, being prepared for class, speaking as much English as possible) are things that should have been noticeable and available for self-assessment. In the survey, 83% of students agreed that self-assessment was easy to do. One potential area of confusion for students is the fact that ‘active listening’ is not clearly defined and may have proved more difficult to score than other items. This criterion could perhaps be better framed in revised versions of the score sheet.

The second key factor with regard to the content validity considers whether lesson objectives are represented in the form of assessment specifications. As this assessment procedure deals with issues of class participation over the course of a semester, it does not lend itself to a structure based on individual lesson or textbook unit objectives. Broader objectives of the Communication Skills course, to develop students’ English spoken fluency and interactive competence, are strongly represented in the six criteria on the SA rubric. Indeed, they are a primary impetus for the creation and implementation of the whole self-assessment procedure.

As discussed in the thesis introduction, subject differences can play a role in how AfL practices, such as self-assessment, are implemented in class. The CS course is an example of a class in which the ‘regulation of learning’ is looser (Wiliam & Leahy, 2007) in light of the broader horizon of goals (such as less use of Japanese, trying to speaking more English) that are part of active class participation.

Considering these elements as a whole (the clear identification and appropriate framing of class participation objectives, and their representation in the score sheet criterion specifications), it may be concluded that the SA procedure does indeed demonstrate content validity. Students are instructed by the SA procedure ‘to perform the behavior that is being measured’ by being active members of the class, speaking as much English as possible and staying focused. A significant challenge facing teachers in devising any SA tool is how to translate learning goals into more transparent categories and easily understood assessment criteria (Oscarson, 1997). The SA procedure in this pilot study demonstrates content validity

by successfully meeting this challenge.

4. Does the procedure demonstrate face validity?

Face validity is the extent to which “students view the assessment as fair, relevant, and useful in improving learning” (Gronlund, 1998, p.210). The SA tool being assessed here can be completed in a timely manner, has clear directions (with L1 translation), six clearly defined criteria to judge active class participation, and criterion-specified behaviors that are directly related to course objectives of developing communicative fluency in English. Student survey responses show that a large number of students understood the SA criteria (95% agreement) and a total of 83% of students thought that SA was easy to do. Also, survey results show that large numbers of students understood why SA was used (91%), believed SA criteria was helpful (88%) and viewed SA of class participation as being fair (83%). As for recommending SA usage in future CS classes, 84% agreed. These responses demonstrate that a majority of students see the SA procedure as ‘fair, relevant and useful’, and, therefore, face valid.

5. Is the assessment authentic?

In this context, the authenticity of an assessment procedure refers to whether tasks included represent or approximate real world tasks. The criterion students are asked to self-assess on the SA rubric include: 1) trying to communicate as much as possible in English by giving lots of information, asking questions and volunteering thoughts and ideas, 2) active listening, and 3) trying hard not to use any Japanese. These are the kinds of behaviors and attitudes that students need to employ when interacting with English speakers outside the classroom environment. In the sampling of students’ comments on Section 2 of the score sheet (from Appendix A), comments by one student, Yoshie, point to possible connections between the assessment criteria and real-world communication:

“Yesterday, some foreign customers came to my part-time job. I could explain in English. I thought my communication skills are up by this class and I am happy about being an active student.”

Being active, involved communicators of English is repeatedly emphasized for students through the SA process, and is also desirable in communicating with English speakers in more natural, outside-class situations. Authenticity, connection to real-world language use,

can indeed be demonstrated by the criteria used on the SA instrument.

6. Does the assessment offer beneficial washback to the learner?

As mentioned, generating positive washback was a primary motive in the creation and design of the self-assessment process used in the CS classes. Considering the potential problem of dealing with unmotivated, passive L2 students, it was hoped that by engaging in self-assessment the washback effect would be, as Nunan (1988) puts it, “for students to develop not only their language but also a critical self-consciousness . . . of their own role as active agents within the learning process” (p.134). Was the implementation of this SA framework effective in promoting students as ‘active agents’ in the classroom community? According to survey responses, 71% of students agreed that SA encouraged their active participation in class (30% expressed strong agreement).

As for whether the procedure caused students to speak more English, a total of 27% expressed strong agreement, while 37% agreed slightly. Some students disagreed that SA made them participate more actively, or speak more English. Indeed they may have been active members of class in any case, without self-assessment. Yet, for a significant number the washback effects were positive and beneficial. The inclusion of a follow-up task on the rubric, the commentary or goal-setting element on Section 2, was intended to help raise the washback potential of the SA procedure.

Only seven of 65 students added any written comments on the student survey. Yet, these give some insights into washback effects of the SA procedure and are presented below in Table 3. While the first two student comments reflect minimal washback, the remainder point to the potential positive effects of using such a SA instrument to impact class participation.

Table 3. Students' written comments on the survey (*verbatim*)

1	<i>"I think this process don't so important maybe".</i>
2	<i>"I think it isn't a bad process, but I don't think it isn't very helpful to improve student's communication skills of English. Therefore it is difficult to see if it is good or bad".</i>
3	<i>"Because of self-assessment, I tried to speak English as possible. I become more active than before".</i>
4	<i>"I think self-assessment is necessary, because it's easy and I can check my skill constantly".</i>
5	<i>"We can know what to do in this class by this SA. So I think this process is a good thing. And we can also find own goal by this SA".</i>
6	<i>"I think it is a good system for students to review their attitude to communicate".</i>
7	<i>"I don't think filling this SA form directly helped my class participation or attitude or my English ability. However, unless you have this SA, we never think back about ourselves so you should continue doing this".</i>

This final comment, referring to the potential for SA to cause students to 'think back about ourselves' is an expression of one of the most important aspects of SA, noted by Bailey (1998):

From a pedagogic point of view, the most intriguing and potentially most useful aspect of self-assessment as an evaluative procedure is probably the consciousness-raising factor. In completing a self-assessment questionnaire (honestly) language learners have to think about their language skills and may presumably become more aware of what they have reported (p.228).

At least for some of the CS students, the SA procedure was successful in providing what Shindler (2003) calls "a concrete and meaningful mechanism for reflection"(p.21). Survey responses, and commentary from Table 3 above shows that beneficial washback may have been negligible for some students. For many others, however, it seems clear that this SA procedure did indeed have a positive impact on students' active class participation and, consequently, perhaps aided the development of their communicative abilities in English. As such, the SA process the students engaged in was indeed, 'assessment for learning'.

Conclusion

This pilot study has assessed the performance and effectiveness of a student self-assessment of class participation framework, using the ‘cardinal criteria’ for evaluating assessment procedures from Brown (2004). Following a simplified evaluation scheme from Brown (2004) which uses three grades (*low, moderate, high*), Table 4 presents a simple assessment of the SA process implemented in this study.

Table 4. *Evaluation of SA of class participation procedure*

<u>Criteria</u>	<u>Assessment</u> (low, moderate, high)
1. Practicality	high
2. Reliability	moderate
3. Content Validity	high
4. Face Validity	high
5. Authenticity	high
6. Washback	moderate - high

Despite areas of weakness, especially in terms of reliability, overall the SA procedure may be judged effective in having a positive impact on a significant number of students’ active class participation. As a consequence, their communicative fluency in English may have also been promoted. Due to its beneficial impact on many of the students involved in the pilot study, this SA of class participation framework was assessed as a valuable, student-centered, assessment tool and it continued to be used in subsequent Communication Skills classes. While there was room for tweaking and improving this SA procedure (for example, improving initial explanation and student training), I am satisfied that it was effective in

increasing students' active class participation. This pilot study shows that self-assessment procedures can be practical, valid, and authentic, while having acceptable levels of reliability and generating positive washback.

According to Black and Wiliam (2006), a primary role of the teacher in classes where formative assessment is implemented is to 'engineer' learning environments which actively involve students in learning tasks. They also remind educators that "serving learning is the first and most important purpose of assessment" (Black & Wiliam, 2006, p. 25). The self-assessment process described here, and the SA instrument used, did help to engineer a learning environment among three classes of freshman students in the CS course. And while the effects of self-assessment may have differed for individual students, the primary purpose of promoting students' learning to be more active, communicative speakers of English was maintained in having students engage in this continuous process of assessing themselves.

We now move on to Part B of this chapter focusing on students as assessors. There we find a different course, and two groups of students who engage in the task of assessing the performance of their peers' slideshow presentations.

Part B. Student perspectives of peer assessment for learning in a Public Speaking course

The student point of view matters because of its affect on learning. From the students' point of view, classroom assessment information is not merely information 'about' himself or herself. Rather, it forms a major part of his or her learning life, becoming part of the lessons he or she is expected to learn, the relationship she or he has with the teacher and the subject matter, and relationships with peers. (Brookhart, 2003, p. 6)

Introduction

Teacher decision-making with regard to the assessment frameworks they use in their courses can be very influential in the degree of student engagement with the subject matter and the degree of student learning which results. While assessment practices can have a profound impact on learning, most assessment is implemented with little or no input from the students themselves (Stefani, 1998). The practice of peer assessment (PA) has been recognized as having possibly enormous benefits in terms of learning gain, and is increasingly being used in higher education to involve students more actively in the assessment process (Race, Brown & Smith, 2005).

Peer assessment has been defined as: “*an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status*”(Topping, 1998, p.250). Use of a PA component in a course can promote student involvement, responsibility and excellence, establish clearer course frameworks, focus attention on skills and learning, and provide increased feedback (Weaver & Cotrell, 1986). PA has a vital role to play in formative assessment by involving students in judging the work of their colleagues, and, with careful implementation, can also be used as a component in summative assessment. While being a way of assessing the products of student learning, PA can also be seen as a process of learning in its own right.

This report investigates student views of a PA framework used within the particular context of a public speaking course for third-year students called *Effective Public Speaking* (EPS). In this research study, peer assessment scores for oral presentations were aggregated into students' overall grades for the course, making up 30% of their final grade. The purpose of this investigation is to explore tertiary EFL students' perceptions of PA, and its impact on students' experience of learning. Primarily based on student responses to an end-of-course PA

survey, the research focused on two key questions:

1. How do students feel about peer assessment, particularly the framework implemented in this Public Speaking course?
2. Do students feel that the PA process (scoring peer presentations, giving/receiving peer feedback) was helpful in their learning to become more effective public speakers? In other words, was this peer assessment process ‘assessment for learning’?

A student presentation in front of a group of peers is a public performance - a showing of skill or talent before an audience. The evaluation of student presentations is a form of *performance-based assessment*, one in which students perform a task and show specific skills and competencies (Stiggins, 1987). In the EPS course, this involves the students demonstrating to an audience their understanding and application of the knowledge and skills of effective public speaking. Basturk (2008) writes that in performance assessments the role of the students in the assessment process is changed from being passive learners to active participants, and notes that it “allows instruction and assessment to be woven together in a way that more traditional approaches fail to accomplish” (p. 13).

PA and Assessment for Learning

The potential for PA to promote students’ learning has a key place in the ideas associated with assessment for learning (AfL). As we have seen, in classroom assessment focused on student learning, the assessment process and its results are turned into instructional interventions which are designed to increase, not just monitor, student learning, motivation and confidence (Stiggins, 2008). In AfL, peer assessment is considered ‘uniquely valuable’ because it motivates students to be more careful in the work they do, it amplifies the student voice in the learning process, and their learning is improved (Black et al., 2003). PA is also a valuable assessment for learning procedure because student learning is promoted as they take on the roles of teachers and examiners of each other, and students find it easier to make sense of assessment criteria if they examine the work of other students alongside their own (Black & Wiliam, 2006).

Black et al. (2003) warn that this learner-centered mode of assessment will only thrive if students are helped by teachers to develop peer assessment skills. They also make the point that “the ultimate aim of peer (and self) assessment is not that students can give each other

levels and grades-these are merely a means to an end... the real purpose-the identification of learning needs and the means of improvement" (p.62). A similar idea is captured in the simple phrase '*learning by assessing*' (Topping, 1998, p.254). Yet, in this research study, as often happens in classroom-based assessment, such formative assessment practices as peer-assessment also must contend with the summative assessment realities that exist.

Formative and Summative tensions

While summative assessment summarizes what students have learnt at the end of a period of instruction, formative assessments are ongoing and occur concurrently with instruction to provide feedback to both teachers and students and serve the purpose of guiding teaching and learning (McTighe & O'Connor, 2005). PA is considered a key formative practice, one which students are, "in the process of forming their competencies and skills with the goal of helping them continue that growth process" (Brown, 2004, p. 6).

However, in this study, peer assessment also serves a summative purpose by aggregating student-generated scores of their peers' performances into final course grades. Because marks and grades may be viewed as threats to valid formative assessment, summative assessment purposes can distort or cancel out any learning benefits for students (Stobart, 2006). Noonan and Duncan (2005) assert that, "based on principles of assessment for learning and formative assessment, it seems that the use of peer-and self-assessment ought to be limited and not used in summative student assessment"(p.6). Limited, yes, but PA does not have to be excluded from summative use. If students learn from them, summative assessments can act formatively (Yorke, 2003). In his review of the PA literature Topping (1998) concluded that studies suggest that feedback of even a simple quantitative nature can result in positive formative affects with regard to improved scores/grades and students subjective perceptions. In their discussion of the potential for assessment to enhance learning, Kennedy, Kin Sang, Wai-ming & Kwan Fok (2006) assert:

Whatever the purpose, there is no reason to prevent these summaries of learning at a point in time from abiding by the principles of formative assessment and assessment for learning. That is, these assessments can be used as feedback to help students move from where they are to where they need to be and they can be designed in ways that reflect the principles of assessment for learning. (p. 8)

The PA framework discussed in this research report was intended to serve the dual role of

using assessment as a formative learning tool, as well as a summative measuring instrument.

Literature Review

An extensive body of research related to the study of peer assessment exists, and a number of reviews and analyses of PA are available (for example, Topping 1998; Falchikov & Goldfinch, 2000; Ballantyne, Hughes & Mylonas, 2002; Bloxham & West, 2004; & Deakin-Crick, Sebba, Harlen, Guoxing & Lawson 2005). Most PA literature is focused on two issues in a particular: evaluating student contributions to group assignments, or the reliability and validity of such types of assessment (Ballantyne et al., 2002).

Student perceptions and experiences with peer-assessment have been little reported in the extensive PA literature. At the turn of the century, Hanrahan & Issacs (2001) noted that there is little in the published literature on how PA and self-assessment are viewed by students, and called for further investigations across subject areas noting that the case-based literature on PA is “still alarmingly sparse” (p.67). While PA has been identified as a key element in formative assessment, there is little research showing the extent to which teachers’ classroom practices utilize this student-centered strategy (Noonan & Duncan, 2005). More recently, the call for further research of student views of PA is echoed in Vu and Alba (2007). With regard to ESL/EFL contexts, PA has not been well-researched, with most of the work been done in peer assessment of writing (Otoshi & Heffernan, 2008).

This review of previous PA research will: 1) briefly summarize literature findings of teacher and student views of using PA, 2) address the issue of using PA scores for summative purposes, and 3) present, in chronological fashion, some relevant research related to the primary issue of this study - student perspectives of PA.

Overview of teacher/student views on PA

Using a criterion of desired performance, peer assessment requires that students closely scrutinize the work of their peers (Vu & Alba, 2007). From the teachers’ perspective, some of the main advantages and disadvantages of using PA that have been identified and described in the peer assessment literature are presented below in Table 1.

Table 1. *Potential advantages and disadvantages of PA* (from Peer Assessment, 2007, University of Technology, Sydney)

Advantages of peer-assessment
1. Helps students to become more autonomous, responsible and involved
2. Encourages students to critically analyze work done by others, rather than simply seeing a mark
3. Helps clarify assessment criteria
4. Gives students a wider range of feedback
5. More closely parallels possible career situations where judgment is made by a group
6. Reduces the marking load on the teacher
7. Several groups can be run at once as not all groups require the teachers' presence
Disadvantages of peer-assessment
1. Students may lack the ability to evaluate each other
2. Students may not take it seriously and allow friendships, entertainment value, etc. to influence their marking
3. Students may not like peer marking because of the possibility of being discriminated against, being misunderstood, etc.
4. Without teacher intervention, students may misinform each other

In his state-of-the-art review of 31 PA studies, Topping (1998) contends that PA is well worth using despite the pitfalls and any difficulties experienced with assessment quality. This is because these potential drawbacks are compensated for by the greater frequency, volume and immediacy of the feedback provided by peers, compared to what the instructor alone can produce.

In their review of student perceptions of assessment in higher education, Struyven, Dochy & Janssens (2005) noted that students hold strong views about assessment methods and that these perceptions play a significant role in how they approach learning. Student concerns about PA that have been identified in the literature include such things as: students being aware of their own shortcomings in subject areas; having doubts about their own objectivity;

feeling the PA process to be unfair; the social effects of PA, such as friendship or hostility; and the view that it is the teachers' 'job' to assess (Cheng & Warren, 1997; Falchikov, 2003).

Using peer assessment for summative grading

While there is general agreement on the potential value of PA to promote learning, the issue of whether peer assessments should form a significant part of student grades is much more contentious (Magin & Helmore, 2001). Table 2 provides a useful summary of the arguments presented in the literature on both sides of the issue.

Table 2. *Arguments against and for the use of PA for summative grades*

Arguments against PA use for summative grades
1. This practice could compromise the main pedagogical intention of PA
2. Peer assessments are too inaccurate for such purposes
3. Reliability and validity concerns (for example, students are ‘poor judges’ of effective communication skills, the potential for bias to influence students marking, variability of marking standards used by peer assessors)
4. The need for universities to have confidence in their assessment practices as these are used for high- stakes certification purposes (and therefore not relying on inexperienced assessors)
Arguments supporting PA use for summative grades
1. Knowing that peer grades will ‘count’ towards final grades will have the washback effect of promoting greater seriousness and commitment from students
2. PA used only formatively may not be taken seriously by students
3. This practice will help develop student autonomy and empower students to make judgments that count
4. While in some contexts fair and valid peer assessments may be difficult to obtain, in other assessment contexts such impediments are minimal or can be overcome
5. It may be possible that assessments based on peer ratings are superior to solely teacher assessments (for example, in oral presentations and communicating to an audience)

(from Magin & Helmore, 2001)

There are valid arguments for both sides of this question, a reflection of the complexity often associated with assessment issues. In this research study I decided that PA would play a significant summative role, finding more convincing the arguments supporting the inclusion of peer scores into the final grade, and concurring with the following views expressed by Vu & Alba (2007):

If peer assessment excludes assigning marks, however, its positive impact on student learning and development is restricted. The act of marking demands that students take greater responsibility, as they are required to base their judgments on thorough examination of the quality of their peers' work. Thus they are more likely to gain deeper understanding of the subject matter. (p. 543)

Addressing the issue of making PA marks meaningful, Race et al. (2005) note the view held by some that PA is only suitable for feedback purposes, but they also advocate that PA should account for something, even if it is a small part of the final grade, if students are to take it seriously. In this study, it was hoped that making peer scores part of the final grade would encourage students to take PA more seriously and more carefully, and subsequently lead to the promotion of learning about the course content and objectives (i.e. 'learning by assessing').

Student perceptions of peer assessment

In his review of peer assessment in tertiary settings, Topping (1998) briefly reports a sampling of student views about PA expressed in the literature. On the positive side these include fairness (being assessed by more people) and the formative usefulness of detailed peer feedback. On the other hand, students expressed a dislike for possible social embarrassment (especially concerning identifying weaknesses in the work of peers) and the fact that PA may be cognitively challenging and straining for students. Hanrahan & Issacs (2001) reported on tertiary students perceptions of self and peer assessment in a health psychology course. Student questionnaire responses about PA included such themes as: motivation to impress peers, difficulties with being objective, discomfort with peers judging work, and gaining a better understanding of marking procedures through PA. Hanrahan & Issacs (2001) reported the generally positive affects peer assessment has on students learning, despite any negative views expressed by individual students.

An important PA study which focused on student perceptions of the process was conducted by Ballantyne, et al. (2002). The authors investigated the implementation of PA in large classes, and they concluded that while there are a number of specific difficulties associated with using PA with larger classes, the learning benefits for students of being involved with PA outweigh any of these drawbacks. Their large study (involving completed questionnaires from 939 students in a variety of classes) conducted at the University of Technology, Australia, obtained a wealth of information about student views of various PA procedures implemented. Their research reported what students liked and disliked about being part of PA. On the positive side, this included such things as: students felt that PA encouraged

them to compare and reflect on their own work; it gave them the opportunity to develop skills useful for future employment. The things students disliked about PA included: questioning peers competency in marking; issues of fairness (feelings that peers were either easy or hard markers); and large numbers of students felt PA was too time-consuming.

Ballantyne et al. (2002) also address the issue of using PA marks for grading purposes. They suggest that, “a ‘reasonable’ number of marks (10-15% of the total) be allocated to student performance in the peer assessment process, as it may boost student engagement and commitment to the task”(p. 435). They also highlight the fact that the clear articulation of assessment criteria is of paramount importance and should be a fundamental aspect of the PA process.

McLaughlin & Simpson (2004) described how first year university students felt about peer assessment. Working in a context of a construction management course, the PA model implemented asked students to assess the group work of their peers. McLaughlin & Simpson (2004) found that in this PA model trialed with freshmen, students were overwhelmingly supportive of the process and viewed PA as a very positive assessment experience. Students perspectives about the PA model used showed they felt they had learned a great deal, enjoyed assessing peers’ work, and a significant portion (43%) preferred PA to lecturer only assessment. Reflecting a fundamental AfL theme, McLaughlin & Simpson (2004) stress the idea that “the assessment process needs to be a learning tool” (p.136).

An investigation of peer assessment with sports studies students in the UK, was conducted by Bloxham & West (2004). In order to encourage students to carry out PA seriously, they awarded 25% of their assignment marks for the quality of peer marking. These researchers noted that actively engaging with the assessment criteria while marking peers is beneficial for assessors in understanding how their own work will be assessed. Bloxham and West (2004) found that with regard to the experience of assessing, and being assessed by peers, two-thirds of students were positive while one-third expressed some disquiet. Some students felt that peer marking resulted in higher or lower grades depending on whether peers were generous or mean assessors. Overall, however, Bloxham & West (2004) concluded that students saw peer assessment as a positive experience that aided in their understanding of the assessment process.

Nigel & Pope (2005) focused on the impact of stress in peer (and self) assessment. They explained that course requirements for students to assess their peers, who will also assess them, can be stress-creating for students. The stress may be caused by inexperience with PA,

the fear of hurting someone, or being hurt by someone. Researching in the context of PA of group projects an undergraduate research methods class, Nigel & Pope (2005) found that peer assessment was more stressful for students, but did lead to improved performance in summative tasks.

Langan and 10 associates (2005) specifically focused on the peer assessment of oral presentations. They compared marks awarded by both students and tutors in environmental or biological courses in the UK. Longman et al. (2005) write: “There seems to be consent that a firm understanding of the assessment criteria, within a study of high design quality, appears to be associated with greater validity of peer assessment” (p. 23). Also, because the teacher is required to manage ‘a group of mostly inexperienced assessors’ PA is more complex and demanding. Longman et al. (2005) concluded that the benefits of PA outweighed any differences between peer and tutor marks. They felt that “the benefits of learner inclusion and the active dimensions of this scheme (e.g. learner empowerment, assessment experience, better understanding of assessment criteria) merit its inclusion in future courses” (p. 31). However they end their article by cautioning that, due to the possibility of bias in PA, there is some doubt as to whether student marks should be used for other than formative purposes.

Wen & Tsai (2006) investigated the views of university students in Taiwan towards peer assessment, particularly the on-line variety. The authors noted that in similar studies students generally showed a liking for PA as it enabled them to compare work with classmates. But students were less appreciative of being criticized by peers, and expressed a lack of self-confidence to peer-assess classmates. Analysis of the student questionnaires in their study led Wen & Tsai (2006) to conclude that, in general, university students had positive attitudes towards PA activities.

Vu and Alba (2007) described students’ experience of PA in a professional course at an Australian university. Students peer assessed a viva voce course component, which consisted of a student interview with the teacher. The PA component was planned and structured so as to “both evaluate and promote student learning” (Vu & Alba, 2007, p. 545). The authors reported that in their case study PA had a positive affect on students’ learning experiences with most students acknowledging learning from both the process and from their peers. An exploration of the power of PA to promote learning in other settings was also called for.

Papinczak, Young & Groves (2007) reported a qualitative study of PA in problem-based learning with freshman medical students at an Australian university. Students in this study took a much more critical view of assessment by peers. In their study the authors noted a

widespread view among students that PA could be corrupted because of bias due to lack of honesty, or ‘friendship marking’.

As a whole, the literature that deals with student perceptions of PA report that students see the positive benefits of having it part of a courses’ assessment framework and the impact it can have on their learning. Student awareness of the potential disadvantages of PA use is also evident in the literature, and, at times a dislike of doing or receiving peer assessment. A review of the literature did not uncover any reports similar to the context of this research project - student perceptions of peer assessment of oral presentations in an EFL context.

This research report takes up the call for additional PA research from Vu & Alba (2007), and investigates PA in the setting of a Public Speaking course at a Japanese university.

Following their example, the PA process discussed here was also designed and structured to both promote and evaluate student learning. In AfL, peer-assessment is regarded as a fundamental component of effective formative practice because, like self-assessment, it is a strategy for “placing the work in the hands of the pupils” (James et al., 2007, p. 11). This part of the thesis sees students doing some of the assessment work, and presents their perceptions of assessing, and being assessed by, their peers.

Methods

Course and students

The Department of English has a one-semester course for junior students entitled Effective Public Speaking (EPS). Lasting from April-July, this course included two classes of third-year students, numbering 55 in total. Student ages were in the 20-21 range. Age considerations may be worthy of note with regard to student views of assessment. The students in this case study have already had a couple of years of tertiary level experience with a range of assessments in a variety of classes. This fact may impact opinions expressed on the PA survey at the end of the course.

The primary learning objectives of the EPS course focused on developing student skills in planning, organizing and delivering effective presentations, supported by a computerized slideshow. Divided into two classes, each student was responsible for delivering two main presentations during the semester. These presentations, both with an informative purpose, were based on news media stories selected by the students. Students were instructed that topics were to be chosen based on their personal interest and interest for the listening

audience. Students selected internet-based stories primarily from such news outlets as The Japan Times, or BBC News and included such topics as ‘World’s costliest cities’, and ‘Convenience stores going green’. Presentations were from 8-10 minutes in length, and all presentations, two per student, were videotaped.

Materials and procedures

Each of the two EPS classes met weekly for 90 minutes, approximately 15 times during the semester. Class time involved such things as: examining and practicing elements of effective public speaking, learning to put together well-designed computer slideshows, and choosing news stories and organizing the information for a presentation format. Students were assigned to planning groups to aid in preparation for their mid-term and final presentations. These planning groups of four students, involved: discussing presentation topic choice, reporting on progress, doing mini-presentations (2-3 minutes) about their topics, and getting feedback from peers. Approximately four of the fifteen classes (two in the middle of the course, two at the end) involved students delivering presentations, assessing their peers, and being assessed by the teacher. There was an approximately six-week gap between mid-term and final presentations. The first class of the semester included an introduction to peer assessment. Students were told about PA, provided with a rationale for why it would be included in the course, and were given the criteria that would be used by both peers and the teacher to assess and score their presentations.

Rather than using a textbook for this relatively short, 14-week course, the theoretical and practical frameworks for the classes were heavily based on a journal article by Yamashiro & Johnson (1997) entitled *Public Speaking in EFL: Elements of Course Design*. In this article, Yamashiro & Johnson introduced a public speaking course which they had developed and used at both secondary and tertiary levels in Japan. A key element of the public speaking course designed by these authors is a reference list (Table 3) of the elements of public speaking covered in the course. Both the peer assessment and teacher assessment rubric criterion used in this case study were based on these 14 points.

Table 3. 14 Points for Public Speaking (from Yamashiro & Johnson, 1997)

	Speaking Area	Comments
Voice Control		
1	<i>Projection</i>	Speaking loud enough (not too loud or too soft)
2	<i>Pace</i>	Speaking at a good rate (not too fast or too slow)
3	<i>Intonation</i>	Speaking using proper pitch patterns and pauses
4	<i>Diction</i>	Speaking clearly (no mumbling or interfering accent)
Body language		
5	<i>Posture</i>	Standing with back straight and looking relaxed
6	<i>Eye Contact</i>	Looking audience members in the eye
7	<i>Gesture</i>	Using few, well-timed gestures, nothing distracting
Contents of Oral Presentation		
8	<i>Introduction</i>	Including attention-getting device, thesis statement
9	<i>Body</i>	Using academic writing structure and transitions
10	<i>Conclusion</i>	Including restatement/summation and closing statement
Effectiveness		
11	<i>Topic</i>	Choice picking a topic that is interesting to the audience
12	<i>Language Use</i>	Varying types of clear and correct sentence forms
13	<i>Vocabulary</i>	Using vocabulary appropriate to the audience
14	<i>Purpose</i>	Fulfilling the purpose of the speaking task

Peer assessment was also an important part of the syllabus designed by Yamashiro & Johnson (1997) and presented in their article. The authors note that students “clarify and deepen their understanding of course objectives by becoming critical evaluators of their peers” (p. 1). They also make the point that through this course material, as well as developing their oral production/ public speaking competencies, students develop their critical thinking skills as they realize they must understand the assessment criteria in order to provide their peers with accurate feedback. A key component of AfL is student understanding of ‘where they need to

get to in their learning' (Assessment Reform Group, 2002). Black et al. (2003) point out the need for structures that are carefully thought out so as to foster effective PA and promote student reflection on their performance. These 14 elements of public speaking provided a series of focus points as students prepared for, practiced and later reflected on their mid-term and final presentations.

The 14 Points also formed the backbone of the course syllabus. It was copied and distributed to the EPS students on the first day of class. Most subsequent classes involved examining these points and completing tasks focusing on each one. Students were also instructed, in their planning groups, to use the 14 points to give feedback to peers after their mini-presentations (practice runs of their presentations, without computer slideshow). The mini-presentations and subsequent group feedback helped students prepare for their performances, but also served as training sessions in the use of the assessment criteria. It was hoped that such use of the 14 points in mini-presentations (and class activities) would serve to help students internalize the key assessment criteria to be used. As the semester progressed, students became very familiar with the differing aspects comprising the key elements of public speaking the course focused on (voice control, body language, content, and effectiveness).

The assessment framework used for the EPS course was comprised of three elements, shown in Table 4.

Table 4. *Assessment breakdown for the EPS course*

Assessor	Percentage of final grade
1. Teacher	60% (30% per presentation)
2. Peers (6-8 students)	30% (15% per presentation)
3. Self	10% (5% per presentation)

While the majority of the weight for grading remained with teacher assessment, almost half (40% in total) of the final course grade was based on student-generated assessment input. Peer assessment made up 30% of students' final grades for the course. Based on Yamashiro and Johnson's (1997) 14 points, a peer-rating (PR) sheet (see **Appendix B1**) was used by students to evaluate and score classmates presentations. Students were rated by from six to eight of their peers, depending on attendance numbers on presentation days. Students were

prohibited from assessing presentations from peers in their planning groups, as this group would already have heard mini-presentations related to their topic, and given feedback advice. Ratings sheets were collected after mid-term and final presentation classes, average scores were determined from peer scoring, and the PR sheets were copied and distributed to provide peer feedback to presenters.

Student self-assessment was also incorporated into the assessment framework for the course. This was comprised of two self-reports, worth 10% of the final grades. After both the mid-term and final presentations, students were given a two-page report sheet with a number of questions asking about their presentation preparation and delivery (for example, “Why did you choose this topic for your presentation?”). For the mid-term report, they were also asked to list three things they hoped to improve in the final presentation. The final report asked them to also write about whether they were able do better in those selected areas. Students were assessed on a 5 point scale (5=excellent, 1=poor) by the teacher based on the depth and degree of analysis in responses to the questions on each report.

Table 5 below describes the PA procedures followed for each of the two presentations.

Table 5. *Peer assessment process*

Peer-assessment procedures for EPS course
1. Prior to mid-term and final presentation classes, peer-rating (PR) sheets were copied and prepared in sets of eight.
2. Students were divided into groups (12-14 per group), and put in separate classrooms. Students were responsible for setting up video recording equipment and recording each presenter.
3. Peer rating sheets were distributed to selected students (who were not members of the presenters' planning group). Peer raters usually numbered six to eight students depending on attendance on the day.
4. During and after each presentation, students were instructed to fill out a PR sheet for each presenter. Presenters were instructed to complete the self-assessment report and submit it in next weeks' class.
5. PR sheets were collected at the end of class and given to the teacher. Video-recordings were brought to the audio-visual center and made available to students, if they wished to see their performance. Video-recordings were also used by the teacher to assess presentations not seen live on presentation days.
6. Prior to the following class, PR sheets were grouped together for each individual presenter and copies of the sheets were made for instructor records. PR scores were recorded for each presenter and an average PA score from 5 (very good) to 1 (poor) was determined.
7. During the following class: <ul style="list-style-type: none"> - self-assessment reports were collected from the previous weeks' presenters, - PR sheets for the previous weeks presenters were returned to students, and - a teacher assessment sheet (using the same peer rating criteria) was also given to students.

It should be noted that, in order to maximize objectivity, peer rating sheets for individual students were not examined by the teacher prior to completing the teacher assessment for each presenter.

Essentially, one week after their presentations, students were able to receive scores and

feedback from six to eight peers, as well as the teachers' assessment.

To gauge student perceptions of the peer assessment process, in the final class a student survey (see **Appendix B2**) was distributed and completed by 53 students. The survey was divided into three sections 1) being a rater/ being rated by peers, 2) the PA process, and 3) additional comments (open-ended). The following four-point Likert scale was used on the survey to gauge opinions: 1=agree, 2=tend to agree, 3=tend to disagree, and 4=disagree. Scale options 2 and 3 gave students the opportunity to express some reservations with the level of agreement or disagreement for each item.

Results

Completed by 53 students in the last class for the EPS course, the student survey was designed to elicit student views of the PA framework implemented. A sample copy of the peer-rating sheet was distributed at the same time in order to remind students of the peer assessment criteria and score sheet structure as they responded to the survey items.

The survey consists of a total of twelve items plus a section for additional student comments. As mentioned, a four-point Likert scale was utilized. Tables 6 and 7 below summarize student responses to survey items. Numbers and percentages for each item are presented, and **bold numbers** indicate largest response choice for each item.

The first section of the survey included eight items focusing on students' perceptions of being both a rater and being rated by peers.

Table 6. Survey Part 1- being a rater/ being rated by peers. (N=53)

Survey item	1.Agree	2. Tend to Agree	3. Tend to Disagree	4.Disagree	Combined totals
1. Assessment items on the sheet (e.g. pace) were easy to understand.	37 (70%)	14 (26%)	1 (2%)	1 (2%)	Agreement = 96% Disagreement = 4%
2. It was difficult to decide the overall score (5 4 3 2 1) for each presenter.	13 (25%)	20 (37%)	17 (32%)	3 (6%)	Agreement = 62% Disagreement = 38%
3. Relationships with presenters (friendships, etc.) may have influenced overall scores and comments I gave.	4 (8%)	15 (28%)	14 (26%)	20 (38%)	Agreement = 36% Disagreement = 64%
4. I was comfortable being a judge and scoring my peers presentations.	14 (26%)	21 (40%)	17 (32%)	1 (2%)	Agreement = 66% Disagreement = 34%
5. I was comfortable having my presentations judged and scored by my peers.	19 (36%)	21 (39%)	11 (21%)	2 (4%)	Agreement = 75% Disagreement = 25%
6. The overall scores my peers gave me were fair and reasonable.	16 (30%)	26 (49%)	10 (19%)	1 (2%)	Agreement = 79% Disagreement = 21%
7. Assessing other students' presentations helped me plan and deliver my own.	32 (60%)	19 (36%)	2 (4%)	0 (0%)	Agreement = 96% Disagreement = 4%
8. PA scores and comments from my first presentation helped me prepare my second presentation.	30 (56%)	20 (38%)	3 (6%)	0 (0%)	Agreement = 94% Disagreement = 6%

Section two of the survey, shown in Table 7 below, focused on the peer assessment process as a whole and the issue of aggregating PA scores into final course grades.

Table 7. Survey Part 2: the peer assessment process (N=53)

Survey item	1. Agree	2. Tend to Agree	3. Tend to Disagree	4. Disagree	Combined totals
9. Students should not be involved in assessing peers; assessment should be solely the teachers' job.	0 (0%)	9 (17%)	27 (51%)	17 (32%)	Agreement = 17% Disagreement = 83%
10. Making PA scores a part of student final grades is a good idea.	14 (26%)	31 (59%)	7 (13%)	1 (2%)	Agreement = 85% Disagreement = 15%
11. Making PA worth 30% of the course's final grade is:		a. Too high 13 (25%)	b. Fair 40 (75%)	c. Too low 0 (0%)	
12. I recommend using PA in future Public Speaking classes.	28 (56%)	21 (38%)	4 (6%)	0 (0%)	Agreement = 94% Disagreement = 6%

Part three of the survey invited additional written comments (in English) about PA. Many students, 36 of 53, wrote further commentary in this section. The 36 written comments were grouped into three categories of student feeling about this PA experience: positive (19 students), negative (10 students), and mixed feelings comments (7 students). A few examples from all three categories provided below in Table 8 will give the reader a taste for some of the feelings expressed.

Table 8. Survey Part 3: Sampling of students written comments (verbatim)

+	<i>PA helped me to make my presentation better. I think it is good system for us to improve our presentation skills.</i>
+	<i>PA is helpful for both presenter and rater. It helps both improve their skills. I hope it will be used in future public speaking classes.</i>
-	<i>I think assessment of the peer is not fair. Therefore I think rate of assessment should be changed between peer and teacher.</i>
+/-	<i>Peer assessment process is a good way to improve our presentations. But it is difficult to evaluate presentations precisely. I don't know whether my evaluation to other students is right or not.</i>

Additional student commentary from the last part of the student survey will be presented below in the Discussion section. Analysis of the data results presented above provide a basis for answering the research questions posed in the introduction to Part B of this student-as-assessor chapter.

Discussion

This discussion re-focuses attention on the two key questions of this research study: how students felt about the PA process implemented in the EPS course, and whether it helped promote student learning in becoming more effective public speakers.

While recognizing that survey research may be criticized for using a blunt instrument to yield only superficial information, survey responses from a group of 53 students, which targets their views and opinions can indeed provide much insight into their perspectives on classroom events and processes. While the students in this case study were forthright in their survey responses and the survey was effective in gauging student feelings about PA, two points should be kept in mind. Firstly, through the realities of professional power relationships in the classroom, teacher views of assessment will influence student attitudes (Sadler, 1998). The fact that PA was advocated by the teacher on a number of occasions throughout the semester, should be considered as a possible source of influence in student survey responses. Secondly, as Fry (1990) noted in his study of peer assessment, student views of the perceived advantages and disadvantages of PA will vary depending on individual values, objectives, and capabilities.

As noted, the Likert scale used in the survey is a four-point, agree/disagree forced response scale. With an even number of responses and no middle choice, students are forced to decide their degree of agreement or disagreement for each item. Twelve declarative statements, written and revised based on a reading of PA literature, were used in the survey.

Overall, survey response data indicate that a majority of students had a positive reaction to the PA format used in the EPS course, yet also expressing some reservations with this type of participatory assessment. While being a positive assessment experience for most, a minority of students expressed a dislike or dissatisfaction with the process. Also, the data shows that for many students the process did indeed serve the purpose of promoting student learning, previously identified as ‘the first priority’ in Afl (Black et. al., 2003). Student perceptions about peer-assessment, both positive and negative, are often congruent with student views expressed in the PA literature discussed earlier.

This discussion will be separated into two parts, following the format of the survey. The beginning and larger section will deal with student views on being a peer assessor and being assessed by peers (survey items 1-8). The second part will deal with the larger issues of student views of peer assessment (items 9-12). Student comments from section three of the survey will at times be used to elucidate points and share student perspectives in their own words.

Part 1: Student views on peer assessing and being assessed

From Yamashiro & Johnson’s (1997) syllabus, the 14 key points for public speaking related to voice control, body language, content and effectiveness, were reproduced for the peer rating sheet used in the course. As mentioned, these key points were also used as a basis for the weekly classes as well as used for students to informally (and formatively) assess and give feedback to group members mini-presentations. These facts are reflected in student responses to **item 1** on the survey, asking whether assessment items were easy to understand. A total of 96% of students agreed that this was the case. Of all items on the survey, this one had the highest ‘agree’ score of 70% (37 out of 53 students). PA is most effective when the criteria is clearly understood by all students, and for presentation assessment the criteria should be made clear from the outset of the course (Race et al., 2005; Papinczak et al., 2007). There seems to be a consensus in the PA literature that a firm understanding of assessment criteria results in greater validity (Langan et. al. 2005). As noted in the thesis introduction, the sharing of such

criteria with students is also a fundamental element of AfL in practice. In making use of the 14 points from the first class and continuing to focus on them in subsequent classes it seems that such familiarity helped students to have a clear understanding of the rating criteria. Using the 14 points in group work to informally assess and provide feedback to group members mini-presentations gave students mark-free opportunities to rehearse. It was also useful in giving students practice to develop the assessment abilities required for peer rating. Liu and Carless (2006) also stress this strategy of embedding peer assessment within regular course processes, asserting that by doing so students are more likely to develop the necessary expertise in making sound judgments.

The peer rating score sheet (**Appendix B1**) uses almost the same scoring system utilized by Yamashiro and Johnson (1997) for rating presenters on individual points, and for arriving at an overall score of: **5** (very good), **4** (good), **3** (average), **2** (weak), and **1** (poor). While providing a range of scoring options, one obvious weakness here is that it is not clear what the differences are between scores on this continuum. This may have impacted responses to survey item 2. A large number (62%) of respondents showed agreement with the survey **item 2** idea that ‘it was difficult to decide overall scores for presenters’. The reason(s) for the expressed difficulty in deciding overall scores is unclear; whether due to the 1 to 5 rating system used, students lack of PA experience, insufficient time, or simply the inherent, often complex, nature of making assessment judgments.

As previously noted, a possible disadvantage of using PA is that reliability of scoring may be affected by student bias caused by the relationship between the assessor and the person being assessed. This point was the focus of survey **item 3**, asking students whether the scores they gave may have been influenced by relationships with presenters. A total of 36% (19 of 53 students) expressed agreement that such influence may have been a factor in their PA scoring. In reviewing the literature on peer feedback, Nilson (2003) concluded that “Apparently most students are loathe to find fault with one another’s products, or at least loath to express these faults. . . . In particular, students do not want to be responsible for lowering a fellow student’s grade” (p. 35). This may take on heightened importance if the fellow student is a friend, and peer scores make up 30% of students’ final grades.

With regard to this EPS course there are also several relevant points to consider. One is that these students had already experienced a year together in a sophomore pre-requisite course and were well known and friendly to each other. An additional factor is the methodology of the peer assessment of oral presentations used in this case study. While

students did not write their name on the peer rater sheet, peer assessment was not anonymous due to the fact that student could directly see the person they were assessing. Additionally, while the rater sheets provided as feedback were anonymous, presenters knew which 6-8 students were assessing them and therefore knew who their judges were. These factors may have injected the ‘relationship factor’ into the scoring process, leading to occasions where, as one student wrote on the survey, ‘*The students and also I tended to be modest in giving scores*’.

Because these student-generated scores were used for summative purposes in determining final grades, reliability becomes a more serious issue. Do student responses to this survey item weaken the reliability of peer scores and consequently students final grades? Perhaps. Yet, a majority of 64% of students (34 of 53) disagreed that their ratings were influenced by relationships with presenters. Prior to the first cycle of presentation classes, and the final performances, students were told of the importance of being honest and fair in their peer ratings. The issue of rater bias may also be connected with a related item on the survey, number 6, which deals with perceived fairness of the scores given by peers.

Survey **item 6** asked students whether ‘The overall scores my peers gave me were fair and reasonable’. During a quick preview of the survey items, it was pointed out to students that the word ‘reasonable’ referred to a sensible judgment based on the content of the presentation and the quality of delivery. Survey results show that 79% of students (42 of 53) thought the peer ratings were fair and reasonable. In a study of peer-assessment of essay writing, Mowl & Pain (1995) concluded:

The research shows that even with subjective methods of assessment . . . students are generally capable and conscientious self-and peer-assessors, as long as they are adequately prepared and reassured about the value of the exercise. (p. 330)

The fact that most students were satisfied that peer scores were generally fair and reasonable indicates that this group of students were, on the whole, ‘capable and conscientious’ assessors of their classmates presentations.

Just one student gave a score of 4 (disagree) for item 6. However, 19% (10 of 53 students) *tended* to disagree, showing they felt that peer scores for their presentations were, at least to some degree, unfair and unreasonable. One student wrote, ‘*I think that PA is sometimes unfair because some people are too strict and some people are too kind*’. Other students expressed similar feelings in their written commentary on the survey. While the influence of

student relationships is a factor here, the simple fact that individuals vary in their perceptions must also be noted. This fact may have led to varying overall peer scores returned to students the week after their presentations. Adherence to the assessment criteria, that is, stricter or looser application of the 14 points when deciding on overall scores, may have varied with individuals in the peer assessment group. Additionally it may simply be the fact that some students may be more skillful, accurate assessors than others. Student dissatisfaction with peer scores due to an elevated evaluation of their own performance, may also have led to determining peer scores to be ‘unfair’, even though they may have been accurate. It is worth remembering the simple truth that because assessment involves making judgments, it will inevitably be subject to some error and bias (Harlen, 2006). Overall, however, almost four out of five students surveyed in this case study were satisfied that peer ratings of their presentations were fair and reasonable.

Peer assessment, and AfL in general, involves a changing role for students; one in which they are “brought into the heart of teaching and learning processes and decision-making” (James & Pedder, 2006, p. 28). Students’ feelings about this change and how comfortable they were in the roles of being an assessment decision-maker, and of being assessed by peers are the focus of items 4 and 5 on the survey.

In **item 4** students were asked to respond to the statement: ‘I was comfortable being a judge and scoring my peers’ presentations’. A significant number, 34% (18 of 53) responded that they were not comfortable in judging their peers presentations. The majority (35 of 53) agreed with the statement but the largest response grouping for this group (40%) selected ‘tend to agree’. These numbers indicate a significant degree of discomfort at judging peers being common in the group as a whole. Such discomfort may be a result of lack of confidence or experience in rating peers, or the stress caused by fear of hurting, or being hurt by, classmates (Wen & Tsai, 2006; Nigel & Pope, 2005). Power relations are also a factor, as students often dislike having power over their classmates or peers exercising power over them (Liu & Carless, 2006). In the context of this EPS class, students are directly looking at and evaluating classmates (sometimes good friends) oral presentations. The nature of this arrangement may have caused some students to be uncomfortable in this peer assessor role.

Students also expressed some discomfort in related survey **item 5**, probing student feelings of being on the receiving end of peer assessment (‘I was comfortable having my presentations judged and scored by my peers’). A total of 75% agreed to being comfortable (40 of 53) with assessment by classmates. One student wrote: *‘I think it is OK that other students judge me.*

*In general I have feedback only from teacher. Therefore, other students' comments help me to make better presentations'. However, a quarter of the students in the survey were uncomfortable with peers assessing them. Such feelings may have been caused by such previously mentioned factors as: worries about peer objectivity, peers capabilities for assessing, and relationships between presenters and assessors. Interestingly, one student commented that the fact of being scored by peers may increase the presenters' stress and feelings of discomfort; '*I think sometimes students cannot feel comfortable while they are presenting because they know many people are assessing them*'.*

The summative element of the PA framework in this research study may have heightened presenter stress, as this comment above perceptively noted. Liu & Carless (2006) observe that in such situations, "the audience for the learners work is no longer just the teacher, but their peers. Learners may resent the pressure, risk or competition peer assessments could easily engender"(287).

Items 4 and 5 clearly show levels of student discomfort in peer assessing classmates and being assessed by them. Previous peer assessment research has shown that students often dislike having some degree of power over their peers or peers having power over them (Falchikov, 2000). Comparing responses to items 5 and 6 on the survey, it seems that students were less comfortable acting as peer assessors and judging classmates (34% disagreeing) than being comfortable having peers assess their performances (25% disagreeing). Scoring peers performances seemed to cause greater student discomfort, perhaps heightened by the summative uses of peer scores. Students may not have wanted to be responsible for possibly lowering the grade of a classmate, and therefore increased student stress may have been caused by doing peer assessment, as compared to receiving it.

According to AfL theory, students learn when they become teachers and examiners of others, and in order for formative assessment to be valid it must lead to further learning (Black et al., 2003; Stobart, 2006). This brings us to responding to the second question focused on in this case study: Did further learning take place as a result of the PA process implemented? We can form a response to this question by examining, in particular, student responses to survey items 7 ('Assessing other students' presentations helped me plan and deliver my own') and 8 ('PA scores and comments from my first presentation helped me prepare my second presentation').

For **item 7**, 60% of students agreed that rating their peers presentations helped their own presentation planning and delivery. This number was the second highest 'agree' score of all

12 items on the survey. A further 36% tended to agree, leaving only two students (out of 53) tending to disagree that PA was beneficial in this way. Student written commentary on part 3 of the survey reflects the positive responses to this item. For example, one student commented: *PA is helpful for both presenter and rater. It helps both improve their skills.* This student view reflects the fact that having the opportunity to apply assessment criteria to the work quality of peers is likely to lead to improvement in the quality of the rater's work also (Gibbs, 1999). Similarly, the experience of commenting on the work of peers helps students develop some degree of objectivity in relation to assessment criteria, which can then be applied to their own work (Nicol & MacFarlane-Dick, 2006). Rust, Price, & O'Donovan (2003) showed that engaging students with marking resulted in a significant improvement on their grades in similar assignments. Totaling 96% of survey respondents in agreement with item 7, clearly students thought that being a peer assessor was helpful in planning and delivering their own presentations.

The formative/summative tension of this PA framework has already been noted, and it was hoped that the summative use of PA scores would not detract from the assessment for learning potential of the processes put in place. According to Black et al. (2003):

. . . it is essential to keep the formative use in the forefront; a new practice might help collect better information about students thinking, but if it stops short of using, or of showing students how to use that information to improve each students learning, it misses the whole point of the exercise. (p. 109)

The formative use of this PA instrument was kept in the forefront, and through students both informally (in mini-presentations to study groups) and formally (in mid-term and final presentations) assessing their peers, the PA process did help improve student learning. If the main validity check for AfL is the successful support of student learning (Stobart, 2006; Gardner, 2006), then student views indicate that this PA process met this crucial marker.

Item 8 on the survey asks students whether feedback from peers (scores and comments) on their first presentation helped them prepare for the second one. More than half (56%) agreed that this had happened, while 38% tended to agree that peer feedback was helpful in this way. One student wrote (verbatim): *The comments in PA are more helpful and effective to make a better presentation. Not only score, but also peer comments are very important, I think.* Thus, the mid-term presentation feedback from peers (and the teacher) feeds forward because it can help improve future student performance on an assessed task (Carless et al., 2006) – the final

presentation.

Because it involved scores being used for final grades, the summative purpose of this PA process is evident. But the formative purpose, and the washback effects promoting, and not just measuring, student learning are also strong here. Stobart (2006) warns teachers to be careful of assessments that are intended to be formative, but in practice are not because they do not generate further learning.

Student responses to items 7 and 8 on the survey help to conclude an affirmative response to the second key question of this investigation; that for some students at least, this PA framework was ‘assessment for learning’ and did indeed promote and encourage student learning of effective public speaking skills. According to Stiggins (2007), when students participate in the thoughtful analysis of quality work: they become better performers; they better understand shortcomings in their own work; take responsibility for improving, and become conscious of their own improvement. This study shows evidence of students engaging in thoughtful analysis of the performance of peers, as well as their own work, and in turn becoming better performers as a result of this process.

Part 2. Student views about the PA process

The four items on part two of the student survey relate to the larger issues of student involvement in peer assessment and using peer markings for summative grading purposes. We will first deal with the issue of student views on the use of PA scores in final grades (in survey items 10 and 11).

As noted, the question of whether assessment by peers should form a significant part of students’ final grades is a contentious issue in the PA literature. Survey **item 10** elicits student perspectives on the issue, stating: ‘Making PA scores a part of student final grades is a good idea’. Overall, students favored this proposition with 85% agreeing (45 of 53). However a majority of these, 59%, selected the ‘tend to agree’ response, indicating some degree of reservation about this summative use of PA scores. It should be remembered that by the time of survey completion, students had gone through two cycles of PA, and their responses reflect student awareness and experience with some of the potential PA problem areas already noted.

Item 11 on the survey asked student to express their opinion about having PA worth 30% of their final grade; whether this number was too high, a fair amount, or too low. A majority

of students, 75%, (40 of 53) viewed this percentage as fair, while 25% thought it too high. Some students indicated that they thought the 10-20% range would have been a better PA contribution to final grades, as the following student comment (verbatim) shows: '*I like PA process. After I judged my peers, it is effective for my presentation. But 30% is a little bit high. 15% - 20% is good I think.*' In hindsight, I would concur with this view, and if repeating this experience with similar classes in the future would keep PA in this percentile range, after considering some of the scoring issues that became apparent in this study.

Items 9 and 12 on the survey seek student views regarding their involvement in the assessment process. **Item 9** presents a negative view of PA for students to respond to ('Students should not be involved in assessing peers; assessment should be solely the teachers responsibility'). A total of 83% of students disagreed with this statement. Responses to this item show student understanding of the potential benefits of their involvement in the assessment process, compared with the traditional teacher-only assessment format. The following student view was also expressed by similar commentary from other students on the survey: '*It is good way to know how my friends or other students think about my presentation. Moreover, I can know many advices or opinion, not only from teacher, but also students'* (sic). One of the key reasons for using PA is that it provides a way of getting much more feedback to students as compared to a sole teacher assessment; "swifter feedback in greater quantity" (Topping, 1998, p. 255).

Some 17% (9 of 53) tended to agree that assessment should be the sole domain of the teacher. The survey does not examine reasons for this view, but presumably these may include some of those previously mentioned. There may also be a cultural element to some students' dissatisfaction with peer assessment. Stobart (1996) writes:

The culture of schooling will also impact on the effectiveness of formative assessment. Entrenched views on teaching and learning may undermine or support formative assessment, as might deeply embedded assessment practices. For example, in a culture where the dominant model of teaching is didactic, moves towards peer and self-assessment by learners may involve radical and managerially unpopular, changes to the classroom ethos (p. 137).

Japan is a culture with such a didactic model of teaching, and assessment has been and remains dominated by teacher-only practices. Yet, ironically, this fact may also be responsible for the positive attitude of some students towards peer-assessment; the refreshing change of being active participants and decision-makers in the assessment process and learning from it.

Finally, **item 12** on the survey asked if students would recommend using PA in future public speaking class. An overwhelming 96% agreed, yet 38% of these chose the 'tend to

agree' response. Despite reservations about some elements of the PA process, responses here show that it was mostly a positive assessment experience for the students in the EPS course, and they feel that future classes also should have similar opportunities to engage with and learn from peer assessment.

Conclusion

While students can learn from being assessed by their classmates, peer-assessment is more about learning than about assessment and the key actor is not the person being assessed but the peer making the judgments (Liu & Carless, 2006). The PA investigation by Langan et al. (2005) concluded that "benefits of learner inclusion and active learning dimensions merit [peer-assessment] inclusion in future courses" (p.31). Considering students perspectives on PA usage in public speaking classes, this research study reaches a similar conclusion. One student had the following comment (verbatim):

Actually I wasn't comfortable to PA, but I think it is really a good system. It is new idea for me, and the score, advice, comments my peers gave me was very helpful and correct. So I want you to continue this system.

The student survey generated a range of student perspectives and commentary about peer assessment, and the particular version of PA used in the EPS course. Overall, student feedback may be viewed as quite positive, with many students expressing a liking and satisfaction with the inclusion of such a participatory assessment model as part of the course. Also, perhaps most importantly, survey responses seem to show that for many of the students involved, the PA process did indeed help support and promote student learning about constructing, delivering and judging effective presentations.

Despite the potential problems that may occur, the pedagogical and practical arguments for incorporating peer assessment into course frameworks are strong and clear, particularly in such performance-based assessment as discussed in this study. With careful attention to design and implementation, the 'learning from assessing' that results for students make up for the efforts made and problems encountered.

While Chapter 3 of this thesis examines more closely the issue of using summative assessment for formative purposes, that is, in fact, what also occurred in this PA research study. Summative assessment was indeed an important purpose of the PA framework used in

this investigation. However, as noted, the formative element was also fore-fronted so that the promotion of learning was just as important as the summative scoring of presentations by peers. Black et al. (2003) observe,

In general, students' learning can be enriched by marking their own or one another's work, whether this be classwork, homework, test scripts or presentations to the class. Students learn by taking the roles of teachers and examiners of others. (p. 51)

This PA process saw students taking on these teacher and examiner roles and this helped to enrich their learning about effective public speaking and presentation construction. The teachers' use of these peer scores for summative purposes does not cancel out the formative benefits that were also incorporated into the assessment framework for the EPS course.

As already noted in the first part of this chapter, achieving student metacognition and developing their power to guide and oversee their own learning is the overall aim of self- and peer-assessment (Black & Jones, 2006). This process of students 'piloting their learning' is evident in the two research studies described in this chapter, as students engage in the construction of understanding and development of skills through what they do in the AfL procedures that they work with, and the classroom roles they take on.

Appendix A1 Six students' commentary on Section 2 of SA score sheet

[These comments were taken from two students in each of the three classes in the study. These particular student commentaries were selected as they reflect the positive potential of the SA framework. They are presented here verbatim.]

	October	November	December (final comments)
Mari -->	<i>I sometimes use Japanese so I will try not to use Japanese from now on. I think my grammar is often wrong, and I sometimes confuse how should I say in English. So I want to speak English more well.</i>	<i>When I confuse how should I say in English, I sometimes use Japanese. I will do my best even if I don't know English from now on.</i>	<i>It was very hard for me to say my own opinions in English, but it is easier to say my ideas than before. I sometimes still confused to say, but I did my best. I enjoyed to communicate with classmates in English.</i>
Haruka -->	<i>I don't like English. But this English class is enjoy and like. So I want to English well and like English!! My aim is AI</i>	<i>I spoke Japanese...so I want to only English speech with everyone. And I like English a little. Until now, I negative English.</i>	<i>So, I wanted to talk many students and get many opinion. So I try, try, try!!</i>
Miho -->	<i>I want to build my vocabulary more, to speak what I want to ask smoothly. I will try hard not to use Japanese more and more.</i>	<i>(Compared with last month) I think I have been used to speaking about me and communicating with class members in English. I would like to communication even more actively.</i>	<i>I realized that the important thing is my attitude of speaking English and listening other student's story. I tried to speak a lot, I tried to listen carefully and ask many questions</i>

Appendix A1 (continued): Student commentary on Section 2 of SA score sheet

	October	November	December (final comments)
Yoshie -->	<i>I try to communicate in English every time but I often don't know English words and I can't explain my thoughts. I think I should study English words.</i>	<i>Recently, I think communicate in English is fun, but I can't talk enough because I know little English words so I think I should study English words.</i>	<i>Yesterday, some foreign customers came to my part-time job. I could explain in English. I thought my communication skills are up by this class and I am happy about being an active student.</i>
Ayako -->	<i>I checked my cell phone during class. I want to stay focused on English more and join actively. I speak English in this class and I can speak fluently in the future.</i>	<i>I enjoy class. Gradually I can ask questions to the partner's in English more. I want to speak only English in this class. Moreover, I'd like to speak actively more.</i>	<i>I think that I can speak and ask questions more than before. Now I enjoy talking to friends in English. I also can say my opinions in English easily, so I want to develop my English skill more.</i>
Miyuki -->	<i>I think I can speak and say my opinion in English a little better than I was former term. But I can't still command English very well. I try to be able to speak English!! I do my best!!</i>	<i>I can speak my opinion and listen to my friends opinions more evenly than last time I wrote this paper. Before I was often embarrassed, so I try to do my best on and on!!</i>	<i>I think I can speak English better than in September. Before when I speak English I need many time to think what I am speaking. But little by little I can make this time shorter. This is my big development from Sept. to Dec.</i>

Appendix A2 Self-Assessment survey (without Japanese translation)

Please choose a number (1-6) which reflects your feelings about the self-assessment (SA) process we have used in CS class this semester. Do not write your name and be as **honest** as possible in your responses. Thank you.

1= **completely agree** 2= **mostly agree** 3= **slightly agree**

4= **slightly disagree** 5= **mostly disagree** 6= **completely disagree**

Feelings about Self-Assessment	I agree	I disagree
1. I understood why we used the SA score sheet in this class.	1 2 3 4 5 6	
2. I understood the six criteria on the SA score sheet	1 2 3 4 5 6	
3. SA criteria are helpful.	1 2 3 4 5 6	
4. SA of class participation is fair.	1 2 3 4 5 6	
5. SA is a waste of time.	1 2 3 4 5 6	
6. SA is easy to do.	1 2 3 4 5 6	
7. My self-assessment is an honest reflection of my efforts to actively participate and speak English.	1 2 3 4 5 6	
8. This SA process encouraged me to actively participate in class.	1 2 3 4 5 6	
9. Because of self-assessment, I spoke more English during class.	1 2 3 4 5 6	
10. I recommend using a self-assessment procedure with future CS classes.	1 2 3 4 5 6	

Other comments about SA (in English):

Appendix B1 Presentation peer-rating sheet (based on Yamashiro & Johnson, 1997)

Public Speaking class		Peer Rating Sheet																																																																																																																																																																							
Speakers Name: _____		Presentation topic: _____																																																																																																																																																																							
Score scale: 5 (very good) 4 (good) 3 (average) 2 (weak) 1 (poor)																																																																																																																																																																									
Circle a number for each category, and then consider the numbers you chose to decide an overall score for the presentation.																																																																																																																																																																									
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th colspan="6">1. Voice Control</th> </tr> </thead> <tbody> <tr> <td>1. Projection (loud/soft)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td>2. Pace (speech rate; fast/slow)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td>3. Intonation (patterns, pauses)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td>4. Diction (clear speaking)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> </tr> <tr> <th colspan="6">2. Body Language</th> </tr> <tr> <td>1. Posture (standing straight, relaxed)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td>2. Eye contact</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td>3. Gestures (well used, not distracting)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> </tr> <tr> <th colspan="6">3. Contents of Presentation</th> </tr> <tr> <td>1. Introduction (grabs attention, has main points)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td>2. Body (focused on main ideas, has transitions)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td>3. Conclusion (summary of main points, closing statement)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> </tr> <tr> <th colspan="6">4. Effectiveness</th> </tr> <tr> <td>1. Topic choice (interesting for audience)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td>2. Language use (clear, correct sentences/slide information)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td>3. Vocabulary (words well-chosen and used)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td>4. Purpose (informative, teaches about topic)</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> </tr> <tr> <th colspan="6">5. Visuals</th> </tr> <tr> <td>1. Effective use of slides to support presentation</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> </tr> <tr> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> </tr> <tr> <td style="text-align: center;">Overall Score</td> <td style="background-color: #e07070;">5</td> <td style="background-color: #e07070;">4</td> <td style="background-color: #e07070;">3</td> <td style="background-color: #e07070;">2</td> <td style="background-color: #e07070;">1</td> </tr> <tr> <td colspan="7"> <p><u>Comments about presentation (optional, in English):</u></p> </td> </tr> </tbody> </table>							1. Voice Control						1. Projection (loud/soft)	5	4	3	2	1	2. Pace (speech rate; fast/slow)	5	4	3	2	1	3. Intonation (patterns, pauses)	5	4	3	2	1	4. Diction (clear speaking)	5	4	3	2	1							2. Body Language						1. Posture (standing straight, relaxed)	5	4	3	2	1	2. Eye contact	5	4	3	2	1	3. Gestures (well used, not distracting)	5	4	3	2	1							3. Contents of Presentation						1. Introduction (grabs attention, has main points)	5	4	3	2	1	2. Body (focused on main ideas, has transitions)	5	4	3	2	1	3. Conclusion (summary of main points, closing statement)	5	4	3	2	1							4. Effectiveness						1. Topic choice (interesting for audience)	5	4	3	2	1	2. Language use (clear, correct sentences/slide information)	5	4	3	2	1	3. Vocabulary (words well-chosen and used)	5	4	3	2	1	4. Purpose (informative, teaches about topic)	5	4	3	2	1							5. Visuals						1. Effective use of slides to support presentation	5	4	3	2	1							Overall Score	5	4	3	2	1	<p><u>Comments about presentation (optional, in English):</u></p>						
1. Voice Control																																																																																																																																																																									
1. Projection (loud/soft)	5	4	3	2	1																																																																																																																																																																				
2. Pace (speech rate; fast/slow)	5	4	3	2	1																																																																																																																																																																				
3. Intonation (patterns, pauses)	5	4	3	2	1																																																																																																																																																																				
4. Diction (clear speaking)	5	4	3	2	1																																																																																																																																																																				
2. Body Language																																																																																																																																																																									
1. Posture (standing straight, relaxed)	5	4	3	2	1																																																																																																																																																																				
2. Eye contact	5	4	3	2	1																																																																																																																																																																				
3. Gestures (well used, not distracting)	5	4	3	2	1																																																																																																																																																																				
3. Contents of Presentation																																																																																																																																																																									
1. Introduction (grabs attention, has main points)	5	4	3	2	1																																																																																																																																																																				
2. Body (focused on main ideas, has transitions)	5	4	3	2	1																																																																																																																																																																				
3. Conclusion (summary of main points, closing statement)	5	4	3	2	1																																																																																																																																																																				
4. Effectiveness																																																																																																																																																																									
1. Topic choice (interesting for audience)	5	4	3	2	1																																																																																																																																																																				
2. Language use (clear, correct sentences/slide information)	5	4	3	2	1																																																																																																																																																																				
3. Vocabulary (words well-chosen and used)	5	4	3	2	1																																																																																																																																																																				
4. Purpose (informative, teaches about topic)	5	4	3	2	1																																																																																																																																																																				
5. Visuals																																																																																																																																																																									
1. Effective use of slides to support presentation	5	4	3	2	1																																																																																																																																																																				
Overall Score	5	4	3	2	1																																																																																																																																																																				
<p><u>Comments about presentation (optional, in English):</u></p>																																																																																																																																																																									

Appendix B2 Student survey: Peer assessment of presentations

(**Note *** A sample copy of the PA sheet was distributed to students while responding to this survey as a reminder of the PA criterion and the process itself.)

During this Public Speaking course, as well as planning, organizing and delivering two presentations, you have also been asked to assess the presentations of your peers. I am interested in student views of this peer assessment (PA) process. Please look at the sample peer-rating sheet again, consider the following statements, and respond in a way that **honestly** reflects your views. **Thank you for your feedback.**

Choose one of the following numbers and write it after each statement:

1 = agree 2 = tend to agree 3 = tend to disagree 4 = disagree

(Note: for item **number 11** below, please circle the letter.)

Part 1: Being a rater/ being rated by my peers

1. Assessment items on the sheet (e.g. pace, language use) were easy to understand. __
2. It was difficult to decide the overall score (5, 4, 3, 2, 1) for each presenter. __
3. Relationships with presenters (friendships, etc.) may have influenced the overall scores and comments I gave. __
4. I was comfortable being a judge of my peers' presentations and giving a score. __
5. I was comfortable having my presentations judged and scored by my peers. __
6. The overall scores my peers gave me were fair and reasonable. __
7. Assessing other students' presentations helped me plan and deliver my own presentations. __
8. PA scores and comments from my first presentation helped me prepare my second presentation. __

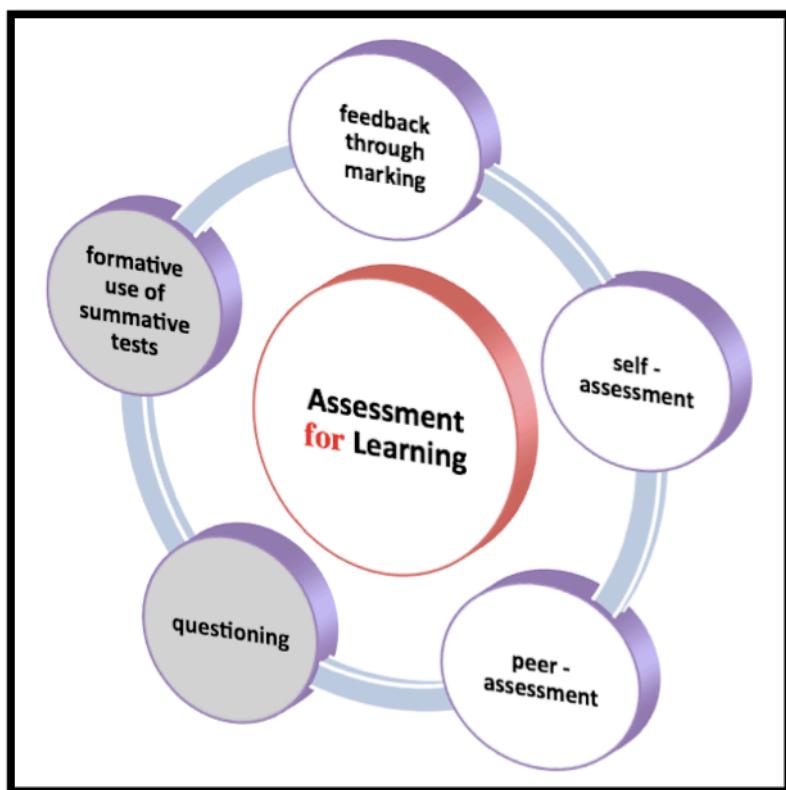
Part 2. The Peer assessment process

9. Students should not be involved with assessing their peers. Assessment should be the sole responsibility of the teacher. __
10. Making PA scores a part of student final grades for the course is a good idea. __
11. Making PA worth 30% of the final course grade is:
 - a) too high
 - b) a fair amount
 - c) too low
12. I recommend using PA in future Public Speaking classes. __

Part 3. Do you have any other comments on the peer assessment process? (English)

Chapter 3: Student as Questioner, and Assessment Synergy

- **Part A:** Student-generated questioning for learning
- **Part B:** Synergizing formative and summative assessment of presentation slide shows



Preview

This final thesis chapter includes research reports focused on the two remaining key components of the AfL framework.

Part A focuses on the issue of questioning as an informal way of assessing student knowledge and understanding about a topic. The research reported on here takes a different approach by looking at student, rather than teacher, question-making. While the topic of teacher questioning has been widely studied and written about in classroom-based research, questions generated by students themselves has received less attention. In ESL/EFL contexts, reports of such research are virtually non-existent. This report describes an investigation into student-generated questioning in a course called Communicative Writing. In a writing class of 22 students, a student questioning process was introduced, based on an Internet documentary film and tape script, for the purpose of determining whether student questioning and subsequent peer discussions promoted understanding of class material.

In Part B of we take a closer look at an issue that was an element of the peer-assessment report in the previous chapter; the integration of formative and summative assessment to maximize student learning. This report of AfL in practice describes one such example with 20 junior students in a public speaking course. Students' computerized slideshows for their final presentations were separately evaluated a total of three times, including a student self-assessment and teacher assessment of slideshow first drafts, prior to the final graded summative assessment. A similar version of the same criterion-referenced rubric was used for the assessment trio making up this process. The synergized assessment regime is described in detail, with student examples from the three assessments provided.

Part A. Student-generated questioning for learning

... what is needed is a classroom culture of questioning and deep thinking, in which pupils will learn from the shared discussions with teachers and from one another. (Black & Wiliam, 1998b, p.13)

Introduction

Classroom dialogue, and the interactions and exchanges that take place during lessons can play a significant role in the learning that takes place, and also reveal the dynamics of the class itself. As Dufficey (2005, p.76) points out, “it could be said that the story of a classroom is told through its talk”. The catalysts for much of this talk are the questions posed and responded to during a lesson. The importance of questioning as a pedagogical tool has long been recognized, researched, and written about (Glassen, 2009). With regard to assessment, teachers use a number of different ways of informally assessing what students know and can do, including questioning, discussion, dialogue and listening to students talk (Harlen, 2007).

Question production in the classroom is not a usual student role (Chin & Brown, 2002), but rather it is the teacher, or the textbook, which asks the questions. Supon and Wolf (1993) describe the typical student and teacher roles in classroom question generation as follows:

Despite nearly universal agreement that students need to do more, think more and be more active in the classroom, most classrooms are still firmly teacher-centered. While teachers use various questioning strategies to develop and enhance critical thinking, historically they are the generators of the questioning process. Thus many students are never taught the skills for generating their own questions. (p. 1)

This thesis section is a report of an investigation into teaching students to produce and ask their own questions. It tells the story of one classroom by examining the student talk that was generated as a result of the question production and peer discussions students engaged in. Supon and Wolf (1993) remind us that encouraging students to ask questions is a skill that teachers must also acquire. In the report that follows, (as in the thesis as a whole) the teacher is also considered a learner in the classroom community. Prior to setting out the research purposes of this study, we begin with a brief look at the role of questioning in AfL, and an explanation of what student-generated questioning entails.

Questioning in the AfL framework

Along with feedback through marking, student self-and peer assessment and the formative use of summative tests, Black et al. (2003) include questioning as one of the four areas for teachers to focus on to improve their formative assessment practices. Reporting on teachers' attempts to improve questioning strategies in their two-year research project in UK schools, Black et al. (2003) write:

Questioning became an essential feature of teachers' classrooms as questions were devised and used to promote thinking. This led to richer discourse, in which the teachers evoked a wealth of information from which to judge the correct understanding of their students. More importantly they had evidence on which to plan the next steps in learning. (p. 41)

The focus of discussion for Professor Black and his colleagues is primarily on improving teacher questioning skills with regard to such things as framing effective questions and increasing 'wait time' for student responses. They also go on to point out that for teachers "the task of improving questioning is a complex and gradual process" (Black et al. 2003, p. 41).

Questioning in AfL, as in the wider pedagogical arena, is also referred to as 'strategic questioning'. In asking about the role of questioning in assessment for learning, a partnership of Education Ministers in Australia, the Curriculum Corporation (2004), provides the following answer shown in Table 1 below.

Table 1: *The role of questioning in AfL* (Curriculum Corporation, 2004, p.1)

What is the role of Questioning in Assessment for Learning?
Strategic questioning is one way in which the teacher can seek evidence to establish where students are in their learning, and is therefore the result of careful planning.
Strategic questioning provides information about student knowledge, understanding and skills that informs the teacher's planning and selection of teaching strategies to move students from where they are to where they need to go.
Specifically, strategic questioning provides teachers with the opportunity to identify and correct misunderstandings and gaps in knowledge, as well as identify the need for extension work for those students whose knowledge and skills base demands it.

Glasson (2009) describes the connection between questioning and formative assessment as “the careful and deliberate use of questioning in order to elicit information from students about what it is that they know and can do, and the formative use of that information to shape future teaching and learning” (p.41). He reports that research has identified a number of key aspects of questioning which contributes to its value as an assessment for learning tool. These include the teachers’ ability to do the following:

- identify, as part of planning, the key questions that are the focus of the lesson
- use open questions, which demand higher-order thinking
- provide for ‘wait time’ or ‘thinking time’ to allow students time to consider a question before they offer their responses
- use prompts to encourage students to produce a response or to elaborate on a response
- make use of answers that display faulty thinking
- model positive listening behavior
- distribute questions around the classroom
- encourage students to ask questions (Glasson, 2009, p. 40)

The reader will note that all of these, except the final one, are teacher-centered actions. This report focuses specifically on the last of these aspects, and the process of making question generation a student-centered activity and responsibility.

While teacher questioning of students is intended to informally assess their level of understanding, questions generated by students and discussed with peers is intended to promote and maximize student understanding; move them to a more complete or clearer understanding of the subject or material at hand. In a classroom assessment guide for teachers produced by provincial education departments in Canada , we are reminded that “Students understanding can be exposed not only through their response to the teachers’ questions, but through the questions they formulate to advance their understanding” (Rethinking Classroom Assessment, 2006, p. 30). Interestingly, the AfL literature focuses very little on student-generated questions, being concerned much more with such teacher-centered issues as identifying key questions and the other items in the Glasson (2009) list above.

In the UK, the Assessment Reform Group (2002) set out ten principles of effective classroom practice. They include the following passage:

Assessment for learning should be recognized as central to classroom practice. Much of what teachers and learners do in the classrooms can be described as assessment. That is, *tasks and*

questions prompt learners to demonstrate their knowledge, understanding and skills. What learners say and do is then observed and interpreted, and judgments are made about how learning can be improved. These assessment processes are an essential part of everyday classroom practice and involve both teachers and learners in reflection, dialogue and decision-making (emphasis added).

These questions that prompt learners to show what they know, understand, and can do may also be created by students themselves. Indeed, the question-generation process itself can aid in the development of learning. The reflection, dialogue and decision-making mentioned above, instigated by student questions, are the subjects of this report.

Black and Wiliam (1998b) concluded that any piece of teaching should include opportunities for students to express their understanding because such opportunities generate the kind of interaction in which formative assessment aids learning. They noted, “Discussions, in which pupils are led to talk about their understanding in their own ways, are important aids to improved knowledge and understanding” (Black & Wiliam, 1998b, p. 11). In examining ways of gathering evidence about students’ ideas and skills, Harlen (2007a) writes:

Pupils' talk is a very valuable source of evidence about their thinking. It is also an important means of helping the development of deep understanding . . . The words that pupils use in talking with one another often give evidence of their ideas. Therefore, it is particularly useful to set up a situation in which pupils converse with each other while the teacher ‘listens in’ without participating in the discussion (p. 118).

As we shall see, setting up such a situation of student-student discourse, intended to assist the development of understanding while the teacher eavesdrops and informally assesses this understanding, is a good description of the investigation engaged in here. The catalyst for this learner interaction is not teacher-generated questioning, but questions constructed and discussed by the students themselves.

Defining student-generated questioning (SGQ) and its uses

Student-generated questioning (SGQ) has been defined as “self-generated requests for information within a topic or domain” (Taboada & Guthrie, 2006, p. 1). In this investigation ‘information’ may also be considered to include requests for opinions and feelings about a

topic/issue. Providing a theoretical framework for student questioning, Ciardello (1998) explains:

It reflects the constructivist view that learning occurs in goal-embedded contexts and is reconstructed through the reciprocal exchange of views and experiences between teachers and students, and students and peers. Training students how to generate questions is an important strategic plan for helping all students think and communicate. (p. 219)

While also noting the constructivist foundations of SGQ, King (1992) writes that “the focus of the strategy is on stimulating students to generate meaning for the material being learned, by making inferences, drawing conclusions, and generally elaborating on the material” (p. 120). While conceptually, the questions generated by students indicate their interests, assumptions and things they are perplexed by, educationally these questions are also *informal diagnostic devices* which can be used by teachers to judge level of understanding and effectiveness of instruction (McDevitt, 1994, p.31). Such ‘informal diagnostic devices’ can also be used by students to assess their own level of understanding, and also that of their peers.

In the context of a reading course, Kiddey and Warring (2001) reported that student generated (SG) questions can serve a number of purposes, including: to improve both the quality and type of questions a reader asks, to promote active meaning-making (predictions, thinking through ideas, substantiations, etc.), to promote student-centered learning and to provide teachers with instant feedback about areas of difficulty. As the teacher is ‘listening/watching in’ on this SGQ process and the subsequent peer discussion, some of the skills and understandings that teachers may observe and assess include:

- the level of understanding of subject content and concepts;
- the quality of the questions students ask;
- the willingness of students to take risks;
- the level of students critical thinking skills;
- the degree to which students are able to refer closely to the text in order to substantiate their interpretation;
- the extent to which students can develop and support their ideas; and
- the extent to which students can use contextual clues to guess meaning

(Kiddy & Warring, 2001, p. 38)

Considering such a list of skills and understandings tied up with SGQ, the process, despite its outward appearance of students simply making and responding to questions, is both dynamic and complex.

SGQ is just the first part of the classroom dialogue process described here. The secondary and complementary move in the process is identified by King (1992) as ‘student-generated peer questioning’ (hereafter known as PQ), in which students in small groups pose the questions they made to their peers and answer each other’s questions. In effect, two activities are involved in this process described here: student-generated questioning (SGQ) followed by peer questioning (PQ) in small groups.

Research purpose

According to Glasson (2009, p. 7), “the experience of teachers who have introduced assessment for learning strategies into their classrooms indicates that it is best undertaken slowly, allowing time for experimentation and consideration”. This investigation into SGQ is considered as an exercise in ‘experimentation and consideration’; it is the first time I have experimented with, indeed carefully considered, the role of questioning in the classroom. In their two-year investigation into implementing AfL into classrooms with UK teachers, Black et al. (2003) reported that as researchers they encouraged the 36 teachers involved to experiment with some of the strategies and techniques suggested by the formative research literature as being effective. As noted in the thesis introduction, individual teachers were then asked to draw up an action plan of the practices they wanted to develop and identify a single focal class in which the strategy would be introduced. A similar procedure was followed here; an experimentation with one particular strategy (SGQ), implemented into one focal class.

While a more detailed account of this investigation will be provided later in the Methodology section, a succinct overview may prove useful here. Very briefly, as part of a writing class students were required, prior to class, to read sections of a documentary tape script which they later watched in class. The tape script and the film sections were used as a basis for student question generation. These questions subsequently formed the basis of classroom dialogue in small PQ discussion groups. The intention was that SGQ, and the following PQ session, would promote student understanding of the material, and would be

helpful to them in writing the follow-up essay based the documentary.

This investigation into the effects of incorporating SGQ into a course curriculum took place over a six-week (and six class) period. It was conducted with one simple research consideration in mind:

- Does student generated questioning promote learning about the course material? (i.e. did further learning take place as a consequence of the SGQ process, and subsequent peer discussions?).

The investigation should provide some evidence, primarily from student feedback on the process, as to whether learning was promoted. This research question is a crucial one because, as Gardner (2006, p. 200) reminds us, “the basic tenet of successfully supporting learning remains the main validity check for assessment for learning”.

Before explaining in detail the nature and context of the investigation reported on here, we first complete this introductory section with a review of the student questioning literature.

Literature review

Questioning is one of the most extensively researched areas of teaching and learning (DfES, 2004). While the literature on teacher questioning is wide-ranging, relatively little research has been done on the issue of student questions; most of which has been conducted related to reading, text-based questions and prose processing (Chin & Brown, 2002). According to King (1994a), question generation by students has also been found to promote learning from classroom lectures, tutoring and problem solving. The research into SGQ in language teaching and the ESL/EFL classroom appears to be virtually non-existent, as an extensive trawl through the related literature has not uncovered any published studies in this particular area. However, much of the research and writing about student questioning in mainstream education, and the ideas and procedures included, help provide a supporting framework for this investigation. The reader is reminded that the research discussed here, representative of the literature as a whole, involves students constructing questions *in their first language*.

In writing about the association between reading comprehension and student questioning, Taboada (2003) summarizes a number of reasons why student questions are educationally important. These are presented here in Table 2.

Table 2. *Reasons why SGQ are important in education* (from Taboada, 2003, pp. 2-3)

Educational importance of student questions
1. Student-generated questions constitute a centerpiece for student learning in a range of content domains. Asking questions within a knowledge domain or in reference to a specific topic is a useful cognitive strategy to facilitate learning.
2. Student questioning signals independence for the learner since questioning is an adaptive action of the student that helps regulate the student's own learning.
3. Student-generated questions constitute a cognitive phenomenon that can reveal important aspects of a learner's knowledge structure.
4. Student questioning relates to complex inquiry-based tasks such as problem solving, . . . Successful reading has many affinities with problem solving, so questions asked before or during reading a text may be informative data for monitoring the reader's understanding of a text.
5. Questions can also be described as intentional acts. A question posed by a student in a learning situation may be indicative of what that student wants to know, as well as of what that student already knows.
6. Student-generated questions are valued in education-at-large because of the influence they have on the learning process.

While students can be question makers and askers in the classroom, it is the teacher who instigates and orchestrates this process of student question generation. In the way they structure classes, teachers have a great influence on when and if students generate questions or ask them (Karabenick, 1994). According to Morgan and Saxton (1994) student questioning should be an integral part of classroom culture, be built into the process of unit planning and development, and be encouraged and supported by teachers as an essential contributor to genuine student engagement.

A UK Department of Education and Skills guidance report (DfES, 2004) for teachers about classroom questioning tells us “Questioning is effective when it allows pupils to engage with the learning process by actively composing responses” (DfES, 2004, p. 18). We may understand that such effective questioning can be student generated. This teachers guide also reports that research shows that lessons including effective questioning will likely have a number of characteristics, including:

- Questions are planned and closely linked to the activities of the lesson
- Closed questions are used to check factual understanding and recall
- Open questions predominate

- Students have opportunities to ask their own questions and seek their own answers, as well as being encouraged to provide feedback to each other
- The classroom climate is one where pupils feel secure enough to take risks and make mistakes (Pedagogy and Practice: Questioning, DfES, 2004, p. 18.)

As we shall see, all of these characteristics of effective questioning use were evident in the SGQ investigation reported on here.

Student questioning is sometimes discussed in the literature as being a *cognitive strategy*, defined as: “a guided learning procedure for internalizing new information and performing higher level thinking operations” (Ciardello, 1998, p. 210). Personal control of cognitive strategies, such as with student questioning, is an essential element in their effectiveness and long-term use (King, 1994a). In a review of studies related to teaching students to ask questions, Roshine, Meister and Chapman (1996) write the following regarding student questioning as a cognitive strategy:

A cognitive strategy is a heuristic. That is, a cognitive strategy is not a direct procedure or an algorithm to be followed precisely but rather a guide that serves to support learners as they develop internal procedures that enable them to perform higher-level operations. Generating questions does not lead directly in a step-by-step manner, to comprehension. Rather, in the process of generating questions, students need to search the text and combine information, and these processes help students comprehend what they read. (p. 182)

Effectively implemented, question generation by students fosters comprehension of the material at hand and promotes active processing of ideas and information. According to Chin and Brown (2002), student questioning plays a significant role in meaningful learning and is an important cognitive strategy because “the act of composing questions focuses the attention of students on content, main ideas, and checking if content is understood” (p. 521).

In summarizing the literature on reading comprehension and student questioning, Taboada and Guthrie (2006) report that studies show that a wide range of students can learn to generate questions about a text and that such questioning does foster reading comprehension. They further explain that the reasons for this relationship between reading comprehension and student questioning have been identified as being due to: (a) active text processing, (b) knowledge use, and (c) attentional focus (Taboada & Guthrie, 2006, p. 2). Similarly, Anderson and Armbruster (1984) posited, “It seems plausible that when student questioning is effective, it is so because students are forced to encode the information more than they might

if they simply read it" (p. 672). They also noted that the actual writing of questions required students to paraphrase or perform some other transformation of the text, thus involving further processing of the material.

The leading researcher and writer on about SGQ in mainstream education, King (1992, 1994a, 1994b, 1999, 2002) points out that in contemporary models of learning, new material is best understood and remembered by individuals when they elaborate on the material in some way. SGQ is one such cognitive strategy for elaborating on material. King (1992) writes:

Students have to think critically about the material just to be able to formulate . . . relevant, thought-provoking questions. Formulating high-level questions based on the presented content forces students to identify the main ideas presented and think about how those ideas relate to each other and to the students' own prior knowledge and experience. Responding to others' (or their own) questions further extends such high-level thinking. When students think about and elaborate on course material in these ways (vs. simply memorizing information as it is presented), they process the ideas more thoroughly and construct extensive cognitive networks connecting the new ideas together and linking them to what they already know. Developing such cognitive representations of the new material facilitates understanding. (p. 114)

King's research on student questioning particularly promotes 'guided student-generated questioning' (1992, 1994a, 1994b) in which students use a set of generic question stems (e.g. "How are --- and --- similar?" or "Why is --- important?") that prompt them to generate specific discussion questions on the material presented. Student pairs or groups then work cooperatively to answer the questions in extended discussions. Through her research, King concluded, "questioning strategies can be used to facilitate the knowledge construction process, which in turn enhances learning" (1994b, p. 364).

Student questioning in the classroom is affected by contextual, social and personal factors. According to Van Der Meij (1994), research suggests that a students' personal characteristics (e.g. self-esteem, motivation) can have a significant effect on their questioning. While also noting that student questioning is dependent on favorable social conditions in the classroom, Van Der Meij (1994) explains, "Questioning often is a very personal affair. It is intimately bound to a student's prior knowledge and skills, and to his or her motivation and volition" (p. 155).

Reflecting perceptions of questioning within the AfL literature, Chin and Brown (2002) also note that student questioning can help guide teachers in the classroom because such

questions can indicate what students have been thinking about the material presented. According to these authors, student questioning can also reveal much related to: the quality of students thinking and conceptual understanding; confusion about various concepts; their reasoning about a topic or subject; and what they want to know. In ending their review of the student-generated questions literature, Chin and Brown (2002) conclude as follows:

The literature review indicates that there is substantial educational potential in student-generated questions beyond that envisaged by research on text-based questions where the focus is on developing comprehension strategies. Rather, students' questions can be used to direct their inquiry and guide construction of knowledge. (p. 525)

In mainstream education in Suffolk, England, teachers carried out action research about encouraging student questioning. They concluded that while students were developing questioning skills they were also:

- developing independence, relying less on the teacher and asking each other more
- taking responsibility for their learning rather than being directed
- working through difficulties rather than automatically asking for help
- able to explain and express themselves more easily
- thinking about what they were trying to achieve by asking questions
- seeking explanations and alternatives more frequently
- reflecting on/evaluating their own understanding and often taking it further.

(Suffolk Advisory Council, 2001, p.21)

This Suffolk report, focused on using questions to help students learn, also makes the important point that if students are to be active questioners they need to practice the skills involved. For teachers, this means providing opportunities for creating questions in various ways and by using a range of stimuli and resources. The report further reminds teachers that it is helpful if there is a clear purpose for the questions generated by the students, either as groups or individuals.

With regard to questioning in the AfL literature, the Suffolk report follows the trend in the larger field, with a focus on teacher questioning and little mention of student questioning in formative assessment. Black et al. (2003) focus on improving the use of question-and-answer dialogue in the classroom by teachers doing such things as improving their own questioning skills, creating a supportive climate for student answers and increasing wait time for student

responses. The main suggestions for teachers arising from the Black et al. (2003) research was to increase wait time, provide rich follow up activities, and more teacher effort on making better questions. They conclude the questioning section of their book by asserting “Put simply, the only point of asking questions is to raise issues about which the teacher needs information or about which the student needs to think” (Black et al., 2003, p. 42). It is the latter part of this assertion, questions about which the student needs to think (to foster understanding) that is of primary importance with regard to SGQ. It was hoped that such questioning would promote student learning about the course material, and subsequently be beneficial in constructing an essay about the topic. We now turn to take a closer look at the context and methodology of the research this report describes.

Methods

This section of the report will focus on three areas in order to provide a full picture of the context and nature of the investigation that was undertaken. The three areas are: the course, the additional materials used on which SGQ was based, and the nature and timeframe of the investigation itself.

The Course

The name of the single focal class chosen to experiment with SGQ was one called Communicative Writing (CW). This is an elective course in the Department of English at TWCU, one open to students from other departments. The two-semester course focused on developing students' abilities in writing in English to communicate their thoughts, feelings and opinions. In the first semester of the CW course focused on here, there were 22 students. These were second and third year students from various departments, such as English, Sociology, Psychology, and had various levels of writing proficiency in English. As usual for TWCU courses, classes were held weekly for 90 minutes, for the 15 weeks of the semester.

The CW course was intended to enable students to improve their writing proficiency in English. Its specific purposes were:

- to improve students' writing skills through writing activities and student interaction
- to increase students' awareness of themselves and to use writing as a means of self-discovery
- to provide a range of readings that serve as springboards to effective writing

The course was thematically organized, focusing on the influences that shape peoples' lives. Students read and wrote about such influences as decisions, other people, events, cultures, media, work and technology. Course assessment was based on attendance and class participation and three short (2-3 page) essays.

The CW course was largely based on the course textbook entitled *The Writers Selection-Shaping Our Lives* (McWhorter, 2006a). The textbook is a thematic reader targeted for 'mid-level writers', and has a series of topical readings with an accompanying apparatus that includes pre-and post writing activities, as well as essay writing assignments. The thematic readings about events, people, decisions, et cetera that shape peoples lives serve as a springboard to the teaching and learning of essay writing. Particularly relevant to this investigation into student-generated questioning, each of the readings in the textbook was accompanied by questions. These textbook-generated questions were separated into two categories, and explained as follows in the accompanying teachers manual:

1. **Finding meaning.** These questions guide the students in grasping the literal content of each essay. Answering these questions enables students to assess their understanding. Students either confirm that they understand the key points or realize that they have not read as carefully as necessary, in which case a closer reading is encouraged.
2. **Thinking critically.** These questions engage students by provoking thought, sparking lively discussion, and fostering critical analysis. They may be used as collaborative activities or as brainstorming sessions to generate ideas about the reading. (McWhorter, 2006b, p. 2)

As we shall see, the questions generated by students were modeled on these two types of textbook-provided questions; finding meaning and thinking critically.

It should be noted that an important part of the CW course was classroom discussions, in pairs and small groups, about the readings and related issues. Such peer interaction in English was intended primarily for pre-writing purposes, to stimulate thinking about the topic, but also served as an opportunity for students to 'exercise their communicative English muscles'. Just as for writing skills, the 22 students in class varied in levels of spoken proficiency in English and ability to discuss the issues presented in the readings.

Additional Materials (online documentary)

After exploring a number of themes and readings in the class text, and having students write two essays, the final part of the course focused on the textbook section related to technology that shapes peoples' lives. After completing and working on two of the textbook readings related to this theme I decided to use a supplementary source, an on-line documentary produced and filmed in the U.S.A entitled *Growing Up Online* (Frontline, 2008). This documentary served as a source for the student-generated questions created and used in class discussions. The classroom used for the CW course is equipped with online access and both a wide-screen TV and individual student monitors for easy viewing of video material. After seeing the documentary, I decided it would fit nicely into the technology theme, and add a supplementary visual element to the end of the first semester of CW. I also decided that the *Growing Up Online* documentary (hereafter GUO) would be the source of the final course essay. Figure 1 below includes a brief summary to the documentary, taken from the online website.

FRONTLINE

WATCH THE FULL PROGRAM ONLINE

PRODUCER'S BLOG: Mom Evan Skinner responds to viewers. And share your ideas for a follow-up report.

□ Introduction

Just how radically is the Internet transforming the experience of childhood?

GROWING UP ONLINE

About the Film

FRONTLINE takes viewers inside the private worlds that kids are creating online, raising important questions about how the Internet is transforming the experience of adolescence. At school, teachers are trying to figure out how to reach a generation that no longer reads books or newspapers. Fear of online predators has led teachers and parents to focus primarily on keeping kids safe online. But many young people think these fears are misplaced. Online media has also intensified the social dimensions of adolescence as teens create and play with identities on sites like MySpace and Facebook and encounter intense peer pressure in a variety of virtual worlds. Parents are confused about how to respond to the increasingly private worlds inhabited by their children, lacking an understanding of both the creative potential and the genuine risks of this new dimension of our cultural environment.

(Frontline, 2008, <http://www.pbs.org/wgbh/pages/frontline/kidsonline/>)

Figure 1. Website introduction to the GUO documentary

The website helpfully provides a teachers guide and a tape script of film. Copies of the tape script were made and provided to the 22 students in class. The reader can get some understanding of the type of reading students were engaged in by looking at **Appendix A1**, which contains an excerpt from one page of the GUO tape script. The online documentary is conveniently divided into seven parts, each of approximately 7-9 minutes in length. The film was watched and discussed over a series of five classes, at the end of which students began to write an essay about it. Students were assigned to read sections of the tape script for homework. These parts, usually two per class, were part of the following classes' viewing and class discussion. During this period of five classes focused on the GUO film, students engaged in question generation about the issues followed by peer discussions using these questions.

While the textbook (and the teacher) provided comprehension and discussion questions for students at the beginning and middle of the CW course, in the remaining classes this online documentary served as the source material for questions generated and discussed by the students themselves. The GUO materials also served as a focus for my experimental investigation of SGQ within this writing class.

The investigation: SGQ

In the teachers' manual accompanying the course textbook, McWhorter (2006b) explains that in the writing classroom, collaborative learning has become an important vehicle, and points out that "recent research indicates that students can learn effectively from one another, especially in structured situations designed to facilitate the exchange of ideas" (p. 10). She also reminds teachers, that student groups need structure and direction to function effectively. After deciding the focal course (CW) and the source material (GUO) for experimenting with SGQ, I needed to make some decisions on how to introduce SGQ to the class, and how to structure the question-generation process and the subsequent peer discussions. Such decisions are important because, as noted by Black and Wiliam (1998b), it is what teachers and students do in the classroom that drives the learning.

If my goal was to utilize student questioning as a basis to promote comprehension and higher-level thinking about the subject matter (Ciardello, 1998), then some degree of classroom instruction and guidance in question generation was required. A number of researchers have developed various question-asking strategies, for example King's (1992)

previously mentioned ‘Guided Student-Generated Questioning’. I did consider using this questioning strategy in which question stems are provided by the teacher for student question-generation. King’s research shows that providing question stems to students (e.g. “Why is . . . important?) promotes higher order thinking and discussion among students. However, while effective in science-related or other more systematic courses, I determined question stems to be less useful in the context of an EFL class with L2 learners because they can be somewhat limiting in the types of questions students construct. I decided to take another approach to guiding students in question generation: skinny and fat questions.

Skinny and Fat questions

Questions, including those made by students, may be classified according to the degree or level of thought that is required for answering them (Chin & Brown, 2002). A simple classification of two types of questions was decided on for use in this SGQ experimentation. I decided that the concept of skinny and fat questions would be a fairly easy distinction for students to conceptualize, as it followed the textbook question distinction mentioned earlier, the division between ‘finding meaning’ questions and ‘thinking critically’ questions. The impetus for experimenting with the fat/skinny dichotomy comes from Sunda (2003), whose explanation of the two question types is presented in Table 3 below.

Table 3: *Defining skinny and fat questions* (from Sunda, 2003, p. 12)

Question type	Definition
Skinny	A skinny question is one that can be answered with a short answer or uses basic recall of factual, literal information. Skinny questions can be answered with a word or two, often the simple recall of a specific detail mentioned in the text.
Fat	A fat question is open-ended; there is no one right answer. The answer requires deeper thought through analysis, interpretation, or evaluation. Fat questions require more mental processing . . . encourage interaction with the text and promote more interesting discussions.

Similar to the distinction between question types used in the class textbook, this simple skinny/fat distinction represents the two cognitive levels of questions, explained by Cotton (2001, pp. 5-6) as follows:

- **Lower cognitive questions** are those which ask the student merely to recall verbatim or in his/her own words material previously read or taught by the teacher. Lower cognitive questions are also referred to in the literature as fact, closed, direct, recall, and knowledge questions.
- **Higher cognitive questions** are defined as those which ask the student to mentally manipulate bits of information previously learned to create an answer or to support an answer with logically reasoned evidence. Higher cognitive questions are also called open-ended, interpretive, evaluative, inquiry, inferential and synthesis questions.

The skinny/fat distinction (reflecting the lower/higher cognitive questions distinction above) seemed a much simpler way of getting students to engage in creating both kinds of questions, while avoiding getting lost in terminology regarding cognitive levels.

At the beginning of the set of five classes devoted to the GUO documentary, I gave to students a handout introducing the idea of student-generated questions, and explained that the types of questions they would make could be considered as either fat or skinny. This handout can be found in **Appendix A2**. During this introductory class, I also included a PowerPoint slideshow to help me explain the upcoming procedures. Excerpts from this slideshow can be found below in Figure 2.

<p>Student Questions</p>  <p>Jupiterimages</p>	<p>Why are student questions important?</p> <ul style="list-style-type: none"> • Get students more involved with things they read, see, hear, etc. • Asking (and discussing) questions should help students have a better understanding of issues, topics, etc. • Students who have a clearer understanding of a topic should be able to write better essays about it.
 <p>3</p>	 <p>Questions</p> <p>4</p>
 <p>5</p>	<p>▪ Skinny questions</p> <p>▪ A question that can be answered with a short answer or uses basic recall of factual information</p> <p>6</p>
 <p>7</p>	<p>▪ Fat questions</p> <p>▪ The question is open-ended; there is no right answer.</p> <p>▪ The answer requires deeper thought, analysis or interpretation.</p> <p>8</p>
 <p>Jupiterimages</p>	<p>▪ Both types of questions are important in understanding and analyzing readings, issues and the world around us.</p> <p>9</p>

Figure 2. Slideshow excerpt used to introduce SGQ and skinny/fat questions to class

Timing and format of SGQ investigation

The SGQ experimentation/investigation involved a mixture of:

- having students read tape script sections (usually for homework),
- watch GUO sections in class,
- compose skinny/fat questions and engage in peer discussions,
- provide feedback via mini-surveys about SGQ and peer discussion,
- write a GUO-based essay, and
- complete a final survey about the GUO process.

All of this happened over a six-class period. Table 4 below summarizes the cycle of events and procedures for the reader.

Table 4. *Time frame, cycle of events in SGQ investigation*

Class 1	Prompted by teacher questions, student discussion groups talk about the Internet and their use of it. GUO introduced and tape script distributed. Homework: read tape script chapters 1 and 2 (approximately 4 pages each).
Class 2	Watch beginning chapters. Issue of SGQ and skinny/fat questions introduced and explained. Some practice asking questions about chapters 1 and 2, and peer discussion. Homework: read chapters 3 and 4.
Class 3	Chapters 3 and 4 watched in class. Teacher explained SGQ activity. Students formed pairs (11 pairs) and were given ten minutes to make Q's together. Half made questions about chapter 3, others about chapter 4. This was followed by a group discussion (4 students, 2 each for chapters 3 and chapter 4), taking turns asking their Skinny/Fat questions. Discussion lasted for about 15 minutes. Last five minutes of class students completed mini-survey 1. Survey and student questions picked up. Homework: read chapter 5.
Class 4	Watch part 5. Students make Skinny/Fat questions <i>alone</i> for 8-10 minutes. Group discussion with 3 or 4 students, for 15 minutes. Mini-survey 2 completed and picked up. Homework: read part 6, make some skinny/fat questions about it
Class 5	Watch part 6. SGQ in pairs or alone. Peer questioning for 15 minutes using SG questions. Read and watch the final part 7. Teacher-generated discussion questions. Teacher instruction about essay requirements. Essay question sheet distributed. Homework: complete 2-3-page essay for next class.
Class 6	Essay submitted. Final student survey re. SGQ distributed and completed by students.

The SGQ process encouraged students to do the reading homework for the course, and it was hoped that the simply structured skinny/fat categorization would help create in the classroom

“a learning context of freedom within structure” (King, 1992, p. 124). Students were free to compose questions as they liked, within the skinny/fat framework.

After considering the CW contents and the investigation format and structure of the SGQ being experimented with, it is perhaps useful to be reminded about the connection with assessment in this whole process. Glassen (2009) explains this connection between questioning and assessment as follows:

Does all of this just sound like 'teaching' rather than 'assessment'? Perhaps that's because everything that students do in the classroom provides teachers with the potential to assess what they know and are able to do. The questions they ask and answer, the written work they engage in, even their facial expressions and body language are all observed and interpreted by the teacher. . . . When teachers see assessment as supporting learning, they see it as central to their teaching. (p.12)

The questions raised by students and their ability to respond to questions in discussion groups can be informally assessed by teachers as they ‘listen in’ on the ensuing interaction taking place in the PQ session. As already noted, SGQ and PQ discussion can also act as an informal type of self-and peer-assessment. Echoing the thoughts of Glassen (2009) expressed above, Marshall (2007) also noted:

. . . the day-to-day business of the classroom . . . this is where most formative assessment takes place - in the cut and thrust of the exchanges between student and student and teacher and student and the relationship this creates. It is during lessons that teachers have the most opportunity to engage with their students and help them progress by developing their thinking and their ability to articulate this within the subject discipline. (p. 149)

SGQ is one approach to using questioning for learning to help students’ progress in their understanding of the material being focused on.

Final essay

The key question in this investigation is whether the SGQ process aided in students understanding of the GUO material, and helped them better complete the final essay for the course. Students were instructed to write a focused, well-organized and clearly written essay on one of the essay topics shown in Box 1.

Box 1. Final essay topics based on GUO documentary

- | |
|--|
| 1. Compare and contrast the experience of Japanese teenagers and the impact of the Internet on their lives with the American adolescents in GUO. What are some similarities and differences? |
| 2. Obviously, there are some positive and negative aspects to the impact that the Internet has on teenagers growing up (and their parents). Do you think the effects of the Internet on teenagers are <u>more positive</u> or <u>more negative</u> ? |

Students submitted a 2-3-page essay on one of these topics in the final class. After handing in their essays, a third and final survey was completed by them about the SGQ process incorporated into the CW course.

Data Collection

The data collected during this investigation is comprised of dozens of SG questions collected at the end of class, and responses to a total of three surveys. Due to their brevity, two surveys are labeled mini-surveys. The third and final survey is slightly longer. I decided to have students complete a short mini-survey at the end of classes 3 and 4, in order to capture student feelings and attitudes literally minutes after engaging in the question generation and peer discussion processes. The mini surveys took about 5 minutes to complete. A slightly longer third and final survey was administered to students at the end of course, after students had submitted their final GUO based essay. These mini-surveys are included below in the results section, while the slightly longer final survey is included in **Appendix A4**.

I had considered arranging for audio equipment and recording student discussion groups. I decided against this as potentially interfering with student interaction and disrupting the relaxed, communicative exchange of information and opinions I was trying to facilitate. In any case, the questions produced and written by students (many of which I was able to collect) and their opinions expressed at the end of classes in the mini surveys provide some indirect, but important, views of the peer discussions that took place and student views about them.

A question-generation worksheet was handed out to students in lessons 3 and 4 of this SGQ cycle. Students used it to compose questions with a partner (class 3) or alone (class 4). These SG questions were collected at the end of class. In this way I was able to obtain dozens

of questions students actually made and used in their discussions. **Appendix A3** shows the question generation sheet distributed for creating questions in pairs. A very similar version of this sheet was distributed in the following week for students working alone to make questions.

The skinny questions were intended primarily to promote understanding of the tape script and film contents, while the fat questions were intended to generate student's opinions and feeling about the material being read and viewed.

We now move on to a presentation of the results of the data collection engaged in through this investigation into whether SGQ was actually 'questioning for learning'.

Results

In this process of 'experimentation and consideration' with student questioning as a key formative assessment practice in AfL, a considerable amount of data was produced. The data set includes: question generation sheets (44, one for each student for both peer and solo question generation), results for two mini surveys, as well as a third final survey. The results will be presented here in three parts:

1. The first part will show data from class 3, including student questions (made in pairs), and mini survey 1 responses and commentary
2. Data from class 4 is presented in this second part, including student questions (made alone), and the second mini survey results and comments
3. This third part will present final survey responses and commentary

Where possible, through the use of scanning original student-completed materials, I include the students' words and views in their original format. As the student peer discussions were not recorded (due to the possible disruptive effects of doing so, as noted) student written commentary on the surveys provide us with an indirect window into the PQ discussions that took place.

Part One (question generation in pairs, mini-survey 1)

We begin here with taking a look at the types of questions students produced when working in pairs. Working with a partner can serve to make question generation a more interactive

process. As noted, in this class, 11 pairs of students were split into two large groups, with one making questions about GUO Chapter 3 ('Self-expression, Trying on New Identities') while the other half of class made questions in pairs about Chapter 4 ('The Child Predator Fear'). Students had about 10 minutes to work together making questions, followed by a 15 minute PQ session. Box 2 below provides some skinny and fat questions made by students from 2 pairs. These are related to two different chapters of the GUO documentary. Prior to making questions, students were again reminded of the distinction between the two types of questions.

Box 2. SG questions from two pairs of students

Student questions: Growing Up On-Line (PBS documentary)	
Question type	Questions
	1. How many kids does she have? Evan Skinner
Skinny	2. What is her position in Chatham High School? 3. Where do parents place one of their computer? 4. Do their children reject or accept messages from strangers? 5.
Skinny questions usually have short answers which can be found in the reading.	?????????????????????????
	1. What do you think about the over-protective parents?
Fat	2. What do you think that the parents got the password to their kid's SNS account? 3. Do you think children can deal with solicitations in right way? 4. 5.
Fat question are deeper and require thinking, analysis and sharing opinions, feelings about something related to the reading.	?????????????????????

Student questions: Growing Up On-Line (PBS documentary)	
Question type	Questions
	1. What does Jessica Long do on the Internet?
Skinny	2. What is her name on the Internet? 3. How did her parents know what she was doing? 4. 5.
Skinny questions usually have short answers which can be found in the reading.	?????????????????????
	1. What do you think Jessica's action?
Fat	2. Why do you think she did it on the Internet? 3. What do you think someone want to new person? 4. 5.
Fat question are deeper and require thinking, analysis and sharing opinions, feelings about something related to the reading.	?????????????????????

Following the question generation time in pairs, students then engaged in a PQ session with two other students for approximately 15 minutes. Mini-survey 1 followed shortly after the discussion. Table 5 gives the survey items and student responses.

Table 5. *Mini-survey 1 items and responses (N = 22)*

A. Making questions	agree	tend to agree	tend to disagree	disagree
1. Making questions about the documentary is basically easy to do.	2 (9%)	10 (45%)	9 (41%)	1 (5%)
2. When asking questions, I thought they were clear and easy to understand for other students.	5 (23%)	11 (50%)	5 (23%)	1 (5%)
B. Answering questions	agree	tend to agree	tend to disagree	disagree
3. When I was asked questions, I thought they were clear and easy to understand	6 (27%)	11 (50%)	4 (18%)	1 (5%)
4. I was able to answer questions from other students and share my ideas/opinions without too much difficulty	9 (41%)	7 (32%)	5 (23%)	1 (5%)
5. Making questions and answering those of other students was helpful to have a clearer understanding of the documentary.	13 (59%)	7 (32%)	2 (9%)	0
C. Extra question	very difficult	difficult	kind of difficult	not difficult
6. As a reader, how would you rate the level of difficulty in reading and understanding the main events and ideas in the <i>Growing Up Online</i> tape script?	0	4 (18%)	16 (73%)	2 (9%)

With regard to making questions, a total of 12 students agreed or tended to agree that it was easy to do, while 10 students disagreed that this was so. It seems clear that question

generation was a challenging task for student. Why that is so (e.g. language proficiency in generating questions, understanding of the GUO content, ability to work with partner in question generation) remains unclear. Survey item 2 shows a combined 73% (16 of 22) of students' thought that the questions they had made were clear and easy to understand in the group discussion.

Section B of the mini-survey contains three items related to answering questions in the PQ session. A similar 77% (17 of 22) of students indicated that they agreed that questions posed by peers in the group discussion were clear and easy to understand. In survey item 4, a majority of students (16 of 22) agreed or tended to agree that they were able to respond to peer questions without too much difficulty. Six students disagreed with this statement showing that they did experience difficulties in expressing their opinion. Survey item 5 sought to determine whether SGQ and PQ were helpful in understanding the documentary better. Thirteen students agreed with this statement, while an additional seven tended to agree. Two students tended to disagree with SGQ being helpful in increasing understanding of the documentary.

A final item on the survey was one inquiring as to student views on the level of difficulty in reading and understanding the GUO material. Only 4 of the students indicated that the GUO material was difficult, with the majority (73%, 16 students) selecting the 'kind of difficult' response. Two students thought the material was not difficult. This information is relevant due to the connection between comprehension of the material, and ability to formulate and respond to questions about it.

Additional commentary from students on mini-survey 1

Students often wrote commentary in the short space provided for doing so at the end of each survey item. Table 6 below provides a sampling of student commentary. As the peer discussions were not recorded, student commentary (written minutes after peer discussions happened) provides a window on the interactions that occurred, and helps deepen understanding of the student perspective on both making and responding to questions. The student commentary samples presented here (and in upcoming sections) were carefully selected as a representative cross-section of all comments provided. The selection process tried to include comments offering insights into students' perceptions of the process, including both positive and negative views. Repetitive or confusing comments were not

included.

Table 6: *Sampling of student comments written on mini-survey 1 (verbatim)*

1. Making questions about the documentary is basically easy to do
It is easy to make opinion about it, but question is not.
I think that is not easy. Sometimes so difficult.
I enjoy it!
It is very difficult! If we don't understand about the story, we can't make questions.
Making skinny questions is easy.
2. Questions we made were clear and easy to understand for others
They can answer clearly.
We had a good discussion.
I think I don't make so clear questions.
Even though we had both fat and skinny questions, when it came to discussion we only had a chance to ask fat questions. However, by making skinny question together with a partner it was easy to understand the documentary.
3. When I was asked questions, I thought they were clear and easy to understand
They made good questions.
Discuss with friends are really good idea. I can know many opinions. Very interesting.
Telling my opinion is difficult!
Yes, but I'm poor at explaining what I want to say.
Sometimes I can't understand them instantly.
Partners explained more detail about questions.
4. I was able to answer questions and share my ideas without too much difficulty
I agree. But I don't know a lot of words. It is sometimes difficult to explain my opinion sometimes.
Making questions with my partner is very easier than to make questions by myself.
I answered the questions with much difficulty.
Their question was simple so I could answer easily.
5. Making questions and answering those of other students was helpful in having a clearer understanding of the documentary
In many times I mistake the meaning of contents so it is good to share ideas with other people.
I think I was able to deepen my understanding for chapters 3, 4.
It can be good feedback to make questions and answer other students ones.
There are some contents I can't understand by myself so it is useful.
6. As a reader how would you rate the level of difficulty of the GUO tape script?
I sometimes feel difficult to understand the meaning.
When I watch the TV, I could understand easily.
It depends on chapters. Chapter 1, 3, 4 was easy to understand, but chapter 2 was kind of difficult, maybe because of vocabulary.

Some student comments in Table 6 are particularly worthy of note at this point. For item one,

a student commented that if students don't understand the contents it can be very difficult to make questions about it. A students' comment for survey item two noted that in the allotted time for the PQ session, there was time only to ask the fat questions. It should be noted here that student were not directed as to which questions they should ask in the PQ discussion, or that they had to try and complete all questions. Typically students began their discussions with skinny questions, but after being told of a 15-minute discussion time limit, they were free to pose and respond to questions as they wished. This style of interaction reflects 'a learning context of freedom within structure' (King, 1992), noted earlier.

Survey item 3 comments remind us of two important points. First, that the students are L2 learners with various levels of proficiency. This may be the major factor responsible for such comments as "Telling my opinion is difficult!" and "I'm poor at explaining what I want to say". The second interesting point is that student comments indicate that in peer discussions there is some negotiation of meaning ("Sometimes I can't understand them instantly"), and further clarification of intended question meaning going on ("Partners explained more detail about questions.").

Student commentary for item 5 in Table 6 is particularly relevant in consideration of the research question that is being focused on in this experimental investigation of SGQ - that of promoting learning and closing gaps in student understanding of the subject matter.

Part Two (solo question generation, mini survey 2)

The subsequent class followed a similar pattern. Students were required to read chapter five of the tape script for homework, and at the beginning of the following class (number 4 in the cycle) we watched this portion of the documentary. After viewing, a new question-generation worksheet was handed out and students were given about 10 minutes to work alone to make questions about chapter five. As a background note, chapter five of the documentary is called '*Private Worlds Outside Parents Reach*', and a key character is a girl called Sara who talks about accessing Internet websites promoting the eating disorder anorexia nervosa. Another key character in this chapter is a mother talking about students uploading video clips to the Internet. Prior to making questions, students were again verbally reminded about the skinny/fat distinction.

With the question formation sheet provided in this class, students were instructed to work alone to ask questions. They were also once again reminded on this sheet to 'try and make

your questions clear and easy to understand'. Box 3 presents the format for this solo questions sheet from one student.

Box 3. Question generation sheet for solo SGQ

Student questions: Growing Up On-Line
Part 5 (Private Worlds Outside Parents Reach)

Question maker : Kyoko

Working alone, make some skinny and fat questions related to Part 5 of GUO. In a few minutes you will get together with other students and talk about your questions, and share ideas, opinions, etc. Try to make your questions clear and easy to understand.

Question type	Questions
	1. what did CAM's mother do when she knew about the rock concert?
Skinny	2. what did the students do at the rock concert which got them to problems.
Skinny questions usually have short answers which can be found in the reading.	3. Does SARA have a eating disorders. say that she has 4.
????????????????????????????????????	
	1. Do you share your problems with your parents.
Fat	2. What do you think about Cam's mother sending the e-mails to other parent? 3. What do you think was the problem between Cam and his mother? 4.
Fat questions require thinking, analysis and sharing opinions, feelings about something related to the reading. Answers will depend on individual students and their views.	

The reader will notice that this question generation sheet is slightly different from the one used for making questions in pairs (e.g. ‘working alone’ directions rather than ‘working with a partner’), but most other features remain identical.

Box 4 below shows three other examples of SG questions sheets, with excerpts of only the

student question section.

Box 4. Three students question sheets (skinny questions on top, fat on bottom)

Yasuna	Midori					
<table border="1"> <thead> <tr> <th data-bbox="255 444 822 485">Questions</th></tr> </thead> <tbody> <tr> <td data-bbox="255 489 822 893"> 1. What is the name of disorder which Sarah struggles with? 2. Do Sarah's parents know her disorder? 3. What did Cam's mother find out on the Internet? 4. </td></tr> </tbody> </table>	Questions	1. What is the name of disorder which Sarah struggles with? 2. Do Sarah's parents know her disorder? 3. What did Cam's mother find out on the Internet? 4.	<table border="1"> <thead> <tr> <th data-bbox="870 444 1527 485">Questions</th></tr> </thead> <tbody> <tr> <td data-bbox="870 489 1527 893"> 1. What problem does Sara have? 2. What is the name of Web sites that Sara goes to? 3. 4. </td></tr> </tbody> </table>	Questions	1. What problem does Sara have? 2. What is the name of Web sites that Sara goes to? 3. 4.	
Questions						
1. What is the name of disorder which Sarah struggles with? 2. Do Sarah's parents know her disorder? 3. What did Cam's mother find out on the Internet? 4.						
Questions						
1. What problem does Sara have? 2. What is the name of Web sites that Sara goes to? 3. 4.						
<table border="1"> <tbody> <tr> <td data-bbox="255 943 822 1028">1. Why does Sarah go on the Internet?</td></tr> <tr> <td data-bbox="255 1033 822 1118">2. What do you think about Sarah's action?</td></tr> <tr> <td data-bbox="255 1118 822 1181">3. What do you think about Cam's mom?</td></tr> <tr> <td data-bbox="255 1181 822 1230">4.</td></tr> </tbody> </table>	1. Why does Sarah go on the Internet?	2. What do you think about Sarah's action?	3. What do you think about Cam's mom?	4.	<table border="1"> <tbody> <tr> <td data-bbox="870 943 1527 1230"> 1. What do you think of the Web sites like "Thinspiration". 2. What do you think of uploading some bad pictures or videos on line? 3. What do you think if your parents check your computer data? (e.g. mixi) 4. Will you involve with children's computer when you become a parent? </td></tr> </tbody> </table>	1. What do you think of the Web sites like "Thinspiration". 2. What do you think of uploading some bad pictures or videos on line? 3. What do you think if your parents check your computer data? (e.g. mixi) 4. Will you involve with children's computer when you become a parent?
1. Why does Sarah go on the Internet?						
2. What do you think about Sarah's action?						
3. What do you think about Cam's mom?						
4.						
1. What do you think of the Web sites like "Thinspiration". 2. What do you think of uploading some bad pictures or videos on line? 3. What do you think if your parents check your computer data? (e.g. mixi) 4. Will you involve with children's computer when you become a parent?						
Anna						
<table border="1"> <thead> <tr> <th data-bbox="255 1280 822 1320">Questions</th></tr> </thead> <tbody> <tr> <td data-bbox="255 1325 822 1545"> 1. How person is SARA in real life? 2. What web site does she go? 3. What did EVAN SKINNER when she knew what happened? 4. </td></tr> </tbody> </table>	Questions	1. How person is SARA in real life? 2. What web site does she go? 3. What did EVAN SKINNER when she knew what happened? 4.	<table border="1"> <tbody> <tr> <td data-bbox="870 1280 1527 2041"> 1. SARA says her real life is fake, and when she is online, she is real person. What do you think about it? 2. What do you think about what EVAN SKINNER? 3. 4. </td></tr> </tbody> </table>	1. SARA says her real life is fake, and when she is online, she is real person. What do you think about it? 2. What do you think about what EVAN SKINNER? 3. 4.		
Questions						
1. How person is SARA in real life? 2. What web site does she go? 3. What did EVAN SKINNER when she knew what happened? 4.						
1. SARA says her real life is fake, and when she is online, she is real person. What do you think about it? 2. What do you think about what EVAN SKINNER? 3. 4.						

As with the previous class, after making questions students formed discussion groups of three or four for a 15-minute PQ session. On this occasion everybody discussed the same GUO chapter. Following the discussion session, students completed the second mini-survey. Student responses and numbers are presented here in Table 7.

Table 7. *Mini-survey 2 items and responses (N = 22)*

A. Making questions					
1. Which do you prefer, making questions with a partner or by yourself?		making Q's with partner	making Q's by myself		
		11 (50%)	11 (50%)		
		agree	tend to agree	tend to disagree	disagree
2. In our group talk, the other students could understand and answer my questions.		12 (55%)	10 (45%)	0	0
B. Answering questions					
3. Questions from other students were clear and easy to understand.		7 (32%)	14 (64%)	1 (5%)	0
4. Today, I was able to answer the questions from other students and share my ideas/opinions with my group without too much difficulty.		8 (36%)	9 (41%)	5 (23%)	0

Student responses showed an even 50-50 split as to whether they preferred working alone or with a partner when making questions. Survey item 2 indicates that in all, more than half of the 22 students (55%) agreed that other students in the discussion group could understand and answer their questions. The remaining students (10) responded ‘tend to agree’ to this statement. As for answering other students’ questions, only one student disagreed that other students’ questions were clear and easy to understand. The majority (64%) tended to agree with this statement, while 7 of 22 students checked the ‘agree’ response. Survey item 4

queried whether students were able to share their opinions without too much difficulty. Almost a quarter (5 of 22) of students disagreed with this statement. Seventeen students agreed that they could express their opinions without too much difficulty, with 9 of 22 responding ‘tend to agree’.

By looking at some of the commentary students wrote on this survey we can get a deeper understanding of what the survey responses indicate. Once again, the responses provide indirect snapshots of the five or six discussions that occurred in the class during the PQ sessions. Table 8 below also provides some understanding of student feelings about solo SGQ and the second peer discussion session. The format and substance of the commentary again remind us that the class is a group of EFL students with various degrees of English proficiency. As noted in the earlier presentation of student comments, samples were selected to provide a representative cross-section of commentary, excluding any repetitive or unclear views expressed.

Table 8. Sampling of student comments written on mini-survey 2 (verbatim)

1. Which do you prefer, making questions alone or with a partner?
- prefer making questions alone
My brain works a lot.
It is easier to make questions by myself because I can think by myself and make questions that I want to ask.
Because I can make a question freely. It took a lot of time to make only one question when I did pair work last week.
Making questions with a partner is confusing. It might arise many questions at the same time.
When we make a group we can have many kinds of questions.
- prefer making questions with a partner
We can make questions after very short discussion so questions can be more creative.
Make questions with a partner come up with many questions more than by myself.
It is easier to make questions with a partner. I'm able to understand more details.
It was easy to make questions by thinking topic together.
2. In our group talk, students could understand and answer my questions
Easy questions like skinny questions had good reaction. But fat questions were difficult.
I make easy questions to answer.
My question is not certain word, but the others try to understand what I said.
Maybe my fat questions was not.
I couldn't make so good questions.
3. Questions from other students were clear and easy to understand.
I think so. But what I didn't understand was explained from other students.
Sometimes I don't know what question means, but I ask them and understand.
The question was so clear that I could say my opinion.
It's easy and comfortable for me. But sometimes difficult. Words I can't understand and long sentences.
4. Today I was able to answer questions and share my opinions without too much difficulty
My listening ability is low. I can't always listen completely the contents. So it is difficult to make opinion and share my idea.
I can know about topic by answer the questions.
I could know a lot of group members opinion and experience today. It was very good discussion.
I could share the opinions. It is very fun.
It is not so easy for me to talk about my opinions for others.
Thinking our opinion is difficult, and telling other people in English is also difficult.
I need more vocabulary.

At the end of this mini survey, in item 5, students were provided with the opportunity to write additional comments. Box 5 provides the responses of three students, which, again, enables us

to get some idea of the peer discussions that occurred. These are representative of similar views expressed by other students.

Box 5. Three students' commentary at the end of mini-survey 2

5. Other comments or feelings about today's Q & A session with your group?

Share my idea with friends are really good.
Also when we make groups, and talk,
we can talk very interesting story. It's
fun for me!

And 7th question is very interesting for answer.

5. Other comments or feelings about today's Q & A session with your group?

Today's topic was very familiar to our generation.
So we discussed so much and we could share our
each opinion. So I could enjoy discussing and
learn some things like I have never known.

5. Other comments or feelings about today's Q & A session with your group?

We tended to think the same or like way
about the question. It was not so interesting.
When I find different ideas, it's more
interesting for me.

The tone of disappointment with this class discussion is obvious in the third comment above, but the other two are much more positive about this second PQ session. It should be noted that students were free to choose the student groupings for the PQ discussions. In survey item 5 commentary, another student wrote, "Discussion with my group was interesting. I know the new idea and information", while a classmate echoed this view by commenting "I can know this topic more and more".

Part Three (final survey about SGQ)

After submitting an essay related to the GUO documentary, students completed a final class survey on the topic of SGQ (see **Appendix A4**). Table 10 below presents student responses to the six survey items. One student was absent.

Table 9. *Results of final class survey about SGQ* (N = 21)

1. In this class, I ---- the reasons why students were asked to make questions about GUO and discuss them in small groups.	understood	kind of understood	did not understand
	13 (62%)	8 (38%)	0
2. The idea of skinny and fat questions and the difference between them was ----	easy to understand	kind of easy to understand	difficult to understand
	16 (76%)	5 (24 %)	0
3. Making questions and using them in small groups was something I ---- doing.	really liked	liked	did not like
	6 (29%)	15 (71%)	0
4. Making Q's and discussing them in small groups was ---- in understanding the GUO documentary.	very useful	useful	a little useful
	12 (57%)	6 (29%)	3 (14%)
5. Making questions and discussing them in small groups was ---- in helping me write the GUO essay.	very useful	useful	a little useful
	9 (43%)	8 (38%)	3 (14%)
6. I would like student questions to be included in the second half of this course.	strongly agree	agree	disagree
	7 (33%)	14 (67%)	0

In survey item 1, students seemed to express some lack of understanding as to why SGQ and PQ were included in this class with 8 (38%) choosing the ‘kind of understood’ response. Responses to item 2 show that most students (16 of 21) thought that the skinny/fat distinction between questions was easy to understand. All students, in item 3 reported that they ‘really liked’ or ‘liked’ making questions and discussing them with their peers. In item 4, 57% (12 of 21) of students reported that engaging in SGQ was ‘very useful’ in better understanding the documentary, with the remaining students choosing the ‘useful’ option. As for writing the essay, a combined total of 81% (17 of 21) reported that SGQ and PQ sessions were either very useful or useful in helping them complete the essay task. All of the students, in item 6, agreed (with 7 students strongly agreeing) that SGQ should be included in the second half of the CW course.

As with the two mini surveys, students were also given the opportunity to make further additional comments about SGQ at the end of the final survey. Six students took the opportunity to make additional comments. These were scanned and collected, and are all presented below in Box 6. The six comments reflect the overall positive student response to this SGQ experiment.

Box 6. Student comments about SGQ on the final survey (N = 6)

7. Any other comments (about student – generated questions)?

I have to know answers to make questions. I have to understand the issues deeply, so it's difficult, but it helps us learn more deeply.

7. Any other comments (about student – generated questions)?

It is very good experiences that students think with other students. You should keep this style.

7. Any other comments (about student – generated questions)?

Making questions and discussing is good for me because I can gather many information. But I can't construct my opinion in a few minutes. It's only hard for me.

7. Any other comments (about student – generated questions)?

Making questions and discuss with friends are very good idea. Just writing essay is boring. So, having disucssion time in class was nice for me.

7. Any other comments (about student – generated questions)?

Making questions and discuss them with 2-3 people, was fun and very good way to understand the contents each other.

7. Any other comments (about student – generated questions)?

Student Questions are very interesting and useful to understand the story.

Discussion

In writing about the importance of questioning in the modern information age, MacKenzie (1997) contends, “Without questioning skills, you are just a passenger on someone else’s tour bus. You may be on the highway, but someone else is doing the driving”. This SGQ experiment is one in letting the students do some of the driving to see whether learning is supported and promoted as a result. As a ‘driver’ in the classroom dialogue that occurs, the student exercises a great deal of control over both the questions asked and the answers that are generated. Such personal autonomy has been identified as a contributing factor to the success of SGQ in mainstream education, as students are creators of their own specific questions rather than simply responding to textbook-supplied or teacher-made questions (King, 1992). The teacher may be seen as the ‘navigator’ in this journey, supervising direction, speed and destination (goals). According to Black and Wiliam (2006) a fundamental role for teachers in AfL-centered classrooms is to ‘engineer’ learning environments so that students are actively involved in learning tasks. They emphasize the need for “students doing the thinking and making that thinking public” (Black & Wiliam, 2006, p. 17.). This is an apt description of the students’ role and behavior in this SGQ experiment.

The key question of this investigation is whether SGQ successfully promoted learning. According to Gardner (2006), if the response to this query is negative, it suggests that the formative assessment procedure is not valid because the primary purpose of promoting learning was not achieved. This investigation concludes that in the CW course, student-generated questioning did indeed support learning about the classroom material (GUO documentary), making this formative assessment procedure valid, effective and beneficial. The process can accurately be described as ‘student questioning for learning’.

According to Van der Meij (2004, p. 141), “questions are seen as an important diagnostic, as a window to the mind”. The responses and peer discussion generated by these student questions can also be seen as a diagnostic practice. Through the questions posed by students and the answers generated, the teacher (as navigator) can informally assess students understanding of the subject matter and proceed or adjust course if necessary. As previously noted, engaging in SGQ and peer discussions also involves an element of self-and peer-assessment by students. King (1999) writes:

On a comprehension-monitoring level, asking and answering these questions is *a form of self and peer testing*, allowing students opportunities to check how well they (and their peers) understand the material as well as to clarify misunderstandings, correct errors, and *fill in gaps in knowledge*. (p. 92, emphasis added)

In addition to this element of self-and peer assessment, closing gaps in student knowledge, skills and understanding (Sadler, 1989) is a key element in formative assessment theory and the AfL framework. Closing gaps in student understanding of course materials is also important in this CW course, and SGQ helped in doing that. That question generation and discussion by students promoted better understanding of the GUO material is reflected in such student comments as: “*In many times I mistake the meaning of contents so it is good to share ideas with other people*”, and “*Making questions and discussing is good for me because I can gather many information*”. The fact that 17 of 21 students in the final survey reported that the process was either ‘very useful’ (12 students) or ‘useful’ (6 students) in helping them understand the documentary also speaks for itself.

Evaluating the efficacy of formative assessment: key ingredients

After noting that the goal of formative assessment is the improvement of student learning and motivation, MacMillian (2007, p. 4) posits that three key ingredients should be looked for to evaluate the efficacy of formative classroom assessment. These are:

1. the extent to which assessments are embedded within instruction;
2. the extent to which additional instructional strategies are employed;
3. the extent of student engagement and learning.

We may use these three ingredients as criteria, particularly the first and third, to evaluate the value or usefulness of the SGQ process in promoting learning.

As an informal assessment mechanism for teachers (as well as students) SGQ was very much embedded with classroom instruction. Generating questions, these ‘informal diagnostic devices’ (McDevitt, 1994), and discussing them with peers formed a significant part (up to 30 minutes) of class time during the GUO cycle of lessons. Students responded positively to the PQ discussions (‘*Just writing essay is boring*’), and lively small group interaction was a typical element of these classes. As one student commented, “*It is very good experiences that students think with other students. You should keep this style*”.

While students were engaging in their discussions, my role was to be a listener and eavesdropper, an informal assessor on the interactions taking place and students’ understanding of the material. I usually did this with a pen and paper in hand taking note of

any common themes or areas of difficulty with the material. Peer discussions were always followed by a short summary by me for the class of important points, or areas of confusion. For example, in discussing the contents of Chapter 6 of the GUO documentary, related to cyber-bullying, I noticed that groups were having difficulty understanding one particular passage. The documentary reports how a boy committed suicide as a result of cyber-bullying, partially due to his relationship with a particular girl online. The tape script excerpt that caused confusion can be found here in Box 7.

Box 7: Tape script excerpt related to the cyber-bullying chapter in GUO

NARRATOR: According to Ryan's friends, the tipping point came when a popular girl at school flirted with Ryan over instant messaging, only to humiliate him later by telling him it had all been a big joke. It was a game she often played with boys on line.

SARAH: I guess the fun is, like, dropping the bomb, you know, "Oh, just kidding!" And then that, like, crushed him. I mean, you wouldn't do that to someone's face, but online, it's completely different. You can do whatever you want and no one can do anything. You're at your house, they're at their house. It's different.

(Growing Up Online tape script, Frontline, 2008)

One group called me over to say they did not understand what had happened. After observing that some of the other groups were also confused about this passage and the girls' role in the suicide, I made a note of this and it became a follow-up teaching point after PQ discussions had ended. As a class, we took a closer look at the vocabulary and intended meaning of this passage, discussed it together, and students were able to develop a better understanding (i.e. close the gap) of what had happened in the documentary.

Student questioning and discussing provided information to assess their understanding and inform my teaching, in order “to move students from where they are to where they need to go” (Curriculum Corporation, 2004, p.1). That is, from a less to more complete understanding of the GUO documentary and the issues of the Internet and adolescence; the subject of the final essay in the CW course. Chin and Brown (2002) comment on the need to engineer SGQ in the classroom, and the teacher’s role in the process, as follows:

... teachers cannot fully rely on students’ spontaneous questioning and must explicitly orient their students towards asking questions, for example, by specifically encouraging them to generate questions, either verbally or written, as part of their class activities. Besides prompting students to think more deeply about what they are doing and encouraging critical

thinking, such questions could also provide feedback to teachers about their students' thinking and puzzlement, and act as a window to the students' minds. (p. 544)

This view of questions as providing feedback to teachers is also one echoed in the AfL perspective of questioning and dialogue in the classroom. James et al. (2007) write that "by promoting thoughtful and sustained dialogue, teachers can explore the knowledge and understanding of pupils" (p. 9).

This cyber-bullying example is just one instance of embedding assessment within instruction. Similar instances occurred in which I was able to assess areas of difficulty or confusion and turn them into teaching points to close gaps in understanding and support learning. In discussing the assessment and instruction connection, Earl (2003) writes

By making connections between curriculum, instruction, assessment, and students' daily lives . . . teachers can engage students and draw them into the learning that assessment encompasses. Assessment does not stand apart; rather it is interwoven with teaching and learning to make connections for students, reinforcing what they know and challenging their thinking (p. 68).

In this SGQ process, assessment is indeed interwoven with teaching and learning, and the classroom dialogue taking place, generated by student questions and peer group discussions, was beneficial in "creating a richer community of learners" in the CW class (Black & Wiliam, 2006, p. 17).

The second key ingredient, according to MacMillian (2007) in evaluating the usefulness of formative classroom assessment is the extent to which additional instructional strategies are employed. Here, MacMillian is referring to following up feedback to students with new strategies and approaches 'that will build on current understandings to broaden and expand learning or correct misunderstandings' (2007, p. 3). In his formative assessment cycle of evaluation of student progress, feedback to students and instructional correctives, MacMillian is primarily concerned with mainstream educational contexts where teachers are meeting class groups on a daily basis. This 'key ingredient' was less relevant in this investigation. SGQ is in itself a new strategy and approach for students in better understanding the material at hand. While some variations were included, for example making questions with a partner or alone, because this was an 'experimentation and consideration' of one particular instructional strategy, the employment of additional ones was not a key consideration here.

On the contrary, the final key ingredient listed by MacMillian (2007) in evaluating formative assessment efficacy, the extent of student engagement and learning, is of

paramount importance. It is also one directly connected to the assertion that the main validity check for any AfL procedure is whether it successfully supported learning (Stobart, 2006; Gardner, 2006).

In terms of student engagement, the SGQ process worked very well in ensuring that all students were involved in both question creation and classroom interaction in PQ discussions. Wix and John-Steiner (2008), in writing about classroom dialogue in higher education discuss what they call ‘dialogic peer inquiry’, which is very similar to the SGQ process described in this report. They describe dialogic peer inquiry as follows:

It requires preparation through both classroom activities and outside reading; the formation of small groups within the class; dialogue and inquiry among peers; and the formulation of questions for one another by peers. Frequent communication among partners is a must. (Wix & John-Steiner, 2008, p.2)

This is also an accurate description of the SGQ process, and helps underline the high degree of student engagement required and generated. Students were kept quite busy in this process through reading the documentary transcript before class, and the SGQ activities engaged in during class time. All 21 students reported in the final survey that making questions and discussing them with peers was something they ‘liked’ (15 students) or ‘really liked’ (6 students) doing. One students’ comment about the degree of cognitive energy expended in this SGQ process, *‘My brain works a lot’*, is a reflection of the level of engagement and involvement that resulted. In their investigation of student-generated questions in a junior high school science class, Chin and Brown (2002) found that “the questions played an important role in engaging the students’ minds more actively, engendering productive discussion, and leading to meaningful construction of knowledge” (p. 540). The same conclusion could be reached in this CW class with regard to student engagement. Black and Wiliam (2006) remind us that it is necessary for learners to be active in creating their own understanding, and note that in the constructivist view of learning a main principle is to begin where the student is and actively involve them in the process. As made clear in the thesis introduction, constructivist learning theory is fundamental to AfL; it is also apparent in the type of student engagement described in this SGQ investigation.

As for the *extent* of student learning caused by the SGQ, that is a more difficult answer to respond to, and especially quantify. However, the fact that learning was supported and promoted by the SGQ process is more readily evident in this investigation. Student survey responses, and in particular student commentary support this conclusion. In the words of one

student, commenting about question-generation, “*I have to know answers to make questions. I have to understand the issues deeply, so it’s difficult, but it helps us learn more deeply.*”

Learning more deeply about the contents of the GUO documentary, to engender better essay writing about it, was the main purpose in incorporating SGQ into the curriculum. Other similar, positive student comments reported in Box 4 of the results section above are further evidence of learning being successfully supported. A student comment about PQ discussions from the end of mini survey 2 noted “*I could enjoy discussing and learn some things like I have never known*”. This quote highlights the point that, as well as learning through interaction with classroom materials, students learn through social interaction (Harlen, 2007).

In her research with SGQ and elementary school children in an American science class, King (1994b, p. 364) concluded, “questioning strategies can be used to facilitate the knowledge construction process, which in turn enhances learning”. This investigation provides evidence that the same claim can be made with EFL students’ question-generation and peer discussion at a Japanese university.

While asking and responding to skinny questions played an important role in understanding the factual content of the GUO documentary, the use of fat questions was perhaps more important, as asking and responding to them required the type of reflection and understanding which drives learning (Chin and Brown, 2002). A few examples of such meaningful fat questions made by students, and recorded above, include:

- *What do you think about the over-protective parents?*
- *Sara says her real life is fake and that online she is real person. What do you think about it?*
- *Do you share your problems with your parents?*
- *What do you think about websites like ‘Thinspiration’?*

The questions made by students, their survey responses, and their commentary about the process leads to an affirmative response to the research question posed in the introduction as to whether SGQ promoted learning.

While we can conclude that SGQ did support and promote student learning of the GUO material in class, whether it was beneficial to them in writing their final essay is suggested by the data, but less certain. Student responses to the final survey do seem to indicate that SGQ helped the essay writing process. Of the 21 students completing the final survey, a total of 17

reported that making questions and discussing them in groups was either ‘very useful’ (9 students) or ‘useful’ (8 students) in helping them write the GUO essay. In terms of essay writing, the SGQ procedures could be viewed as an extended pre-writing process. While such speaking activities as pre-writing discussions are popular in ESL writing classes, little is known about whether it affects the quality of the compositions students write (Shi, 1998). In summarizing the effects of pre-writing discussions on ESL students writing Shi (1998, p. 322) reports the following findings from the research literature:

- Pre-writing discussions seemed to lead to better drafts than write-only conditions;
- Such discussions produce a variety of positive effects in different writing tasks; and
- Pre-writing discussions do contribute to students’ better understanding of the complexity of topics, especially with peer groups.

While an examination of *how* such pre-writing talk as SGQ affected the quality of student compositions is outside the scope of the research reported on here, that it did have a positive effect is suggested. Again, the majority of students certainly seemed to think that the composition of their final essay was indeed helped by the SGQ process.

As an assessor of students’ final essays, there is no basis for postulating that student essay writing abilities did improve in this final essay as a result of SGQ. This lack of triangulated evidence showing the impact of SGQ more clearly is, however, not seen as a particular weakness as showing such an impact was not the goal of this research. Also there is no particular reason to doubt or disregard student views that the process did help them write a better final essay. Reflecting a key philosophy underlying AfL, Black et al. (2003) note that an important element in engineering effective learning environments is “making students’ voices louder and making the teachers hearing’ better” (p. 59). As well as making students voices louder in the classroom through SGQ, a conscious effort was made to amplify student voices in this report of the process. Any perceived lack of supporting evidence here as to improvements in writing, should not unduly distort how these voices are heard.

SGQ and lifelong learning

Aside from dealing with the course curriculum and content at hand, AfL also promotes the learning of skills, competencies and understanding that will benefit students beyond the classroom as lifelong learners. After determining that SGQ does successfully promote student

learning in the CW class, I would like to end this discussion section by connecting the question generation skills explored and discussed in this report with life outside the classroom.

In New Zealand, the Ministry of Education has identified and compiled a list of five key student competencies they have determined to be important capabilities for living and lifelong learning. The ministry contends that these competencies can and should be used by people to live, learn, work and contribute as active members of their communities (The New Zealand Curriculum). The five competencies are listed below in Table 10. As the reader will note, all of them are evident in the SGQ process described in this report.

Table 10. *Five key student competencies for living and lifelong learning*

Key Competency	Explanation
1. Thinking	Using creative, critical, and metacognitive processes to make sense of information, experiences, and ideas.
2. Using language, symbols and text	Working with and making meaning of the codes in which knowledge is expressed. Languages and symbols are systems for representing and communicating information, experiences, and ideas.
3. Managing self	This competency is associated with self-motivation, a “can-do” attitude, and with students seeing themselves as capable learners. It is integral to self-assessment.
4. Relating to others	Interacting effectively with a diverse range of people in a variety of contexts. This competency includes the ability to listen actively, recognize different points of view, negotiate, and share ideas.
5. Participating and contributing	Being actively involved in communities, which are drawn together for purposes such as learning or work. This includes a capacity to contribute appropriately as a group member, to make connections with others, and to create opportunities for others in the group.

(from The New Zealand Curriculum, NZ Ministry of Education)

According to the New Zealand Curriculum, students need to be challenged and supported in developing these competencies over time, as they interact with people, places, ideas and things. All five competencies (thinking, using language symbols and text, self-management, relating to others, and participating and contributing) are demonstrated and developed by

students in the SGQ process described here. Indeed, they could all be used as a list of important competencies that AfL theory and practice seeks to promote among students in (and beyond) the classroom community. Hence, not only can SGQ successfully promote student learning in the classroom, it can also develop the kind of competencies AfL seeks to promote for lifelong student (and teacher) learning.

Conclusion

True learning is characterized not so much by the answering of questions as by the asking of them. (UNESCO, 1980)

Student generated questioning has an important, if under-emphasized and under-researched, role to play in AfL- centered classrooms. This role is widely applicable across subjects and settings, including in an EFL context as this research report shows. Ciardello (1998) is right in his simple assertion that “All students have the potential to learn how to think, reflect, and question in a competent manner” (p. 219). This is no less true in an L2 learner context, especially in higher education with older and more mature students. The teacher’s role is to help develop this potential and engineer the type of classroom environment where student questioning is embedded in the curriculum and is part of the formative assessment framework of a course; that is, student questioning for learning.

As with earlier research reports in this thesis, this report describes a process of using a particular assessment procedure in a course, “to put classroom flesh on the conceptual bones of the idea of assessment for learning” (Black & Wiliam, 2006, p. 25). It describes a process of ‘experimentation and consideration’ with questioning which focuses on those generated and discussed by students, rather than those coming from the teacher or the textbook. The guiding question here was whether SGQ successfully promoted student learning. Over a cycle of six lessons, this study generated a significant amount of data in the form of copies of student-made questions, three surveys (two mini-surveys, followed by a final survey at the end of the cycle) and a large number of students written comments about the process. This experiment was deemed a successful one in that it did promote student learning about the class materials and the strategy was, on the whole, positively viewed by students. ‘Putting the spoon in the students hand’ did indeed help generate student enthusiasm, effort and efficacy in the CW class. The results of this preliminary investigation led to the continued use of

student-generated questions with the CW class in the second semester. The positive results also encourage the further use of this formative assessment strategy in other classes and courses. In addition, the benefits of promoting student-questioning skills for lifelong learning beyond the classroom have also been noted in this report.

While teacher-generated questions are of much importance, as evidenced by the amount of research and writing devoted to them, what is being reported on and advocated here is the ‘question-driven classroom’ (Chin & Brown, 2002) in which each student is put in the active role of questioner. By putting the student in the questioner role, teachers can help engineer a classroom culture of questioning in which students learn from both the question-generation process itself, and the resulting shared discussions with peers. At the same time, teachers can assess students’ understanding and abilities, and make any needed course or speed adjustments as navigator, and provide guidance along the learning pathways traveled.

We now move on to the last of the five AfL research reports included in this thesis, and a more detailed look at an issue that has appeared earlier in this thesis - making formative use of summative assessment.

Part B. Synergizing formative and summative assessment of students' presentation slide shows

Simply knowing the final score of the game after it is over is not very useful. What we need is a vivid rendering of how that game is played. (Eisner, 1985, p.131)

Introduction

According to the online version of the Oxford English dictionary, the word *synergy* comes from the Greek term ‘sunergos’ meaning ‘working together’, and refers to the ‘*interaction or cooperation of two or more agents to produce a combined effect greater than the sum of their separate effects*’ (AskOxford.com, 2009). This fifth and final research report describes a process where formative and summative assessment interact and cooperate; working together to improve student learning and classroom teaching. In their book reporting efforts to put formative assessment procedures into practice in schools in the UK, Black et al. (2003) found that the formative use of summative tests had an important role to play, and teachers sought to achieve a positive relationship between the two, sometimes conflicting, forms of assessment. This paper also reports on an attempt to forge a positive relationship between the two, creating a synergized assessment process in the context of a public speaking course and assessing students’ presentation slide shows.

The assessment framework implemented will be discussed in more detail below, but a useful introductory point notes that while summative assessment (SA) has the purpose of reporting on student learning achieved at a certain time, formative assessment (FA) has the single purpose of informing both learning and teaching (ASF, 2005). The concept of assessment *purpose* is to be noted in this gloss, because, “with assessment, purpose is everything” (Stiggins, 2008, p.3).

As evidenced by previous literature reviews and source citation references used throughout this thesis, formative assessment has been much discussed and researched in the decade since Black & Wiliams’ (1998a) seminal review of its uses and benefits. Yet, for teachers at all levels of education, the word ‘assessment’ is primarily associated with the end-of-unit or end-of-term summative version when grades are calculated and distributed to students. This association is a reflection of the persistent view that assessment signifies making judgments rather than helping learning (Harlen, 2007). Yet, because of the importance of both formative and summative purposes in promoting successful learning, and the need for balance between

them, assessment needs to be seen as an instructional tool for use while learning is happening, as well as an accountability tool to check if learning has occurred (Balanced Assessment, 2003). Stiggins and Chappuis (2005) neatly capture the essence of this FA challenge as follows: “The teaching challenge is to use the assessment in advance of the graded event, as a vehicle to deepen the learning and to reveal to students their developing proficiencies” (p. 5).

This report discusses one response to this challenge of synergizing FA and SA to make learning and teaching more effective. Unlike the other reports in this thesis, it does not have a specific research question to answer, but is a more straight-forward descriptive account which explains how this AfL component was put into classroom practice. It will describe an assessment regime comprising a series of three assessments, two formative and one summative in a TWCU course called ‘Public Presentations’ (PP). Students’ computerized slide shows for their end of term presentations were twice assessed formatively prior to the summative grading event. All slide shows were constructed using the most common presentation software, Microsoft PowerPoint.

An oral presentation supported by a set of slides (or ‘deck’), is a communicative event with two main elements: the performance (the speaker) and the slide show (visual aids) (Farkas, 2005). In this report, these two main components of students’ final presentations were assessed separately. In effect, each students’ final presentation was comprised of two separate ‘tests’, one for the performance, and a separate assessment for the slide show. This investigation deals only with the slide show assessment, in particular, the two separate formative assessments leading up to the final summative assessment of student slide shows. During the semester, students submitted a working version of their slide show, which was formatively assessed by both themselves and the teacher, prior to submitting a revised version of the deck on the day of their final presentation. These final versions of students’ slide shows were then summatively graded, and this score made up 25% of their overall course grade.

In the process described here the same assessment instrument, a slide show scoring rubric, was used three times, resulting in “the interweaving of formative assessment tasks towards a summative event” (Dunn et al., 2004, p.18). The purpose of the assessment regime implemented was for the trio of assessments to synergize into “a balanced and integrated assessment system, with all parts working together in the service of student success” (Stiggins, 2006, p.17).

This report will describe how FA and SA were synergized together in a one-semester TWCU course, with the focus on the formative use of summative assessment. The assessment

tool and the procedures for its various uses in the PP course will be introduced and delineated. This will be followed by a discussion and analysis of the procedures, with some particular student examples of the different assessments generated. We begin by framing this report with a look at the relevant literature connected with integrating FA and SA, and using summative assessment for formative purposes.

Literature Review

In an article about understanding and using PowerPoint software, Farkas (2005) notes that while this medium has generated much casual commentary, very little careful analysis or empirical research has been done. Published research on the assessment of students' slide show presentations is also extremely limited. Only one other example (Dobson, 2006) has been identified in the literature reviewed for this research study. Dobson (2006) acknowledges that examples of assessing students through their Microsoft PowerPoint presentations in the academic community are rare. He reported on such an assessment with undergraduates at a Norwegian university and focused on the validity of assessing PowerPoint slide shows. On the basis of his case study, Dobson concluded that presentations can be set as student assignments and used for assessment purposes, in particular for formative purposes. He also called for further case studies of PowerPoint presentations used for assessment within an academic setting.

In order to provide a supporting framework for the procedures later described and discussed in this report, this review will focus on discussion of the following two areas: a brief clarification of the formative and summative assessment distinction, and source literature related to achieving synergy between both assessment purposes.

1. The FA/SA distinction

It has been noted that summative and formative assessment functions can be seen as the “the ends of a continuum along which assessment can be located” (William & Black, 1996, p.542). Sadler (1989) explains that the meanings attached to the word formative relates to the idea of “forming or molding something, usually to achieve a desired end”, while summative “is concerned with summing up or summarizing the achievement status of a student and is geared towards reporting at the end of a course of study” (p.120). Sadler (1989) was one of the first

to note that the primary distinction between FA and SA was not related to the timing of the assessment, but rather to assessment purpose and effect. However, timing is a consideration of note, and connects with a key difference between SA and FA; while summative assessment generally signifies the end of instruction (or part thereof), formative assessment anticipates that further action will be taken (Baroudi, 2007).

What determines whether an assessment may be labeled formative or summative is dependent on how the results are used; SA documents how much learning has occurred at a point in time and its results are used to make some sort of judgment (typically a grade), while FA is more dynamic and ongoing and its results are used by students and teachers to decide next steps in the learning process (Chappuis & Chappuis, 2008). A key point here is that an assessment may not be inherently labeled formative or summative, but application of these labels will be dependent on purposes, and particularly use of results. Wiliam & Leahy (2007) are explicit about this stating:

... the terms formative and summative apply not to assessments themselves, but to the functions they serve. As a result, the same assessment can be both formative and summative. Assessment is formative when the information arising from the assessment is fed back within the system and is actually used to improve the performance of the system. (p. 39)

The crucial idea here is that formative assessment is focused on the improvement of student performance. The peer-assessment report from Chapter 2 provided an example of an assessment framework serving both formative and summative functions.

It should also be noted that while all assessments have the potential to serve a summative function, only some have the additional capability of serving formative functions (Wiliam & Black, 1996). After their review of the literature, Wiliam & Leahy (2007) point out that the empirical evidence suggests that the assessment which has the greatest impact on student achievement is short cycle FA (i.e. within and between lessons). Yet, teaching also involves determining and assigning grades which monitor and record student achievement, and somehow AfL principles must be reconciled with this summative-purposed assessment of learning.

2. Synergizing FA and SA

In the language assessment literature, some attention has been paid to the issue of bringing

formative and summative assessment together for a more synergized relationship (for example, Rea-Dickins', 2001; Lantolf & Poehner, 2004; Gottlieb, 2006). Here, in particular, we pay some attention to two other pertinent sources dealing with the use of summative assessment for formative purposes: Black et al. (2003) and Maxwell (2004).

Black et al.'s (2003) book, the catalyst and primary source work for this thesis, which reported on putting assessment for learning ideas into practice involved 36 Math, Science and English teachers in secondary schools in England. In that context, the giving of regular tests was a familiar part of classroom practice and attempts were made to convert these summative tests into more formative assessment practices. The investigation conducted at these UK schools found that teachers struggled to reconcile formative assessment practices with the pressures of external high-stakes summative testing. Black et al. (2003) noted that it was unrealistic to expect teachers to practice separation between assessment for and assessment of learning and that the challenge was to achieve a more positive relationship between them. In relation to the formative use of summative tests, the teachers devised three main ways of using classroom tests, beyond just assessing attainment, to develop students understanding:

1. Helping students to prepare for tests by reviewing their work and screening past test questions to identify areas that could be improved. This reflection of their areas of weakness enabled them to focus their revision.
2. Asking students to set test questions and devise marking schemes.
3. Using the outcome of tests diagnostically and to involve students in marking each other's tests, in some cases devising a marking scheme.

The teachers involved in Black et al.' (2003) research, "used their creativity to graft formative value on to summative procedures" (Harlen, 2005, p. 217). This grafting of formative value on to a summative assessment is also a good description of what was attempted in the procedures reported in this report. As will be seen, the assessment framework discussed here is similar to the first approach mentioned above; preparing for the final presentation (the 'test') to identify weak areas and focus on revisions that would improve the students' slide shows.

Another relevant source describing the use of summative assessment for formative purposes is by Maxwell (2004), where it was used for certification purposes in secondary schools in Queensland, Australia. Maxwell describes an assessment approach where student portfolios are used to collect evidence of learning over time. Maxwell (2004) defines this

process as ‘progressive assessment’ and writes:

... progressive assessment blurs the boundary between formative and summative assessment. All progressive assessment necessarily involves feedback to the student about the quality of their performance. This can be expressed in terms of the student’s progress towards desired learning outcomes and suggested steps for further development and improvement. (pp. 2-3)

In order for such an assessment process to work, according to Maxwell, it is necessary for learning expectations to be clearly expressed for students in terms of criteria showing common dimensions of learning. Only when this is done can students be engaged as to whether they are on-target with regard to the learning objectives, and know what is required for them to improve their performance in future assessments using the same criteria. The process discussed in this report takes a similar approach to Maxwell (2004) in using the same common dimensions of learning (criteria) for both formative and summative assessment events.

Other related sources from the assessment literature also provide useful reference points and sources of information for teachers and researchers exploring the issue of formative and summative assessment synergy.

John Biggs has been an influential voice in the area of teaching and assessment in higher education, and has focused some attention on formative and summative assessment working together to assist the learning process. Biggs (1998, p.105) argued that, “sensible educational models make effective use of both FA (formative assessment) and SA (summative assessment)”. He noted “there is a powerful interaction between FA and SA that could usefully be incorporated in an overall synthesis, so that both backwash (from SA) and feedback (from FA) are conceptualized within the same framework” (Biggs, 1998, 106). The ‘overall synthesis’ referred to here was also a goal of this research project, an attempt to conceptualize FA and SA within the same assessment framework.

In writing about assessment formats and their potential to enhance learning, Kennedy et al. (2005) contend that the coalescence of FA and internal summative assessment is possible. While noting that SA is different in form and function from formative assessment, these authors note that a synthesis between the two can occur if SA can be more inclusive of the basic principles of formative assessment, and this will have a positive effect on the learning and teaching that occurs in a course.

Two reports focused on secondary school assessment, in the U.S.A and in the U.K, provide useful sources of information and analysis for implementing a balanced, integrative assessment framework.

Published by the National Education Association in the U.S., *Balanced Assessment: The Key to Accountability and Improved Student Learning* (2003) offers a number of useful points for consideration when attempting to make formative use of summative assessment. This report asserts that in a balanced assessment system, FA and SA will complement each other in using assessment to both increase student achievement and then to document that new achievement status. It noted that, “Although they are different, both assessment of and for learning are important. While they are not interchangeable, they must be compatible” (Balanced Assessment, 2003, p. 7).

The second report was published by the Quality Assurance Agency of Higher Education in the U.K. and is entitled *Integrative Assessment: Balancing assessment of and assessment for learning* (2007). This guide discusses a variety of ways of striking an optimal balance between FA and SA in order to help students reach their fullest potential as learners. One such strategy is termed ‘feedforward assessment’, and is described as follows:

This strategy aims to improve the balance between assessment for and assessment of learning by interlinking the twin functions more directly, yet not confounding them. . . . What this particular strategy seeks to do is to convert feedback into feedforward, by interconnecting assignment and assessment tasks and creating a recursive cycle, or 'feedback loop', in which feedback comments on one task, draft or set of questions can be fed directly into a subsequent task or draft, or will aid preparation for an exam. . . . students therefore have the opportunity for 'low-stakes' practice on assessable work, and to benefit directly from the feedback in a way that can also contribute to a subsequent formal mark or grade. (Integrative Assessment, 2007, p.4)

As will be seen, such a feedforward framework and the use of a ‘low-stakes practice’ is an apt description of what this formative use of summative assessment report describes. The Integrative Assessment (2007) report also notes that the same assessment can be used for both formative and summative purposes.

An article explicitly focused on the uses of formative and summative assessment is Harlen’s (2005) *Teachers summative practices and assessment for learning—tensions and synergies*. In it, Harlen discusses how FA and SA can affect one another in both positive and negative ways. Harlen makes it clear that for an assessment to have a formative purpose it is essential to report to students the things that need to be improved to raise their performance to a higher level. Harlen (2005) contends that while formative and summative assessment are

needed for different educational purposes, they can exist in synergy. However, she stipulates that to achieve some degree of synergy between FA and SA, teachers need to be involved in planning and developing assessment criteria.

In an article concerned with the perceptions and realities of formative and summative assessment, Taras (2008) notes that discussion about the tension between SA and FA is prevalent in the assessment literature. However, she also points out that the two are more easily reconciled in a higher education context, due to the absence of external summative testing, and the fact that all assessment is the responsibility of instructors and controlled by them. Her article reports on a small-scale study of lecturers at an English university, and found that they had an incomplete and fragmented understanding of FA and SA and the relationship between them. Taras (2008) writes that future research needs to address the issue of reconciling FA and SA so that they are ‘mutually supportive’, and notes that reports of learners involvement in the assessment process has been a neglected area of research.

Methods

Public Presentations (PP) is a two-semester course made up of 20 third-year students. Classes met weekly, for 90 minutes, during each 15-week semester. Essentially, the PP course was designed to develop in students the communication skills necessary to analyze verbal discourse and to perform effectively in public speaking situations. Course objectives included students gaining the skills, experience, and self-confidence for speaking in public settings, including a thorough technological competence for multimedia presentations. The course textbook was *The Essential Elements of Public Speaking, 2nd Edition* (DeVito, 2006).

In the PP course, topics covered included methods of organizing and delivering a speech, the types and uses of evidence as supporting material, and the effective use of visual aids, particularly computerized slide shows. In addition to regularly occurring mini-presentations during class time, students were required to prepare and deliver a main final presentation at the end of each semester. In the first semester students delivered an informative speech, while the focus shifted to a persuasive presentation for the second semester. These two main presentations were required to be between 8-10 minutes in length, and supported by a computerized slide show (typically using Microsoft PowerPoint software).

PP was divided into Part 1, in the first semester (April-July) and Part 2 for the second semester (Sept. – Jan.). The assessment framework for Part 1 was as follows: three chapter

tests (45%), final presentation (40%), and a final presentation self-report (15%). This self-report involved answering a series of questions about the final presentation after viewing a video recording of their performance. As mentioned, the final presentation for Part 1 was an informative speech, which, according to the course textbook, “seeks to create understanding: to clarify, enlighten, correct misunderstandings, or demonstrate how something works” (DeVito, 2006, p.18).

Chapter tests, used in both semesters, were taken from the *Instructors Manual and Test Bank* accompanying the course textbook and were included to ensure that students were reading and engaging with the textbook content. Chapter tests were comprised of multiple-choice and true-false items. Scores were recorded and used for summative grading purposes.

In the second semester, changes were made to the assessment framework of the course. The biggest change was that I decided that the presentation slide show and the performance (speech delivery) would be assessed separately. This was primarily due to the problematic nature of student slide shows for their informative presentations at the end of the first semester. Problems with slide shows (such as excessive textual information, unclear progression of ideas, poor layout, and lack of source citation) were common in students’ informative presentations, and I decided to pay special attention to improving student slide shows in the second semester. An added benefit of making two separate assessments for students’ final presentations was that it also lightened the teachers’ assessment load on presentation days. As students were presenting, a primary focus could be on assessing students’ spoken delivery and the slide shows could be separately evaluated at a later point.

The assessment framework for the second part of Public Presentations was as follows: three chapter tests (45%), slide show (25%) and final presentation delivery (30%). Students were required to submit a final version (color copy) of their presentation slides on the day of their final presentation and it was these sets of slides which were summatively graded. As mentioned, in the second semester, students were required to deliver a persuasive presentation, one which “seeks to influence attitudes or behaviors: to strengthen or change audience attitudes or to inspire hearers to take some specific action” (De Vito, 2006, p.18). Persuasive topics selected by students included those opposing TV advertising for kids, cosmetic surgery, vegetarianism, and those supporting school uniforms, gay marriage and abortion.

The class textbook, and the course syllabus based on it, takes students through ten steps in preparing and presenting a public speech, including such stages as developing a thesis and

main points, and constructing the speech. Approximately a month before doing their final presentation, students reached this construction step in the process, and it included working on the presentation slide show. At this point students were required to bring to class a first draft of their slide show. It should be noted that in the two classes prior to students submitting their first drafts, classroom instruction specifically paid attention to the slide show criteria used on the assessment rubric (introduction, content, text elements, layout, graphics/graphics, writing mechanics and citations). In addition, the course textbook chapter on ‘Using Supporting Materials and Presentation Aids’ (De Vito, 2006) was reviewed, with particular emphasis on the extensive section related to computer-assisted presentations. Students were told to use this textbook section to help plan and organize their slide shows, and classroom instruction focused on it.

On first draft submission day, a slide show assessment rubric was distributed and time was spent reviewing and explaining the criteria. Students were reminded that the exact same criteria would be used to assess the final draft of their slides accounting for 25% of their final grade. They then were instructed to take 20 minutes to self-assess the first draft of their slide show.

A student self-assessment version of the slide show evaluation rubric is shown below in Table 1. This rubric is a modified version of the PowerPoint score sheet made by Vandervelde (2001), and available online. The teacher versions, both formative and summative, were very similar to this, with the same criterion specified.

Table 1. Slide show evaluation rubric, Part 1 (student self-evaluation version)

Presentation Slideshow (first draft)		<u>Self-evaluation</u>
Student name:	Topic:	
<u>Part 1</u>		
		4 = very good 3 = good 2 = average/OK 1 = weak
Criteria	Rating	
1. Introduction	Introduction presents the overall topic and draws the audience into the presentation. It is clear, coherent and interesting.	
2. Content (most important)	Information is accurate and serves the purpose of persuading the audience. The content is clear and concise with a logical progression of ideas and supporting information.	
3. Text elements	Fonts, point size and colors make text easy to read. Text is easy to follow and understand.	
4. Layout	The layout is visually pleasing and contributes to the overall message with appropriate use of headings, sub-headings and white space. Layout is not cluttered or confusing.	
5. Graphics/visuals	Graphics/visuals assist in presenting an overall theme and enhance understanding of concepts, ideas, etc.	
6. Writing mechanics	The text is clearly written with no/few errors in grammar, vocabulary, and spelling. Easy to read and understand.	
7. Citations	Sources of information are properly cited so audience can determine credibility and authority of information.	
Slideshow: Overall Evaluation	<u>first draft</u>	
	4 3 2 1	

During self-assessment, students were required to carefully examine the first draft of their slide show according to the seven different criteria and mark each element from 4 (very good)

to 1 (weak). An overall mark for the draft of the slide show deck was also required, for a more holistic view of the deck they had constructed thus far. On Part 2 of this evaluation, shown on the reverse side of the student version, students were instructed to make notes in three boxes related to (1) strong points of their slide show, (2) weak areas, and (3) things to work on for the final draft.

This rubric, and the three versions of it used in the assessment regime described in this report, makes concrete for students the idea of *task clarity* in that they can clearly understand the slide show learning goals and how their learning will be evaluated (McTighe & O'Connor, 2005). In writing about rubrics, Arter (2004, p. 2) states:

The best use of rubrics in the service of student learning occurs when the rubric or scoring guide really covers what is essential for a high quality performance— nothing of importance is left out and nothing trivial or off the mark is left in. By using this type of rubric, doing well on the assignment and learning what quality looks like will occur together.

As with other rubrics used in the research studies in this thesis, the slide show rubric became a learning (and teaching) tool as well as being an assessment instrument. It also incorporates the key AfL idea of sharing assessment criteria with students.

After student self-assessment (formative assessment 1), the decks were then given to the teacher for a second formative assessment, using a teachers' version of this rubric. The teachers version used in assessing the first draft of student decks included an added space at the bottom for written commentary. Notes were also made on the slides themselves indicating confusing parts, or suggesting changes. For example, awkward phraseology was marked "R" for 're-write'. While problems or weak areas on slide shows were identified by the teacher, students were required to make their own revisions (just as in Chapter 1's feedback research). Student self-evaluations were not viewed by the teacher prior to formative assessment, and only given a cursory viewing later. While providing formative feedback on individual slide shows, notes were made by the teacher about recurring problems (for example, lack of citations) and these notes were used to plan for re-teaching in subsequent classes.

First drafts of slide shows with teacher notations, and the teachers' version of the rubric (formative assessment 2) were returned in the following week's class. Students were given time to work on revisions during class and the teacher was able to talk briefly and individually with most students. Students were able to ask questions, and time was also made in the final classes for re-teaching or reminding students about weak areas noticed in the first

drafts. After this class, students had approximately two weeks to further edit and revise their slides and prepare for their final presentation.

Final presentations for the 20 students were completed over a three-day period during the final exam week, with approximately 7 presentations delivered in a 90 minute session on each day. On those days, the performance of student presentations was graded using a different assessment rubric, which included such criteria as vocal and physical delivery, language use and effective conclusion. In addition to evaluating presentation delivery, final versions of the slide shows were picked up from students for summative assessment, using the same criteria included in both formative assessments.

One week later students were able to pick up both grades (one each for presentation delivery and slide show). Students' final grade sheets contained two summative assessments (with feedback and scores), one on each side of the grade sheet. One side contained the summative slide show rubric, while the other contained an assessment of the performance, worth, respectively, 25% and 30% of their final grades.

The assessment tools and procedures described above show how synergy of this trio of assessments was designed in order to use the assessment process not just to check for learning, but to promote more learning regarding effective presentation slide show construction. This process of making formative use of summative assessment will be further delineated and analyzed in the following section, which also examines some student examples for each of the three assessments.

Discussion

The discussion section of this report takes a closer look at how the summative assessment of student slide shows was used for formative purposes in order to achieve a more synergized assessment framework, and promote student learning. It will be divided into three parts: first an analysis of the two formative assessments (student self-assessment, teacher assessment), followed by a discussion of the summative assessment and grading of the final version of student slide shows, and finally a discussion of the synergy demonstrated in integration of the three assessments. Readers will be able to have a clearer picture of how the assessment process played out in practice through the use of five examples from students' slide shows. The complete assessment documentation of one of these students' slide shows is included as an exemplar in the appendices to this report.

At the beginning of this section it may be useful for the reader to briefly turn to the Appendices section, as it shows readers an example of the assessment process used for each student, and will help clarify the discussion to follow. The following four appendices are included for one student (Yuka) and her slide show supporting the controversial topic of abortion: **Appendix B1** (formative assessment 1: student self-assessment, part 1), **Appendix B2** (part 2 of student self-assessment), **Appendix B3** (formative assessment 2: teacher assessment and feedback) and **Appendix B4** (summative assessment and score for final draft of slide show). These appendices may be found on pages 276 to 279.

The reader is reminded here that the assessment procedures used in this process stem from the teachers' practices described by Black et al. (2003) in UK secondary schools as they tried to work out useful strategies in making formative use of summative tests. Two of the three practices used by the teachers in that research report were made use of here: engaging students in a reflective review of their work to plan effective revisions, and using self-assessment and applying specific assessment criteria to show students how their slide shows could be improved. In this case, as already noted, a student's final presentation may be considered as two separate tests - the delivery of the presentation, and the slideshow. The formative assessments of slideshows used in this course were strongly influenced by, indeed a preparation for, the summative assessment of the final version of the slideshow.

Part A. The Two Formative Assessments

1. Student self-assessment

The requirements needed for assessment to serve a formative function were delineated by Wiliam and Black (1996) as follows:

In order to serve a formative function, an assessment must yield evidence that, with appropriate construct-referenced interpretations, indicates the existence of a gap between actual and desired levels of performance and suggests actions that are in fact successful in closing the gap. (p. 542)

As will be seen, the slide show assessment rubric used here notifies students of this gap between actual and desired performance levels of their work, and points them towards the gap-closing actions they need to take prior to the final graded assessment.

As noted in the methods section, students' were required to self-assess their slide shows according to seven criterion (introduction, content, text elements, layout, graphics/visuals, writing mechanics, citations) prior to giving an overall evaluation of their first draft (4= very good, 3= good, 2= average/OK, 1= weak). On the reverse side of the rubric, for part 2, students made notes related to strong and weak points and things to work on for the final version of the slide show. Table 2 below shows how four students completed the self-assessment (SA) of their first draft of their decks. Next to the student name, the table provides the topic of their persuasive presentation, and their self-assessment mark. Student commentary (related to strong points, weak points and things to work on) has been copied verbatim from the original documents.

Table 2. *Four students self-assessment of slide show first drafts*

Student	Persuasive topic	SA overall mark	Slide show: strong points	Slide show: weak points	Things to work on for final version
Ikuko	Against GM foods	1 (weak)	- sources	- too many words on slides - few visuals - same titles for some slides - not enough slides	- I'd like to put many visuals, graphs and so on. I want to increase the number of slides because I want to find more information
Emi	In favor of School Uniforms	2 (average)	- good visuals	- few slides - point size is small - not enough evidence for support - a little confusing	- more, better evidence - more slides - content is clear
Sayaka	(Being) Conscious Consumers	2 (average)	-'Buy Nothing Day' slides are fine	- Not enough explanation - Not enough visuals - Introduction is not interesting for audience - I need to explain the negative effects of huge consumerism - lack of citations and references page	- More on negative effects of consumerism - find more visuals - get more information to explain and introduce topic
Risa	Supporting Gay Adoption	3 (good)	- Chart and map are effective to explain topic and help audience to understand topic easily. -Basic information is also enough to explain topic.	- Have to analyze situation after adoption and focus more on adoptive children's' issues	- I need more research about adoptive children, and I'd like to make my slides better in terms of visuals.

The assessment literature is replete with ideas on the positive benefits of student self-assessment, a topic explored in detail in Chapter 2. Suffice it to say here that Black and Wiliam (1998b) contend that self-assessment is ‘an essential component’ of FA but they also importantly point out that students “can only assess themselves when they have a sufficiently clear picture of the targets that their learning is meant to attain” (p.9). In this case, the rubric criterion identifies the key elements for an effective slide show and a sufficiently clear target picture is presented for students to compare their slide shows against. The specification of the assessment criteria should enhance student understanding of the assessment task (Yorke, 2003), and clearly points to the goals they are working towards. The rubric criteria also provide information in how their work may be improved (for example, ‘sources of information are properly cited’).

Students were explicitly reminded that exactly the same criterion would be used to grade the final version of their slides, as used here in this ‘low-stakes practice’. The importance of such formative use of student self-assessment is captured by Sadler (1989); “A key premise is that for students to be able to improve, they must develop the capacity to monitor the quality of their own work *during actual production*” (p.119, emphasis in original). This first formative assessment encourages such monitoring as students begin to work toward producing the final version of their slide shows.

2. Teacher assessment

After completing their self-assessments, the first drafts of the slide shows were collected for teacher assessment (TA) and the provision of feedback. The feedback included returned slide shows with margin notations, questions or suggestions, including the use of a limited correction code (for example, ?= not understood, R=rewrite). Attached to the returned first draft was the teachers’ version of the slide show rubric, with the added ‘comments’ section.

Table 3 below shows teacher marks and feedback comments for the same four students (Ikuko, Emi, Sayaka, and Risa) as in Table 2 above. The overall marks in Table 3 are intended for formative purposes only, simply to give students a general idea of how this draft of the slide show looks from the teachers’ perspective.

Table 3. Teacher formative assessment of four students' slide show first drafts

Student/ Topic	TA overall mark	Rubric comments
Ikuko - Against GM food	2 (average)	Your position on the issue is clearly given Ikuko, and you do present the other side also. You provide three main reasons but this could be more effectively done. Editing and revising should make the slides stronger and clearer. Pay attention to the items I have scored 2 and 1 (citations) above.
Emi - In favor of School Uniforms	2 (average)	A good start Emi, but the content needs to be a bit stronger and deeper. Are you focusing on the U.S.A? More factual information about the situation there would be useful (e.g. costs). More pictures of uniform types would also help, and better organization of your content (reasons why you are in favor of uniforms).
Sayaka - Being Conscious Consumers	1 (weak)	Content needs to be stronger in order to persuade people to be aware of/care about this issue. Lots of work to do to turn this into an effective slide show, Sayaka. Too much about 'Buy Nothing Day' – a third of your slides.
Risa - Supporting Gay Adoption	2 (average)	Some good parts, Risa (e.g. visuals), but the problem is the content is too informative and not persuasive enough. What are your arguments supporting gay adoption? Make them stronger and clearer for the audience.

Readers will note some discrepancies between teacher and student self-evaluations, including a higher teacher assessment for Ikuko, and lower teacher assessments for both Sayaka and Risa. This may be explained by the fact that, while being familiar with the key components in an effective slide show, the students' self-assessment was the first time for them to see the criteria together in one rubric. It may be that the teacher has a firmer grasp of the key components and looked at slide shows with a more critical eye. Perhaps these students do not yet have a concept of quality roughly similar to that held by the teacher (Sadler 1989).

The key consideration here is directing formative feedback towards closing the gap between the learning aimed for and students' present understanding and performance (James, 2006). Students need to see what needs to be improved in order to raise their performance and produce better slide shows prior to submitting the final version for summative grading.

While the teacher commentary provides holistic feedback, assessment on individual criteria in the rubric indicates deck strengths and weaknesses to students, and where they need to focus their attention in order to be scored higher in the summative assessment to follow.

Guskey (2003) writes: “Teachers who develop useful assessments, provide corrective instruction, and give students second chances to demonstrate success can improve their instruction and help students learn” (p.6). Making note of areas that scored 1 (weak) or 2 (average) in teacher assessment of student slides showed that the key areas that students needed to improve were content, writing mechanics and citation of sources. In particular, the first draft of student slides often contained a lot of informative content but lacked in persuasive reasons for taking a position and making these reasons explicit for the audience on the slides. These areas were re-focused on as learning objectives in the two subsequent lessons following the return of the first drafts and the teacher feedback. In this way the progress of learning was being assessed as it happened, enabling adjustments to be made for things not working as planned (Integrative Assessment, 2007). Making adjustments to class content and instruction focused on enabling students to strengthen the weaknesses in their slide shows prior to the final presentation.

The two formative assessments described here are an attempt to promote a situation described by Wood (1987, p.242) in which “the teacher/tester and the student *collaborate* actively to produce the best performance” (emphasis in the original). Collaborating and working together to ensure that students could create the best possible slide show to support their presentation is evident through the formative component of the process described here.

Based on the findings from Black & Wiliams’ (1998a) seminal formative assessment review, Clarke (2001) identified a number of key factors for effective formative assessment. The FA procedures engaged in here in this research study are assessed, and shown in Table 4 below, as to whether they include (✓) or exclude (✗) these elements.

Table 4. Assessing the formative assessment process used

Formative assessment - key elements (Clarkee, 2001)	✓ / ✗
1. Providing effective feedback to students	✓
2. The active involvement of students in their own learning	✓
3. Adjusting teaching to take account of assessment results	✓
4. Recognizing the influence of assessment on student motivation and self-esteem	✓
5. Students assess themselves	✓
6. Students understand how to improve	✓

Regarding FA, Black (2003) makes a further critical assertion that: “For formative assessment, the learning caused by the assessment is paramount . . . What really matters is whether the result of the assessment is successful learning” (p. 14). In this case, did the two formative assessments, and subsequent classroom instruction, promote successful learning of slide show construction? Examining the summative assessment of students’ slide shows should help answer this question.

Part B. The Summative Assessment

As students’ final presentation sessions arrive, the teacher must disengage from the ‘active collaboration’ process, and put some distance between themselves and the students while putting on the judge/summative assessor hat. This can be a very challenging transition to make after spending weeks or months in helping, guiding and supporting student learning. Nevertheless,

... formative assessment must at some point - or points - give rise to a summative assessment of achievement at a predetermined stage of the learning process. Eventually formative assessment, which forms the platform for teaching in a course, must give rise to measurement of desired learning outcomes. (Dunn, et al., 2004, p.74).

During final presentations, the focus was on students’ 8-10 minute performance and, as mentioned, they were separately graded on this most important element (worth 30% of the

final grade). After their performance, students submitted the final drafts of their slides, those used during the persuasive presentation, to be summatively graded. Table 5 below gives a summary of the final grades and teacher commentary for the slide shows of the four students previously mentioned. While the same evaluation criteria was used as in the formative assessment rubric, a different overall scoring range (from 3 to 10) was used for each slide show. This broader overall score range provided a finer scale of options for the grading of the final edition of the slides. The slide show score was used to comprise 25% of the overall course grade.

Table 5. Summative assessment of slide show final drafts

Scoring framework: **Weak (3-4), Average/OK (5-6), Good (7-8), Very Good (9-10)**

Student/topic	Presentation slide show score	Teacher comments
Ikuko - Against GM food	8 (good)	A persuasive argument is effectively presented on the slides, with a good selection of visuals. Better editing and organizing of content and better writing mechanics would have made it even stronger. A good slide show, Ikuko.
Emi - In favor of School Uniforms	8 (good)	A good slide show Emi, especially the visual elements. Organization of information and reasons for your position could have been a bit stronger and better.
Sayaka - Being Conscious Consumers	9 (very good)	Very good slide show Sayaka, which effectively supports your persuasive message to care about being conscious consumers. Good choice of supporting visuals also. Some textual elements could have been a bit clearer and easier to read.
Risa - Supporting Gay Adoption	9 (very good)	A strong slide show Risa, with great visuals. Reasons for your position are organized and easy to follow. But a little more background information about the issue would have made the slide show even stronger.

Many of the 20 students in the PP class (for example, Sayaka and Risa above) made substantial improvements in their slide shows after the formative assessments of their first drafts. In order for an assessment system to function formatively as well as summatively, it must be responsive in a timely fashion to the information that is produced (Wiliam & Leahy, 2007). In the assessment system reported on here, students had time to assess their current level of performance, prior to the graded event, and make the necessary adjustments while it still counted. As such, an affirmative response can be made to the paramount FA concern identified by Black (2003) as to whether the assessment led to further learning taking place. The assessment regime implemented did result in better slide show production for the final presentations.

As with the evaluation of the FA process above by using ideas from Clarke (2001), we may also briefly evaluate the summative assessment process. According to Stiggins et al. (2004), to produce accurate results, four standards of quality must be met by assessment procedures. They must:

- be designed to serve a specific, pre-determined purpose;
- arise from a specific, predetermined definition of achievement success;
- be designed specifically to fit into each particular purpose and target context;
- communicate their results effectively.

The summative dimension of the assessment process described here meets each of these standards of quality.

According to Black (2003): “Summative tests should be, and be seen to be, a positive part of the learning process in which pupils come to see that tests are helpful to them, not merely, not even principally, ways of judging them” (p.16). The summative assessment discussed here, evaluation of the final draft of student slide shows, had such an intention to be a positive, helpful part of the learning process. While it did serve as an exercise in judging students’ work, it was principally intended to help students produce a better slide show for their final presentation. I believe it was successful in doing that. Whether students saw this summative assessment as a positive, helpful part of the learning process is unclear and perhaps less important than the fact that it did in fact play such a beneficial role.

Part C. Synergy

The fact that the same slide show criteria were used for all three versions of the assessment is the glue that bonds them together into a synergized whole. Biggs (1998) asserted that FA and SA are best connected when they are ‘deeply criterion-referenced’, and specify what counts for quality work.

In this situation . . . where reflective learning takes place, the backlash from the summative assessment tool can be very positive. A condition is that that assessment is deeply criterion-referenced, incorporating the intended curriculum, which should be clearly salient in the perceived assessment demands. When that happens you get aligned instruction, where teaching to the test is exactly what you want because it is teaching the intended curriculum. . . . the summative assessment is defining the parameters for the formative assessment. (Biggs, 1998, p. 32)

That is what happened in this case, where the summative assessment rubric clearly defined, indeed was almost identical to, the two formative assessments used. Classroom instruction did indeed involve ‘teaching to the test’ as students constructed and adjusted their slide shows according to the assessment criteria specified.

The assessment framework described here shows formative and summative assessment integrating and complementing each other, and the synergy created did result in both improved student learning and teacher instruction about effective slide show construction. Dunn et al. (2004) write:

Ideally, summative assessment comes at the end of a systematic and incremental series of learning activities that have formative assessment tasks set at key points during the course. The formative assessment tasks are interwoven with the teaching and learning activities and then a summative task is set (p. 18).

This is a reflection of what happened in the process described in this research study, as the various assessment strands were set at key points, and tied together to form a coherent whole that had a positive effect on student learning.

Conclusion

The assessment process described in this report had its impetus in a desire for more student attention being paid to the improvement of slide shows in the second semester of the Public

Presentations course. This resulted in the modification of the final presentation assessment into two components, one for the slide show and another for the performance. It would have been simpler to use only a summative assessment of the final presentation slide show. But that would not have necessarily led to better slide show production from students. To make this learning outcome more likely, FA was injected into the process. A new assessment framework was constructed, one aimed at improving students' performance through a series of three assessments steps: self-assessment, teacher assessment (both formative) and the final summative assessment. In this way, assessment originally designed for a summative purpose was converted into assessment *for* learning (James, 2006).

Chappuis (2005) is perceptive in describing assessment for learning as 'a human process', one where students and teachers work together to generate information about what students have learned and then using this knowledge to promote even more learning. The synergized FA and SA process described here was indeed such a human process, with the combined effect of the three assessments leading to better slideshows and promoting student learning. As Brookhart (2007, p.45) contends, "Formative and summative assessment can be seen as parts of the same whole (the whole being student learning)".

In order to be supportive of student learning, assessments should not be just isolated, one-shot, graded events but, wherever possible, be part of a series which is ongoing and interconnected (Stiggins, 2006). This report describes one such interconnected example of grafting formative value onto an end-of-term summative assessment. It required a little more time and effort on everybody's part, but the payoff in better student work was tangible, with noticeably better presentation slide shows in the second semester of the course.

The end of this fifth and final report brings us, following the Appendices section, to the concluding chapter of the thesis where we take a more holistic view of this research project putting AfL into practice in a higher education EFL context.

Appendix A1 *Growing Up Online* tape script excerpt (from Frontline, 2008)

Part 4. The Child Predator Fear [page 1 of 4]

EVAN SKINNER: My fear isn't that I have bad kids, my fear is that my good kids will make a bad decision, one bad judgment, and pay for it permanently. If it's on the Net, it's open to anyone. There are no safeguards. Someone can always find everything.

NARRATOR: Evan Skinner is the-stay-at-home mother of four teenagers in Chatham, New Jersey, and president of Chatham High School's parent teacher organization.

EVAN SKINNER: There is no lack of parent involvement in this community. It is what is socially expected here. There is a culture of involvement. There's a culture of participation, and so people do.

NARRATOR: Chatham, New Jersey, is less than an hour from Manhattan by train, but has the look and feel of a small town. Parents here have worked hard to create a haven for their children.

EVAN SKINNER: Our kids are very much the children of a small town in a protected environment. Kids walk to school. There are crossing guards. It's incredibly friendly. You know, it's safe.

NARRATOR: But the Internet, and social networking in particular, has punctured that sense of safety.

EVAN SKINNER: The scariest, worst part for me is stalkers, is somebody becoming obsessed with one of my children. I have two very attractive daughters. You know, some guy that all of a sudden decides that, really, my daughter was meant for him- that kind of stuff scares me. Kids think, "I'm in my home. How could anything bad happen to me?" They don't realize that when they're sharing on that keyboard, it's, like, Let 'em on in baby, because they're right here.

GREG ABBOT, Texas Attorney General: Parents need to understand there are predators on the Internet who are more vicious than those who used to lurk in parks or playgrounds.

NARRATOR: Media coverage of on-line predators has been building in the last year.

Fat and Skinny Questions

In the coming weeks, we will spend some time on *student-generated questions*. These are questions that students make about the things they read and issues related to the readings.

Being able to ask good questions is a skill, one that will be useful as a university student, and throughout your life. One simple but effective way to think about questions is to divide them into fat questions and skinny questions.

Question type	Definition	Examples
Skinny	<p>Skinny questions often recall exactly, or in your own words, information you read about. The answer can usually be found in the reading. Or they can be simple, factual questions to another student that can be answered in a few words (e.g. yes, no).</p> <ul style="list-style-type: none"> - Skinny questions are factual, closed (only one right answer), and direct. 	<p>What is... (the reading about?)</p> <p>What ...?</p> <p>Can ... ?</p> <p>Who is...?</p> <p>When did...?</p> <p>How many...?</p>
Fat	<p>Fat questions are open-ended and ask for a longer, more thoughtful answer. Such questions are deeper and require students to think deeper about a reading or an issue/idea. There is no 'correct' answer.</p> <ul style="list-style-type: none"> - The answer to a fat question won't be found stated in a reading, and will involve opinions, judgments, feelings etc. 	<p>Why...?</p> <p>Do you think...?</p> <p>How do you feel about....?</p> <p>What if...?</p> <p>Would you...?</p> <p>Do you agree . . .</p>

Both types of questions are important, but fat questions are the ones that will require more critical thinking about readings and issues. In the coming weeks we will work on students making questions about the material we are working on in class. The hope is that by making questions and discussing them together, you will develop a better understanding of the readings/issues and this will help you when you later write your final essay for the course.

Appendix A3 Question generation sheet used by students for SGQ (pairs version)

Student questions: **Growing Up On-Line (PBS documentary)**

Question maker (s): _____

GUO chapter number: _____ GUO chapter name: _____

Working with another student, make some skinny and fat questions related to one chapter of GUO tape script. In a few minutes you will get together with other students and talk about your questions (be sure they are clear and easy to understand). They will also ask you questions about a different GUO chapter.

Question type	Questions
	1.
	2.
Skinny	
Skinny questions usually have short answers which can be found in the reading.	3. 4.
	5.
	? ? ? ? ? ? ? ?
	1.
Fat	2.
Fat question are deeper and require thinking, analysis and sharing opinions, feelings about something related to the reading.	3. 4.
	5.

Appendix A4 Final class survey regarding SGQ

Student generated questions

For some classes at the end of the course, we spent time on student-generated questions. These were questions (skinny and fat) that you made about the things you read and saw, and issues related to Growing Up On-line (GUO) documentary. I am very interested in your feelings about student-generated questions. Think about the following survey items, and provide an honest response to each one. Thank you for your feedback.

1. I _____ the reasons why students were asked to make questions about GUO and use them in small group discussions in this class.

- a) understood b) kind of understood c) did not understand

2. The idea of skinny and fat questions and the difference between them was

- a) easy to understand b) kind of easy to understand c) difficult to understand

3. Making questions and using them in small group discussion was something I _____ doing.

- 1) really liked 2) liked 3) did not like 4) really did not like

4. Making questions and discussing them in small groups was _____ in understanding the GUO documentary and the issues related to teenagers and the Internet.

- a) very useful b) useful c) a little useful d) not useful

5. Making questions and discussing them in small groups was _____ in helping me write the GUO essay.

- a) very useful b) useful c) a little useful d) not useful

6) I would like student questions to be included in the second half of this course.

- a) strongly agree b) agree c) disagree d) strongly disagree

7. Any other comments (about student-generated questions)?

Appendix B1 Formative Assessment 1- Student Self-Assessment (Part 1)

Student Name: <u>Yuka</u>		Topic: <u>Abortion</u>			
4 = very good 3 = good 2 = average 1 = weak					
Criteria	Rating				
Introduction	Introduction presents the overall topic and draws the audience into the presentation. It is clear, coherent and interesting.				4 3 <u>2</u> 1
Content (<i>most important</i>)	Information is accurate and serves the purpose of persuading the audience. The content is clear and concise with a logical progression of ideas and supporting information.				4 3 <u>2</u> 1
Text elements	Fonts, point size and colors make text easy to read. Text is easy to follow and understand.				4 <u>3</u> 2 1
Layout	The layout is visually pleasing and contributes to the overall message with appropriate use of headings, sub-headings and white-space. Not cluttered or confusing.				4 <u>3</u> 2 1
Graphics/visuals	Graphics/visuals assist in presenting an overall theme and enhance understanding of concepts, ideas, etc.				4 <u>3</u> 2 1
Writing Mechanics	The text is clearly written with no/few errors in grammar, vocabulary, spelling. Easy to read and understand.				4 3 <u>2</u> 1
Citations	Sources of information are properly cited so audience can determine credibility and authority of the information				4 3 2 <u>1</u>
	Slideshow: Overall evaluation (<i>first draft</i>)				4 3 <u>2</u> 1



Appendix B2 Formative Assessment 1-Student Self-Assessment (Part 2)

Part 2

Make some notes (bullet points are fine) in the following boxes about your slideshow (what looks good, weaknesses, things to work on).

Strong points:

- Font size
- Photos (visual aids). ← but I should include other photos.
-

Weak Points:

- Introduction is not enough to gain more attention and interests from audience.
- Contents are not so clear.
- Organization is not easy to follow and not easy to understand.
 ↳ not effective.
- No work consulted page and No in-text citation

Things to work on for the final version:

- Put more interesting introduction before "What is abortion?" page.
- I found other useful information and material, so I change all my contents page. and make it more effective by using examples.
- I'll finish to make work consulted page and citation.

Appendix B3 Formative Assessment 2- Instructor Feedback

Slideshow (First Draft)		Instructor Evaluation		
Student Name:	<u>YuKa</u>	Topic: <u>Abortion</u>		
[4 = very good 3 = good 2 = average 1 = weak]				
Criteria	Rating			
Introduction	Introduction presents the overall topic and draws the audience into the presentation. It is clear, coherent and interesting.			
Content (most important)	Information is accurate and serves the purpose of <u>persuading</u> the audience. The content is clear and concise with a logical progression of ideas and supporting information.			
Text elements	Fonts, point size and colors make text easy to read. Text is easy to follow and understand.			
Layout	The layout is visually pleasing and contributes to the overall message-with appropriate use of headings, sub-headings and white-space. Not cluttered or confusing.			
Graphics/visuals	Graphics/visuals assist in presenting an overall theme and enhance understanding of concepts, ideas, etc.			
Writing Mechanics	The text is clearly written with no/few errors in grammar, vocabulary, spelling. Easy to read and understand.			
Citations	Sources of information are properly cited so audience can determine credibility and authority of the information			
	Slideshow: Overall evaluation (<i>first draft</i>)			
	4	3	2	1
	4	3	2	1
	4	3	2	1
	4	3	2	1
	4	3	2	1
	4	3	2	1

Comments: A good first draft, YuKa. It looks like an interesting presentation and the visuals work well. Stronger, a bit deeper content in favor of a woman's right to choose would make it even better. Work on citations also.

Appendix B4 Summative Assessment- Slide show final draft

Persuasive speech: **Slideshow evaluation**

[4 = very good 3 = good 2 = average 1 = weak]

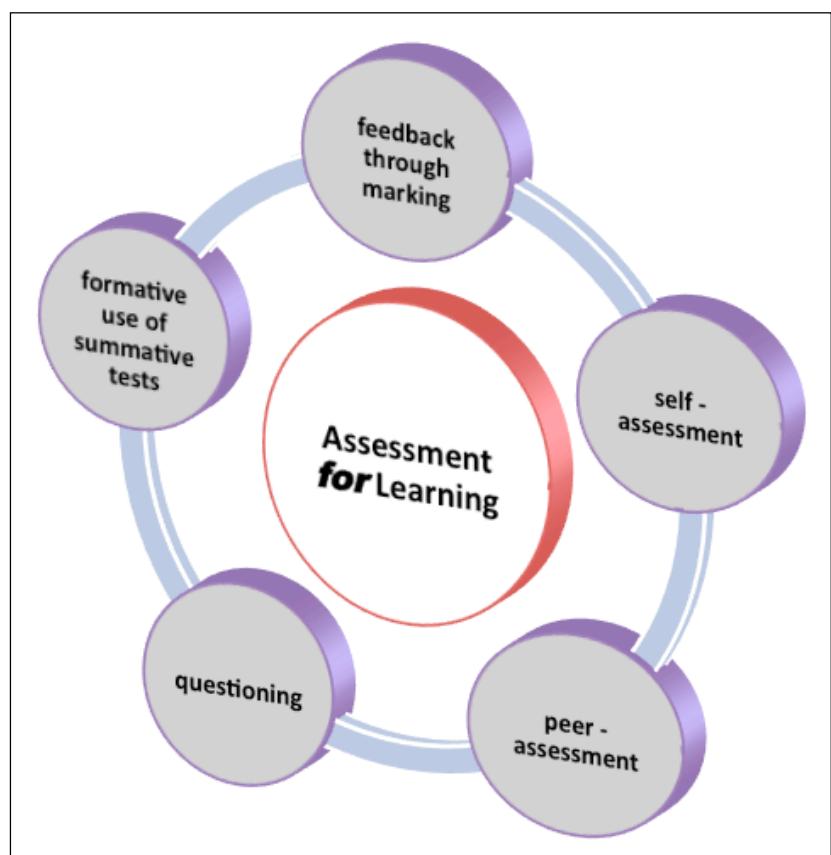
Criteria	Rating
1. Introduction	Introduction presents the overall topic and draws the audience into the presentation. It is clear, coherent and interesting. 4 3 2 1
2. Content (<i>most important element</i>)	Information is accurate and serves the purpose of <u>persuading</u> the audience. The content is clear and concise with a logical progression of ideas and supporting information. 4 3 2 1
3. Text elements	Fonts, point size and colors make text easy to read. <u>Text is</u> easy to follow and understand 4 <u>3.5</u> 2 1
4. Layout	The layout is visually pleasing and contributes to the overall message with appropriate use of headings, sub-headings and white-space. Not cluttered or confusing. 4 3 2 1
5. Graphics/visuals	Graphics/visuals assist in presenting an overall theme and enhance understanding of concepts, ideas, etc. 4 3 2 1
6. Writing Mechanics	The text is clearly written with no/few errors in grammar, vocabulary, spelling. Easy to read and understand. 4 <u>3</u> 2 1
7. Citations	Sources of information are properly cited so audience can determine credibility and authority of the information 4 3 2 1

Weak (3-4) Average/OK (5-6) Good (7-8) Very Good (9-10)

Overall Slideshow score: 9 / 10

Comments: A very good slideshow Yuka, with strong persuasive content and good choice of text elements and layout. The slideshow works well to effectively support the presentation. Some language use could have been smoother and clearer (e.g. title slide, "approval for").

Conclusion



“Formative assessment is not new. But what is important is . . . to take note of the research that has gone on before, and to do further research to show not just this is a good idea, but it actually does work in classrooms”. (Chris Harrison, Teachers’ TV, 2006)

This thesis presents a series of research reports in which the key AfL procedures were put into practice in a higher education setting in Japan with adult EFL learners. Following the KMOFAP research conducted by the Kings College London research team in England, and reported in Black et al. (2003), this research project has a similar overall aim: “to develop the implementation of formative assessment in the normal professional practice of teachers” (p. 17). While the majority of AfL research has occurred mostly with, and is geared towards, young learners in primary and secondary schooling in Anglophone cultures (Stobart, 2006), this thesis ‘puts classroom flesh on AfL bones’ in an EFL context with adult Japanese students through a series of five ‘living examples of implementation’. In brief, all five of the research studies concluded, based on the research data generated and analyzed, that the procedures implemented in each case could be positively evaluated as assessment *for* learning - they helped support and promote student learning. According to Gardner (2006):

The extent of existing knowledge and understanding of such a complex process and set of techniques [AfL] is still in its early stages. . . . What we can say categorically about assessment for learning, however, is that it is more often than not a fundamental element of any successful learning context. (p. 203)

My EFL research in Japan in a HE context supports this categorical assertion. An important element in the learning students experienced in the courses that provided the setting for the five research studies was influenced by, and enhanced by, the AfL practices implemented.

In bringing this thesis to a close, this concluding chapter is separated into two sections:

1. **Part A** provides an assessment of the key AfL procedures implemented to determine if they followed recommended practice, and responds to the broader research question posed in the introduction. (i.e., What does AfL practice look like in a higher education context with EFL students?)
2. **Part B** points out some differences or distinctions regarding putting AfL into practice in a HE context with adult students, particularly, but not solely, related to considerations of age, culture, and EFL. It ends with a commentary about research strengths, limitations and future possibilities.

It is perhaps useful at this point to also review the rationale for this research, presented in the introductory chapter. The research conducted and reported on in this thesis seeks to make an original contribution to the AfL/formative assessment literature by addressing the issues of: lack of understanding how formative assessment may be developed at a higher education level (Murphy, 2006); the need for more studies of AfL practice in varying contexts (Willis, 2006); providing more information about the assessment practices of tertiary English language teachers (Brindley, 2007); and, in particular, Black and Wiliam's (2006) call for extending AfL research into tertiary, non-statutory assessment contexts. The thesis provides one teachers' response to Yorke's (2003) challenge to make formative assessment part of good teaching practice in higher education.

Part A: Assessing the AfL research studies

An assessment undercurrent

In the thesis introduction, it was pointed out that the studies presented here feature a variety of research characteristics which may be found throughout the thesis: the fact that it could be categorized as 'qualitatively oriented inquiry' (Cumming, 2004); the extensive use of student perspectives, feelings and commentary about the AfL practices implemented; the presence of a self-assessment element in all five studies; and the extensive use of rubrics as both teaching and learning tools. One further distinguishing characteristic of this thesis, not previously mentioned but which may have become apparent to the reader, is the use of some type of 'external assessment criterion' for evaluating the AfL procedures focused on in each chapter. Four of the five reports included such an 'outside measuring stick', except for the peer-assessment report in Part B of Chapter 2. Such reference to external benchmarking criteria is important in terms of grounding my research's claims to validity. Each of the four reports also uses different sources from the formative assessment literature as lenses to scrutinize and analyze the assessment frameworks put into practice. The external assessment tools used in four of the reports are as follows:

1. In the teacher self-assessment of written feedback (Chapter 1, Part A), the 'Seven-Seven framework for feedback assessment' used criteria from two important articles (Gibbs & Simpson, 2004-2005; Nicol & McFarlane-Dick, 2006) to analyze and assess feedback practices used. In addition, Part B of the feedback chapter presents Lee's (2007) four essential conditions required in order to use feedback to promote

assessment for learning, and, in the conclusion, briefly assesses the written feedback provided against these conditions.

2. In the self-assessment study (Chapter 2, Part A), the assessment process was itself assessed according to the principles of practicality, reliability, validity, authenticity, and washback; the ‘five cardinal criteria’ from Brown (2004).
3. From Chapter 3, and the student-questioning report, three key ingredients for evaluating the efficacy of formative classroom assessment (from MacMillian 2007) were used as an assessment analysis tool.
4. Based on the findings of Black and Wiliam (1998), Clarke (2001) identified six key elements for effective formative assessment, and these were used to evaluate the assessment procedures of the final report - synergizing formative and summative assessment practices (Chapter 3, Part B).

Such external sources were used for the sake of validity. They provide a reference point (or points) in the related literature to assess the formative procedures implemented against, and also serve to anchor the related analysis and discussion in each report. This underlying assessment theme continues in this conclusion. An overall assessment of the processes and procedures implemented in these five reports is made against an important criterion of what AfL should look like in practice presented at the beginning of this thesis.

Applying the seven key AfL characteristics (ARG, 1999) to the five research studies

In the thesis introduction a point was made about the dangers of teachers potentially misunderstanding AfL, or misapplying it in actual classroom practice. As explained, in order to prevent such teacher missteps the ARG (1999) identified seven key characteristics of AfL practice so as to make them explicit for teachers, and show how this formative assessment framework can relate to effective teaching and learning. The seven key criteria included: embedding assessment into the teaching/learning process; sharing learning goals with students; helping them recognize standards they are aiming for; involving them in self-assessment; providing effective feedback; having the underpinning idea that all students can improve; and enabling both teachers and students to review and consider assessment data. To help conclude this research project of putting AfL into practice in a tertiary EFL setting we now return to these criteria and use them to assess the procedures implemented in each of the five reports.

Table 1 below presents a summary evaluation of all five research studies using the seven key characteristics of AfL practice, as determined by the ARG (1999). Each of the five studies is assessed using a very simple scheme asserting whether the characteristic in question is evident in the study (Y=yes), not evident (N=no) or partially evident (S=somewhat). While this exercise is one of self-assessment, as I use the key criteria to make judgments about how well each study matches recommended practice, it is based on the data produced, analyzed and reported for each study and therefore the seven characteristics should be evident (or not, as the case may be) to the objective reader also. In considering the information presented in Table 1, the reader is reminded of the influence of particular course contexts and purposes on the particular formative assessment procedures. This fact is noted by Black and Wiliam (2006) when writing that “the subject being taught at the time exerts a strong influence on the way that formative practices are implemented” (p.84). Black and Wiliam’s (1998b) assertion that individual teachers need to find their own ways of incorporating AfL ideas and practices “into his or her own patterns of classroom work” (p. 15) should also be kept in mind. In Table 1 the original ARG (1999) statements are reformulated as questions. As detailed explanations and analysis has already been provided in each of the five research reports, just a short commentary will follow the table presented here.

Table 1. Assessing the five research studies using seven key AfL characteristics

AfL characteristics (from ARG, 1999)	Y = yes, evident	S = somewhat evident	N = not evident		
Seven key questions	1. Feedback through marking study	2. Self-assessment study	3. Peer-assessment study	4. Questioning study	5. Formative use of summative tests study
1. Is the assessment embedded in, and part of classroom teaching and learning?	Y	Y	Y	Y	Y
2. Does it involve sharing learning goals with students?	Y	S	Y	S	Y
3. Does it help students to know and to recognize the standard they are aiming for?	Y	Y	Y	S	Y
4. Does it involve students in self-assessment?	Y	Y	Y	Y	Y
5. Does it provide feedback which leads to students recognizing their next steps and how to take them?	Y	N	S	N	Y
6. Is it underpinned by confidence that every student can improve?	Y	Y	Y	Y	Y
7. Does it involve both teacher and students reviewing and reflecting on assessment data?	Y	Y	Y	S	Y

It should be remembered that each individual report focused on one particular key component of the AfL framework rather than trying to ensure all elements were implemented. For example, the self-assessment study sought to highlight that particular aspect of AfL, and not the feedback component. However, the reader is again reminded of the point made by

Marshall (2007), included in the thesis introduction, that while each strategy can be considered discretely (as each of the five studies does), they also overlap (as evident throughout the thesis, and reflected in the table above).

The feedback report, (Chapter 1) and the assessment procedures used match up well with the seven key characteristics. All of them are evident in the assessment framework put in place for the Junior Composition writing course. As it particularly looked at the feedback provided to students, this key element of AfL was concluded in the report to indeed be ‘feedback for learning’.

In the self-assessment study (Chapter 2, Part A), the Communication Skills course students self-assessed their class participation three times throughout the semester. A ‘yes’ assessment is determined for most of the seven key ARG (1999) criteria, except two. In terms of the assessment ‘sharing learning goals with students’, this is less clear in an oral communication course where the goal is a broader ‘horizon’ of improving communication skills in English. Much less clear than the writing goals for the Junior Composition class, for example, the improvement of communicative competence is less well defined and much looser, as is the corresponding regulation of learning (Wiliam & Leahy, 2007). However, students were repeatedly made aware of the broader goal - the idea that more active participation in class would enhance their communication abilities. As such, a ‘somewhat evident’ response is perhaps most appropriate to this question of whether the self-assessment framework did share learning goals with students. In fact, students were asked to set their own goals in terms of active class participation. This SA framework also did not provide teacher feedback to individual students for reasons already noted (primarily due to the practicalities of giving feedback in the existing course circumstances) and as such received the ‘N’ rating for this criterion. However, in considering the context and purposes of the self-assessment procedure implemented in this course, the lack of teacher feedback to individual students is not viewed as problematic or a misapplication of AfL principles in practice.

The third research project evaluated in Table 1, peer-assessment of student presentations, was assessed as having a ‘yes’ response to all questions except one; also related to feedback. Like the SA study, no formative teacher feedback was provided to students about their two presentations for the *Essential Public Speaking* course. While teacher feedback was not a part of this course, at least not in a formal sense, teacher comments about topic choice and discussions with individual students during class time would be considered informal formative feedback (hence the “S” rating for this criterion). Students did also receive some

formative feedback from their peers after mini-presentations in their planning groups. Peers did make use of the key points that were included in the final summative assessment rubric to help provide formative feedback and suggestions for improvement. As such, there was some formative feedback provided to students as to the next steps in preparing for the final version of their presentations.

Of the five research reports, at first glance the questioning study seems to be the weakest in terms of having the range of seven key AfL characteristics as part of the assessment framework implemented. The reader is reminded that in a writing class of 22 students, a student questioning process was introduced, based on an Internet documentary film and tape script, for the purpose of determining whether student questioning and subsequent peer discussions promoted understanding of class material. The report concluded that the student-generated questioning process did have the desired result of deepening student understanding of the online documentary. But, as with the SA study and the purpose of developing students' communication skills, some of the seven key characteristics are not clearly evident in the questioning study. This study received an 'S' (somewhat evident) assessment in the following areas for the following reasons:

- It does not have clearly defined learning goals (item 2), but students were made aware of the overall purpose of student-generated questioning and peer discussion leading to better understanding of the documentary in question.
- The formative assessment procedure also did not clearly point out to students the standards they were aiming for, as these were not specified in this particular context. The distinction between skinny and fat questions did provide some guidance on question creation.
- There was some reviewing and reflection of the assessment data (item 7), in this case the particular questions generated by students but, again, this was not a strong component of the framework in place. Students did review their questions (and the peer discussions) as they completed the surveys used in this course.

In addition to these three characteristics receiving an 'S' evaluation, no feedback was provided to individual students as to recognizing next steps and how to take them (item 5). Considering the purpose of engineering a 'questions-driven classroom', and the research report conclusion that SGQ did help successfully support learning, these seemingly weak areas are not considered significant in terms of the course context and purposes of the

assessment.

Like the first feedback report, Table 1 shows that the final report (the formative use of summative assessment) also received all ‘Y’ evaluations for the seven characteristics. In both cases this is due to inclusion in the curriculum/assessment framework of such elements as clearly defined goals, and both self-assessment and individual teacher feedback (formative and summative) being connected to a criterion-referenced assessment rubric.

In glancing horizontally across Table 1 the reader will note that question five, related to the provision of feedback enabling students to recognize next steps in their learning, has a weaker rating than the other characteristics. These particular items receive a ‘somewhat evident’ or ‘not evident’ assessment for three of the five reports (self-assessment, peer-assessment and questioning). In the thesis introduction it was noted that much of the AfL literature and research comes from science and math subjects. It is useful to be reminded of Marshall’s (2007) assertion that in other subjects ‘paths of progression’ are less clear, including subjects like English, “where progression is a much messier business” (p. 136). Perhaps this is even more true in contexts where English is a foreign language, and in considering the complexity of teachers making judgments about the development of student understanding and proficiency in the target language.

Overall, Table 1 above reveals that, as a body of research, each of the assessment frameworks set up in these five courses at TWCU incorporate and include many, if not all, of the seven key characteristics of AfL in practice recommended by the ARG (1999).

The broad research question

The introduction to the thesis explains that while each of the five studies included in this thesis has a corresponding research focus related to it, the overall thrust of the thesis concerns this central research question:

◆ **What does AfL practice look like in a HE context with adult EFL students?**

A simple response is that it looks very much like AfL practiced in the context of a primary or secondary school setting with students and teachers sharing a first language. AfL is conducive to optimizing learning, and this fact remains true in a higher education EFL context with adult students. The research reports in this thesis help support Glassen’s (2009) far-reaching claim that AfL strategies can be used to improve learning and performance, “with all adults, no

matter what the context is” (p. 126).

This research also supports the Black and Jones (2006) assertion that AfL has *generic* features that are applicable across all stages and school subjects. They were referring, in particular, to mainstream education in the UK and its related subjects and stages. The research conducted here broadens this claim to say that these generic AfL features (i.e. the five key components) are applicable across broader educational contexts, such as in an Asian setting with older, post-secondary EFL students. In this thesis, pedagogical procedures that prove effective in one context of learning (AfL in mainstream primary and secondary education) are shown to be effective in other learning contexts also, just as Tomlinson (2005) asserted. This thesis also supports the corresponding Black and Jones (2006) assertion that some AfL features are *specific* to particular stages and school subjects. For example, the feedback chapter notes that Junior Composition writing students were permitted to give peer feedback in their first language (Japanese) in the interests of maximizing the amount and clarity of feedback provided. Again, as Tomlinson (2005) noted, some modifications from context to context may be required to use pedagogical procedures effectively.

It has been asserted that the educators’ primary purpose is: “to bring as many students as possible to their highest possible level of achievement” (Chappuis, Stiggins, Arter & Chappuis, 2004, p. 14). This is true for teachers at all educational levels, including tertiary education, and the fact that the students are adult EFL students does not change this primary purpose. AfL theory and practice is beneficial in helping students reach their maximum levels of achievement and learning, a fact that remains true in the context of a Japanese university with adult EFL students. It also develops the skills and attitudes to enable students to continue in their learning endeavors beyond the formal educational setting in their individual daily lives. In talking about formative assessment online on *Teachers TV*, Professor Paul Black observes:

“It opens up a large number of much deeper issues than one might imagine. And I think those issues are right at the heart of respecting pupils and helping them grow into adults who are confident about learning and about collaborating with one another - which is what we all need.” (Black 2006)

Clearly, the Japanese women in the research studies reported here should already be considered as adults. Yet respect for, and collaboration with others, as well as confidence in learning are attitudes and abilities that extend beyond an age range or educational setting; they

are the kind of ‘life skills’ which AfL promotes and helps develop.

Part B: Distinctions and considerations - applying AfL to a HE context

While this thesis supports the universal application of the generic features of AfL in educational settings, there are some differences and distinctions that became apparent when taking these formative assessment ideas for a walk and implementing them in a higher educational setting in Japan. These additional points of consideration are related in particular to the three areas of student age, culture, and EFL. Each point will be briefly addressed here in turn, along with further considerations. Perhaps it should be briefly noted that the issue of gender is not a consideration in this research. All participants were female and there were no inter-gender interactions, apart from with myself as teacher.

1. Age considerations

The fact that higher education students are older learners has a number of implications for putting AfL into practice. As Hilles and Sutton (2001) explain, whether they are native speakers or ESL students, adults are different from ‘typical’ or ‘traditional’ primary or secondary students in that they bring more life experience and cognitive maturity to the classroom. This experience and maturity, as with the 18-21 year old women in these research studies, may affect how teachers choose to implement or work with formative assessment practices, as well as how students respond to them. For example, I briefly experimented with the commonly-referred to AfL practice of ‘traffic lighting’ - having students self-assess their work by using green, amber or red colors to indicate their level of comprehension or ability. Using colored markers in a self-assessment activity, I experimented with this for three classes in my third-year Junior Composition course (20-21 year old students), but discontinued it after noticing that some students indicated that the use of colored markers in class to be somewhat juvenile or childish. Specific AfL practices that work well with younger learners may be less effective with adults.

On the other hand, adult learners may have an understanding of priorities and a maturity that many younger students do not possess, enabling them to direct their own learning agendas (Hilles & Sutton, 2001). This can be a particularly important advantage with putting AfL into practice with adult students in light of the fact that, “this ability to monitor one’s

own learning may be one of the most important benefits of formative assessment” (Black et al., 2003, p. 67).

Another important age-related point is connected to the psychology of adult students. Hilles and Sutton (2001, p. 387) explain that, “adult learners are also psychologically vulnerable, perhaps in a way that children are not, precisely because they are adults and have already formed a strong sense of who they are”. This can make implementing AfL more challenging, perhaps, especially in considering procedures like self-and peer-assessment, and student response to feedback. As we saw in the thesis introduction, one of the key factors in using assessment to improve learning is the recognition of the profound influence assessment can have on student self-esteem and motivation; both crucial influences on learning (ARG, 1999). One of the 10 principles of AFL (ARG, 2002) also notes that assessment needs to be sensitive and constructive because of the emotional impact it may have. To cite just one related example from this research, of 23 students in the JC writing class, 21% reported that they ‘sometimes’ or ‘usually’ felt a bit upset or hurt by any negative teacher comments about their writing. If adult learners are indeed more ‘psychologically vulnerable’, teachers need to keep this in mind when considering putting AfL procedures into practice, especially with adult EFL students who may have limited communicative ability in the target language, while being proficient and confident communicators in their first language.

2. Cultural considerations

This thesis involves ‘taking a set of AfL ideas around the world’ and implementing them in a far east Asian HE context. The issue of cultural differences and possible affects on pedagogy and classroom processes has been noted at times in this thesis. For example, in the feedback chapter the point was made that in some cultures students may expect teachers to notice and comment on their errors; and in the self-assessment report the observation was made that learner autonomy is not esteemed in some cultures. These cultural factors may have influenced student perspectives and views on the AfL procedures implemented, and expressed in the surveys conducted in the research, as well as affecting some of the procedures themselves (e.g. cultural norms regarding modesty and its impact on self-assessment). However, while acknowledging the possible impacts cultural factors and considerations can have, there is no evidence in these reports that they had an excessive or distorting influence on the implementation of AfL ideas put into practice in this Japanese context. In writing

about teaching EFL and matching procedures to learning contexts, Tomlinson (2005) asserts,

. . . as a teacher trainer in many cultures, I have found that students and teachers are willing to experiment with methodological change if the changes are justified to them and if the changes are seen as potentially valuable. There are also many examples in the literature of such willingness to change and of pedagogic procedures being valuable irrespective of the cultures they are used in. (p. 145)

The AfL practices implemented in these research studies were justified to students in each case, and many of the students' responses in the surveys show that they considered such practices as peer assessment and first draft teacher feedback (on both essays and slideshows) to be valuable in promoting learning. While paying attention to assessment purposes and being sensitive to local considerations, properly implemented and justified to students, the key AfL ideas and practices may also be considered as pedagogic procedures which are valuable irrespective of the cultures and classroom contexts they are used in. The effective provision of feedback to students, and their active involvement in their own learning are just two key AfL factors in improving learning which are universally applicable. In any case, social anthropologist Brian McVeigh's comment regarding culture and learning is quite appropriate here: "Teaching and learning should not be a "cultural issue" or one about one's national origins, but rather one of quality, standards and evidence that a student has actually learned something" (in Spencer, 2003, p. 3). By keeping the focus on student learning, 'the progressive development of understanding' (Harlen, 2007), and seeking to facilitate and maximize that learning, AfL practice can transcend potential cultural limitations.

3. EFL considerations

The fact that the approximately 190 students involved in these five research reports were EFL students had some impact on the implementation of these ideas. For example, a student's level of language proficiency impacts on how well they can formulate and respond to student-generated questions in the classroom. Also, teacher-student interaction, of primary importance in AfL, is obviously impacted if they do not share a common language and cannot interact communicatively together. Other types of classroom discourse will also be affected by a students' English proficiency, for example in peer-assessment. Gardner (2006) notes that PA can be very demanding on students' communication skills, "in particular, listening, turn-taking, clear and concise verbal and written expression" (p. 191). This will be even more

demanding of EFL students, especially those struggling to express and comprehend meaning in a foreign language. Yet, as with the issue of cultural considerations, in this thesis the EFL context itself of a woman's university in Tokyo was not seen to devalue or distort the implementation of AfL practice.

According to the ARG (2008) the ultimate goal of assessment in language teaching is improved student learning through the implementation of effective assessment practices. It seems a truism to say that languages are learned by using them for communicative purposes. Edwards (2004) poses and responds to the following language learning question thus:

What is the most effective, scientifically proven way to learn English and other foreign languages, to achieve true fluency? The pendulum of language learning has swung in recent years towards learner autonomy and student - centered teaching as the most effective means to address the language learning needs of the next generation, equipping them at the same time with the critical thinking skills necessary to meet the challenges of an increasingly complex world. (p. 19)

As we have seen, learner autonomy and student-centered teaching, as well as developing critical thinking skills (for example, in assessing their own work and that of their peers) are all important elements of AfL practice. In such an increasingly complex world, the claim has been made that "proficiency in the English language now relates directly to the life chances of students entering an intensely integrated international order" (Porcaro, 2004, p. 79). This research conducted at TWCU in Japan concludes that AfL in classroom practice can help contribute to such English proficiency development, and may also help develop the attitudes and skills to benefit students 'life chances' beyond the classroom.

4. Further considerations

Perhaps one of the key differences in the implementing of AfL in a HE context is the frequency of classes. University classes, as is the case at TWCU, are often weekly, or perhaps twice weekly, and not everyday like the primary and secondary classes from mainstream education and discussed in the AfL literature. It seems fair to say that AfL practices may be more easily and concretely implemented in classes where students meet four or five times a week, rather than once or twice weekly. This is one potential drawback of implementing AfL practice in HE contexts, perhaps requiring more effort by the teacher to ensure that the procedures are well understood and practiced by students.

On the other hand, it could be argued that AfL could be more easily implemented in HE contexts for two reasons:

1. Instructors typically do not have to deal with the externally administered summative assessments found in secondary education, those which can distort AfL practices or make them more difficult to implement.
2. In tertiary settings instructors perhaps have more control over the courses, curricula and assessment frameworks used. As in this thesis, they can often be the sole decision-makers as to the type of AfL procedure to implement, and the focal class to do so with.

These two points were also noted by Taras (2008), and previously mentioned in the final research report in Chapter 3 above.

Thus, the practicalities and timing of higher educational contexts will have some impact, both positive and negative, when putting AfL into classroom practice.

AfL practices implemented by instructors can help a higher education institute achieve one of its primary purposes. As noted in the thesis introduction, a key purpose of higher education should be to provide a foundation for a lifetime of learning in work and other social situations that students find themselves in after graduation (Boud & Falchikov, 2006). AfL encourages learners to be “reflective, strategic, intentional and collaborative” (James et al., 2007, p. 28); all are important attributes in constructing a foundation for lifelong learning.

Looking backwards and forwards: strengths, limitations, and future possibilities

The introduction to this thesis described it as ‘a report of a university instructor finding his own way of embedding these AfL ideas into ‘patterns of classroom work’ with the guiding aim of promoting student learning’. It has been, aside from brief discussions with colleagues, very much a solo venture of understanding, implementing, analyzing, reflecting and writing. In a very tangible way it has been a personal experience of learning through assessment for this teacher/researcher. Through the process of conducting doctoral research, and managing the interaction between AfL, curriculum and pedagogy in the five studies, I was engaged in assessment for learning with regard to my own professional development as an educator. As teacher and researcher in these classes at TWCU over a three-year period, I was also engaged

in a construction of knowledge process as I was actively engaged in ‘a progressive development of understanding’ (Harlen, 2007b) about aligning pedagogy and AfL in order to promote and maximize student learning.

The qualitative orientation of these research reports, seeking to illuminate local perspectives in rich detail (Creswell, 1998), also promotes the development of *reflexivity* on the part of the researcher. According to Schwandt (1997) reflexivity involves, “an acknowledgement of the inquirer’s place in the setting, context, and social phenomenon he or she seeks to understand and a means for a critical examination of the entire research process” (p. 164). Now, reaching the end of this research road, I am developing such a critical examination of the whole research process. I suspect this understanding will sharpen and grow with the perspective and understanding that the passing of time often gives. However, from this time and place (approximately one year after completing the data collection process, and after moving to Canada), what strengths, limitations and future possibilities do I, as a qualitative researcher engaging in reflexivity, perceive in this body of research and the process of producing it?

Summary of strengths of this research:

- with publication of two of the reports included in this thesis - peer-assessment, (White, 2009a), and self-assessment (White, 2009b) - this research has already made some contributions to the body of applied linguistics knowledge related to classroom-based assessment. If the researcher’s primary goal is to add knowledge (Burns, 2000), then this research has achieved some degree of success;
- this thesis is the first of its kind, and original in researching the full range of AfL procedures in a tertiary classroom setting;
- it provides support, in new contexts and with adult learners, to the applicability of AfL procedures across a wider range of ages and subject matters (i.e. EFL courses). This may be positively viewed as broadening the conceptualization of AfL beyond just dealing with students as ‘children’ or young people, but to include older, more mature learners;
- the research provides a view of AfL principles and practices being implemented in a new cultural context (Japan), one that is not usually associated with formative assessment practices in education;
- the voices and views of students are widely incorporated into this student-and-

- learning-focused series of reports;
- this thesis shows that individual teachers can incorporate meaningful and practical AfL procedures in their classes without the need for support networks, or dependence on colleagues or other external groups.

Like any ‘systematic and principled inquiry’ that comprises research, limitations are also a part of the process.

Summary of limitations:

- while having its advantages, having a sole researcher as part of a research program also has drawbacks such as the lack of investigator triangulation previously mentioned;
- the research purposely adopts a ‘primarily pragmatic approach’ (Black, 2007) into investigating AfL in the classroom. However, the lack of exploration and analysis of the more theoretical aspects of formative assessment, or the related debates that exist, may be seen as a limitation;
- while the research has some breadth in that it covers five different courses and student groupings, the total sample of students involved (approximately 190) is rather small, and all research is conducted at one institution;
- use of a wider range of research methods (e.g. interviews, control groups, additional teachers/classes; statistical analysis) may have led to a more robust research regimen and provided a richer source of data for analysis.

Future possibilities

In terms of AfL principles and practices, the research described here could be developed into a larger research project drawing together a number of different teachers and institutions, following the KMOFAP model implemented in England. Of the five reports in this thesis, the one focused on student generated questions is perhaps the least researched and experimented with. There are a number of possible avenues for research in this area in particular (e.g. affects of language proficiency on question-generation, investigating student question generation and impact on essay writing). With regard to the topic of feedback through marking, the use of AfL procedures and computer-mediated technology to improve student writing is a possible research area. Whether feedback for learning can be made stronger in EFL/ESL writing classes and even more effective with more than one round of essay drafting

is also a possible avenue to explore. Investigating the cultural considerations that can impact, or possibly distort, the effects of AfL in practice seems another possibility. A variety of other research possibilities exist in this relatively new formative assessment/assessment for learning corner of applied linguistics. As noted by Black (2005) in the thesis introduction, this is a new field which will continue to be offer rich possibilities for research. I think this is particularly true in second or foreign language learning contexts, where AfL has been much less implemented and researched.

Personally speaking, my AfL journey continues. Assessment for learning has had a deep and positive impact on my work as a language educator and classroom instructor. Aside from working to complete an advanced academic degree, reading, researching and writing about this topic has made me more assessment literate, and without a doubt a better language educator and facilitator of student learning.

Gardner (2006) records the following comment from a secondary teacher in the UK who was involved in implementing AfL in her classes:

“Assessment for learning has been a joy. It is intellectually profound, yet eminently practical and accessible. [It] has enhanced the learning of all of us. . . . It has been the best educational development of my career.” (p. 42)

This is one of the most memorable and fitting comments about AfL I have read in the hundreds of sources examined for this research. It is indeed practical, accessible and profound. The power of AfL theory and practice to enhance the learning of all those involved in these HE classes and courses has become abundantly obvious to me over the past three years. In addition to supporting and promoting student learning, this formative assessment framework is sure to impact my continued educational development as both teacher and learner in the various classroom communities I find myself a part of. According to Black and Jones (2006), the implementation of formative assessment methods “is best seen as a voyage of discovery, a journey into new territories of teaching and learning” (p. 9). This thesis describes just one, three-year episode of that AfL voyage; a journey that will continue in the coming years as new teaching and learning territories are navigated and explored.

.

References

- ACTFL proficiency guidelines* (1999). Retrieved March 21, 2008, from
<http://www.actfl.org/i4a/pages/index.cfm?pageid=3325>
- Airasan, P. & Gullickson, A. (1997). *Teacher Self-Evaluation Tool Kit*. London: Corwin Press, Inc.
- Akbari, R. (2007). A critical appraisal of reflective practice in L2 teacher education. *System*, 35, 192-207.
- Alderson, J. & Banerjee, J. (2001). Language testing and assessment (part 1). *Language Testing*, 18(4), 213-236.
- Allwright, D. (2000). Interaction and negotiation in the language classroom: Their role in learner development. *CRILE Working Papers*, Lancaster University. Retrieved December 4, 2006, from <http://www.ling.lancs.ac.uk/groups/crile/docs/crile50allrigh.pdf>
- Allright, D. (1999). Research methodology for language learning. In K. Johnson and H. Johnson (eds.) *Encyclopedic Dictionary of Applied Linguistics*. Oxford: Blackwell. 274-280.
- Amidon, E. & Flanders, N. (1967). *The Role of the Teacher in the Classroom: A Manual for Understanding and Improving Teacher Classroom Behavior*. Minneapolis: Association for Productive Teaching.
- Anderson, T.H., & Armbruster, B.B. (1984). Studying. In P.D. Pearson (Ed.), *Handbook of Reading Research* (657- 679). New York: Longman.
- Andrade, H. & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment and Evaluation in Higher Education*, 32(2), 159-181.
- ARG. (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: Cambridge University Press.
- ARG. (2002). *Testing, motivation and learning*. University of Cambridge: Assessment Reform Group. Retrieved April 2, 2008 from
<http://gtcni.openrepository.com/gtcni/bitstream/2428/4624/1/Testing,%20Motivation%20and%20Learning.pdf>
- ARG. (2008). *Changing assessment practice: Process, principles and standards*. Assessment Reform Group. Retrieved February 21, 2009 from www.aria.qub.ac.uk
- Arter, J. (2004). Why use rubrics anyway? Assessment Training Institute. Retrieved July 2, 2008 from www.assessmentinst.com/forms/Article-Dec%202004.pdf
- AskOxford.com. (2009). Synergy. Retrieved January 22, 2009 from
http://www.askoxford.com/concise_oed/synergy?view=uk
- ASF (2005). Assessment systems for the future project: Working paper 3. Retrieved July 14, 2008 from www.assessment-reform-group.org

- Ashby, E. (1984). Forward, in I.M. Brewer, *Learning More and Teaching Less*. Guildford: Society for Research into Higher Education.
- Atkins, M. (1995). What should we be assessing? In P. Knight (Ed.) *Assessment for Learning in Higher Education* (pp. 25-34). London: Kogan Page Limited.
- Bachman, L. (2000). Forward. In G. Ekbatani & H. Pierson (Eds.). *Learner Directed Assessment in ESL* (pp. ix-xiii). New Jersey: Lawerance Erlbaum Associates.
- Bailey, K. (1998). *Learning About Language Assessment*. New York: Heinle and Heinle.
- Baker, T. L. (1994). *Doing Social Research*. New York: McGraw-Hill.
- Balanced Assessment: The Key to Accountability and Improved Student Learning. (2003). National Educational Association. Retrieved June 10, 2007 from www.assessmentinst.com/forms/nea-balancedassess.pdf
- Ballantyne, K., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education*, 27(5), 427-441.
- Barudi, Z. (2007). Formative assessment: definition, elements and role in instructional practice. *Post-Script: Postgraduate Journal of Education Research*, Vol. 8(1), 37-48. Retrieved July 24, 2008 from www.edfac.melbourne.edu/research/resources/student_res/Ziad_Baroudi.pdf
- Basturk, R. (2008). Applying the many-facet Rasch model to evaluate PowerPoint presentation performance in higher education. *Assessment & Evaluation in Higher Education*, 2008, First Article, 1-14.
- Benson, P. (2001). *Teaching and Researching Autonomy in Language Learning*. Essex, England: Pearson Education.
- Biggs J. (1998). Assessment and classroom learning: a role for summative assessment? *Assessment in Education: Principles, Policy and Practice*, 5(1), 103–10.
- Biggs, J. (2003). *Teaching for Quality Learning at University*. New York: McGraw-Hill.
- Black, P. (2001). Formative assessment and curriculum consequences, In D. Scott (Ed.) *Curriculum and Assessment: Institutional Perspectives on Curriculum Studies*, 1, (7-23). Westport CT: Ablex Publishing.
- Black, P. (2003). *Formative and Summative assessment: Can they serve learning together?* Paper presented at the annual convention of the American Educational Research Association, Chicago.
- Black, P. (2004). *The Nature and Value of Formative Assessment for Learning*. Kings College London Assessment Reform Group. Retrieved on August 5, 2008 from www.kcl.ac.uk/content/1/c4/73/57/formative.pdf
- Black, P. (2005). Formative assessment: views through different lenses. *The Curriculum Journal*, 16(2), 133-135.

- Black, P. (2006). Secondary assessment-Formative assessment. Teachers TV video. Accessed April 12, 2009 from <http://www.teachers.tv/video/565>
- Black, P. (2009). Formative assessment issues across the curriculum: the theory and the practice. *TESOL Quarterly*, 43(3), 519-525.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working Inside the Black Box*. London: nferNelson Publishing Company.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning: Putting it Into Practice*. Berkshire, England: Open University Press.
- Black, P., & Jones, J. (2006). Formative assessment and the learning and teaching of MFL: sharing the language learning road map with the learners. *Language Learning Journal*, 34, 4-9.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (1998b). *Inside the Black Box: Raising Standards through Classroom Assessment*. London: King's College.
- Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.) *Assessment and Learning* (pp. 9-26). London: Sage Publications.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31.
- Bloxham, S. & West, A. (2004). Understanding the rules of the game: marking peer assessment as a medium for developing students' conceptions of assessment. *Assessment & Evaluation in Higher Education*, 29(60), 721-723.
- Boud, D. (1995). Assessment and learning: contradictory or complimentary? In P. Knight (Ed.), *Assessment for Learning in Higher Education* (pp. 35-48). London: Kogan Page Limited.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22, 2: 151-167.
- Boud, D., & Falchikov, N. (Eds.) (2007). *Rethinking Assessment in Higher Education*. New York: Routledge.
- Brindley, G. (1989). *Assessing Achievement in the Learner-centered Curriculum*. Sydney: National Centre for English Language Teaching and Research, Macquarie University.
- Brindley, G. (2007). Editorial. *Language Assessment Quarterly*, 4(1), 1-5.
- Broadfoot, P., & Black, P. (2004) Redefining assessment? The first ten years of Assessment in Education. *Assessment in Education*, 11(1), 7-27.
- Brookhart, S. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practices*, 22, 5-12.

Brookhart, S. (2005). *Research on formative classroom assessment*. Paper Presented at the Annual Meeting of the American Educational Research Association: Montreal, Canada.

Brookhart, S. (2007). Expanding views about formative classroom assessment: a review of the literature. In James McMillan (Ed). *Formative Classroom Assessment: Theory into Practice*. New York: Teachers College Press, Colombia University. pp. 43-62.

Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. New York: Pearson Education.

Brown, J. D. (1998). (Ed.) *New Ways of Classroom Assessment*. Alexandria, VA: TESOL Incorporated.

Brown, E. & Glover, C. (2006). Evaluating written feedback. In C. Bryan & K. Clegg (Eds.), *Innovative Assessment in Higher Education* (pp. 81-91). New York: Routledge.

Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32, 653-675.

Brown, J. D. & Rodgers, T. S. (2002). *Doing Second Language Research*. Oxford: Oxford University Press.

Brualdi, A. (1998). Implementing performance assessment in the classroom. *Practical Assessment, Research & Evaluation*, 6(2). Retrieved June 28, 2008 from <http://PAREonline.net/getvn.asp?v=6&n=2>

Bryan, C. & Clegg, K. (Eds.) (2006). *Innovative Assessment in Higher Education*. New York: Routledge.

Burns, R. (2000). *Introducing Research, 4th edition*. Frenchs Forest, NSW: Pearson.

Butler, Y. (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.

Campbell, A., MacNamara, O., & Gilroy, P. (2004). *Practitioner Research and Professional Development in Education*. London: Paul Chapman Publishing.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219-223.

Carless, D. (2007). Conceptualizing pre-emptive formative assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 171-184.

Carless, D., Joughin, G., Ngar-Fun, L., and Associates. (2006). *How Assessment Supports Learning*. Hong Kong: Hong Kong University Press.

Cassidy, S. (2007). Assessing 'inexperienced' students' ability to self-assess: Exploring links with learning style and academic personal control. *Assessment & Evaluation in Higher Education*, 32(3), 313-330.

Chappelle, C. & Brindley, G. (2002). Assessment. In N. Schmitt (Ed). *Introduction to Applied Linguistics* (pp. 267-288). New York: Hodder Arnold.

- Chappuis, S., & Chappuis, J. (2008). The best value in formative assessment. *Educational Leadership*, 65(4), 14-19.
- Chappuis, S., Stiggins, R.J., Arter, J., & Chappuis, J. (2004). *Assessment for Learning: An Action Guide for School Leaders*. Portland, OR: Assessment Training Institute.
- Cheng, W., & Warren, M. (1997) Having second thoughts: student perceptions before and after a peer assessment exercise. *Studies in Higher Education* 22, 233-239.
- Cheng, L., Rodgers, W., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, 33(1), 9-32.
- Chin, C., Brown, D. (2002). Student-generated questions: a meaningful aspect of learning in science. *International Journal of Science*, 24(5), 521-549.
- Ciardello, A. (1998). Did you ask a good question today? Alternative cognitive and metacognitive strategies. *Journal of Adolescent & Adult Literacy*, 42(3). 210-19.
- Clarke, S. (2001). *Unlocking Formative Assessment*. London: Hoddon and Stoughton.
- Cohen, R., Boud, D., and Sampson, J. (2001). Dealing with problems encountered in assessment of peer learning, in Falchikov, *Learning Together: Peer Tutoring in Higher Education*, London: Routledge Falmer.
- Cook, V. (1996). *Second Language Learning and Teaching*. (2nd ed.). London: St. Martins Press.
- Cook, A. (2001) Assessing the use of flexible assessment, *Assessment and Evaluation in Higher Education*, 26(6), pp. 539-549.
- Cooper, N. (2000). Facilitating learning from formative feedback in level 3 assessment. *Assessment and Evaluation in Higher Education*, 25(3), pp. 279-291.
- Cotton, K. (2001). Classroom questioning. North West Regional Educational Laboratory. Retrieved Nov.15, 2008 from www.nwrel.org/scpd/sirs/3/cu5.html
- Creswell, J. W. (1998). *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Thousand Oaks, Calif.: Sage.
- Croker, R. (2009). An introduction to qualitative research. In J. Heigham and R. Croker (eds.) *Qualitative Research in Applied Linguistics: A Practical Introduction*. Hampshire, UK: Macmillan Publishers Limited. 3-24.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.
- Cumming, A. (2004). Broadening, deepening, and consolidating. *Language Assessment Quarterly*, 1(1), 5-18.
- Cumming, A. (2009). What needs to be developed to facilitate classroom-based assessment? *TESOL Quarterly*, 43(3), 515-519.

Curriculum Corporation (2004). Assessment for Learning: Strategic Questioning. Facilitators Notes: Strategic Questioning Module. Retrieved December 14, 2008 from <http://cms.curriculum.edu.au/assessment/pd/stratques.asp>

Daugherty, R., & Ecclestone, K. (2006). Constructing assessment for learning in the UK policy environment. In J. Gardner (Ed.) *Assessment and Learning* (pp. 149-168). London: Sage Publications.

Davison, C., & Leung, C. (2009). Content issues in English language teacher based assessment. *TESOL Quarterly*, 43(3), 393-415.

Deakin-Crick, R., Sebba, J., Harlen, W., Guozing, Y., & Lawson, H. (2005). Systematic review of research evidence on the impact on students of self-and peer-assessment. Protocol. In *Research Evidence in Education Library*. London: EPPI-Center, Social Science Research Unit, Institute of Education, University of London.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.

DeVito, J. (2006). *The Essential Elements of Public Speaking, 2nd Edition*. Boston: Pearson Education Inc.

DfES, (2004). Department of Education and Skills, UK, *Pedagogy and Practice: Teaching and Learning in Secondary Schools, Unit 7: Questioning*. Retrieved Oct 11. 2008 from http://www.standards.dfes.gov.uk/secondary/keystage3/all/respub/sec_pptl0

Dobson, S. (2006). The assessment of student PowerPoint presentations-attempting the impossible? *Assessment & Evaluation in Higher Education*, 31(1), 109-119.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer, and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350.

Duff, Patricia. (2001). Research approaches in applied linguistics. In R. Kaplan (ed.) *The Oxford Handbook of Applied Linguistics*. London: Oxford University Press. 14-23.

Dufficy, P. (2005). 'Becoming' in classroom talk. *Prospect*, 20(1), 59-81.

Dunleavy, P. (2003). *Authoring a PhD*. New York: Palgrave MacMillan.

Dunn, L., Morgan, C., O'Rielly, M., & Parry, S. (2004). *The Student Assessment Handbook*. London: RoutledgeFalmer.

Earl, L. (2003). *Assessment as Learning*. Thousand Oaks, California: Corwin Press.

Edwards, N. (2004). Rediscovering the creative heart of Japanese education: Fostering intrinsic motivation through a love of language. *The Language Teacher*, 28(1), 19-23. Retrieved February 1, 2009 from <http://www.jalt-publications.org/tlt/articles/2004/01/edwards>

- Eisner, E.M. (1985). *The Art of Educational Evaluation*. Philadelphia: Falmer Press.
- Ellery, K. (2008). Assessment for learning: a case study using feedback effectively in an essay-style test. *Assessment & Evaluation in Higher Education*, 33(4), 421-429.
- Falchikov, N. (2003). Involving students in assessment. *Psychology Learning and Teaching*, 3(2), 102-108.
- Falchikov,N., & Goldfinch, J. (2000). Student peer assessment in higher education: a meta-analysis. *Review of Educational Research*, 70, 287-322.
- Farkas, D. (2005). Understanding and using Powerpoint. Retrieved June 10, 2007 from faculty.washington.edu/farkas/TC510/Farkas-STC-05-UnderstandingPowerPoint.pdf
- Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, 31(3), 15-39.
- Ferris, D. (2002). *Treatment of Error in Second Language Student Writing*. Ann Arbor, MI: The University of Michigan Press.
- Ferris, D. (2006). Does error feedback help student writers? New evidence on the short-and long-term effects of written error correction. In K. Hyland & F. Hyland, (Eds.) *Feedback in Second Language Writing: Contexts and Issues* (pp. 81-104). Cambridge: Cambridge University Press.
- Frankland, S. (2007). (Ed.) Enhancing Teaching and Learning through Assessment. Dordrecht, Netherlands: Springer.
- Frary R. B. (2003). *A Brief Guide to Questionnaire Development*. Virginia Polytechnic Institute & State University. Retrieved December 20, 2009 from <http://www.ericae.net/ft/tamu/vpiques3.htm>
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27-32.
- Frontline (2008). *Growing Up Online*. Public Broadcasting Station (PBS). Available online at <http://www.pbs.org/wgbh/pages/frontline/kidsonline/>
- Fry, S.A. (1990). Implementation and evaluation of peer marking in higher education. *Assessment & Evaluation in Higher Education*, 15, 177-189.
- Gardner, J. (Ed.) (2006). *Assessment and Learning*. London: Sage Publications.
- Gay, L. R., & Airasian, P. (2003). *Educational Research: Competencies for Analysis and Application (7th Edition)*. New Jersey: Prentice Hall.
- Gibbs, G. (1999). Using assessment strategically to change the way students learn. In S.Brown & A.Glasner (Eds.) *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*. Buckingham & Philadelphia: SRHE & Open University Press, pp.41-53.
- Gibbs, G. (2006). How assessment frames learning. In C. Bryan & Clegg, K (Eds.) *Innovative*

- Assessment in Higher Education* (pp. 23-36). London: Routledge.
- Gibbs, G. & Simpson, C. (2004-05). Conditions under which assessment supports learning. *Learning and Teaching in Higher Education, Issue 1*, 3-29.
- Glasson, T. (2009). *Improving Student Achievement: A Practical Guide to Assessment for Learning*. Curriculum Corporation: Victoria, Australia.
- Goldstein, L. (2006). Feedback and revision in second language writing: contextual, teacher, and student variables. In K. Hyland & F. Hyland, (eds.) *Feedback in Second Language Writing: Contexts and Issues* (pp . 185-205). Cambridge: Cambridge University Press.
- Gottlieb, M. (2006). *Assessing English Language Learners*. Thousand Oaks, CA:Corwin Press.
- Gronlund, N. (1998). *Assessment of Student Achievement*. Boston: Allyn and Bacon.
- Grotjahn, R. (1987). On the methodological basis of introspective methods. In C. Faerch and G. Kapser (eds.) *Introspection in Second Language Research*. Clevedon: Multilingual Matters. 230-250.
- Grundy, S. (1989). Beyond Professionalism. In Carr,W. (Ed.), *Quality in Teaching: Arguments for a Reflective Profession* (79-100). London: Falmer Press.
- Guskey, T. (2003). How classroom assessments improve learning. *Educational Leadership*, 60(5), 6-11.
- Haines, C. (2004). *Assessing Students Written Work: Marking Essays and Reports*. London: RoutledgeFalmer.
- Hancock, C. (1994). Alternative assessment and second language study: What and why? ERIC Digest, ED376695. Retrieved March 21, 2008, from <http://www.ericdigests.org/1995-2/language.htm>
- Hanrahan, S. & Issacs, G. (2001). Assessing self-and peer-assessment: the students' views. *Higher Education Research and Development*, 20(1), pp.53-70.
- Harford, J. & MacRuairc, G. (2008). Engaging student teachers in meaningful reflective practice. *Teaching and Teacher Education*, 24, 1884-1892.
- 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 formative and summative purposes. In J. Gardner (Ed.) *Assessment and Learning* (pp. 103-119). London: Sage Publications.
- Harlen, W. (2007a). *Assessment of Learning*. London: Sage Publications.
- Harlen, W. (2007b). Formative classroom assessment: The key to improving student achievement. In J. MacMillian (Ed.). *Formative Classroom Assessment: Theory into*

- Practice.*(pp.116-136) New York: Teachers College Press.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education*, 4, 365-380.
- Harlen, W., & Winter, J. (2004). The development of assessment for learning: learning from the case of science and mathematics. *Language Testing*, 21(3), 390-408.
- Harrington, K., Elander, J., Lusher, J., Aigegbayo, O., Pitt, E., Norton, L., Robinson, H., Reddy, P. (2006). Using core assessment criteria to improve essay writing. In Bryan, C. & Clegg, K. (Eds.). *Innovative Assessment in Higher Education* (pp. 110-120). New York: Routledge.
- Harris, M. (1997). Self-assessment of language learning in formal settings. *ELT Journal*, 51(1), 12-20.
- Harrison, C. (2006). Secondary assessment-Formative assessment. *Teachers TV*. Accessed April 12, 2009 from <http://www.teachers.tv/video/565>
- Harrison, M., Head, E., Haugh, D. & Sanderson, R. (2005). Self-evaluation as motivation for active participation. *Learner Development: Context, Curricula, Content*. Japan Association of Language Teaching College and University Educators, 37-59.
- Hattie, J.A. (1987) Identifying the salient facets of a model of student learning: a synthesis of meta-analyses, *International Journal of Educational Research*, 11, 187-212.
- Hattie, J., & Timperly, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hedge, T. (2000). *Teaching and Learning in the Language Classroom*. Oxford: Oxford University Press.
- Hilles, S. & Sutton, A. (2001). Teaching adults. In M. Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language* (pp. 385-401). Boston: Heinle & Heinle.
- Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and mathematics: a comparison. *The Curriculum Journal*, 16(2), 153-176.
- Hounsell, D. (1995). Marking and commenting on essays. In F. Foster, D. Hounsell & S. Thomson (Eds.) *Tutoring and Demonstrating: A Handbook* (pp. 51-64). Edinburgh: The Center for Teaching, Learning and Assessment, University of Edinburgh.
- Hounsell, D. (2005). Understanding teaching and teaching for understanding. In: Marton, F., Hounsell, D. and Entwistle, N., (eds.) *The Experience of Learning: Implications for Teaching and Studying in Higher Education. 3rd (Internet) Edition* (pp. 238-257). Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment.
- Hyland, K. (2002). *Teaching and Researching Writing*. London: Pearson Education Limited.
- Hyland, K. (2003). *Second Language Writing*. Cambridge: Cambridge University Press.
- Hyland, K. (2006). Feedback on second language students' writing. *Language Teaching*, 39, 83-101.

Hyland, K., Hyland, F. (Eds.) (2006). *Feedback in Second Language Writing: Contexts and Issues*. Cambridge: Cambridge University Press.

information. (n.d.). The American Heritage® Dictionary of the English Language, Fourth Edition. Retrieved September 10, 2008, from Dictionary.com website:
<http://dictionary.reference.com/browse/information>

Integrative Assessment: Balancing Assessment of and Assessment for Learning. (2007) Guide No. 2. The Quality Assurance Agency for Higher Education. Retrieved July 4, 2008 from www.enhancementthemes.ac.uk/documents/IntegrativeAssessment/IA%20Balancing%20assessment.pdf

James, M. (2006). Assessment for learning: what is it and what does research say about it? In M. James and 14 Others. *Learning how to learn: Tools for Schools* (pp. 7-14). London: Routledge.

James, M., & Pedder, D. (2006a). Beyond method: assessment and learning practices and values. *The Curriculum Journal*, 17(2), 109-138.

James, M. & Pedder, D. (2006b). Professional learning as a condition for assessment for learning. In J. Gardner (Ed.) *Assessment and Learning* (pp. 27-44). London: Sage Publications.

James, M., Black, P., McCormick, R., & Pedder, D. (2007). Promoting learning how to learn through assessment for learning. In M. James and 12 Others. *Improving Learning How to Learn: Classrooms, Schools and Networks* (pp. 3-30). London: Routledge.

James, R., McInnis, C., & Devlin, M. (2002). *Assessing learning at Australian universities*. Center for the Study of Higher Education, The University of Melbourne, Australia. Retrieved May 17, 2008 from <http://www.cshe.unimelb.edu.au/assessinglearning/>

James, M. and 14 Others. (2006). *Learning how to learn: Tools for schools*. London: Routledge.

James, M. and 12 Others. (2007). *Improving Learning How to Learn: Classrooms, Schools and Networks*. London: Routledge.

Karabenick, S. (1994). Relation of perceived teacher support of student questioning to students' beliefs about teacher attributions for questioning and perceived classroom learning environment. *Learning and Individual Differences*, 6(2), 187-204.

Kennedy, K., Kin Sang, J., Wai-ming, F., & Kwan Fok, P. (2006). Assessment for productive learning: forms of assessment and their potential for enhancing learning. Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore, 2006. Retrieved August 10, 2007 from www.iaeap2006.seab.gov.sg/conference/download/abstracts/Assessment for productive learning - Forms of assessment and their potential for enhancing learning.pdf

Kiddey, P., & Waring, F. (2001). *Success for All: Selecting Appropriate Learning Strategies*. Education Department of Western Australia: Curriculum Corporation.

- King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist*, 27(1), 111-126.
- King, A. (1994a). Autonomy and question asking: The role of personal control in guided student-generated questioning. *Learning and Individual Differences*, 6(2), 162-185.
- King, A. (1994b). Guiding knowledge construction in the classroom: effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31(2), 338-368.
- King, A. (1999). Discourse patterns for mediating peer learning. In *Cognitive Perspectives on Peer Learning* (pp. 87-115). Edited by Angela O'Donnell, Alison King. New York: Lawerance Erlbaum
- King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice*. Accessed Nov. 10, 2008 from http://findarticles.com/p/articles/mi_m0NQM/is_1_41/ai_90190487
- Kirschner, L. & Mandell, S. (2006). *The Pocket Wadsworth Handbook*, 3rd Edition. Boston: Thomson Wadsworth.
- Klenowski, V. (2009). Assessment for learning revisited: an Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice* 16(3). 263–268.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- Knight, P. (Ed.) (1995). *Assessment for learning in higher education*. London: Kogan Page Limited.
- Knoblauch, C.H. and Brannon, L. (1984). *Rhetorical Traditions and the Teaching of Writing*. New Jersey: Boynton Cook Publishers.
- Kroll, B. (2001). Considerations for teaching and ESL/EFL writing course. In M. Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language* (pp. 219-223). Boston: Heinle & Heinle.
- Krueger, R. A., & Casey, M. A. (2000). *Focus groups: A Practical Guide for Applied Research* (3rd edition.) Thousand Oaks, Calif.: Sage.
- Lambert, N. & Coombs, B. (1998). How students learn: *Reforming Schools through Learner-centered Education*. Washington: American Psychological Association.
- Langan, M. and 10 Associates (2005). Peer assessment of oral presentations: effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment & Evaluation in Higher Education*, 30(1), 21-34.
- Lantolf, J., & Pavlenko, A. (2001). (S)econd (L)anguage (A)ctivity theory: Understanding second language learners as people. In M. Breen (Ed.), *Learner Contributions to Language Learning* (pp. 172-182). London: Longman.

- Lantolf, J. P. & Poehner, M.E. (2007). Dynamic assessment: Bringing the past into the future. *Journal of Applied Linguistics*, 1, 49-74.
- Lea, M., & Stierer, B. (2000). (Eds.) *Student Writing in Higher Education*. Open University Press: Buckingham, U.K.
- Lee, I. (2007). Feedback in Hong Kong secondary writing classrooms: Assessment for learning or assessment of learning. *Assessing Writing*, 12, 180-198.
- Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing*, 17(3), 144-164.
- Lee, G. & Schallert, D. (2008). Meeting in the margins: Effects of the teacher student relationship on revision processes of EFL college students taking a composition course. *Journal of Second Language Writing*, 17, 165-182.
- Leung, C., & Lewkovicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly*, 40(1), 211-234.
- Liu, N. F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment, *Teaching in Higher Education*, 11(3), 279-290.
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.). *Handbook of Second Language Acquisition*. San Diego: Academic Press.
- Lucas, P. (1991) Reflection, new practices and the need for flexibility in supervising student-teachers, *Journal of Further and Higher Education*, 15(2), 84-93.
- MacMillian, J. (2007). (Ed.) *Formative Classroom Assessment: Theory into Practice*. New York: Columbia University Teachers' College Press.
- Magin, D. & Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: how reliable are they? *Studies in Higher Education*, 26(3), pp. 288-297.
- Maki, P. (2004). *Assessing for learning*. Sterling, VA: Stylus Publishing.
- Marsh, C. (2007). A critical analysis of the use of formative assessment in schools. *Educational Research for Policy and Practice*, 6(1), 25-29.
- Marshall, B. (2007). Formative classroom assessment in English, the Humanities, and Social Sciences. In J. MacMillian (Ed.). *Formative Classroom Assessment: Theory into Practice*. (pp.136-153) New York: Teachers College Press.
- Marshall, B. & Drummond, M. J. (2006). How teachers engage with assessment for learning: lessons from the classroom. *Research Papers in Education*, 21(2), 133-149.
- Marshall, B., & Wilam, D. (2006). *English Inside the Black Box*. London: nferNelson Publishing Company Ltd.

Maxwell, G. S. (2004) Progressive assessment for learning and certification: some lessons from school-based assessment in Queensland. Paper presented at the Third Conference of the Association of Commonwealth Examination and Assessment Boards, March 2004, Nadi, Fiji. Retrieved July 22, 2008 from www.spbea.org.fj/aceab/GMaxwell.pdf

McColskay, W. & Egelson, P. (1993). *Designing Teacher Evaluation Systems that Support Professional Growth*. Greensboro, N.C: University of North Carolina at Greensboro, SouthEastern Regional Vision for Education.

McCormack, J. & Slaght, J. (2005). *Extended Writing and Research Skills, Teachers Book*. Reading, UK: Garnet Publishing Ltd.

McDevitt, T. (1994). Introduction: Individual differences in question asking and strategic listening processes. *Learning and Individual Differences*, 6(2), 131-135.

McGarrell, H. & Verbeem, J. (2007). Motivating revision of drafts through formative feedback. *ELT Journal*, 61(3), 228-236.

McKenzie, J. (1997). The question is the answer. *From Now On*, 7(2). Accessed online November 15, 2008 at <http://questioning.org/Q6/question.html>

McLaughlin, P. & Simpson, N. (2004). Peer assessment in first year university: how the students feel. *Studies in Educational Evaluation*, 30, 135-149.

McTighe, J. & O'Connor, K. (2005). Seven practices for effective teaching. *Educational Leadership*, November, 10-17.

McWhorter, K. (2006a). *The Writers Selection: Shaping Our Lives, 4th Edition*. Houghton Mifflin Company: Boston.

McWhorter, K. (2006b). *Instructors Resource Manual, The Writers Selection: Shaping Our Lives, 4th Edition*. Houghton Mifflin Company: Boston.

McVeigh, B. (2002). *Japanese Higher Education as Myth*. New York: M. E. Sharpe Inc.

Montgomery, J. & Baker, W. (2007). Teacher written feedback: student perceptions, teacher self-assessment and actual teacher performance. *Journal of Second Language Writing*, 16, 82-99.

Morgan, N. & Saxton, J. (1994). *Asking Better Questions Models, Techniques and Classroom Activities for Engaging Student Learning*. Ontario, Canada: Pembroke Publishers.

Mowl, G. & Pain, R. (1995). Using self and peer assessment to improve students' essay writing; a case study from Geography. *Innovations in Education and Training International*, 32, 324-335.

Murphy, S. (2000). A sociocultural perspective on teacher response: Is there a student in the room? *Assessing Writing*, 7, 79-90.

Murphy, J. (2001). Reflective teaching in ELT. In D. Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language* (pp. 499-514). Boston: Heinle & Heinle.

- Murphy, R. (2006). Evaluating new priorities for assessment in higher education. In C. Bryan & Clegg, K (Eds.) *Innovative Assessment in Higher Education* (pp. 37-46). London: Routledge.
- Murphy, Sandra. (1994). Portfolios and curriculum reform: Patterns in practice. *Assessing Writing* 1, 175-206.
- Nasser-Abu Alhija, Fadia. (2006). Book review. *Studies in Educational Evaluation*, 32, 401-409.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22, 155-175.
- Nichols, G. (2004). *Professional Development in Higher Education*. New York: Routledge.
- Nicol, D. & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Nicol, D., & Milligan, C. (2006). In Bryan, C. & Clegg, K. (Eds.). *Innovative Assessment in Higher Education* (pp. 64-78). New York: Routledge.
- Nigel, K. & Pope, N. (2005). The impact of stress in self- and peer assessment. *Assessment & Evaluation in Higher Education*, 30(1), pp. 51-63.
- Nilson, L. (2003). Improving student feedback. *College Teaching*, 51(1), 34-39.
- Noonan, B. & Duncan, R. (2005). Peer and self-assessment in high schools. *Practical Assessment Research & Evaluation*. Retrieved on May 5, 2008 from pareonline.net/genpare.asp?wh=0&abt=10 - 11k
- Norton, L., & Norton, J. (2001). Essay feedback: how can it help students improve their academic writing? Paper presented at *First International Conference of the European Association for the Teaching of Academic Writing across Europe (EATAW)*. Gronigen: 18-20, June.
- Nunan, D. (1988). *The Learner-centered Curriculum*. Cambridge: Cambridge University Press.
- Nunan, D. (2004). Research methods. In M. Bryan (ed.) *Routledge Encyclopedia of Language Teaching and Learning*. London: Routledge. 515-520.
- Nyquist, J. B. (2003). The benefits of reconstruing feedback as a larger system of formative assessment: A meta-analysis. Master's thesis, Vanderbilt University, 2003.
- O'Brian, T. (2004). Writing in a foreign language: teaching and learning. *Language Teaching*, 37, 1-28.
- OECD. (2005). *Formative Assessment: Improving Learning in Secondary Classrooms*. Paris: OECD.
- Ofsted. (2003). Good assessment practices in modern foreign languages (MFL). The Office of Standards in Education. Retrieved February 20, 2009 from http://www.qca.org.uk/qca_4369.aspx

- O'Neill, P. (2000). Introduction. *Assessing Writing*, 4, 1-4.
- Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham & D. Corson (eds.). *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment* (pp.175-187). Netherlands: Kluwer Academic Publishers.
- Otoshi, J. & Heffernan, N. (2008). Factors predicting effective oral presentations in EFL classrooms. *Asian EFL Journal*, 10(1). Retrieved May 15, 2008 from http://www.asian-efl-journal.com/March_08_jo.php
- Oxford, R. (2001). Language learning styles and strategies. In M. Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language* (359-367). Boston: Heinle & Heinle.
- Papinczak, T., Young, L., & Groves, M., (2007). Peer assessment in problem-based learning: a qualitative study. *Advances in Health Science Education*, 12, 169-186.
- Pedder, D., James, M., & MacBeath, J. (2005). How teachers value and practice professional learning. *Research Papers in Education*, 20(3), 209-243.
- Peer Assessment. (2007). University of Technology Sydney. Institute for Interactive Media and Learning. Retrieved May 22, 2008 from <http://www.iml.uts.edu.au/assessment/students/peer.html>
- Phillips, E. (2000). *Self-assessment of class participation*. Unpublished paper, Department of English, San Francisco State University.
- Pintrich, P.R., & DeGroot, E.V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33-40.
- Polos, A., & Mahoney, M. (2008). Effectiveness of feedback: the students perspective. *Assessment & Evaluation in Higher Education*, 33(2), 143-154.
- Porcaro, J. (2004). Promoting progressive change in the work of secondary school JTEs. Toyama University of International Studies, *Journal of the Faculty of Humanities and Social Sciences*, 4, 79-87.
- Price, M., O'Donovan, B. (2006). Improving performance through enhancing student understanding of criteria and feedback. In Bryan, C. & Clegg, K. (Eds.). *Innovative Assessment in Higher Education* (pp. 100-109). New York: Routledge.
- Prosser, M. and Trigwell, K. (1999). *Understanding Learning and Teaching: the Experience in Higher Education*. Buckingham: Society for Research in Higher Education, Open University Press.
- Pryor, J. & Crossouard, B. (2005). *A sociocultural theorization of formative assessment*, Paper presented as the Sociocultural Theory in Educational Research and Practice Conference, University of Manchester, 8-9, September.
- QCA (2005). Qualifications and Curriculum Authority. Characteristics of AfL. Retrieved Sept. 10, 2008 from http://www.qca.org.uk/qca_4337.aspx

- Race, P., Brown, S. & Smith, B. (2005). *500 Tips on Assessment*. London: Routledge
- Ramaprasad, A. (1983) On the definition of feedback. *Behavioural Science*, 28, 4-13.
- Rea-Dickins, P. (2001). Mirror, Mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462
- Reddy, M. (2007). Effect of Rubrics on Enhancement of Student Learning. *Educate*, 7(1). 3-17. Retrieved February 6, 2009 from
<http://www.educatejournal.org/index.php?journal=educate&page=issue&op=view&path%5B%5D=23>
- Rethinking Classroom Assessment with Purpose in Mind. (2006). Western and Northern Canadian Protocol for Collaboration in Education. Retrieved August 16, 2008 from www.wncp.ca/assessment/rethink.pdf
- Richards, J. & Lockhart, C. (1994). Reflective Teaching in Second Language Classrooms. New York: Cambridge University Press.
- Roshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, 66(2), 181-221.
- Ross, J. (2006). The reliability, validity and utility of self-assessment. *Practical Research, Assessment & Evaluation*, 11(10). Retrieved March 1, 2006 from <http://pareonline.net/pdf/v11n10.pdf>
- Ross, S., Jordan, S., & Butcher, P. (2006). Online instantaneous and targeted feedback for remote learners. In Bryan, C. & Clegg, K. (Eds.). *Innovative Assessment in Higher Education* (pp. 123-131). New York: Routledge.
- Rust, C., Price, M. and O'Donovan, B. (2003) Improving students learning by developing their understanding of assessment criteria and processes', *Assessment and Evaluation in Higher Education*, 28(2), 7-164.
- Sadler, R. (1989). Formative Assessment and the Design of Instructional Systems. *Instructional Science*, 18, 119-144.
- Sadler, R. (1998) Formative assessment: Revisiting the territory. *Assessment in Education*, 5(1), 77-84.
- Schon, D. A. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Schon, D. A. (1987). *Educating the Reflective Practitioner*. San Francisco: Jossey-Bass.
- Schwandt, T.A. (1997). *Qualitative Inquiry: A Dictionary of Terms*. Thousand Oaks, CA. Sage Publications.
- Scott, V. (1996). Rethinking Foreign Language Writing. Boston: Newbury House.
- Seliger, H.W. & Shohamy, E. (1989). *Second language research methods*. Oxford: Oxford University Press.

- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Shi, L. (1998). Effects of prewriting discussions on adult ESL students' compositions. *Journal of Second Language Writing*, 7(3), 319-345.
- Shindler, J. (2003). Creating a more peaceful classroom community by assessing student participation and process. *OJPCR: The Online Journal of Peace and Conflict Resolution* 5(1). Retrieved January 10, 2006, from http://www.trinstitute.org/ojpcr/5_1shindler.htm
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Skinner, E., & Belmont, M. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, 85(4), 571-581.
- Spencer, S. (2003). An interview with Dr. Brian J. McVeigh. *The Language Teacher*, August, 1-8. Retrieved January 22, 2009 from <http://www.jaltpublications.org/tlt/articles/2003/08/spencer>
- Stefani, L. (1998). Assessment in partnership with learners. *Assessment and Evaluation in Higher Education*, 23(4), 339-350.
- Stern, L., & Solomon, A. (2006). Effective faculty feedback: the road less traveled. *Assessing Writing*, 11, 22-41.
- Stiggins, R. (1987). Design and development of performance assessment. *Educational Measurement: Issues and Practice*, 6(3), 33-42.
- Stiggins, R. (2002). Assessment Crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, 83, 758-765. Retrieved February 5, 2008 from <http://www.pdkintl.org/kappan/k0206sti.htm>
- Stiggins, R. (2006). Assessment for learning: a key to motivation and achievement. *Edge*, 2(2). 1-19. Phi Delta Kappa International. Retrieved July 28, 2008 from www.assessmentinst.com/forms/KappanEdgeArticle.pdf
- Stiggins, R. (2007). Conquering the formative assessment frontier. In J. McMillan (Ed.), *Formative Classroom Assessment: Theory into Practice* (pp. 8-27). New York: Teachers College Press, Columbia University.
- Stiggins, R. (2008). *Assessment Manifesto*. Educational Testing Service. Retreived on June 4, 2008 from www.assessmentinst.com/forms/AssessmentManifesto.pdf
- Stiggins, R., Arter, J., Chappuis, J., & Chappuis, S. (2004). Classroom assessment for student learning: doing it right-using it well. ETS Assessment Training Institute.
- Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory Into Practice*, 44(1), 1-18.

Stobart, G. (2006). The validity of formative assessment. In Gardner, J. (Ed.), *Assessment and Learning* (pp. 133-146). London: Sage Publications.

Straub, R. (2000). The student, the text, and the classroom context: A case study of teacher response. *Academic Writing*, 7, 23-55.

Struyven, K., Dochy,F., and Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341.

Stubbs, M. (2000). Society, education and language: The last 2000 (and the next 20?) years of language teaching. In H. Trappes-Lomax (Ed.), *Change and Continuity in Applied Linguistics* (pp. 15-37). Clevedon, UK: British Association of Applied Linguistics/Multilingual Matters.

Suffolk Advisory Council, 'How do they walk on hot sand?'. Retrieved Nov. 20, 2008 from www.slamnet.org.uk/assessment/Hengrave%20site/HOTSAND.PDF

Sunda, R. (2003). Thinking about thinking: What makes a good question? *Learning and Leading with Technology*, 30(5), 10-15.

Supon, V., & Wolf, P. (1993). Tearing down walls to promote student-generated questions. (Eric Document Reproduction Service No. ED 361 336). Retrieved Dec. 15, 2008 from http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED361336&ERICExtSearch_SearchType_0=no&accno=ED361336

Taboada, A. (2003). The Association of Student Questioning with Reading Comprehension. Retrieved Nov. 20, 2008 from www.lib.umd.edu/drum/bitstream/1903/44/1/dissertation.pdf

Taboada, A., & Guthrie, J.T. (2006). Contributions of student questioning and prior knowledge to construction of knowledge from reading information text. *Journal of Literacy Research*, 38(1), 1-35.

Taras, M. (2002). Using assessment for learning and learning from assessment, *Assessment and Evaluation in Higher Education*, 26, 501-510.

Taras, M. (2008). Summative and formative assessment. *Active Learning in Higher Education*, 9(2), 172-192.

The New Zealand Curriculum, Key Competencies. Ministry of Education, the New Zealand government. Retrieved Dec. 28, 2008 from http://nzcurriculum.tki.org.nz/the_new_zealand_curriculum/key_competencies

Tomlinson, B. (2005). English as a foreign language: matching procedures to the context of learning. In E. Hinkel (Ed.). *Handbook of Research in Second Language Teaching and Learning*. (137-153). New Jersey: Lawerance Erlbaum Associates, Inc.

Topping, K. (1998). Peer-assessment between students in colleges and universities. *Review of Educational Research* 68, 249-276.

UNESCO (1980). *UNESCO Handbook for Science*. Paris/London: Heinemann

- Van Der Meij (1994). Student questioning: a componential analysis. *Learning and Individual Differences*, 6(2), 137-161.
- Vandervelde, J. (2001). Powerpoint Rubric. University of Wisconsin-Stout. Retrieved June 10, 2007 from <http://www.uwstout.edu/soe/profdev/pptrubric.html>
- van Teijlingen, E., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update*, Issue 35, Winter.
- Vu, T. & Alba, G. (2007). Students' experience of peer assessment in a professional course. *Assessment & Evaluation in Higher Education*, 32(5), 541-556.
- Weaver, W. & Cotrell, H.W. (1986). Peer evaluation: A case study. *Innovative Higher Education*, 11,25-39.
- Weaver, R. (2006). Do students value feedback? Students perceptions of tutors written responses. *Assessment & Evaluation in Higher Education*, 31(3), 379-394.
- Wen, M. & Tsai, C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education*, 51, 27-44.
- Wiggins, G. (1993). *Assessing Student Performance*. San Francisco: Jossey-Bass.
- White, E. (2009a). Student perspectives on peer assessment for learning in a Public Speaking course. *Asian EFL Journal*, 33, 1-55.
- White, E. (2009b). Assessing the assessment: an evaluation of a self-assessment of class participation procedure. *Asian EFL Journal* 11(3), 75-109.
- Wiggins, G. (1997). Feedback: How learning occurs. In: E. E. Chaffee (Ed.), *Assessing Impact: Evidence and Action* (pp. 31–39). Washington, DC: American Association for Higher Education.
- Wiliam, D. (2007). What does research say the benefits of formative assessment are? *Assessment Research Brief*. National Council of Teachers of Mathematics, 1-3. Retrieved on February 24, 2009 from <http://www.nctm.org/news/content.aspx?id=11466>
- Wiliam, D. (2008). Improving learning in science with formative assessment. In Janet Coffey, R. Douglas and Carole Stevens (Eds). *Assessing Science Learning: Perspectives from Research and Practice* (pp. 29-42). Arlington, VA: NSTA Press.
- Wiliam, D. & Black, P. (1996). Meaning and Consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, 22(5).
- Wiliam, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. In J. McMillan (Ed.), *Formative Classroom Assessment: Theory into Practice* (pp. 8-27). New York: Teachers College Press, Colombia University.
- Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. Dwyer (ed.) *The Future of Assessment: Shaping Teaching and Learning*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004) Teachers developing assessment for learning: impact on student achievement. *Assessment in Education*, 11(1), 49-66.
- Willis, J. (2007). Assessment for learning-why the theory needs the practice. *International Journal of Pedagogies and Learning*, 3(2), 52-59.
- Wix, L., & John-Steiner, V. (2008). Peer inquiry: Discovering what you know through dialogue. *Thinking Skills and Creativity*, article in press, 1-9.
- Wojtas, O. (1998) Feedback? No, just give us the answers, *Times Higher Education Supplement*. September 25.
- Wood, R. (1987). *Measurement and Assessment in Education and Psychology: Collected Papers 1967-1987*. London: Falmer.
- Yamashiro, A. & Johnson, J. (1997). Public Speaking in EFL: elements for course design. *The Language Teacher*. Retrieved February 10, 2006 from <http://www.jalt-publications.org/tlt/files/97/apr/yamashiro.html>
- Yorke, M. (2003) Formative assessment in higher education: moves towards theory and enhancement of pedagogic practice, *Higher Education*, 45, 477-501.