Using "Assessment for Learning" practices with pre-university level ESL students: A mixed methods study of teacher and student performance and beliefs

D. Christian Colby

Department of Integrated Studies in Education

McGill University, Montreal

August, 2010

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements of the degree of PhD

© D. Christian Colby, 2010

ABSTRACT

The use of assessment to foster learning has become established in classroom settings in recent years, where it has drawn considerable research interest, as learners have come to take more responsibility for their learning. The Language Testing (LT) community has recently called for more research into advances in alternative assessment practices (Brookhart, 2005; Fox 2009; Harlen & Winter, 2004; McNamara 2001a, 2001b; Pellegrino et al., 2001; Poehner and Lantolf 2005; Rea-Dickins 2004; Shohamy, 2004; Turner, 2009). The present research reports on an exploratory study incorporating treatment and control groups, in which *assessment for learning (*AFL) principles were applied in two pre-university English for academic purposes (EAP) classes. The study focussed on student learning of a grammatical feature (the use of *would* and *will* in contingent use contexts) as a vehicle for investigating AFL. The study has sought to (a) interpret AFL by developing AFL procedures appropriate to a second language (L2) classroom, (b) apply these AFL procedures in an L2 classroom setting, and (c) investigate their effect on learning, and in addition, to investigate for evidence of the *assessment bridge* (AB), the area of classroom practice linking assessment, teaching, and learning. An AFL methodology for L2 settings was developed for the study in the form of teacher training. The AFL pedagogical materials included computer-assisted language learning (CALL), an online individual, group and teacher-class concept mapping exercises. The data collection instruments included the concept maps produced, classroom observation field notes, transcribed group and class discourse, teacher and student survey questionnaires, and pre- and post-treatment tests to indicate trends. The data were analyzed by mixed methods and the results triangulated. The results found evidence of several instances of the AB and suggest that the application of AFL procedures may have enhanced student learning of the modal usage in question. This study reporting concludes with a call for a research agenda in the LT community for further study of applications of an AFL approach in EAP classroom settings.

RÉSUMÉ

Dans les dernières années, le recours à l'évaluation pour favoriser l'apprentissage est devenu une pratique courante dans les salles de classe. Cela a eu pour effet de créer un intérêt grandissant pour la recherche, puisque les apprenants prennent davantage leur apprentissage en main. La communauté des chercheurs en évaluation des langues a récemment demandé que plus de recherches soient faites sur les progrès dans le domaine des pratiques d'évaluation alternatives. (Brookhart, 2005; Fox 2009; Harlen & Winter, 2004; McNamara 2001a, 2001b; Pellegrino et al., 2001; Poehner et Lantolf 2005; Rea-Dickins 2004; Shohamy, 2004; Turner, 2009). La présente recherche fait état d'une étude exploratoire qui incorpore des groupes expérimentaux et contrôles, dans lesquels les principes de l'*évaluation pour l'apprentissage (*EPA) ont été appliqués et ce, dans deux cours d'anglais pour des études au niveau préuniversitaire. L'étude s'est appuyée sur l'apprentissage par les étudiants d'un trait grammatical (l'utilisation de *would* et *will* dans un contexte hypothétique) comme véhicule pour étudier l'EPA. Cette étude a cherché à (a) interpréter l'EPA en développant des procédures d'EPA appropriées pour une classe de langue seconde, (b) appliquer ces procédures dans une classe de langue seconde, et (c) étudier leur effet sur l'apprentissage, en plus de chercher des cas de pont évaluatif (PE), cette zone de la pratique pédagogique faisant le lien entre l'évaluation, l'enseignement et l'apprentissage. Dans le cadre de cette étude, une méthodologie de l'EPA dans un contexte de langue seconde a été développée sous la forme d'une formation des enseignants et le matériel pédagogique qui a été utilisé, incluait l'enseignement assisté par ordinateur (EAO), ainsi que 3 exercices de schématisation conceptuelle: individuel en-ligne, en petits groupes et en classe avec la participation de l'enseignant. Les instruments de collecte de données incluaient les schémas conceptuels produits, les notes d'observation prises en classe, la transcription des discussions de groupe et de classe, les questionnaires de sondages menés auprès des étudiants et des enseignants, ainsi que les prétests et les post-tests afin de démontrer certaines tendances. Les données ont été analysées utilisant une méthodologie mixte et les résultats triangulés. Ces derniers ont mis en évidence plusieurs occurrences du PE et

ont suggéré que l'application des procédures d'EPA aurait aidé les étudiants dans leur apprentissage de cette forme grammaticale. Ce rapport d'étude recommande donc à la communauté des chercheurs en évaluation des langues de mener des recherches plus exhaustives au sujet des applications d'une approche de l'EPA dans le domaine des cours d'anglais.

TABLE OF CONTENTS

TABLES

FIGURES

ABBREVIATIONS AND ACRONYMS

AB	assessment bridge
AFL	Assessment for Learning
AiFL	Assessment is for Learning (Scotland)
ALTE	Association of Language Testers in Europe
AM	acquisition metaphor (of learning)
ARG	Assessment Reform Group
AUA	assessment use argument
CALL	computer-assisted language learning
CBA	classroom-based assessment
CDA	cognitive diagnostic approaches
CGI	cognitively guided instruction
CM	concept map
CMC	computer-mediated communication
CET	College English Test (China)
CGI	cognitively guided instruction
CG1, CG2	control group 1, control group 2
CLIO	culture and learning in organisations (UK)
DA	dynamic assessment
DELNA	Diagnostic English Language Needs Assessment
EAL	English as an additional language
EAP	English for academic purposes
EFL	English as a foreign language
ESL	English as a second language
FITB	fill-in-the-blanks
IL	interlanguage
LEA	local education authority (UK)
LEARN project	Learners' Expectations of Assessment for Learning Nationally project (UK)
LT	language testing
L1	first language
L2	second language

OHP	overhead projector
PM	Participation metaphor (of learning)
QCA	Qualifications and Curriculum Authority (UK)
QUAL	qualitative
quan	quantitative
MM	mixed methods
MMR	mixed methods research
NCTM	National Council of Teachers of Mathematics (US)
NRC	National Research Council (US)
SBA	School-based assessment (Hong Kong)
SLA	second language acquisition
TESL	teaching English as a second language
TG1, TG2	treatment group 1, treatment group 2
ZPD	zone of proximal development

ACKNOWLEDGEMENTS

So many people have helped me in this project in large and generous ways and in small ways that had large effects. First, I would like to give my sincere and heartfelt thanks to my supervisor, Dr. Carolyn Turner, who gave and gave with endless empathy and certainly with grace under pressure, and for her initial encouragement that I should continue my studies following my M.A. degree. I would also like to thank the professors, support staff and librarians at McGill who were ever so kind and helpful. I have been blessed, in addition, with having had an exceptional and generous Advisory Committee, from whom I have learned so much, as each member has contributed meaningfully to the design and flavour of my study. They are Dr. Janna Fox, Dr. Joanna White, and Dr. Lise Winer. My very sincere thanks also go to the PhD support group of Beverly Baker, Candace Ferris, Pamela Gunning, Talia Isaacs, Heike Neumann, Dr. May Tan, and Lily Wang (Jing). Many others have helped me in many ways, with advice and help in software applications, graphic design and translating, and included in that group are Jean Villeneuve, Véronique Boucher, Josette Sergerie, Rose Marie Smith and Jim Howden. I would like to thank the administrators as well as the teachers and students of the school involved in my study for many acts of kindness, generosity and patience. A special thanks to the teacher who called her substitute teacher from an unexpected trip to a hospital emergency room to say, "Be sure to collect Christian's data tomorrow!" Finally, my deep gratitude will always be with my little family, my dear and supportive husband Peter (who used his teacher-knowledge to suggest I look at *will* and *would*) and my little curious son Seamus, future researcher.

CHAPTER 1 – INTRODUCTION TO THE STUDY

I see, Prometheus, and I would commend
To thee, so quick of wit, the wisest course

Aeschylus[1]

1.1 The thesis inspiration

Picture this, a classroom of learners and one teacher presiding. The teacher asks, "Who understood the reading I've just read aloud?" Several hands shoot up enthusiastically and a chorus of inquisitive and friendly learner responses is heard to say, "I understood a lot but I didn't understand the main idea," "I'm having trouble with some of the terms," "My comprehension is good but my expression is weak," and "I'm not sure about this reading, can you please help me work out what you want us to learn from it?" This may sound like a teacher's "stuff as dreams are made on" (Shakespeare, trans. 1961) and rather unlike real-life classroom encounters. Wouldn't it be wonderful if learners were so interested in taking charge of their learning, if learning could occur in such a climate of teacher-learner cooperation and even more, if learners felt no reluctance in admitting they had not yet understood or learned content in a classroom setting? Questions of this nature have inspired me to study the second language (L2) application of a formative assessment methodology intended to promote just such a classroom learning climate, the *Assessment for Learning* (AFL) approach. AFL is an attempt to use state-of-the art formative assessment techniques to help teachers and learners work together to advance classroom learning. Before proceeding, it would be useful to first ask, 'how did the idea of AFL come to be?'

1.2 Assessment for Learning introduced

AFL grew out of Black and Wiliam's (1998a) landmark survey of 250 studies of successful classroom-based learning, in which formative assessment practices were identified as the key common element where there was

[1] This quote, of Oceanus speaking to Prometheus from *Prometheus Bound* is intended to illustrate an example of the use of *would*, which is the grammatical feature employed in this study, thus evoking a wishful state as well as wisdom herein as a wished-for state of affairs.

evidence of successful learning. Drawing on the work of Black and Wiliam, the Assessment Reform Group (ARG) of King's College, London, developed the AFL approach to formative assessment.

Formative assessment had been used by teachers as a common teaching tool to help individual learners identify areas in their learning progress needing improvement. The AFL approach has afforded teachers and learners a more refined version of traditional formative assessment, which has been designed to promote increased learner responsibility for learning and has a cooperative teacher-learner focus. AFL has attempted to situate learners in a position of control of their learning outcomes, and to help teachers and learners share the goals of assessment in order to motivate learners and advance the learning process. In so doing, AFL has sought to diminish the negative associations learners may have when assessment outcomes indicate gaps in their learning.

AFL was originally used in general education, where it has been widely reported to be instrumental in advancing classroom-based learning (Black & Wiliam, 2005b; 2006; Cowie, 2005; Harlen & Winter, 2004). My own experience of coordinating a research project examining formative assessment practices in an L2 classroom setting (Colby-Kelly & Turner, 2007), has inspired me to investigate how AFL might also help the learners I know best: second language (L2) learners in a classroom setting.

1.3 AFL and L2 classroom settings: A rare combination

My interest in L2 AFL was also inspired by the work of Harlen and Winter (2004), who reported on the how AFL was applied in the disciplines of mathematics and science, and who argued for AFL applications in other learning contexts, such as in L2 learning. Historically, AFL has mostly been applied in general education in subjects as diverse as science, geography and English language arts, disciplines which by their varied nature require that the AFL approach be implemented in different ways. But how might these applications differ in actual practice? Actually, Hodgen and Marshall (2005) discussed the question in their comparison and contrast of AFL, in applications in mathematics and English Language Arts classes. Hodgen and Marshall demonstrated that it was possible to vary methodologies in interpreting AFL in the two disciplines, while illustrating similarities and

differences in how the AFL approach was conceptualized and put into classroom practice in each. The work of Harlen and Winter and that of Hodgen and Marshall have set precedents in the literature as to the breadth of potential AFL applications. Their research suggested to me that L2 learners might also benefit from an AFL approach.

In 2005, when I began conceptualizing a study of AFL in an L2 classroom setting, I found that there was little application of the approach in those settings. I also found that research into L2 AFL was rare. In fact, despite my having conducted a thorough search, I was unable to find *any* pedagogical materials for L2 learners using the AFL approach. Thus, the idea of applying AFL in an L2 classroom setting would take hold, and remain a concern of mine until the present research realized this application of AFL for L2 learners.

This thesis tells how that AFL application was accomplished, how I developed AFL teaching materials specifically for L2 learners, how AFL procedures were applied in an L2 classroom setting, and how I conducted a mixed methods study to examine the effect of AFL on learning a specific language feature, in two L2 classes of adult learners. The results of this innovative research contribute to the literature by providing a picture of the effect of this unique application of AFL in an L2 classroom setting.

1.4 Definition of terms

Before going further in this discussion, it is important to establish some definitions of terms as they will be used in the present research.

Assessment could be said to be an umbrella term in the field of language testing (LT), since it covers many varied kinds of evaluation, including formative assessment. Generally *assessment* refers to evaluation that is informal in nature, while *testing* more often refers to more formal, mostly summative measurement of student performance or production. In the LT literature the two terms are often used interchangeably. For the purposes of the present study that practice will be adopted, so that unless there is a specific need to differentiate between the two, the terms assessment and testing will be used interchangeably.

In the research literature the terms formative assessment and AFL have also frequently been used interchangeably. In fact, Gardner (2006) has said "there is little of substance to distinguish the two terms 'formative assessment' and 'assessment for learning'…" (p. 2). As noted above, AFL is a methodological *approach* to formative assessment procedures; AFL can be said to encompass evolving formative assessment. Thus, following the work of Gardener (2006), in the present study I will use the terms formative assessment and AFL interchangeably, except when a clear distinction is required to ensure clarity.

The group responsible for initiating and spearheading the AFL approach is the Assessment Reform Group (ARG), a group of researchers in assessment policy and practice in the UK, dating from 1989. Drawing on the ARG (2002), Bachman (2005), and Lynch (2001), Colby-Kelly and Turner (2007) defined some key terms which I will use, and which are central to this study.

Thus, (a) *assessment* is seen as the process of seeking and interpreting evidence for making substantively grounded decisions or judgments about the product of a learning task; (b) *formative assessment* is taken to mean the process of seeking and interpreting evidence for making substantively grounded decisions or judgments about the product of a learning task in order to decide where the learners are in their learning, where they need to go and how best to get there; and (c) *feedback* is understood to mean comments or information learners receive from a teacher, from other learners, or from *themselves* upon reflection, on the product of a learning task (including self-assessment; peer-assessment; and teacher-student, teacher-group, and teacher-class feedback).

Similarly, the *assessment bridge* (AB) is understood to be the area of classroom-based assessment (CBA) encompassing *assessment* (where learners are in their learning), *teaching* (where they need to go and how best to get there) and *learning* (action on the part of the learner). Thus, feedback is included in the assessment bridge. The assessment bridge is the place where assessment, teaching and learning interweave in the classroom, and it is the place where the fields of language testing and second language acquisition interface, with the desired result of learning.

I will cite Larsen-Freeman's (2009) definition of *grammar*, who has defined grammar as "a system of meaningful structures and patterns that are

governed by particular pragmatic constraints", adding that inherent in her definition are the concepts of grammar as form, meaning and use (p. 521). Finally, I will use a definition of *grammatical ability for assessment purposes* put forth by Purpura (2004), which is that it is "the capacity to realize grammatical knowledge accurately and meaningfully in test-taking or other language-use contexts" (p. 89).

1.5 The study rationale, the Assessment Bridge, and original contribution

To return to this pursuit of applying AFL in an L2 classroom setting, the present research accords with recent calls in the language testing literature for study into innovation through the means of alternative assessment in second language learning (Garner, 2006; McNamara 2001; Poehner & Lantolf 2005; Rea-Dickins 2001, 2004, 2006b).

As the area of classroom practice linking assessment, teaching and learning, the AB has also increasingly become a source of research interest as several (Turner, 2006, 2009; Solomon, 2002; Young, 2005) in SLA and in LT have called for a greater interface between the disciplines. Colby-Kelly and Turner (2007) found evidence of the AB in their reporting of various instances of formative assessment in their study of classroom assessment practices and called for "taking L2 testing research to the next level, investigating the 'usefulness' (Bachman & Palmer, 1996; Bachman, 2005) of formative assessment and AFL" (p. 9).

That said, the focus of this study is not on detailed learning outcomes as might be expected in an SLA-oriented investigation. The focus of this study is, however, to investigate the effect of introducing an AFL methodology in an L2 classroom setting on teacher and student perceptions of learning from the perspective of LT. Thus, the main focus of the study is on the assessment methods applied and their perceived effect on learning.

Thus, overall the present research is an exploratory study investigating the effects of an application of AFL practices in an L2 classroom setting, and examining if there is evidence of the AB there. On a macro level, I have sought in the present study to (a) interpret the tenets of AFL by developing AFL procedures appropriate to an L2 classroom, (b) apply these AFL

procedures in an L2 classroom setting, and (c) investigate their effect on learning.

There has been much study of AFL in general educational research, but with some exceptions, little research thus far in the realm of L2 classroom-based studies (Rea-Dickins, 2004). This study contributes to the field of language testing and assessment research in helping to close the gap between AFL research in general education and that of L2. This study will inform the research literature and make an important contribution to the second language research literature, due to its innovation through the development and application of:

- AFL procedures for an L2 classroom setting
- AFL teacher training in that setting
- AFL pedagogical materials for L2

The results of this study will yield a profile of teacher and student behaviours and perceptions about the effects of an application of AFL techniques and procedures in this second language classroom setting.

1.6 The thesis content

This first chapter has provided an introduction to this innovative study of an application of L2 AFL. The second chapter will review the theories of learning that have led to the development of formative assessment and subsequent AFL practices in classroom settings, and how the theory and practice of AFL have resulted in the choice of pedagogical activities used in the present research. A description of the study's mixed-methods design as well as an explanation of the steps involved in the development of the pedagogical and data collection instruments are provided in Chapter 3. Chapter 4 explains in detail the methodological and data analysis procedures used for the data collection of the study. The results of the qualitative, quantitative, and mixed-methods analyses are presented in Chapter 5. In Chapter 6 these results are discussed and triangulated to provide a picture of all of the study results in concert. Chapter 7 provides concluding remarks, mention of the study limitations, contributions to the literature, and several recommendations for future research.

CHAPTER 2 – A REVIEW OF THE LITERATURE

Don't worry about the mistakes – there aren't any.

Chet Atkins[2]

In this chapter I will review the development of the AFL approach from its inception to the present time, arising as it has in general education from motivational studies, mastery learning, and then to formative assessment techniques, and finally culminating in the updated version of formative assessment, AFL. I will discuss the groundbreaking work of Black and Wiliam (1998a, 1998b), which has largely served as the inspiration for AFL. The principles defining AFL will also be examined. Recent innovations in assessment, such as diagnostic, dynamic, and alternative assessments and their relation to AFL will be discussed. Following that the AFL research to date in general education studies, as well as a comparison study of AFL in application in mathematics and in English language arts classrooms (bridging science and language studies) will be examined. AFL research in L2 classroom settings will be reviewed, and the current need for more research in these settings as well as calls for more research will be discussed. Finally, an answer to these calls will be proposed in the present study.

2.1 To start at the beginning: Theories of learning

The story of how to apply a new methodology of assessment to help classroom-based students learn, must by virtue of its goal originate in an understanding of a theory of learning in classroom settings. In fact, the AFL approach asks that teachers understand *how* as well as *what* their students are learning. There are three basic educational theories about how learning takes place (James, 2006), each differentiated by its view on the processing, storage, and retrieval of knowledge. They are the behaviourist, cognitive (or constructivist), and socio-cultural (or situated) theories of learning. Each will be discussed with a view to how it pertains to the present research.

[2] This quote is in line with the philosophy of AFL, in that error is not viewed as a bad thing. In fact, the AFL approach views error as a neutral or positive occurrence, or as a means leading to the end goal of learning.

2.1.1 Behaviourist theory.

The behaviourist theory of learning is based on the observation that learning often occurs when certain behaviours are positively reinforced (Harlen, 2006). Thus, behaviourists view learning as a conditioned response to external stimuli. To behaviourists, skills are conceived on a skill hierarchy continuum, and are acquired in sequence. In the behaviourist approach, observable behaviours are seen to be indicators of acquired knowledge (James, 2006).

In an AFL or indeed in any classroom setting, teachers are taking a behaviourist teaching approach when they give positive reinforcement to learners who demonstrate target behaviour. Moreover, in an *L2* classroom setting, teachers take a behaviourist approach when they encourage students to recite and retrieve target language forms.

Thus, in behaviourist theory, conditioning in the learning environment determines learning success. However, it is important to point out that a drawback of the behaviourist approach is that student performance may *not* reflect learning, since a behaviourist methodology encourages, and may only stimulate, short-term memorization rather than actual learning. In addition, behaviourism works on the assumption that learning difficulties may be remedied by more practice, which may not be the case, particularly in the face of problems with complex cognitive causes of the difficulties.

The hierarchical skills continuum of the behaviourist approach has value for the teaching of some subjects, such as mathematics and science. It is reflected in the AFL practices of encouraging learning by scaffolding knowledge, step-by-step. The behaviourist approach also informs the AFL approach in its positivist aspect. But, it will be recalled that AFL asks that teachers understand *how* and *what* their students are learning, questions better answered by taking a more cognitive approach to learning.

2.1.2 Cognitive theory.

Another layer of understanding learning can be seen in the cognitive (or constructivist) theory of learning, which is concerned with how learners' active engagement, construction of meaning, mental schema, and understanding of knowledge, may contribute to learning (James, 2006).

Included in this theory are several concepts that have been incorporated into AFL practice.

First, is the view of the teacher-as-learning-guide, leading students towards their learning goal. Second, is a focus on using problem-solving to advance learning. Third, the use of deductive and inductive reasoning is also important in the cognitive theory of learning. The knowledge (usually self-knowledge) of the mental processes used in learning have been defined as metacognition (Richards, Platt & Platt, 1992). Also, the self-monitoring and self-regulation of metacognition are also key factors in the cognitive theory, as it is in AFL practice. Finally, background or prior knowledge is considered to be an important basis for building new mental schema (or scaffolding knowledge) in learning, as also emphasized in AFL practice. The latter accords with the AFL focus on prior knowledge with regard to the learning checks inherent in formative assessments. Many more layers of the concept of learning are revealed in the cognitive theories of learning, but there are more again to be discovered, in the socio-cultural theory of learning.

2.1.3 Socio-cultural theory.

The socio-cultural theory of learning (also known as situated theory), draws on several influences, most notably on the work of Lev Vygostsky (1962, 1978), in which learning is seen as 'situated' in social interaction between the individual and the social environment surrounding that individual. The socio-cultural theory emphasizes group learning effects, shared learning, and co-constructed learning which would be of benefit to the group as well as to individual learner. From a socio-cultural perspective, learning cannot occur in isolation.

The socio-cultural theory has also drawn on Vygotsky's (1969, 1978) theory of the learner *zone of proximal development* (ZPD), wherein learners' *actual* versus *future* or *potential* development was defined by Vygotsky (1978) in the ZPD, as "…*the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers*" (p. 86, italics in original). Though Vygotsky attributed the ZPD to learning development in children, the concept

has been widely applied to any learners (Black & Wiliam, 2006). The ZPD is highly social in that it specifies that interaction with a more capable other (peer or adult) is required in order for the learner to move beyond the actual to the potential competence.

In the classroom setting, Pellegrino, Chudowsky, and Glaser (2001) found that learning in a socio-cultural environment could enhance formative assessment practices by allowing evidence of individual progress to be more readily accessible to learner peers and teachers.

In this highly socialized theory, both cultural artefacts and mediation influence the development of learning outcomes. James (2006) discussed how the socio-cultural approach related to knowledge scaffolding, in the following:

> It is important to find activities that learners can complete with assistance but not alone so that the 'more expert other', in some cases the teacher but often a peer, can 'scaffold' their learning (a concept shared with cognitivist approaches) and remove the scaffold when they can cope on their own (p. 57).

I would add that the concept of knowledge scaffolding is also shared by the behaviourists.

Thus, the socio-cultural approach to learning reveals more layers of the learning paradigm, allowing for a deeper understanding and a clearer picture of how learning as an ethnographic phenomenon may take place in a classroom setting.

2.1.4 Theories of learning, classroom settings, and AFL

James (2006) discussed ways in which the behaviourist and cognitive approaches might be best used in classroom application. She suggested that behaviourist approaches would be well suited to the development of habitual behaviours or basic skills, and that cognitive approaches would work well in cases where a deep understanding of conceptual structures was a learning goal. Moreover, a socio-cultural approach is exemplary in ESL classrooms where group and paired interactions as well as class discussions are essential elements. It would make sense to draw from all three approaches to best advantage where appropriate, in classroom learning situations.

In the past learning has been viewed as an acquisitional process, and its holders could be said to have enjoyed the affluence of its possession. A more recent view of learning as an experiential process has come to be seen as a valid alternative viewpoint. With this in mind, Sfard (1998) classed learning concepts as either *acquisition* or *participation* metaphors, which she called the AM and the PM (p. 5), noting that with the latter "ongoing learning activities are never considered separately from the context within which they take place…the learner should be viewed as a person interested in participation in certain kinds of activities rather than in accumulating private possessions" (p. 6). This may lead us to ask, is learning a possession to be obtained by filling up minds much as cars may be filled with gasoline, or alternately, is learning an experience which by its nature is constantly shifting in character as the principle players interact with their environment?

It would seem that both the acquisition and the participation metaphors have merit. The AM works well as a basis for conceptualizing incremental and scaffolded learning activities, such as are often used in mathematics lessons or in instruction in learning to play a musical instrument. On the other hand, the PM would be more appropriate in all other learning situations (and in fact occurs in the AM as well, though less obviously), due to the socio-cultural and co-constructed, interactional nature of learning, which cannot occur in isolation.

In a study by the pro-AFL Teaching and Learning Research Programme (TLRP) in the UK, it appears the teachers under investigation had a strong grasp of effective ways to use learning theories such as the AM and PM in classroom teaching application, since James and Brown (2005), reported that the teachers were confident about their practices for improving classroom learning. However, the teachers were "much less confident" about gathering evidence demonstrating that learning (p. 8). This suggests a need for more research into effective formative assessment applications, AFL and, as it is sometimes called, classroom based assessment (CBA), that have been made by several (Davison, 2007; Rea-Dickins, 2001, 2004, 2006b; McNamara, 2001a; Turner, in press). It is significant that while there are L2 teachers using formative assessment techniques in their classes, there is very little research to shed light on the effectiveness of those practices. Some teachers have

expressed difficulty in reconciling theory and practice (Turner et al., 2010; Turner & Tan, 2009).

2.1.5 Learning and the challenge of assessing the 'unobservable'

LT and assessment practices must constantly respond to the requirement that adequate and fair judgements of L2 ability be made by making inferences based on examinee or learner behaviour. One of the challenges involved in making accurate assessment decisions is that learning is not always observable nor does it always result in observable behaviour. In a discussion of learning and interlanguage (IL), Lightbown (2000) noted that "the research on developmental sequences makes it clear that *progress* [sic] in a learner's IL will not necessarily show up as greater accuracy" (pg. 442). In reference to Lightbown, Larsen-Freeman also commented on the dilemma this phenomenon presents to assessing learning, in the following:

> Often learners go through developmental sequences where they progress, but their production is not target-like and therefore still very inaccurate … a lot of development will take place without its being manifest in learner performance. I believe this to be true because of the nonlinearity of the learning process. This is, of course, a problem for language testers as well as researchers (D. Larsen-Freeman, personal communication, Dec. 5, 2006).

Others (Turner, in press; Turner & Upshur, 1995; Young, 1995) have remarked on the difficulty of assessing unobservable progress made in language learning. In fact, Young discussed learner IL and some evidence that learner IL has been shown to restructure and reorganize, with the result that "intermediate stages may be further from the target than either beginning or advanced stages - a pattern commonly called U-shaped behavior" (p. 18). Turner and Upshur, as well as Turner have also argued that the complex phenomenon of learning may not be evident in learner performance. Clearly, the variability of IL performance in classroom settings has implications for teachers in formative assessment situations, who have a need to use learner performance to inform on the direction of subsequent teaching.

At this point the three prevalent theories of learning have been presented in relation to AFL in this discussion. Having discussed the behaviourist,

cognitive, and socio-cultural theories of learning, the discussion will now move on to review the advances in educational research over time that have influenced the development of the AFL approach in general education, in L2 settings, and moreover, that have influenced the conceptualization of the present study. Therefore at this point, the chronicle of how the present application of L2 AFL came into being will resume its course.

2.2 Using assessment to improve learning: Early research influences

2.2.1 Classroom learning and motivational studies.

The beginnings of the AFL approach can be seen in the motivational research of the 1980's, when it was suggested that learner motivation was an essential element in successful learning (Ames, 1992; Blumenfeld, 1992; Dweck, 1986). The effect of assessments on motivation was also investigated and acknowledged by Crooks (1988) and Natriello (1987). In fact, Crooks's investigation into the effect of classroom assessment on learner motivation, achievement and strategy use, led to the development of assessment practices geared to increase learner autonomy and control. Natriello's model of classroom-based evaluation included the following stages:

1. Establishing the purpose for evaluating students
2. Assigning tasks to students
3. Setting criteria for student performance
4. Setting standards for student performance
5. Sampling information on student performance
6. Appraising student performance
7. Providing feedback to student performers
8. Monitoring outcomes of the evaluation of students (p. 156).

Natriello's (1987) research influenced and foreshadowed later pedagogy towards including more formative assessment in classroom settings. He found that where 30% of the students surveyed reported they had not understood evaluation criteria, those students qualified their evaluations as unsatisfactory (p. 163).

The work of Crooks and Natriello led to new practices towards skills mastery in classroom settings.

2.2.2 Mastery learning.

Mastery learning allows learners to repeat identical assessments until the mastery of a particular skill has been achieved, and in so doing it treats assessment as a positive force of encouragement. Mastery learning thus increases learner autonomy. In comparing mastery learning to more traditional assessment methods Dweck (1986), signalled a change in the thinking of the time, noting that the approach was better suited to children who appeared to enjoy pursuing *mastery-oriented* tasks, as opposed to the *helpless* behaviour of those who tended to display avoidance and low persistence in tasks (p. 1040).

As was the case with mastery learning, formative assessment sought to promote successful learning using assessment to 'hone' skills throughout a course of study and prior to final, summative assessments. Unlike mastery learning, formative assessment uses various kinds of assessment procedures to achieve learning. Black and Wiliam (1998a) found mastery learning to be useful in that it provided timely, diagnostic feedback to students, and because it afforded the opportunity for peer reflection on learning goals. As in the case with AFL, mastery learning gives less weight to the number of times a learner attempts learning but instead focuses on a positive eventual learning outcome. Today mastery learning is used much less frequently than is formative assessment.

2.2.3 Formative assessment.

Formative assessment approaches assessment with the goal of fostering or *forming* the acquisition of learning through the use of ongoing assessments intended only to enhance learning and which do not contribute to final grade calculations. The term *formative evaluation* was first coined by Scriven (1967) and was later defined by Bloom, Hastings and Madaus (1971) as "another type of evaluation which all who are involved – student, teacher, curriculum maker – would welcome because they find it so useful in helping them improve" (cited in Black & Wiliam, 2003). Formative assessment includes learner self- and peer-assessment, teacher feedback conferencing and other ongoing, learner-empowering methods which seek to foster greater learner responsibility and self-awareness.

2.2.4 The influence of Sadler and others

The idea that CBA should serve learning has evolved over time, although it has more recently culminated in the AFL approach. For example, Sadler's (1989) theoretical article on learning and assessment has been credited by Brookhart (2005) as being responsible for first promoting the idea of modifying and expanding concepts of formative assessment towards those more serving of learning. Sadler noted the following three important conditions for improving learning, in his suggestions that learners:

1. understand the criteria and standards that were aimed for
2. be able to assess their own performance and to compare it with the standards
3. and be able to engage in action in order to close the gap.

Sadler cautioned 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).

On a related note, Glaser (1990) also argued that "…we must foster an environment in which tests are seen as measures of those forms of human competence that are essential to future learning rather than merely as indices to current or past achievement" (p. 480). Others have voiced similar views. For example, Genesee and Upshur (1996) advocated for effective classroom-based evaluation as part of instructional monitoring and modifying to enhance learning. In the same year that Black and Wiliam's (1998a, 1998b) studies were published, the journal *Assessment in Education* similarly proposed that "…assessment should advance learning rather than simply measure it…" (Anonymous, 1998 p. 301). More recently, Leung and Mohan (2004) were able to report that "There is now widely recognized support for classroom-based formative teacher assessment of student performance as a pedagogically desirable approach to assessment which is capable of promoting learning" (p. 335).

2.3 Black and Wiliam: Successful learning and formative assessment

Black and Wiliam (1998a, 1998b) reviewed 250 studies of successful learning in various classroom settings and in various disciplines in general education. They concluded that both learning and motivational benefits

occurred when formative assessment was used, and particularly as self- and peer-assessment practices. While Black and Wiliam's landmark review did not include studies of L2 learning, a review of AFL would be incomplete without a discussion of their work, since that work has served as the key inspiration for the AFL method. In their survey review, Black and Wiliam found that the one consistent feature across the studies they examined was that attention to classroom-based formative assessment led to significant gains in learning. Therefore in their conclusions, Black and Wiliam (1998b) declared formative assessment to be the key factor contributing to successful learning.

In their study, Black and Wiliam (1998a) discussed learner feedback and formative assessment, and categorized the latter in terms of *kinds of formative assessment*, *kinds of feedback*, *quality of feedback*, and the *impact of feedback on learner motivation*, as well as *teacher and student perspectives on formative assessment*. Following the publication of their work, Black and Wiliam's (1998a. 1998b) results served as the fundamental inspiration for mandating AFL in the educational policy of primary and secondary classroom practice in general educational disciplines in the UK.

Black and Wiliam (1998a) reported results from eight studies which they said were representative of the 250 they studied overall. These results will be briefly presented here, in order to illustrate the impact these findings had as Black & Wiliam's review gave rise to the AFL approach to formative assessment.

2.3.1 Fontana and Fernandes (1994) and learner self-assessment.

Fontana and Fernandes (as cited in Black & Wiliam, 1998a) studied the effects of learner self-assessments and reported that a treatment group of elementary students trained in learning objectives, assessment criteria, and self-assessment, showed significant improvement in pre- and post- tests of mathematical achievement.

2.3.2 Whiting, Van Burgh and Render (1995) and mastery learning.

A study by Whiting, Van Burgh and Render (as cited in Black & Wiliam, 1998a) investigated mastery learning use, in a case study of one teacher over a period of 18 years of teaching business to approximately 7, 000 students, who

the authors reported had highly successful learning outcomes, consistently high grade point averages and final test scores, less learner time before moving on to successive units, and diminished instances of test retakes. Black and Wiliam concluded that "the students' learning styles were changed as a result of the method of teaching" (p. 11).

2.3.3 Martinez and Martinez (1992) and frequent assessments.

In contrast to mastery learning's allowing retaking of *identical* tests to influence learning, Martinez and Martinez (as cited in Black & Wiliam, 1998a) investigated whether the increased frequency of *various* tests might have an effect on American college-level students' learning of algebra. The post-test results showed significant gains for those students tested more frequently, particularly for those students assigned to an inexperienced teacher.

2.3.4 Butler (1988) and the learning effect of feedback.

Butler (as cited in Black & Wiliam, 1998a) looked at intrinsic motivation and the effect of using three methods of assessing two written tasks, a) individual written comments, b) grades alone, and c) written comments and grades combined in a study of 48 Israeli elementary school mathematics students. Butler reported that a low achiever group's interest was undermined when grades were present in the feedback, whereas in contrast, the high achieving group maintained a level of interest in all three feedback situations. The study results support the premise that students may benefit more from commentary feedback than from grades.

Black and Wiliam (1998a) found this study to have implications for teachers' use of formative assessment, and particularly on its quality, as well as teachers' need to attend to formative assessment feedback and differential effects between high and low achievers. They concluded these results a) were consistent with findings that task-involving evaluation is more effective than ego-involving evaluation, b) supported the premise that especially on divergent tasks, a focus on grade achievement can lower task performance quality, and c) even when feedback comments are helpful for learner

performance, this effect may be undercut by the effects of normative feedback (grades).

2.3.5 Schunk (1996) and focussing on learning vs. performance.

Schunk (as cited in Black & Wiliam, 1998a) reported results that were similar to those of Butler (1998), in his study of 44 American elementary school students who were divided into groups focussing on learning how to solve problems (*learning goals*), and on simply solving problems (*performance goals*).

The results indicated that the effect of frequent self-evaluations out-weighed the effect of the two goal types; the learning goal group achieved higher results on achievement and motivation than did the performance-oriented group. In addition, the performance-oriented students who did not participate in self-evaluations had the poorest results of all. These results suggest a strong effect on motivation when the learners focussed on learning rather than performance.

2.3.6 Frederiksen and White (1997) and reflective assessments.

The objective of Frederiksen & White's (as cited in Black & Wiliam, 1998a) study was to develop an American middle school science curriculum based on practical questioning, on the subject of force and motion. The work of an experimental group peer discussions were structured to promote systematic, reasoned inquiry and reflective evaluations, including peer- and self-assessments. A control group lessons included general discussions of the modules.

A pattern of significant overall gains in the experimental group was reported and in each case the lowest baseline test achievers showed the greatest gains in the three results outcomes. In addition, in the experimental group, those students who demonstrated the best understanding of the assessment process went on to achieve the highest outcome scores.
Black and Wiliam reported the study's distinctive use of 1) different assessments directly related to teaching objectives, and 2) superior gains of those students heretofore labelled to be of 'low' and 'high' ability.

2.3.7 Fuchs and Fuchs (1986) and graphing learning progress.

Black and Wiliam cited Fuchs and Fuchs' overview meta-analysis of 21 studies (as cited in Black & Wiliam, 1998a) and found that the teachers who produced graphs of student progress reported greater learner gains than those who did not do so.

2.3.8 Bergan, Sladeczek, Schwarz and Smith (1991) and proactive diagnostic assessment.

The results of Bergan, Sladeczek, Schwarz and Smith's research results (as cited in Black & Wiliam, 1998a) will be reported in more detail because they demonstrate the potential positive effects of formative assessment practices, in a dramatic way. In fact, it would not be an exaggeration to say that the treatment results in Bergen et al.'s study had significant implications for the future of the participant children involved. Their study explored the premise that attention to the early acquisition of basic skills was essential for later academic success in Reading, Mathematics and Science. The participants were 838 five year-old kindergarten students from disadvantaged homes in six regions of the USA. Of these, 428 constituted a treatment and 410 a control group, having 29 and 56 teachers, respectively. A priori, expertise and experience were ascertained to have been similar in both teacher groups. The treatment consisted of the following focus on individual students by teachers through:

- observations to assess individuals' progress;
- differentiated tasks within activities to determine learning needs;
- locating students on a criterion-referenced model of the development of understanding;
- implementation of an assessment and planning system to inform on learning and teaching needs (as cited in Black & Wiliam, 1998a, p. 12).

The teachers engaged in consultations on learner progress after the first two weeks, and after four weeks they administered new assessments for further learner diagnostic information and to re-assess learner needs. Pre- and post-tests were also administered.

Bergan et al. reported that there were significantly higher post-test scores in Reading, Mathematics and Science in the experimental group than in the control group, and that initial outcomes were strong determiners of later outcomes.

Furthermore, they found that in the control group, 1 in 3.7 children were referred as having particular learning needs, but that ratio fell to 1 in 17 in the experimental group. Similarly, 1 in 5 of the control group children was referred to special education classes, where only 1 in 71 was so referred in the experimental group of children.

Bergan et al. concluded that children's capacities were under-developed in conventional teaching practices and that the result of this that many children would have unnecessarily disadvantaged futures. In citing this study, Black and Wiliam concluded that the study was an example of how rigorously implemented formative assessments may empower teachers to make better educational decisions. Bergan et al.'s work stands as an important reference foreshadowing the development of the AFL approach, depending, as it does, on frequent and carefully constructed formative assessments to foster learning gains. In addition, this study supported the premise put forth by Black and Wiliam, that formative assessment practices could positively influence learning in weak as well as in strong learners.

2.3.9 Black and Wiliam's contribution summarized

Black and Wiliam (1998a) drew on the work of Crooks (1988) and Black (1993), and found four major weaknesses in conventional classroom practice, which were that:

- CBA was geared towards superficial and rote learning;
- teachers did not review assessment questions or consult sufficiently with peers, leading to a lack of *reflection* in their assessments;
- grading was given precedence over learning;
- a normative over a criterion approach to assessment was favoured (Black & Wiliam, 1998a, p. 17).

Black and Wiliam (1998a) cautioned that ecological validity would be necessary in order to draw inferences to actual classroom experience, however they asserted the field could learn much from the kind of studies they cited.

They also advised caution when research designs included situations in which "those teaching any experimental groups are not the same teachers as those for any control groups" (p. 16). Black & Wiliam found the consistent feature influencing learning gains across the studies surveyed, was attention to formative assessment. Additionally, they declared that assessment qualifies as formative only if feedback information has been used. Following the publication of Black and Wiliam's (1998a; 1998b) review, the study became instrumental in changing British national educational policy towards one of classroom based formative assessment and AFL.

Other researcher has also influenced the re-focussing of CBA towards more formative functions, some of which will be reviewed in the following sections.

2.4 Other research contributions to AFL

Tunstall and Gipps (1996) analyzed the feedback of eight teachers in eight primary school classes in six schools of five local education authorities (LEAs) in London. Their research objective was to collect evidence of individual, whole class and across the curriculum feedback, by means of interviews, classroom observations and field notes.

Tunstall and Gipps (1996) identified two principal types of feedback as either *evaluative* (including rewarding, punishing, approving and disapproving) or *descriptive* (including specifying attainment, specifying improvement, constructing achievement, and constructing progress). They suggested feedback changed in style, purpose, meaning and processes when moving from evaluation to description and they suggested that their typology, widely cited following the publication of this work, could inform assessment in support of learning.

As did Black and Wiliam (1998a, 1998b), many (Black, 2005; Blumfeld, 1992; Cowie, 2005; James & Gipps, 1998; Leung, 2005; McNamara, 1996) in general education *and* second language education have come to view assessment as an essential component of learning. Others (Ali & Ho, 2007) from diverse fields have added to these calls to "optimise learning opportunities through assessment…[with assessment]…an integral part of course design" (p. 273).

McNamara observed the change towards using assessment for learning purposes, in noting that it "represents a radical challenge to the tradition of institutional language testing: it is not about measurement; it does not serve administrative functions; it does not adhere to psychometric canons. Its aim is to improve pedagogy" (cited in Hinkel, 2005, p. 778). Others (Black, 2005; Blumfeld, 1992; Cowie, 2005; James & Gipps, 1998: Leung, 2005) have echoed his sentiment, resulting in the refinement of formative assessment practices as AFL, the cornerstone of CBA whose raison d'être is to drive learning.

At this point in the discussion AFL will be examined in greater depth, to arrive at a better understanding of the approach.

2.5 AFL characterized

AFL is essentially formative assessment used in a dynamic, interactive, cooperative and learner-teacher-focused approach. It has learner motivation as its cornerstone. The Qualifications and Curriculum Authority (QCA, 2007) of Great Britain have defined AFL as "…using assessment in the classroom to raise pupils' achievement [ ] based on the idea that pupils will improve most if they understand the aim of their learning, where they are in relation to this aim and how they can achieve the aim (or close the gap in their knowledge)" (p.1). Simply put, AFL seeks to help teachers better identify what learners know and what they need to learn.

AFL proposes teachers ask questions of learners geared to provoke a thoughtful attitude on the part of learners. For example, its proponents recommend teachers ask "What do you think I want you to learn from this lesson?" in order to stimulate curiosity and enhance learner autonomy and responsibility for learning (Harlen & Winter, 2004). In essence, the AFL methodology proposes using effective questioning techniques and feedback, sharing learning goals, and employing learner self- and peer-assessments. This questioning technique resonated in Pellegrino et al. (2001), who encouraged teachers to rethink established ways of assessing competence. These authors used the following illustration of two lines of teacher-learner questioning to assess competence, the first discouraged and the second highly recommended:

Assessment #1

Question: What was the date of the battle of the Spanish Armada?

Answer: 1588 [correct].

Question: What can you tell me about what this meant?

Answer: Not much. It was one of the dates I memorized for the exam. Want to hear the others?

Assessment #2

Question: What was the date of the battle of the Spanish Armada?

Answer: It must have been around 1590.

Question: Why do you say that?

Answer: I know the English began to settle in Virginia just after 1600, not sure of the exact date. They wouldn't have dared start overseas explorations if Spain still had control of the seas. It would take a little while to get expeditions organized, so England must have gained supremacy somewhere in the late 1500s. (p. 28)

This is a compelling example of effective teacher assessment by means of an AFL guided questioning technique, such as is recommended throughout this book. The exchange shows that the *why* question form elicited a more effective response than did the *what* question format. However, part of the reason for that is dependent on the kind of answer provided by the learner in each case. The second response lent itself very well to an illustration of an effective teacher questioning episode, but unfortunately this does not always occur in day-to-day classroom reality. Nonetheless, asking learners *Why do you say that?* as opposed to *What can you tell me about what this meant?* would, by its open structure be more likely to effect a richer response. Furthermore, *why* questions do generally oblige the interlocutor to more considered responses than do other question formats (although both questions could potentially have been answered by that bane of a teacher-learner catechism: *"I don't know.")*

Pellegrino et al. (2001) have provided an informative overview of recent and proposed innovations, and have made recommendation for CBA in support of learning. Their work is informative concerning various aspects of formative assessment in classroom use, among which are the timing and effectiveness of guided teacher questioning, the alignment of curriculum and

instruction, the quality of teacher feedback, as well as learner and teacher roles. The authors also discuss Cognitively Guided Instruction (CGI), an approach to "continuous formative assessment … that borrows from cognitive science to characterize the semantic structure of word problems, along with typical strategies children use for their solution" (p. 231).

AFL has been shown to be a powerful learner motivator positively affecting achievement in mathematics, science and other subject matters (Harlen & Winter, 2004; Rea-Dickins & Gardner, 2000; Weeden & Winter, 1999, 2000). Essentially, AFL is interested in (1) how classroom learning is monitored, (2) on the perceptions of students and teachers regarding assessment, and (3) on how both (1) and (2) may enhance the learning process.

2.5.1 The 10 principles of AFL

AFL may be deemed a mostly learner-directed approach. This is illustrated in the ARG's underlying ten AFL principles, which are:

1. *AFL should be part of effective planning of teaching and learning.*
2. *AFL should focus on how students learn.*
3. *AFL should be recognized as central to classroom practice.*
4. *AFL should be regarded as a key professional skill for teachers.*
5. *AFL should be sensitive and constructive because any assessment has an emotional impact*.
6. *AFL should take account of the importance of learner motivation.*
7. *AFL should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed.*
8. *Learners should receive constructive guidance about how to improve.*
9. *AFL develops learners' capacity for self-assessment so that they can become reflective and self-managing.*
10. *AFL should recognize the full range of achievements of all learners.* (ARG, p. 1)

The teaching and learning philosophy of the ten principles will be explained in greater detail in the following:

1. *AFL should be part of effective planning of teaching and learning*. Teacher planning should a) allow learners to be familiarized with the criteria and goals of assessment, b) include opportunities for learners and teachers to obtain and use information about learner progress to help learners reach their goals, c) be flexible in order to accommodate emerging skills, and d) incorporate learner strategies to help them accomplish this.
2. *AFL should focus on how students learn*. Learners should be aware of *how* they are learning as well as *what.* In addition, teachers and learners should a) be mindful of this process when assessment is planned, and b) when assessment outcomes are interpreted.
3. *AFL should be recognized as central to classroom practice.* In classroom teaching practice, assigned tasks and questions prompting learners and encouraging demonstration of knowledge, understanding and skills, and the judgements teachers make of these learner responses (and the resulting decision-making), can all be described as assessment.
4. *AFL should be regarded as a key professional skill for teachers.* Formal teacher training should consist of pre-service and ongoing instruction in assessment practices including the knowledge and skills required to a) plan for assessment, b) observe and analyze evidence of learning, c) give feedback to learners, and d) support learners in self-assessment.
5. *AFL should be sensitive and constructive because any assessment has an emotional impact.* Teachers should be aware of the potential negative impact of grades; accordingly, their feedback should a) be as constructive as possible, and b) focus on the work and not on the individual learner in order to foster both positive motivation and learning.
6. *AFL should take account of the importance of learner motivation.* Feedback that includes comparing some learners to more successful

learners is likely to a) demotivate learners, and b) lead to learner withdrawal from the learning process due to feelings of low self-esteem. Rather, positively motivating feedback that encourages learning through a focus on progress and achievement, is constructive and provides learners with choice and self-direction (to "protect learner autonomy"), and is recommended.

7. *AFL should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed.* It is considered essential for effective learning that learners play a part in deciding learning goals and identifying criteria for progress toward these goals so as to develop both an understanding of and a commitment to their own learning. Teachers should communicate assessment criteria to learners by a) discussing them with learners using appropriate language, b) using illustrative examples, and c) encouraging learner self- and peer-assessments.

8. *Learners should receive constructive guidance about how to improve*. In giving feedback, teachers should a) identify learner strengths and advise how to develop them, b) be clear and constructive regarding weaknesses, and c) provide opportunity for improvement to occur. This is intended to help guide learners in the planning of their next steps in their progress.

9. *AFL develops learners' capacity for self-assessment so that they can become reflective and self-managing.* AFL suggests "independent learners have the ability to seek out and gain new skills, new knowledge and new understandings … to engage in self-reflection and to identify the next steps in their learning" (p.2). In view of this, teachers are encouraged to foster learner desire for, and the required skills for achieving learner autonomy, achievement, as well as best-level learner accomplishment.

10. *AFL should recognize the full range of achievements of all learners*. AFL proposes its methodology be used in all learning opportunities and contexts, (presumably in all subjects), and it seeks to promote

recognition of learner achievement as well as best-level learner accomplishment.

2.6 AFL approaches in classroom application

In order to understand how to develop my study of L2 AFL, it was necessary to go back to examining how AFL has functioned in classroom settings. AFL was first implemented in general education, so this will be discussed first. A strong trend towards learner self-assessment is reflected in the studies reported. A comparative study of how AFL has been applied in a mathematics and a English language arts classroom follows, ending with a discussion of recent research, which will help to answer the question of how AFL might best be applied in an L2 classroom setting.

2.6.1 AFL approaches applied in general education.

Weeden & Winter (1999) conducted Phase 1 of *LEARN, Learners' Expectations of Assessment for Learning Nationally*, a project of the Culture and Learning in Organisations (CLIO) Centre for Assessment Studies of the University of Bristol, and reported to the British Qualifications and Curriculum Authority (QCA). Interviews were conducted with more than 200 children from school years 3-13. Initially, a semi-structured interview was trialled in three schools. The interview's intent was to explore students' perceptions of their own learning processes, focusing on mathematics, English (L1) and science subject matters. Following the trials, the interview content was modified to reflect four areas of interest, which were *learners*, *expectations*, *self-assessment*, and *feedback*. After that, interviews were conducted on more than 200 students from 20 diversely socioeconomic school contexts, in southwest England. Weeden & Winter (1999) used interview questions for example, in the *expectations* section, such as "What did you do? What did the teacher tell you to do? Did you know what the teacher wanted you to do before you did the work?" (p. 14).

Secondly, in each of the 20 schools and for each year level represented, teachers were asked to identify six students of differing ability levels for further interviews. These students were interviewed for about half an hour each, while field notes were made and the interviews audio-recorded. The

students were asked to bring to the interview samples of recent assessed and unassessed work. The students were asked to choose among five possible choices of statements which best described their feelings regarding 1) learning and 2) assessment. An example among the choices of the latter was "I like it best when the feedback suggests how I can improve my work" (p. 18). Perhaps not surprisingly, most students (58%) chose this statement as representative of their views, and the majority increased with the ages of the children.

The results showed there was little self assessment in the school programme and most students understood assessment to mean summative assessment. Feedback quality was negatively perceived by most students; there was evidence that "much feedback was either unfocused or of little use in improving work" (p. 11). The report recommended teachers be advised on how to 1) motivate students, and in particular those with low self-esteem, through assessment, 2) communicate to students what is expected of them as well as programme learning requirements, 3) develop opportunities for self- and peer-assessments, and 4) improve feedback quality.

Triggs, Weeden, Winter & Broadfoot (2000) conducted Phase 2 of the *LEARN, Learners' Expectations of Assessment for Learning Nationally* project. This phase sought to 1) develop students' understanding of the goals of learning, 2) develop student self-assessment skills, and 3) close a perceived gap in performance and desired standards.

Two year 1 and three year 5 classes in two primary schools were investigated over the space of one term. In the two schools, the five classes were made up of 18-31 students each, and the total student participants in the schools equalled 50 and 68 respectively. In each class, the teacher participants identified three above-average and three low-to-average attainment-level students, with the result that these 45 students became the focus of data collection. A symbol system was introduced to the teachers and students to help codify self-assessments. Project manager meetings with the schools numbered two days. The data consisted of students' views taken from conference interviews, from student work, and teacher research diaries.

Triggs et al. (2000) reported six principal findings. They were that 1) students were initially positive about the symbol system but appeared to lose interest with time; 2) students sometimes had difficulty distinguishing between

amount of effort and the successful or unsuccessful attainment of learning objectives; 3) the year 5 students were the most positive about having an opportunity to share their views on their efforts, 4) students generally were confused about task assignment and learning objectives ('what I have to do: what I have to learn'); 5) in self-assessments students generally showed a marked preference for assessment criteria of neatness, penmanship, quantity produced and task completion, while the higher-attaining students tended to refer to learning intentions; 6) in teacher-student conferences there was much evidence that students were able to understand corrections and improvements where they had been unable to do so alone; and 7) students did not appear to be making any use of written teacher feedback.

Triggs et al. (2000) made several recommendations about the use of feedback, for example that time for "good and useful written feedback" be allotted in the classroom (p. 5). In addition, they called for the classroom use of more articulated and specific feedback than that intended solely for student motivational purposes.

Brookhart (2001) examined 50 academically successful Grades 10 and 11 American high school students of (L1) English and Anatomy and found these students regularly used on-going self-assessment practices. Brookhart (2005) reviewed over 100 largely theoretical studies on the development and practice of formative assessment. She found that over time many studies of formative assessment had engendered an evolved conceptualisation of formative assessment, as in the following definition:

> Expanding concepts in the definition of formative assessment: Information about the learning process (Scriven, 1967); that teachers can use for instructional decisions (Bloom et al., 1971); and students can use for improving their own performance (Sadler, 1983, 1989); which motivates students (Black & Wiliam, 1998; Brookhart, 1997; Crooks, 1988; Natriello, 1987). (p. 1)

In the conclusions to her theoretical overview, Brookhart (2005) called for more research into three areas yet to be investigated. They were 1) validity and reliability theory applied to classroom-based formative assessment; 2) specific ways to improve teacher assessments; and 3) specific ways to improve student (particularly self-) assessments.

Ross, Hogaboam-Gray & Rolheiser (2002) examined the effect of self-evaluations on mathematics achievement in 259 Grades 5-6 students in Ontario over a treatment period of 12 weeks. A control group of students numbered 257 participants. The students were given a performance pre- and post-test. Teacher instruments found no significant difference between participant treatment and control group teachers in terms of their efficacy, self-reported assessment practices and beliefs about teaching. The treatment teachers attended four brief teacher training/team meetings. In six 30-minute lessons, the treatment students were taught self-assessment techniques. This instruction followed the authors' theoretical framework incorporating the following self-assessment elements: self-observation, self-judgement and self-reaction. In addition, there were four instructional stages intended to 1) involve students in defining the criteria of assessment, 2) teach students how to apply them, 3) give students feedback on their self-assessments, and 4) help students use assessment data to develop action plans.

The results showed that the treatment students outperformed the control group, with an effect size of .40. The authors concluded self-evaluation in mathematics had a positive effect on mathematics achievement. This finding lends support to an AFL approach to classroom assessment and learning.

In order to better understand a learner perspective of assessment, Brookhart & Bronowicz (2003) investigated 161 students' assignment of interest and importance to tasks, efficacy, and goal orientations in seven elementary and high school classes in four American schools. Semi-directed interviews were administered to the students. Brookhart & Bronowicz (2003) found developmental differences in student comments about studying, and they suggested there may be a developmental progression in learner understanding of academic success. However, the results showed there were more similarities than differences in student perspectives surrounding assessment; assessment perceptions were found to reflect personal connections with assessment and the consequences of assessments.

McDonald & Boud (2003) conducted a large-scale study involving a quarter of the final year high school students of one year in Barbados. A treatment group of 256 students were trained in self-assessment techniques for use in general curriculum subjects over the academic year. A matched group

of classes served as a control group. The treatment group students were chosen from 10 high-, middle- and low-achieving schools in order to incorporate students of all ability levels. The results showed significant differences overall and in each subject matter for those students trained in self-assessment however, McDonald & Boud (2003) questioned whether these results might be repeated in less favourable conditions.

The AFL approach has been adopted in state educational departments in England, Northern Ireland, and Wales, as well as in Scotland, Australia, New Zealand and in Hong Kong. The AFL influence has been felt in the U.S. as well, and in some Canadian provinces (Alberta Education, 2008; Krakow, 2005; Legendre, 2001; Ministère de l'Education du Québec, 2003; Western and Northern Canadian Protocol for Collaboration in Education, 2006). In the United States, the National Council of Teachers of Mathematics (NCTM) has supported the AFL framework of Wiliam and Thompson (as cited in Wiliam, 2007, p. 1), for application in teaching mathematics. At this point, the ways in which these AFL applications have been realized, will be examined.

Maxwell (as cited in Harlen, 2005) reported on an approach to CBA in Queensland, Australia. The assessment was called 'progressive assessment,' and consisted of formative assessments given over time in a portfolio assignment, which contributed to later summative assessments. Significant in this example is the fact that teachers were free to plan their own assessments, while the formal leaving Senior Certificates were separately administered by state officials. (A similar situation, in which teachers were 'unencumbered' by leaving examination administration, was approvingly reported by Black (2004) in the case of the French school system.)

Similarly, in a British study, Harlen (2005), found the approach of using the same evidence for both formative and summative purposes, to be useful in promoting what she termed "the synergy of formative and summative assessment" (p. 220). She called for practice in which a teacher should be the person responsible for using the evidence for both kinds of assessments. According to Harlen, "The procedures that will most help both the effectiveness of formative assessment and the reliability of summative assessment are those that involve teachers in planning assessment and developing criteria" (p. 221). Of interest to the present discussion, she also

declared “If we fuse, or confuse, formative and summative purposes, experience strongly suggests that ‘good assessment’ will mean good assessment *of* learning, not *for* learning” (p. 220).

2.6.4 AFL in mathematics and science classrooms

Wiliam, Lee, Harrison & Black (2004) investigated the effect on achievement of introducing formative assessment in mathematics and science classes in England. Twenty-four teachers participated in the study, and each had a class size of approximately 24 students. Each treatment class was matched with a control group class. Pre-tests were administered. The treatment group teachers were trained in formative assessment techniques and strategies over a six-month period prior to their using these methods in class. For example, self-assessment strategies included students’ use of red, amber or green ‘traffic lights’ to show when material had been understood and the learner could move on, as well as information for students such as “One of these twenty answers is wrong: find it and fix it!” (Wiliam et al., p. 55). The results showed a mean effect size of .32 in favour of the treatment.

Harlen and Winter (2004) also reported on the application of the AFL approach in mathematics and science classrooms in schools in England. Their overview was highly detailed, underscoring CBA practices in that setting. Thus, it included an AFL framework for ‘quality teacher assessment’ including 1) information gathering techniques, 2) using progress indicators such as progress maps, (also known as ‘developmental indicators’), 3) questioning and listening strategies, (the former intended to elicit thoughtful responses), 4) effective feedback use, and making sure learners understand and share teaching goals with teachers, and 6) self- and peer-assessment techniques. They concluded AFL was a powerful learning tool, particularly in the case of low-achieving learners.

All of the above research points to favourable learning outcomes in many contexts when classroom-based AFL was introduced in general education. One question that remains is, would these results translate into success in L2 classroom settings as well? To investigate that question further, the discussion will consider a comparison study of AFL in mathematics and English arts

classroom settings, in order to more fully understand how AFL was applied there, and the relative success of the applications in each context.

2.6.5 AFL in mathematics and in English arts: A study in contrasts.

In illustration of how AFL in the exact sciences might be applied to the liberal arts, and hence from there to the L2 classroom, Hodgen & Marshall (2005) studied and compared how AFL and formative assessment operated in a mathematics and in an English literature class. Moreover, some parallels can be drawn from the teaching approach used in an English literature class, to that of the second language classroom, thus informing the SLA and LT fields.

Hodgen and Marshall (2005) noted similarities in the AFL applications of the two disciplines in their use of paired, group work and in whole class discussions; in scaffolding knowledge; and in providing formative assessment in "rich and challenging activities".

They found important differences in application as well. The mathematics classroom employed a discrete focus, and used AFL and formative assessment to draw learner attention to an underlying mathematical misconception, which was useful in then bringing learner attention to the desired, correct mathematical application.

On the other hand, in the English literature classroom, a broader approach was taken to AFL and formative assessment, through both directed and freely creative activities, subtly directed learners to a desired point of view. Hodgen and Marshall concluded that both the cognitive (math class) and socio-cultural (English class) approaches can learn from each other. Interestingly, as in the findings of Black & Wiliam (1998a), they found the pedagogical excellence of the teachers appeared to be an influential determiner of the successful application of AFL and formative assessment in these classrooms.

2.6.6 AFL in second language education.

This leads to a review of some of the few studies of formative assessment and an AFL approach that have been reported in L2 classroom based research. In a three-year study of L2 learning, Cheng, Rogers, & Hu (2004) investigated the CBA methods of 267 teachers in Canadian (ESL)[3], Hong Kong (ESL/EFL)

[3] ESL refers to English as a second language and EFL to English as a foreign language.

and Chinese (EFL) settings, and asked the research question, "What assessment purposes, methods and procedures do ESL/EFL instructors report using in their student assessment in three different tertiary contexts?" (p. 363).[4]

The instructors were surveyed by means of a questionnaire intended to elicit their views on the purposes (student-centred, instructional or administrative), methods (*instructor-made*, *student-conducted* or *non-instructor developed*), and procedures (*sources of assessment, methods of providing feedback and reporting*, and *time spent*) relative to assessment and evaluation. The questionnaires had previously been piloted on a small sample of teachers from these regions. Following trials, the actual study questionnaire responses numbered 95 from Canada, 44 from Hong Kong, and 124 in Beijing, for a sum total of 263. The means of class size were 15 students in Canada, 19 in Hong Kong and 45 in Beijing,

Analysis of the questionnaire data showed the reported methods of providing feedback and reporting were not significantly different across the regions. Cheng et al. (2004a) however, did find some differences in instruction in the three situations. They interpreted the results in the following way:

> The differences in instruction we found in the three settings are derived from the nature of the courses themselves, the teaching experiences of ESL/EFL instructors, their knowledge in assessment, the needs and levels of the students, the teaching and learning environment (e.g. size of the classes), and the role and impact of external testing on teaching and learning. (p. 378)

In the three settings the overall pattern of differences was between Beijing and Canada/Hong Kong, or Hong Kong and Canada/Beijing. For example, the instructors in Hong Kong reported using significantly fewer student-centred, instruction-centred, and administration-based purposes for assessments and

[4] Cheng et al. differentiate between *assessment* and *evaluation*, as follows "…*assessment* is defined as the process of collecting information about a student to aid in decision-making about the progress and language development of the student. *Evaluation* is defined as the interpretation of assessment results that describes the worth or merit of a student's performance in relation to a set of learner expectation or standards of performance" (italics added, p. 363).

used significantly fewer objective methods of assessment methods than did those in Canada and Beijing. Cheng et al. (2004a) attributed this finding to the reported higher level of teaching experience and better qualifications of the Hong Kong teachers, and their greater knowledge of assessment, as well as the fact that in Hong Kong classroom assessments were not influenced by the presence of external testing, as was the case in Beijing where students were required to succeed in the College English Test (CET) prior to university graduation. Cheng et al. concluded that assessment and evaluation across tertiary-level ESL/EFL courses is complex and variable across different settings. This study argues for the need for teacher training in assessment methods such as AFL, engineered to help drive student learning.

Leung (2004) examined and discussed the conceptualization of classroom-based formative assessment. Drawing on evidence in Rea-Dickins (2001), he supported her view that formative assessment occurs along a continuum from formal to the informal applications. This is in contrast to what Leung terms the somewhat idealized vision of formative assessment promoted by the British QCA (2001).

Leung (2005) overviewed the construct and practice of classroom-based teacher assessment, (which he observed has also been characterized as *alternative, authentic, classroom, educational, formative, informal* or *statutory*) and concluded the following:

> …one should not assume that just because teacher assessment takes place in natural and real classroom contexts, it will automatically yield pedagogically useful information or lead to higher quality learning or teaching…The research agenda for teacher assessment, seen in this light, should investigate what teachers look for and do when they carry out classroom assessment, how they use the information yielded for teaching purposes, and how this kind of research knowledge can be used in professional development. (p. 885)

Rea-Dickins and Gardner (2000) conducted a case study of nine inner-city schools in England where learners spoke *English as an Additional Language*

(EAL)[5] and where low levels of English achievement had been encountered. Their work was part of the state-sponsored *Early Years Intervention* project. Their research objectives were to 1) determine the kind of assessment practices in use, 2) what was considered the most important issues in both academic and ESL achievement by those professionals who acted as support for the language learning process, and 3) assessment representations and how assessment supported curricular decision-making and language learning in classroom practice.

The study consisted of four parts. First, 29 teachers, educational assistants and language support co-ordinators completed short, open-ended questionnaires to elicit their perceptions of learner assessment. The questionnaire results informed the rest of the study. Secondly, classroom language assessment was quantified; thirdly, the questionnaires were followed-up by 10 interviews of 28 people, and fourthly, there were classroom observations in three schools and further teacher and professional interviews.

The authors concluded that while CBA has generally been regarded as low-stakes, they found crucially important instances of teacher assessment based on decisions regarding learner performances in class. They also questioned the adequacy of teacher training relative to the practice. Rea-Dickins and Gardner found their study raised many more questions for further research. Among their queries, they asked "What does it mean to sample comprehensively a learner's language development over time?…If teachers are required to monitor the language development of their learners, then how do they acquire the skills to do this?" (p. 239). These findings have implications for both pre-service and ongoing and teacher training in valid and reliable educational decision-making.

Rea-Dickins, (2001) conducting research in England, used teacher interviews, classroom observations, audio- and video-recordings and lesson transcripts to investigate CBA in an ESL setting. In this descriptive study, she

[5] In this article and elsewhere, Rea-Dickins has equated the British term, EAL with the North American term, ESL. Hence in the interests of consistency I will hereafter refer to this subject matter as ESL.

set out to understand different classroom assessment requirements, practices and purposes.

In the discussion of the data analysis, Rea-Dickins (2001) termed the following three assessment 'identities': 1) *bureaucratic*, to fulfil obligations to an external agency/ies; 2) *pedagogic*, to inform teachers and support professionals on learner progress; and 3) *learning*, to support as opposed to measuring learning. Rea-Dickins qualified the latter in the following way:

> Good 'assessment for learning' thus motivates learners to become engaged in the interaction through which they are enabled to develop skills of reflection (as a basis for self- and peer-monitoring), as well as providing them with an ability to reflect meta-cognitively on their own learning. (p. 452)

In the conclusions of the above study, Rea-Dickins argued that formative assessment in the classroom required further detailed analysis, and proposed questions to be answered in future research should include a) what constitutes quality in formative assessment, b) analysis of whether formative assessment creates opportunities for learning to take place, c) what evidence there is of language learning, and d) whether ESL teachers can differentiate between learning, special educational and curriculum content needs in the classroom.

Rea-Dickins (2006a) investigated interaction in assessment in order to examine the learner's role and the nature of teacher scaffolding as well as orientations in the assessment process. Primary school students were tracked in classroom observations for one week in each of three terms of the school year. Ecological validity was enhanced by observing students over a full week, in all their school subjects while in fact focussing on literacy and science lessons.

Data was collected using classroom observations including audio- and video-recordings; field notes; pre- and post-assessment interviews with language-support teachers; and post-lesson observation interviews with classroom teachers. Five assessments were recorded and subsequently transcribed. The participants were two language-support teachers, one mainstream class teacher and their classes of about 24 students each, aged 6-7.

In addition to this, two of the learners in each class observed were observed over a longitudinal, 12-month period.

Analyses suggested that different teacher orientations, regarding a formative or summative aspect, occurred within assessment instances and might unwittingly limit learning potential. Rea-Dickins (2006a) also situated the interaction of instruction-embedded assessment at the juncture of LT and SLA, a concept resonated in Bachman (2006) as the "interface of LT and SLA" and in Colby-Kelly and Turner (2006; 2007) as the "assessment bridge." Rea-Dickins here cited an example of this in a literacy lesson wherein "…through a range of different teacher prompts and learner initiations, illustrates an orientation towards developing learner language awareness, understanding and knowledge, as opposed to measuring language knowledge" (p. 179).

Finally, Rea-Dickins (2006a) raised the following questions for further research investigation:

- *How exactly* do learners get to know what they need to know about *what* is being assessed, and *why* and *how* they are being assessed and graded? (cr. Clarke 1998)
- How do they know or how do they learn what it is they have to do in order to achieve and adequate or good performance? (cf. Gipps 1994)
- How do they know which criteria teachers are using on different occasions for different assessments?
- How do learners actually interpret teacher feedback on their performance? (p. 182).

Rea-Dickins (2006b) took a sociocultural approach to case studies of three primary schools in England. The participants were 6-7 year-old students in six classes. The data collection included 3 complete weeks of classroom observations in each of the three schools over the period of the school year, audio- and video-recordings, pre- and post-observation interviews, pre- and

post-assessment activity questionnaires as well as teacher workshops and interviews, and samples taken of learners' written work.

In her conclusions, Rea-Dickins (2006b) found teachers have been called to support language development while measuring attainment, supporting subject learning and measuring subject learning, and she characterized teachers' conflicting roles as a tug-of-war. Others (Rea-Dickins, 2001; McNamara, 2001b; Brindley, 2001) have also reported conflicting teacher roles in assessment. Leung and Rea-Dickins (2007) found the orientation pull in different directions to occur in the absence of adequate policy and teacher training in England.

In a recent adoption of the AFL approach in L2 classrooms, Davison (2007) and Hamp-Lyons (1999, 2006) discussed a 2002 change of policy in the Hong Kong Education and Manpower Bureau, from a traditional focus on summative to formative assessment and AFL, called school-based assessment (SBA). Davison reported that when SBA was implemented, teachers and students surveyed expressed positive impressions of the approach, citing the increased opportunities for using information from assessments to help in learning (p. 54).

The L2 studies cited up to this point all report on a trend in LT towards the inclusion and application of more learner-centered, formative, and AFL-styled approaches to classroom-based instruction and assessment.

2.7 AFL and the classroom interface of assessment and learning

For the considerations of pedagogical instrument development, it is necessary at this point to refine some of the ideas presented pertaining to learning and AFL, with a view to determining a sound basis for interpreting AFL in an L2 classroom setting.

2.7.1 The Assessment Bridge in the classroom

As noted earlier in this discussion, the AB is understood to be the area of CBA encompassing *assessment* (where learners *are* in their learning), *teaching* (where they need to go and how best to get there) and *learning* (action on the part of the learner). In other words, the AB is the place where language assessment and second language acquisition interface, with the hoped-for

result of learning (Colby-Kelly & Turner, 2007, p. 12). Rea-Dickins (2006a) alluded to this idea when she discussed the embedded nature of formative assessment activities in the classroom setting, and underscored the element of spontaneity in the AB, noting "…it is the very nature of formative assessment that the learners' language development is mediated through teacher feedback in the ongoing interaction, most of which can be neither planned in advance" (p. 167). Further, Rea-Dickins (2006a) reported findings which can be said to be in favour of the usefulness of the AB: "The analysis suggest that there are different teacher orientations within assessment and highlight the potential that assessment dialogues might offer for assessment as a resource for language learning, thus situating this work at the interface between assessment and second language acquisition" (p. 163).

In a baseline study of formative assessment practices, Colby-Kelly and Turner (2007) investigated "how classroom learning is monitored, how students and teachers perceive assessment, how the two may be useful and enhance learning, and how assessment, curriculum, teaching, and learning link to form an assessment bridge" (p. 32). They reported on evidence supporting observed teacher-engagement in the AB, and also on teacher justification for ongoing assessment as a requirement for learner progress.

Given that the consequences of teacher feedback is not always readily discernable to teachers, Colby-Kelly and Turner (2006; 2007) applied a validity argument in the form of an AUA used to determine if there was evidence that teacher feedback in AB instances benefitted learning. Through student interview data they found teacher feedback did not always have the positive effect intended in the absence of a learning component in the feedback. Their findings point to the need for a more thorough understanding of the effects of feedback, and echo objectives of AFL, which seek a deeper understanding of the effects of feedback and of formative assessment in general (see Pellegrino et al., 2001 for an example of teacher feedback adjusted to fit an AFL bent in order to enhance learning in a class discussion, p. 226-227).

In fact in recent time many calls have been made for an alignment of curriculum, instruction and assessment in classroom settings (Black, 2004, 2005; Black & Wiliam, 1998a, 1998b, 2003, 2005b; Brookhart, 2003; Colby-

Kelly, 2008, 2009a, 2009b, 2010; Colby-Kelly & Turner, 2006, 2007; James, 2000, 2005, 2006; Leung, 2004, 2007; Leung & Mohan, 2004; Solomon, 2002; Pellegrino et al. 2001; Turner, 2009).

Moreover, in view of Turner's (2009) discussion of the influence of washback on CBA, in ideal settings, one could even posit that positive washback may be viewed as a precursor to the AB. A discussion of washback is beyond the realm of this review, however for a thorough discussion of washback, see Cheng et al. (2004b).

2.7.2 'Noticing'.

The discussion will now focus on how 'noticing' may play into the pedagogy of an AFL classroom setting.

Many researchers of SLA have agreed that learners' *noticing* of linguistic forms is necessary for acquisition to occur (Schmidt, 1990; Robinson, 1995). According to Swain and Lapkin "…a learner will on occasion become aware of (i.e., notice) a linguistic problem (brought to his/her attention either by external feedback (e.g., clarification requests) or internal feedback. Noticing a problem 'pushes' the learner to modify his/her output" (p. 372). More recently, Swain (2001) stated "…I am engaged in reworking the notion of output to incorporate it within a view that focuses on language learning and use as dialogue – dialogue with the self – serving both communicative and cognitive functions" (p. 279).

It is noteworthy that formative assessment, as defined by the proponents of AFL (determining where learners are in their learning and where they want to go), incorporates the SLA notion of fostering acquisition through learners' 'noticing' the difference between their interlanguage (IL) and target language production. Thus, a large thrust of the AFL approach involves helping learners become cognizant of the gap between learner and target performance. In fact, Fulcher (2010) characterized AFL, noting that in the AFL classroom, "all assessment is descriptive rather than evaluative, helping learners to 'notice the gap'" (p. 1). Thus, in developing AFL pedagogical materials for an L2 classroom setting, I felt that a component intended to foster learner noticing of that gap would be essential.

2.8 Validity and reliability concerns of AFL approaches

Messick (1989) has defined validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment (p. 13, italics in original). In a discussion of washback, Messick (1966) held that washback "refers to the extent to which the introducation and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning" (p. 241). In light of Messick's definition, we can see positive washback as a precursor to AFL in its use of assessment to drive and positively influence learning. In a discussion of washback, Turner (2009), recalled Messick's (1996) caveat, noting "…the effects are only washback evidence if they can be linked to the introduction and use of the test" (p. 56).

Thus to move on, validity can be said to refer to whether an assessment instrument or method is measuring what it purports to measure. On the other hand reliability is concerned with knowing if the same results would be found if replicated. Proponents of alternative assessments have maintained that AFL and formative assessments must be inherently valid since they look at learner performance on authentic tasks, and are reliable because teachers engage in repeated measures (Huerta-Macías, 1995).

Others (Black and Wiliam 2005; Brown & Hudson, 1998; Harlen, 2005) have disagreed with these claims, finding them too limited, particularly in view of high-stakes assessments affecting students' lives. Brown and Hudson asserted "precedents exist for clearly demonstrating the reliability and validity of such procedures in the long-extant performance assessment branch of the educational testing literature" (p. 656). Wilde, Del Vecchio and Gustke (cited in Huerta-Macías) proposed ensuring reliability in alternative assessments by 1) designing many tasks leading to the same outcome, 2) using trained raters, and 3) periodically monitoring raters. However, Brown and Hudson found these proposals to be too simplistic; instead they suggested ensuring validity and reliability by adapting established procedures in order to develop sound alternative assessment practices.

In addition, Messick (1996) cautioned that while authenticity and directness of assessment were widely assumed to create more positive consequences for teaching and learning (positive washback), these assumptions would require investigation before their truth could be asserted (p. 243). On a similar note, Norris, Brown, Hudson, & Yoshioka (1998) found "The issues of reliability and validity must be dealt with for "alternative assessments" just as they are for any alternative in assessment - in an open, honest, clear, demonstrable, and convincing way" (p. 5).

Another concern in alternative or formative assessment practices has been that while tasks have been designed for authenticity and therefore greater validity, less attention has been paid to reliability. This includes reliability between different teacher-raters (inter-rater reliability) and from individual teacher-rater's different assessment episodes (intra-rater reliability).

Rea-Dickins and Gardner (2000) investigated formative assessment and found it could be entangled with summative assessments, and as well they identified sources of inconsistency in its reliability and validity. They found "The validity of inferences made for individual learners depends upon their reliability, and we have shown that this reliability cannot be entirely assumed" (p. 238). Others (Harlen, 2005; Harlen & James, 1997; Rea-Dickins, 2006a) have also found formative and summative assessments in teacher practice were confounded.

These concerns have been under scrutiny for some time. Balliro (1993) sought to offset them by recommending teachers be trained to agree on what to look for in learner work and on how to analyze it, so that improved validity and reliability in assessments would result. However, more recently Leung (2005) took a newer approach to traditional concerns of validity and reliability, with reference to what he termed *construct-referenced assessment*, the logic of which he illustrated in the following:

> Instead of asking "How valid and reliable is teacher assessment?" as a form of testing, the questions are now "What do teachers look for in their classroom assessment?" and "How useful is the information produced by this kind of assessment for thinking about learning and teaching?" (p. 885)

In her landmark article about changing perceptions about validity in CBA, Moss (2003) observed that in classroom settings, validity must largely be

about consequences. Moss illustrated this, arguing that "Assuming interpretations are intended to inform instructional decisions and that instructional decisions entail interpretations about students' learning, it is on evidence of their (immediate, long-range, and cumulative) effects on which their validity primarily rests" (p. 19). Fulcher and Davidson (2007) have taken a similar view to using classroom-based formative assessment, noting the constraints of large-scale standardized testing don't apply in classroom situations. They argue that "the usefulness of assessment, the validity of interpretation of evidence, is meaningful only if it results in improved learning" (p. 35).

The need for a separate paradigm for assessments used to drive learning has been recently made as the requirements of large-scale psychometric testing have been seen to be in disaccord with the needs of CBA (Turner, in press). As such, the validity and reliability standards of psychometric testing has been seen to be inappropriate for application in classroom use of the sister practices of formative assessment, AFL and CBA.

2.9 Calls for L2 AFL and alternative assessment research

McNamara (cited in Hinkel, 2005) has noted the underrepresentation in LT of classroom assessment and has called for a research shift in that direction**.** At the same time the LT community has recently called for more research into advances in alternative assessment practices for application in L2 learning situations (Brookhart, 2005; Harlen & Winter, 2004; McNamara 2001a, 2001b; Pellegrino et al., 2001; Poehner and Lantolf 2005; Rea-Dickins 2004; Shohamy, 2004).

In a similar vein, Pellegrino et al. (2001) made twelve recommendations for research into assessment to drive learning and for a shift in focus towards more classroom-based formative assessment. Leung (2005) has argued that research into CBA "should investigate what teachers look for and do when they carry out classroom assessment, how they use the information yielded for teaching purposes, and how this kind of research knowledge can be used in professional development" (p. 885). In addition, Leung (2004) underscored the need for empirical research to adapt a socially co-constructed view of

formative teacher assessment, while proposing the following research questions:

1. Do we know enough about formative teacher assessment as practice?
2. Have we developed epistemologically and methodologically justifiable approaches to studying formative teacher assessment empirically?
3. Do we know how to ask questions about the soundness of teacher assessment in terms of meeting student learning needs in the classroom (as opposed to in theory)? Would we need to find different ways of asking questions about complex issues traditionally associated with formal testing such as reliability (e.g., rephrased as "can we rely on teacher assessment to identify learning in a consistent way?") and validity ("what is the basis of decision-making when teachers choose a focal point in their assessment?")
4. Are we taking enough notice of teacher practices in the way we characterize formative teacher assessment?
5. How can we learn form the work in teacher development with reference to formative teacher assessment? (p. 38)

In addition to all of these calls, Rea-Dickins (2004) has noted with accuracy that "assessment, with specific reference to teaching and learning in the language classroom, has remained, until recently, relatively unresearched" (p. 249). These words of Rea-Dickins have served to contribute to inspiring the focus of the present research. Other influences which have also been included in the final research design (see Chapter 3) will be discussed at this point.

2.10 Theory and research influencing L2 AFL in the present study

In addition, other kinds of formative assessment practices that have evolved in recent time are diagnostic assessment and dynamic assessment. These assessment practices will be discussed in the sections that follow, with a view to how they have influenced the present research.

2.10.1 AFL, diagnostic assessment, and the present study

In striving to identify where learners are in their learning, AFL reflects a facet of LT, diagnostic assessment, which has been defined by the *Association of Language Testers in Europe* (ALTE) as "A test which is used for the purpose of discovering a learner's specific strengths or weaknesses [the results of which] may be used in making decisions on future training, learning or teaching" (as cited in Alderson, 2005). Diagnostic testing has usually been used in large-scale testing, often for placement purposes, but recently diagnostic testing has taken on a more active role in formative learning, particularly in online self-assessment. In fact, Alderson characterized diagnostic tests as "The type of test that comes closest to being central to learning" (p. 4).

Online diagnostic assessments have the advantage over most classroom-based teacher assessments in that they generate immediate feedback. An example of a widely used online diagnostic test is *Dialang*, developed by a team headed by Charles Alderson in 1995 under the auspices of the European Commission, and available in most European languages. Alderson (2007) discussed diagnostic assessment in relation to CBA, noting:

> Until recently, there has been very little research into teacher-based assessment or any form of formative assessment in foreign language learning. However, Cheng et al. (2004a) show that teachers do indeed claim to use assessment to diagnose their learners' strengths and weaknesses, and thus looking at teachers' assessment practices could be a profitable way forward. (p.33)

Alderson's remarks lend credence to research such as the present study, seeking to better understand how assessment may be used to further learning.

In recent time a lively debate has recently ensued across the pages of a special issue of *Language Assessment Quarterly* amongst LT researchers (Alderson, 2010; Lee & Sawaki, 2009a, 2009b, 2009c, 2010; Sawaki, Kim, & Gentile, 2009, 2010; Jang, 2009, 2010) concerning the application in L2 learning of diagnostic testing, and a newer branch of it, Cognitive Diagnostic

Approaches (CDA), as well as Q-matrices[6]. In particular, Alderson (2010), while citing the special issue as a diagnostic assessment as a landmark, yet questioned the appropriacy of using proficiency tests for diagnostic purposes (p. 96). In answer, Jang (2010) agreed with Alderson's view that "the lack of an adequate theory that guides the development of diagnostic assessment or the technical complexity of CDM should not dissuade us from exploring the opportunity to problematize and expand research on CDA in fruitful directions" (p. 117).

Jang (2009) has summed up the discussion in a sense, in noting that cognitive diagnostic assessment provides learners with formative diagnostic assessment information about learner competencies in specific skill sets. A detailed examination of CDM can be found in the special issue of *Language Assessment Quarterly*, Volume 7, 2009. An in-depth discussion of this topic is beyond the scope of this review, having as its focus the AFL approach to CBA.

AFL focuses on knowing exactly where learners are in their learning in order to help them progress. Similarly, research into diagnostic assessment to situate learning has increased in recent time (Elder et al., 2007; Fox, 2008, 2009; Fox & Hartwick, in press; Read, 2008). For example, Read reports on a diagnostic EAP assessment tool for undergraduate students at a New Zealand university, which is used to identify problems and offer students remedial solutions, from the start of their academic studies. The test, the Diagnostic English Language Needs Assessment (DELNA), is given to both student speakers of English as an L1 and L2, in order to ensure uniformity of treatment within the student population. DELNA takes a positive approach, not unlike AFL in that the test has no punitive component affecting university admission, it is free to students, remedial help is encouraged and not mandated, and it has deliberately been termed an assessment, and not a test. Similarly, Elder et al. reported on a study aimed at improving inter-rater reliability in DELNA writing tasks.

In a study of students of EAP at a Canadian university, Fox (2009) found diagnostic testing to be an effective for instruction by focussing attention on

[6] A Q-matrix is "an interface to specify the construct being assessed with a set of operational cognitive skills" according to Jang (2009, p. 210).

learner needs and strengths. Fox and Harwick (in press) reported that diagnostic assessment can inform learner awareness of their learning needs, and thus increase motivation and learning potential.

Despite the above, more research is needed to garner a better understanding of how diagnostic assessments might be best employed in L2 classroom settings that apply an AFL methodology. Online diagnostic testing can offer learners immediate feedback. That feature of diagnostic testing has influenced the decision to include diagnostic self-assessments in the CALL pedagogical materials developed for the present research, which will be discussed further in Chapter 3.

2.10.2 AFL, dynamic assessment, and the present study

The co-constructed, social-interactional approach harmonizes well with the formative assessment practices of AFL with their emphasis on learner input. For example, social interactionalists Poehner and Lantolf (2005) have reported on mediated formative assessment, termed *dynamic assessment* (DA).

DA is grounded in Vygotsky's (1969, 1978) theory of the learner *zone of proximal development* (ZPD), pertaining to learners' *actual* versus *future* or potential development. DA views mediation as a naturally-occurring social-interactional phenomena, which is indispensible in fostering future development in learners.[7] Pellegrino et al. (2001) alluded to the absence of ZPD, making reference to the danger of teachers "subscribing to the belief that learning ability or intelligence is fixed", which is the antithesis to the idea of the ZPD (p. 240).

By way of a definition of sorts, of DA, Sternberg and Grigorenko (2002: vii) have said:

> [Its outcome] takes into account the results of an intervention. In this intervention, the examiner teaches the examinee how to perform better on individual items or on the test as a whole. (cited in Poehner & Lantolf, p. 234)

In actual classroom practice, the mediation element of DA may function in concert with AFL, which also gives due importance to learner feedback.

---

[7] In contrast to DA and its basis in the ZPD, Lantolf and Poehner (2004) have called assessment practices which are not sensitive to the ZPD, *static assessment*, or SA. (p. 51)

Lantolf and Poehner (2004), however, argued that teacher feedback was intuitive rather than based on learning theory principles. They also provided insight into the process of learning and the mediation involved in DA (and in fact, in AFL), suggesting:

> Researchers are beginning to recognize that formative assessment can provide more than feedback into the instruction cycle and opportunities for learners and teachers to reflect on learner performance. Indeed, as the work of Rea-Dickins (2001) demonstrates, in some types of formative assessment, especially in what Ellis (2003: 313) refers to as 'incidental' formative assessment, teachers are able to guide learners through dialogic interaction toward enhanced performance and learning. (p. 68)

Leung (2007) compared DA and AFL and found similarities between them, particularly in relation to four of the 10 principles of AFL, 1) *Assessment for learning should be part of effective planning of teaching and learning,* 2) *Assessment for learning should focus on how students learn,* 3) *Assessment for learning should be recognised as central to classroom practice,* and 8) *Learners should receive constructive guidance about how to improve,* which he found accorded with the DA practice of supporting learning through mediation.

Interest in DA and alternative assessment in classroom-based second language acquisition (SLA) and LT research remains as yet infrequent. Nevertheless, in recent time some LT researchers (Cheng et al.., 2004a; Leung, 2004, 2007; Leung & Mohan, 2004; Rea-Dickins, 2001, 2004, 2006a, 2006b; Rea-Dickins & Gardener, 2000) have begun investigating in this area.

Nevertheless, there is some confusion in the LT literature concerning the terms AFL, DA and formative assessment. This is unfortunate and needs to be addressed, particularly in view of the fact that these approaches offer much to researchers, learners and teachers. For the purposes of the present study, as mentioned in Chapter 1, it will be recalled that AFL is herein considered to be a variety of 'evolving formative assessment.'

2.11 Research influences to pedagogical materials development

At this time, it would be opportune to recall the suggestion of Colby-Kelly and Turner (2007) in relation to their participant school, which was that "We would challenge them to go further and to take an AFL approach, making assessment a primary factor in learning, to capitalize on the already formidable component of formative assessment in their programme" (p. 33). This call accords with the objective of the present research, since it has sought to develop and apply an L2 AFL approach in the same setting.

As a sound basis on which to develop the pedagogical materials necessary for the L2 application, the discussion will now return for a moment to the AFL 10 principles, previously introduced (see section 2.5.1). Since the 10 principles represent the grounded methodology of AFL, it would follow that they should be examined and the characteristics of each considered prior to the development of any L2 AFL pedagogical materials (or data collection instruments, for that matter). Thus, in the next section the 10 principles will be examined in light of how they may be applied in an L2 AFL classroom setting.

2.11.1 The 10 principles of AFL categorized for pedagogy

The 10 principles above (numbers 1-10) touch on various facets of teaching and learning. While there is some overlap between them, the principles effectively address several key issues which need to be considered for AFL pedagogical material development to proceed. These issues are: (a) *the learner-focus*, (b) *AFL assessment objectives*, (c) *teacher and learner goal sharing* and (d) *learner motivation*[8].

The 10 principles are categorized within 4 issues, which are illustrated in Table 1. First, (a) the learner focus (principle nine), suggests learners develop self-assessment techniques in order to more accurately gauge where they are in the learning process, gain autonomy in their learning, and become reflective about the process. In a similar vein (b), the AFL assessment objectives

[8] There are other Principles pertain mostly to teacher training and practice. They are that AFL '*should be part of effective planning of teaching and learning* (Principle one), that it '*be recognized as central to classroom practice* (Principle three), and that it '*be regarded as a key professional skill for teachers* (Principle four) (QCA, 2007). These matters of teacher training and practice will not be included in this discussion, since they do not contribute to the task at hand, the development of the classroom-based pedagogical materials used in this study.

(principle two, eight and ten) posit a learner- centered focus on *how* students learn, recognize the need for constructive teacher guidance to foster this and build on knowledge already attained (scaffolding), as well as to recommend students' learning efforts in all areas be appreciated as a means of helping students achieve their best. Another issue, that of (c) teacher and learner goal sharing (principle seven), espouses a cooperative learning environment in which teachers are encouraged to 'hold nothing back' in imparting their knowledge of the goals of assessment with students. Finally, as to (d) learner motivation, (principles five and six) the AFL approach suggests teachers appropriate a heightened sensitivity to the emotional impact of assessment, and to the importance of student motivation in driving the learning process.

Given the above, pedagogical materials had to be developed that included these AFL features. To recapitulate, these materials needed to focus on learners and encourage learner autonomy, take into account how students learn, incorporate constructive and sensitive feedback to guide the learning process, motivate the learners in all of their learning, and include self- and peer-assessments, all to help learners achieve their best outcomes. An element of learner reflection, teacher guided questioning and knowledge scaffolding had to be built into the materials. Certainly the materials needed to be appropriate for an L2 classroom setting.

2.11.2 Choosing Computer-Assisted Language Learning

Chapelle, (2001, 2005) in discussing Computer-Mediated Communication (CMC, in which language learning is aided by computerized simulations of interlocutor interactions using Multimedia programs), has argued that these computer environments may promote learner attention to IL and target language forms and help learners advance in their learning. I would posit that a CALL environment based on AFL principles, could similarly present a potentially useful language learning tool.

Table 1: The 10 principles categorized

Key issue addressed	*Principle expressed*	*Principle no.*
Learner-focussed approach	AFL develops learners' capacity for self-assessment so that they can become reflective and self-managing	9
AFL assessment objectives	AFL should focus on how students learn.	2
	Learners should receive constructive guidance about how to improve.	8
	AFL should recognise the full range of achievements of all learners.	10
Teacher and learner goal sharing	ALF should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed.	7
Learner motivation	AFL should be sensitive and constructive because any assessment has an emotional impact.	5
	AFL should take account of the importance of learner motivation.	6

Vygotsky's (1962, 1978) sociocultural theory situates learning as a product of social interaction with interlocutors. Herring, (1966) in giving CMC a broader definition than that which (more usually), refers only to multimedia environments, has argued for "communication that takes place between human beings via the instrumentality of computers" (p. 1). Choosing CALL has the added advantage of allowing learners access to this instrumentality, in interaction with the computer as 'nonjudgmental target language producer' wherein non-target 'errors' may be committed continually without compunction.

It was reasoned that these exercises would be an essential preparatory requirement prior to learners' undertaking a concept mapping exercise, since the exercises were intended to holding learners' interest and stimulate their background knowledge. A CALL module could serve this purpose well. In addition, by virtue of the fact that the CALL exercises and the individual student concept mapping (CM) exercises were both computerized, they would provide learners with a symmetry of practice environment media.

2.11.3 Choosing Concept Mapping

'*Deep learning'*

James and Gipps (1998) have argued for the kind of active learning and learner understanding of pedagogical materials which I feel a concept mapping activity could entail. They call this *deep learning*. James and Gipps have recommended a deep learning approach that would include the following:

1. An intention to develop personal understanding
2. Active interaction with the content, particularly in relating new ideas to previous knowledge and experience
3. Linking ideas together using integrating principles
4. Relating evidence to conclusions. (p. 287)

Similarly, James (2006) suggested that cognitive, constructivist learning theories focus on "how people construct meaning and make sense of the world through organizing structures, concepts and principles in schema (mental models)" (p. 55).

Recently several (Bevan, 2007; Daley, 2004; James, 2006; Jonassen et al, 2008; Novak & Cañas, 2008) have found CALL, CMC and other computer-

based activities to be effective in supporting learning. For example, in a discussion of Mindtools (CM, databases, spreadsheets and other computer applications), Jonassen et al. noted "students cannot use Mindtools without thinking deeply about the content they are learning, and if they choose to use these tools to help them learn, the tools will facilitate the learning and meaning-making process" (p. 83). Elsewhere Jonassen et al. (2008) decried the classroom practice whereby "too often, teachers present formulas used to calculate without teaching the process conceptually" (p. 85).

The decision to develop a CM activity to help students attempt to conceptualize a feature of English grammar, and in so doing apply AFL in an L2 setting, was largely inspired by the work of Lev Vygotsky, (1969, 1978) and in particular through an interest in his theory of the learner *zone of proximal development* (ZPD). Many (Poehner & Lantolf, 2005; Nyikos & Hashimoto, 1997; Lantolf & Aljaafreh, 1995) have argued for applying the ZPD concept in L2 learning situations, citing the ZPD in terms of *actual* and *potential* development. I felt that a CM activity could help learners conceptualize a language feature at the same time as presenting a challenge intended to bring learners closer to an understanding of target form usage, helping them go further towards their full learning potential. I propose a CM activity could be a useful means of guiding adult EAP learners to develop schema to foster their understanding of an English grammatical form. Others (Conlan & Bird, 2004; James, 2006; Luckin & du Boulay, 1999) have also suggested that CM activities can be an effective learning tool.

2.11.4 CALL as preparation for Concept Mapping

Since it was felt that the CM exercises could not proceed successfully without sufficient learner motivation to perform them, the CALL module exercises were also designed with a view to increasing learner engagement, reflective thinking and hypothesizing, and background knowledge, as well as to incorporate self-assessments, in preparation for the CM exercises.

2.12 Chapter summary

In this chapter, the evolution of formative assessment and the AFL approach from its early inception in motivational studies and in mastery

learning has been discussed. Particular attention has been paid to Black and Wiliam's (1998a, 1998b) seminal research leading to the development of the AFL approach to classroom based formative assessment procedures. Following this, applications of formative assessment geared to enhancing learning and AFL in general education settings and in L2 research have been reviewed.

In comparison to studies in general education, research into the application of AFL in L2 settings is rare; pedagogical materials for L2 classrooms are rare; teacher training in AFL procedures for L2 teachers is rare. It is clear that more research is needed about the application of AFL in L2 classroom settings.

The present research will address this imbalance by developing AFL pedagogical materials appropriate for an L2 classroom setting, by applying those materials in an L2 classroom, and by teacher-training for the application of AFL in an L2 classroom.

The present research also seeks to answer the challenge of Colby-Kelly and Turner (2007) noted above, of applying AFL where teachers and school administrators have requested it be done, at the same participant school. The study rationale, and the development of specific pedagogical instruments for achieving it will be explained in Chapter 3. In addition, the theoretical frameworks on which they are based, and the development of the AFL instruments used in the study will also be explained in Chapter 3. The methodology and study procedures will be explained later, in Chapter 4.

CHAPTER 3 – PREPARING FOR AFL IN AN L2 CLASSROOM: THE RESEARCH DESIGN AND INSTRUMENT DEVELOPMENT

Break, break, break,
on thy cold grey stones, O Sea!
And I would that my tongue could utter
the thoughts that arise in me.

Tennyson[9]

Having described in the previous chapter the research influences supporting the conceptualization of the L2 AFL pedagogical materials developed in the present research, in this chapter I will begin with the impetus for this study, the research questions. The evolution of the study will then be described, from the selection of lessons focusing on the modal form *would* used a vehicle for investigating the AFL approach in an L2 classroom setting, to the mixed methods design of the study, the study context and participants, the pedagogical and data collection instruments, and the data analysis procedures.

This chapter will centre on explaining how new pedagogical materials and data collection instruments were developed for the study, and given that considerable preparation was required to accomplish this, the study methodology will be explained in the following chapter. This discussion will begin by introducing the focus of this research, the research questions.

3.1 The research questions

As mentioned in the previous chapter, this study was inspired by a call for more research into alternative assessments in L2 settings in general, and in view of my questioning whether an AFL approach might be suited to an L2 classroom setting. Beyond that lay the question of how to best realize L2 AFL in that setting, given that existent pedagogical materials for L2 AFL could not be found, and making it necessary to develop new materials. There was also the uncertainty as to whether the often-reported success of AFL in other

[9] This quote is a purposely chosen complex example of a *would* form, accessible to the reader who has grasped the full meaning of the form, which in this case evokes an other-worldly context.

disciplines would in fact transfer to an L2 classroom setting, with its particular character, requirements and learning challenges.

I had decided on a Vygotskian theoretical standpoint which is reflected in the study design and research questions. I had also decided to use CALL and CM in the L2 AFL treatments. All of these considerations would lead to the research questions of the study, the central of which was:

- ***Is there evidence that AFL practices in a second language classroom enhance the learning of a specific language feature?***

Underlying the central research question were seven secondary questions:

1. *Is there evidence that a **computer assisted language learning** exercise in a second language classroom enhances the learning of a specific language feature?*

2. *Is there evidence that an **online concept mapping** exercise in a second language classroom enhances the learning of a specific language feature?*

3. *Is there evidence that a **group concept mapping** exercise in a second language classroom enhances the learning of a specific language feature?*

4. *Is there evidence that a **teacher-class concept mapping** exercise in a second language classroom enhances the learning of a specific language feature?*

5. *Is there evidence of the* **Assessment Bridge** *when* Assessment for Learning *practices are employed in a second language classroom?*

6. *Is there evidence that **students** in a North American second language classroom demonstrate agreement with the 10 principles of AFL proposed by the ARG?*

7. *Is there evidence that **teachers** in a North American second language classroom demonstrate agreement with the 10 principles of AFL proposed by the ARG?*

3.2 The research design

3.2.1 The research paradigm

As mentioned in Chapter 2, Sadler (1989) noted three conditions for improving learning, that students could 1) understand the aimed for criteria and standards, 2) assess their own performance and compare it with the standards, and 3) engage in action in order to close the gap between the two. Drawing on Sadler, the AFL approach seeks to help learners notice the gaps between their own and target competence. AFL incorporates Sadler's suggestions in recommending learners understand assessment criteria, notice gaps in their own and target performance, and take steps to achieve the goals of learning. Keeping the three conditions for improving learning in view, the present research proposes to apply AFL practices in an L2 classroom setting, and to investigate the effect of that approach on student learning.

3.2.2 The statement of inquiry

The purpose of this study is to investigate the effect on learning of an application of AFL procedures in a second language classroom setting. In order to apply these procedures, AFL pedagogical materials were developed. Thus, this study has sought to apply an AFL methodology in a second language classroom setting, and to investigate the effect of that method on learning. It is interested in teacher and learner perceptions of AFL, and also whether there is evidence of the AB in that setting.

3.2.3 The rationale for using mixed methods

A mixed methods paradigm was selected as the best methodology to follow in order to adequately investigate the research questions driving the study. That is, in order to get a full picture of teacher and student perceptions of the effects of the L2 AFL treatment on learning, instruments to be analyzed qualitatively were anticipated in the form of questionnaires with open questions. Quantitative instrument analysis alone could indicate trends, but could not contribute to explaining the central research question as to whether learning may have been enhanced by the L2 AFL. On the other hand, qualitatively analyzed instruments could help elucidate such factors as *why* the participants held certain views, *how* they approached teaching or learning,

what strategies they chose to employ in their teaching or learning, and even whether they believed there were any effects of the treatment possibly in the absence of quantitative evidence of such effects. In essence, qualitative investigation would be the means of gathering elusive information pertaining to teacher and learner thought processes. Thus qualitative investigation married with quantitative trend-seeking analysis would work effectively in answering the research questions and inform the study as a whole.

With these issues in mind, Creswell and Plano Clark (2007) cited pragmatism as the paradigm for mixed methods research, noting that that mixed methods "is premised on the idea that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone" (p. 18). In fact, the approach of the study, being of an exploratory character and taking a Vygotskian stance towards the co-constructed nature of learning, led to the decision to give weight to qualitative over quantitative data to best answer the research questions.

3.2.4 The mixed methods methodology

This study employs a sequential, exploratory mixed methods research design following Creswell & Plano Clark (2007). The quantitative and qualitative data collected were analyzed and triangulated (Tashakkori & Teddlie, 2003; Creswell & Plano Clark, 2007). Qualitative measures were predominant in the study, while the quantitative measures took a secondary place. The quantitative measures were taken pre-, post- and delayed-post treatment. Therefore, the mixed methods notation system describing this research is the following:

quan → QUAL → quan.

The mixed methods study design has been illustrated in Figure 1. The long black arrow on the left side of Figure 1 represents the study timeframe. The boxes show the sequencing of the study treatment phases over time, as well as pre-, post- and delayed-post data collections before and after the treatment period. Arrows point to another box at the bottom of the figure, to indicate the triangulation of all data, which occurred following all data collection. The grey rectangle represents the four phases of the study treatment.

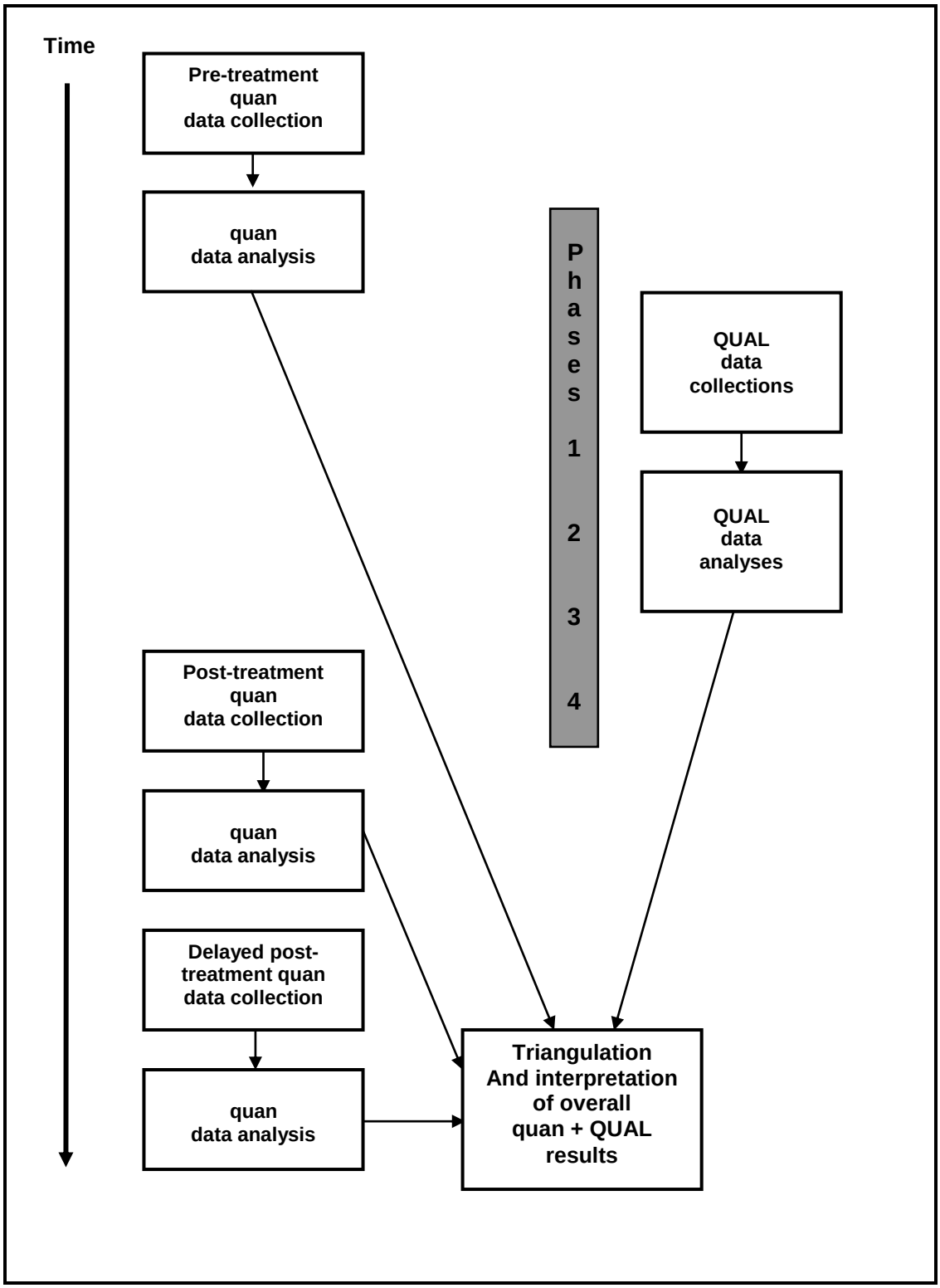


Figure 1 Mixed methods research design of the sequential exploratory study

3.2.5 The quasi-experimental nature of the study

The design of the study was quasi-experimental. That is, the students in the study had been selected to be in their class groups by the participant school. Thus the groups constituted already existent, intact groups. Several (Brown & Rodgers, 2002; Hatch and Lazaraton, 1991) have observed that in educational research it is usually not possible to randomly select participant students. In a discussion of this, Brown and Rodgers noted that "…most studies in our field tend to be quasi-experimental rather than truly experimental" (212).

Brown and Rodgers (2002) categorized four types of quasi-experimental second language research. The present study falls into category *D,* described as having "two treatments [and] two groups…classic method studies involving a control group and an experimental group, each taught by a different method" (212). In this study, the assignment of the four participant classes into two treatment and two control groups was done randomly.

3.2.6 The split-plot study design

A split-plot design was selected as the design for this study. Normally, a split plot design takes place over two time periods wherein the treatment and control groups and their teachers reverse roles in Times 1 and 2 (Hatch & Lazaraton, 1991). However in this study an amended split-plot design was used to accord with the participant availability and the operational constraints of the school under investigation. This was done in order to get the richest data possible under challenging circumstances. Thus, three teachers and four different classes participated in this study, which took place in two time periods. In Time 1, teacher *A* and her class participated as the 'treatment' teacher and class group, and teacher *B* participated as the 'control' teacher and class group. In Time 2, teacher B was the treatment teacher with a different class group and a different teacher, teacher *C* was the control group teacher with a different class group. Times 1 and 2 took place over two subsequent school sessions, with different students comprising each of the four participant classes. Using this design made it possible to make comparisons between the teachers and students in Times 1 and 2, and between the treatment and control group classes. The split-plot design is illustrated in Figure 2.

Data collection		
Time 1	**TEACHER A Treatment group**	**TEACHER B Control group**
Time 2	**TEACHER B Treatment group**	**TEACHER C Control group**

Figure 2 Teachers and classes in the split-plot design

3.2.7 The treatment timeframe

The study took place over a period of 6 weeks in May and June of 2009; Times 1 and 2 lasted 3 weeks each. Each study period comprised 3 weeks in mid-term of the school's 10-week term sessions. On the first and last days of Times 1 and 2 the essay and FITB tests were administered, as were the AFL 10 principles questionnaires. The treatment followed thereafter over a 3-week period.

The study treatment took place in four sequential phases, which occurred over each 3-week timeframe in this approximate sequence:

- Week 1 - Phase 1. Individual learner self-directed online CALL learning module
- Week 2 - Phase 2. Individual learner online CM production
- Week 3 - Phase 3. Peer, co-constructed group CM production using paper-and-pencil
  - Phase 4. Teacher-class co-constructed CM production using acetate sheets and an overhead projector (OHP)

The study timeframe is illustrated in more detail in Figure 3.

	Mon.	Tues.	Wed.	Thurs.	Fri.
TIME 1 week 1	***PHASE 1 : CALL*** intro. to CALL	CALL	CALL	CALL	-
week 2	***PHASE 2: CM*** Introduction to online CM software	Online CM	Online CM	Online CM	-
week 3	**(Holiday)**	***PHASE 3:*** Group paper-and-pencil CM production	-	***PHASE 4:*** Teacher-class CM production, guided questions	-
TIME 2 week 1	***PHASE 1 : CALL*** intro. to CALL	CALL	**(Holiday)**	CALL	CALL
week 2	***PHASE 2: CM*** Introduction to online CM software	Online CM	**(Holiday)**	Online CM	-
week 3	Online CM	***PHASE 3:*** Group paper-and-pencil CM production	-	***PHASE 4:*** Teacher-class CM production, guided questions	-

Figure 3 Study design schedule for three-week treatment period

3.3 The research context

Following Colby-Kelly and Turner's (2006, 2007) reporting of formative assessment practices in their baseline descriptive study of pre-university English for Academic Purposes (EAP) classroom setting in Quebec, the teachers and administrators of the participant school expressed an interest in a further study in which AFL procedures would be applied in that setting.

The present research is in response to that request. Thus, it will investigate an application of AFL in a second language classroom and the effectiveness of the approach in contributing to learning.

3.3.1 The participants

Sixteen students from one class and their teacher participated in a piloting of the study. The main study participants included 55 students across four classes and their three teachers, comprising individual treatment and control groups investigated twice in sequence (in this split-plot design). Figure 4 shows the codes for all participant students and their placement in treatment and control group classes. To preserve participant anonymity, the names of the 55 students have been omitted from the data and replaced with numerical codes. The class selection had been done prior to the study by the participant school. The selection of the classes in treatment or control groups was random.

Student Participants (N=55)			
Time 1		**Time 2**	
Treatment n=14	**Control n=16**	**Treatment n=14**	**Control n=11**
S1	S15	S31	S45
S2	S16	S32	S46
S3	S17	S33	S47
S4	S18	S34	S48
S5	S19	S35	S49
S6	S20	S36	S50
S7	S21	S37	S51
S8	S22	S38	S52
S9	S23	S39	S53
S10	S24	S40	S54
S11	S25	S41	S55
S12	S26	S42	
S13	S27	S43	
S14	S28	S44	
	S29		
	S30		

Figure 4 Application of the split plot design

All of the pilot study and main study student participants were from various mother tongue backgrounds and all were enrolled in the pre-university Intensive EAP programme, at Advanced 1 level. Of the 55 student participants, 33 were male and 22 were female. The exact number of participants for each instrument varied. All of the participant students had previously been placed in the Advanced I level by the school either by means of pre-testing, or following advancement from the Intermediate level.

The goal of all but one of the 55 students was admission to an English-language university. In order to gain university admission, the students were required to succeed either in the subsequent Intensive EAP Advanced 2 level with a grade of 70% or more, or in the university admissions ESL examination.[10]

The participant teachers and students were recruited from the advanced level pre-university EAP programme. The cohort included three teachers and four classes (one teacher participated as a control group teacher in the first treatment, and as the treatment group teacher of the second treatment as well). Two additional teachers participated peripherally, by administering two delayed post-tests.

The participant teachers were all female and reported English as their mother tongue. All had 10 or more years of experience at the school and all but one had MA degrees in second language education; one had a certificate in teaching English as a second language (TESL). All reported having taken courses in second language evaluation. In addition, a secondary school teacher of advanced level ESL students participated in the scoring of the essay tests. That teacher reported having 34 years of teaching experience.

All ethical procedures were followed and consent forms were administered and signed by all participants.

---

[10] Part One of the admissions ESL test examines knowledge of formal aspects of language and Part Two examines student performance on an academic essay writing task. The test is 2 ½ hours long and is described on the university website as being in some parts similar to the Michigan MELAB test. Examinees are encouraged to consult existing MELAB and TOEFL Test of Written English (TWE) preparatory materials prior to taking the test.

3.3.2 The school curriculum introduced

The teachers involved in the study had previously developed and written the school's curriculum documents in a group project. The school used a communicative teaching method, with a grammatical form-focussed component. While the participant classes also use a traditional grammar-based textbook, the teachers reported that it was seldom, if ever used.

In their study of assessment practices at the school, Colby-Kelly and Turner ((2006, 2007, 2008) found that the teachers employed a strong component of formative assessment procedures; the majority of assessment instances observed were formative.

3.3.3 The school curriculum analyzed

Colby-Kelly and Turner (2006, 2007, 2008) conducted an extensive analysis of the curriculum of the participant school in their investigation of formative assessment practices there. Drawing on their work, a new analysis of the curriculum used in Advanced Level 1 of the school was done a priori to enhance the ecological validity of the present research. Thus, the pedagogical materials already in use at the school were examined prior to beginning the development of L2 AFL pedagogical materials.

3.3.4 The AFL procedural framework requirements

An AFL methodology was applied in the L2 classroom setting through an interpretation of AFL tenets through the use of *computer assisted language learning* (CALL); *concept mapping* (CM) exercises; self- and peer-assessments; and through the application of teacher-guided questioning and knowledge-scaffolding techniques.

3.3.5 Using a grammar point to investigate AFL

A point of grammar was chosen as a means for investigating the AFL application in an L2 classroom setting. In order to enhance the content validity of the investigation it was felt the grammar point should be:

1. Challenging for advanced learners to master
2. Part of the learning agenda at the students' level
3. Important to master for academic purposes

First, if the grammar point had proven to be relatively easy for the students to master, any suppositions as to the effectiveness of the AFL treatment might be called into question. Therefore, for the purposes of the study it was felt that a difficult concept would pose the best challenge and be more appropriate. Furthermore, it was hoped that a difficult grammatical point might engender more learner discussion about the nature of its usage in reflective exercises than might a usage that was more straightforward.

Second, it was very important that the grammar point already be part of the school's agenda for the level under investigation for the purposes of comparison with control groups. Again, there would be greater content validity if the treatment *and* control group teachers were already accustomed to teaching the grammar point, so that they (and the student participants) might more readily regard the teaching and learning of the point within the realm of AFL as a normal part of the course content.

Third, in order to preserve ecological validity, the grammar point should be one that would be important for students to master to help them achieve academic success in the course requirements.

English modal usage has presented difficulties for the students of Advanced level 1at the participant school, as the teachers attested in incidental comments prior to the study. The results of piloting of the study also substantiated their impressions (see below for pilot study information). These forms have been challenging for ESL learners from diverse L1 backgrounds in general. In particular, the *contingent use modal* form 'would' in obligatory contexts (Quirk and Greenbaum, 1973) is often confused with 'will' in learner written and spoken production. This modal form signals to language users an as-yet unreal or intangible state. Quirk and Greenbaum's example of this form is, "We would love to go abroad if we had the chance" (55). The contingent use modal form is often omitted by second language learners in favour of *will*, as in *We will love to go abroad if we had the chance.

Celce-Murcia and Larsen-Freeman (1999) have termed this a *modal-like* form for expressing preference and desire, as in "Sarah would like to travel around the world" (p. 147). It is often produced by learners as **Sarah will like to travel around the world*. Similarly, Dixon (1994) has called it the present-

unreal, as in “If I were you, I would study” (143). It may be incorrectly produced as “*If I were you, I will study” (143).

Thus, for the reasons stated above, the contingent use modal form was chosen to be used as a vehicle for investigating the effect of AFL procedures in this setting. It fulfilled the three requirements noted above, since it had proven challenging for this population of learners, it was part of their learning curriculum, and they needed to master target usage to succeed at academic production once they began their undergraduate university studies.

3.4 The treatment instrument development

For the purposes of the present research, it was decided that pedagogical materials would be developed to apply AFL procedures in this L2 classroom setting. The following section will illustrate how this was accomplished.

3.4.1 Pedagogical materials development

A strong element of AFL is that of enhanced learner motivation and learner engagement. Thus it was essential that the pedagogical materials developed for this study would engage the young adult learner participants. Due to their youth, experience, and socio-economic background, (see Participants, above), it was expected that most or all of the participant learners would be computer-savvy and comfortable with e-learning. Therefore, the learners might be more interested and more readily apply themselves to Computer-Assisted Language Learning (CALL) than they would to learning through traditional paper-and-pencil methods.

A point of grammar had been chosen as the medium for this application of the AFL approach. The point would be conceptualized by individual learners. Then the point would be conceptualized and discussed amongst peer groups of learners, with their conceptions modified and adjusted in self-, peer-, and class-teacher co-constructed reflections including teacher-guided questioning and knowledge scaffolding.

3.4.1.1 Teacher training instructional materials.

The study rationale and design, the modal grammar feature to be used in the investigation, AFL procedures, CALL, CM and the data collection schedule comprised the content of the AFL training sessions given to the treatment

teachers. An agenda specifying the content was given to the teachers (see Appendix C). The study instruments and pedagogical materials were reviewed to familiarize the teachers with the overall design of the study. The data collection tests, questionnaires and teacher and student surveys (of perceptions following interventions) were discussed with the teachers at that time as well. Drawing on Hodgen & Marshall's (2005) reporting of AFL procedures observed in an English language arts class and on the literature review, AFL guided questioning techniques, group work, whole class discussions, scaffolding of knowledge and provision of challenging and rich formative assessment activities in a second language classroom setting were discussed.

Concept Mapping theory was explained and illustrated with concept maps (for example Venn drawings, diagrams, and charts), and printouts from the online tutorials taken from the *Inspiration* educational CM software used in the study (this is available from the researcher).

3.4.1.2 AFL pedagogical material development for an L2 classroom.

AFL was originally applied in mathematics and science classrooms and the approach has since been used in general education. Hodgen & Marshall (2005) have reported on and compared AFL procedures in English arts and mathematics classes. Yet AFL has until very recently rarely been employed in the L2 classroom setting. In fact a comprehensive search of pedagogical materials for L2 classroom use conducted in the initial stages of the present study was unsuccessful and thus failed to reveal any.

A classroom-based application of AFL must begin with the principles of AFL (QCA, 2007). To revisit them, they are that:

1. *AFL should be part of effective planning of teaching and learning*.
2. *AFL should focus on how students learn*.
3. *AFL should be recognized as central to classroom practice.*
4. *AFL should be regarded as a key professional skill for teachers.*
5. *AFL should be sensitive and constructive because any assessment has an emotional impact.*
6. *AFL should take account of the importance of learner motivation.*
7. *AFL should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed.*

8. *Learners should receive constructive guidance about how to improve*.
9. *AFL develops learners' capacity for self-assessment so that they can become reflective and self-managing.*
10. *AFL should recognize the full range of achievements of all learners*. (QCA, p. 1)

3.4.1.3 The categorized 10 principles of AFL.

As has been noted in some detail in Chapter 2 (section 2.11.1), the AFL 10 principles are the keystone of the AFL approach and as such were categorized for the purposes of developing new L2 AFL pedagogical materials in the study. To call this to mind, the 10 principles were categorized in terms of four key issues they address, which are AFL's (a) *learner-focus*, (b) *AFL assessment objectives*, (c) *teacher and learner goal sharing* and (d) *learner motivation.*

In view of categories above, pedagogical materials were developed that included their foci. Therefore, the materials needed to focus on learners and encourage learner autonomy, take account of how students learn, incorporate constructive and sensitive feedback, and motivate the learners in their learning progress, incorporating self- and peer-assessments. Additionally, learner reflection, teacher guided questioning and knowledge scaffolding would appear in the materials.

3.4.1.4 The English Mystery Module.

Ten online pages were developed for use as the CALL exercises and called *The English Mystery Module*. The pages were created using Microsoft Word software, upon the recommendation of the director of information technology of the participant school, as being the most accessible and user-friendly software to the purpose. The first page of the module is illustrated in Figure 5. (The complete module pages are available from the researcher).

Hyperlinks to all of the pages were provided in order to afford students easy access to the features on different pages and to give them self-directed practice and learning opportunity. The online page titles and content were the following:

1. ‘Welcome to the English Mystery Page!’ (including introductory and some guidance text.)
2. ‘Modal Mystery Clues’ (including five clues of varying levels of complexity)
3. ‘DiaModal Mystery Diagnostic Quiz’ (including a fill-in-the-blanks quiz requiring the user to choose to fill in *would* or *will*.)

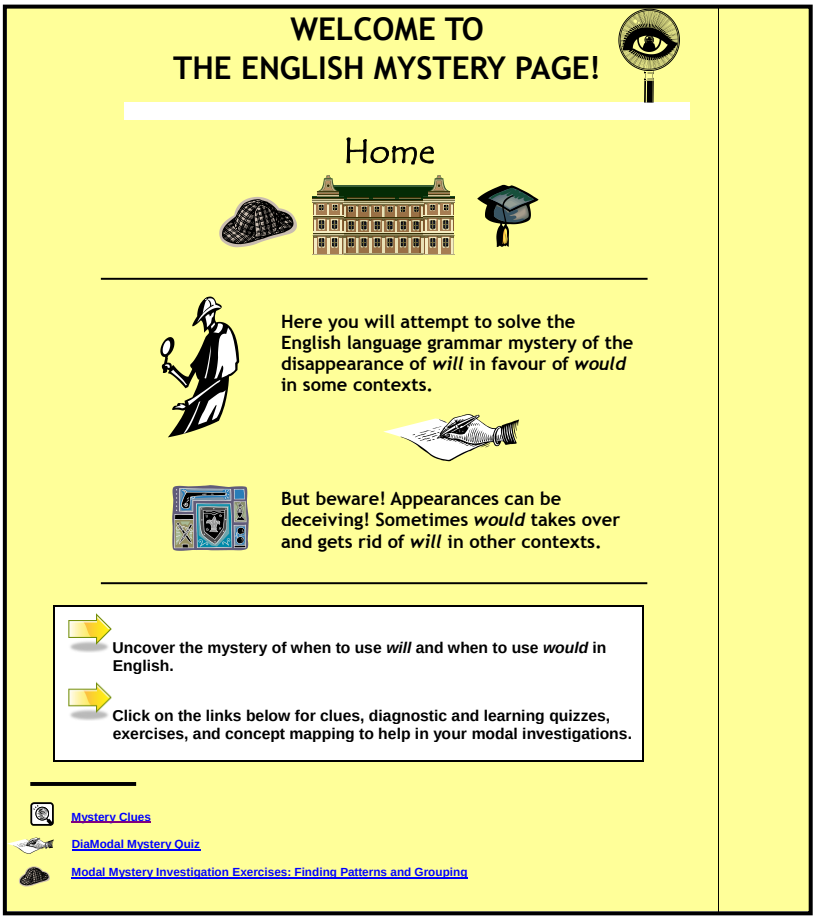


Figure 5 Page one of the English Mystery Module

4. ‘DiaModal Mystery Diagnostic Quiz Answer Key’ (including answers to the page three quiz, a chart for noting number of correct and incorrect responses.)
5. ‘DiaModal Mystery Inspector Report: Self-Assessment’ (including a chart for reporting into which range of correct responses learners found themselves after some practice, which were one to two, three to four, five to six, seven to eight, and nine to ten. Page five also included a

self-reflective survey of learner perceptions as to how challenged they felt by the material, in order to situate the learners as to where they were in their learning progress.)

6. 'Modal Mystery Investigation Exercise: Finding Patterns' (including the same quiz questions as those found in the quiz on page three. On page six, however, all of those questions for which the correct response was *would* and all of those for which the correct response was *will* were grouped together, followed by the self-reflective question, 'What do you think is the same about these sentences?')
7. 'Modal Mystery Investigation Exercise: Finding Patterns Answer Key' (including the page six quiz with the answers indicated.)
8. 'Modal Investigation Exercise: Grouping' (including a new quiz in which the items share a common situational theme and developed to guide learners to a better understanding of the contingency or unreal context of the contingency modal use by guiding users to group their quiz responses in one of three categories – '*this is certain to happen; this may happen*; or *this may happen if something else happens*.')
9. 'Modal Grouping Exercise Answer Key' (including the correct responses and categorical groupings of those responses, as explained above.)
10. 'Closing the Case: Modal Reflections…' (including a short and easy two-item quiz where the correct response to both is *would* and which most learners would be expected to succeed, and reflective questions asking however, what is different about the two instances, as well as asking '*what does this tell you about* will *and* would?' A hyperlink to the interactive *A Quizzical Mystery Quiz* was also present.)

Moreover, the Module and hyperlinks were developed to offer learners immediate access to computerized practice pages for a) practice producing target forms through online 'interactions' with the computer, b) answer grid pages intended to promote learners' noticing of differences between their IL and target forms, c) a metalinguistic component in the form of links to a 'clues' page to foster understanding of target form usage in different contexts, d) self-assessment checklists to help learners track their target production

progress, and e) checklists to help learners track their metalinguistic conceptualization progress.

3.4.1.5 A Quizzical Mystery Quiz.

The CALL materials also included a link to an interactive self-assessment quiz called *A Quizzical Mystery (LearnModal) Quiz* (see Appendix A). The quiz included a learning component in the form of access to optional hints intended to guide learners towards an understanding of the contingent modal form usage. The page is shown in Figure 6.

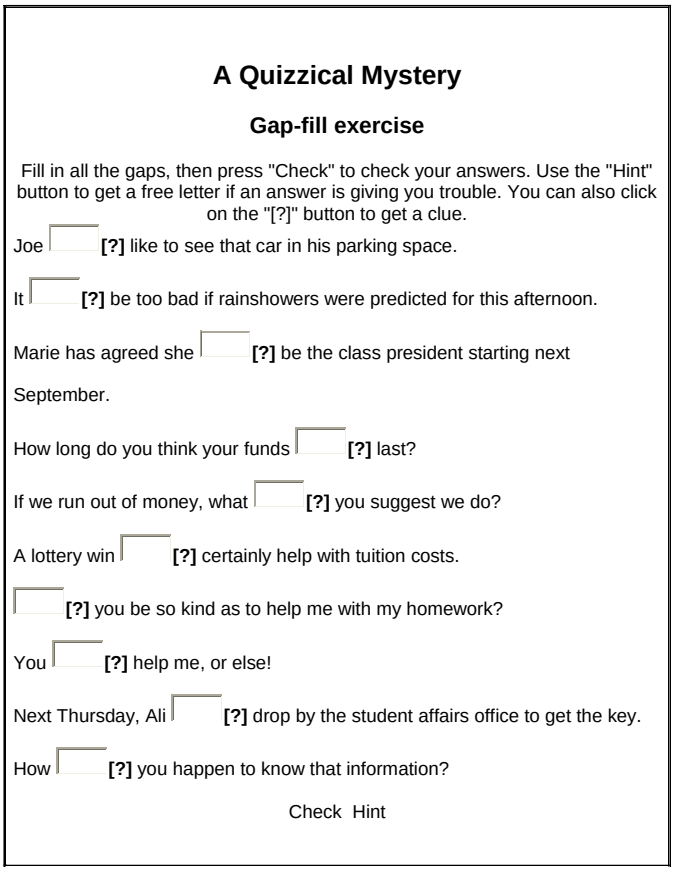


Figure 6 The interactive CALL page with guided hints

This quiz featured a drop-down menu which provided users with a choice of *would* or *will*, as well as the option of clicking on a usage 'hint' by means of a guided question intended to challenge the learners to discover grammar usage patterns themselves. The learners could click on one of the forms and receive immediate feedback as to whether that answer was correct. If the choice was incorrect, the learners had the option of clicking on a question

mark symbol (**[?]**), and the drop-down menu would provide the hint requested (see Appendix B). For example, the first item and hint are the following:

Item: Joe _____ like to see that car in his own parking spot.
Hint: What does Joe mean by this? Is he talking about his own car?

Therefore, the Quizzical Mystery Quiz was developed to give CALL learner-users pop-up metalinguistic 'hints' and 'guided questions' to encourage learner reflection, and to provide guidance about the target form usage.

In conclusion, these CALL activities follow the AFL approach, incorporating a) a learner-focus which encourages learner autonomy and self-reflection, b) a focus on how students learn and by providing constructive guidance, and by c) attempting to engage and motivate the learners. In addition, the CALL exercises incorporated a self-diagnostic element. For all of the reasons stated above, a CALL module was chosen as the means by which AFL would be applied in this L2 classroom setting.

3.4.1.6 Concept mapping educational software.

In order to allow students a variety of conceptual options in the online concept mapping exercise, *Inspiration* educational CM software (available at http://www.inspiration.com/) was downloaded in the school computer lab for the students to use in the study. Figure 7 illustrates an example of CM production tools taken from a tutorial on the *Inspiration* software website.

3.4.1.7 Pedagogical materials development summarized

Earlier in this chapter I noted that in order to develop AFL pedagogical materials for the L2 classroom, several components would need to be present. They are: (a) *the learner-focus*, (b) *AFL assessment objectives*, (c) *teacher and learner goal sharing* and (d) *learner motivation*. In order to respect this, it was decided that individual learners would each complete an online practice CALL module and an online CM production exercise. This would be followed by a group co-constructed CM production exercise, followed in turn by a teacher-class co-constructed CM production exercise.

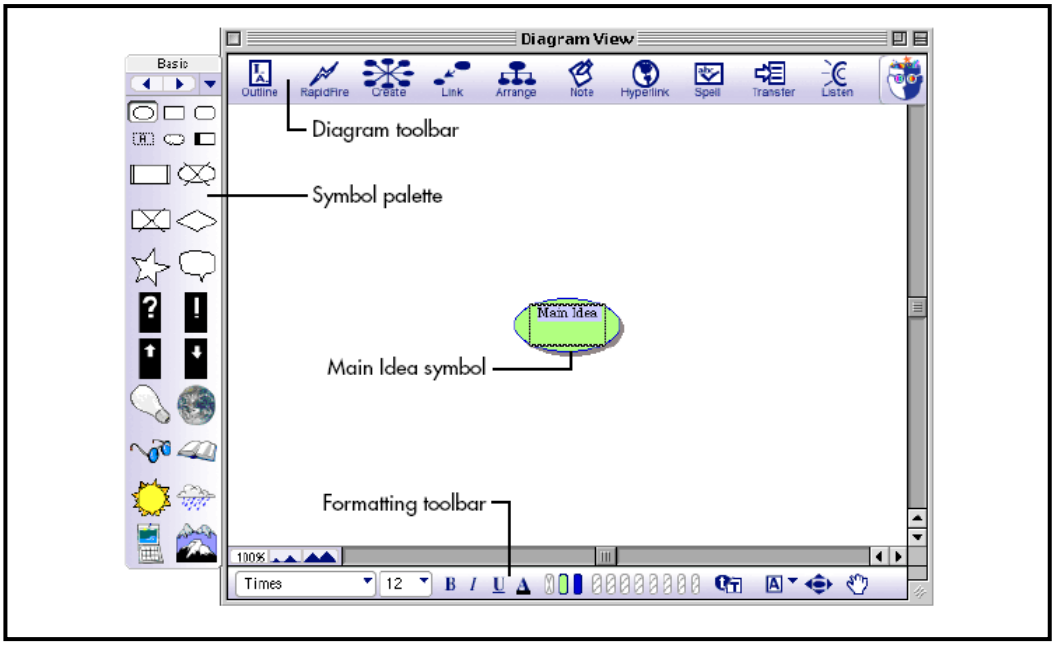


Figure 7 Concept map production software instrument

In keeping with the AFL components noted above, the individual online CALL and CM exercises were designed a) to be learner-focused, and to encourage learner autonomy and self-reflection. In addition, all of the CM exercises would reflect AFL assessment objectives by b) focusing on the learning process rather than on the final product, and by encouraging learners' efforts throughout the process. The group and teacher-class co-constructed CM production exercises would reflect AFL assessment objectives by incorporating peer feedback and teacher guidance. Moreover, the CM production would c) promote teacher and learner goal sharing, since the learners would be well informed about the goals of the activities beforehand. Finally, the CM exercise would d) attempt to engage learners to increase their motivation, and would stress (as in b, above) that errors should not be considered to be inherently 'bad,' but helpful elements in the advance towards final learning goals.

Therefore, with each step building on the previous one, the pedagogical materials used in this study would proceed in the following sequential phases:

- Phase 1. Individual learner self-directed online CALL learning module
- Phase 2. Individual learner online CM production
- Phase 3. Peer, co-constructed paper-and-pencil CM production in groups.

- Phase 4. Teacher-class co-constructed CM production using acetate sheets and an overhead projector (OHP)

The four phases, applied sequentially, operationalized in classroom practice the knowledge scaffolding feature of AFL.

3.4.2 Data collection instrument development

The pedagogical materials development has been described up to this point. The discussion will turn now to other materials developed for the study, which are of equal importance, they are the data collection instruments.

3.4.2.1 The English Mystery Module pages.

Online pages five and ten from The English Mystery Module were printed by the students after their practice with the module, and collected as data in the study. To revisit, these pages contained the following materials:

5. 'DiaModal Mystery Inspector Report: Self-Assessment' (including a chart for reporting into which range of correct responses learners found themselves after some practice, which were one to two, three to four, five to six, seven to eight, and nine to ten. Page five also included a self-reflective survey of learner perceptions as to how challenged they felt by the material, in order to situate the learners as to where they were in their learning progress.)
10. 'Closing the Case: Modal Reflections…' (including a short and easy two-item quiz where the correct response to both is *would* and which most learners would be expected to succeed, and reflective questions asking however, what is different about the two instances, as well as asking '*what does this tell you about* will *and* would?' A hyperlink to the interactive A Quizzical Mystery Quiz was also present.)

3.4.2.2 A Quizzical Mystery Quiz.

Following their practice with the online page of A Quizzical Mystery Quiz, the student printed paper copies of the completed online quiz were submitted as data in the study.

The purpose of the collection of the pages of The English Mystery Module and A Quizzical Mystery Quiz was to encourage students and motivate them

in their learning, to investigate the extent to which the students had participated in the module, and to qualitatively analyze the data at a later date.

3.4.2.3 Student surveys on CALL.

A survey comprised of two questions was prepared in order to collect data on student perceptions of the CALL experience and their learning, after their CALL experience with The English Mystery Module and A Quizzical Mystery Quiz (see Appendix F). The questions asked were the following:

1. Did this help you learn?
2. In what way?

3.4.2.4 Individual student online CM production.

Concept maps produced by individual students using *Inspiration* educational CM software were a source of data used to investigate their effect on learning.

3.4.2.5 Student surveys on individual online CM exercise.

A two-question survey was prepared in order to collect data on student perceptions of their individual CM experience and learning (see Appendix G). The questions asked were the following:

1. Did making a concept map help you learn?
2. In what way?

3.4.2.6 Student group CM production exercise.

The paper-and-pencil concept maps produced by groups in the group CM production exercise were collected and enlarged to poster size for display later on the blackboard and on walls around the classroom (to be used in the teacher-class CM production exercises). Additionally, the concept maps were reproduced on acetate pages for use on an OHP at the same time, (also for use in the subsequent teacher-class CM production exercises).

3.4.2.7 Teacher-class CM production exercise.

The enlarged group concept maps produced were displayed on the blackboard and on walls around the classroom so that the student groups could refer to them and show them to their classmates when the group CM presentations took place as the first part of the teacher-class CM production

exercise. The group concept maps were also reproduced on acetate pages for use on an OHP at the same time, in order to ensure that all students could see the details in the maps. In addition, blank acetate pages were used for the treatment teachers to produce teacher-class concept maps.

3.4.2.8 Classroom observation field notes.

Drawing on Colby-Kelly and Turner (2007) and Tesch (1990) field notes made in classroom observations of the co-constructed discussions focused on interaction categorizations (see Chapter 4, Procedures).

3.4.2.9 Audio recordings of classroom observations.

Digital audio recordings of the group CM production, and teacher-class CM production exercises were made and the electronic files saved.

3.4.2.10 Transcriptions of audio recorded classroom observations.

Selected transcriptions were made from the audio recordings of the group CM production, and teacher-class CM production exercises, in order to illustrate trends, categorize discourse, and contribute to answering the research questions.

3.4.2.11 Student post-treatment survey questionnaire.

A student questionnaire, called the Mystery Module and Concept Mapping Exercises Post-Questionnaire, comprised of 12 open and 15 closed questions was developed in order to survey student perceptions of their CALL and CM individual, group, and teacher-class production experience, and these students' perceptions of their own learning (see Appendix I).

3.4.2.12 Teacher post-treatment survey questionnaire.

A teacher questionnaire, called the Mystery Module and Concept Mapping Exercises Teacher Post-Questionnaire, was developed and comprised of 10 open and 7 closed questions. The questionnaire surveyed teacher perceptions of their students' CALL experiences and individual, group, and teacher-class CM production, and student learning (see Appendix H).

3.5 The data collection instruments of the treatment and control classes

The instruments that were administered commonly to the treatment as well as the control classes will be described in the sections that follow.

3.5.1 Pre- and post-treatment questionnaire administrations

3.5.1.1 Teacher questionnaires on AFL 10 principles.

In order to survey the participant teachers on their opinions on the tenets of the AFL approach, a questionnaire was developed (see Appendix K). The questionnaire included a brief definition of formative assessment, and asked the extent to which the respondents agreed with each of ten statements, which comprised the 10 principles of AFL. The statements were not identified as the 10 principles in order to glean respondent perceptions devoid of any possible previous bias. A Likert scale was used to rank the respondents' levels of agreement for each of the 10 principles, (where 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree). The teacher and student questionnaires were identical, with the exception of their identifying titles, and as noted below in the next section.

3.5.1.2 Student questionnaires on AFL 10 principles.

To survey the participant students on their perceptions of the 10 principles of AFL, the students were given a questionnaire that was identical to that of the teachers, except that the definition of formative assessment was simplified and directed to students (see Appendix L).

3.5.2 Pre-, post- and delayed post-treatment test administrations

While quantitative measures were not the focus of the study, pre-, post- and delayed-post quantitative measures in the form of FITBs and essay production tests were administered to the treatment and the control classes in order to identify trends. Additionally, these instruments contributed to the ecological validity of the study, since they were similar to the kinds of assessments these students could expect to encounter in the course of the school's usual formative assessment procedures. More information about these tests will follow in the next sections.

3.5.2.1 Essay tests.

Paper-and-pencil tests requiring all student participants to produce one-page essays were developed for the study in order to assess student writing competence in using the target contingency use modal when prompted using a conceptual rather than a form-focused prompt (see Appendix D). Nine parallel tests were developed in all in order to collect pre- and post-treatment data from the two treatment groups, the two control groups. The ninth test was developed at the request of the first treatment teacher, in case of need due to student absence for a period of less than a day. The tests instructions prompted the students to imagine an agreeable unreal situation, such as having unexpectedly won a prize. The essay instructions began with phrases such as 'Imagine….' or 'Suppose…', clearly marking the proposed situations as unreal and asking the students to write a one page essay indicating what they planned to do in the situation. Modals were intentionally excluded from the instructions so as to avoid providing the students with target language forms in advance of their writing.

3.5.2.2 Fill-in-the-blanks tests.

Parallel paper-and-pencil fill-in-the-blanks (FITB) tests were developed to assess student writing competence in producing correct forms when choosing between the provided forms *would* and *will*, in ten sentences (see Appendix E). Some of the sentences described unreal situations (requiring an answer of *would*) and the others described real situations (requiring an answer of *will*).

One FITB test was designed for administration in the pre-treatment phase of the pilot study. There was no post-treatment test required in the pilot study. In the actual study, four FITB tests were developed for pre- and post-administration to the Times 1 and 2 treatment and control groups. In addition, one FITB test was developed for Time 1 latecomers (in the case of absences of two days or less). No latecomer tests were developed or required in Time 2. One FITB test was developed for use as a delayed post-test for administration to all student participants concurrently. Therefore, a total of six FITB tests were developed for the study.

The item sentences included more *would* than *will* form answers to avoid students guessing answers should they assume there must be equal numbers of

forms in the test item answers. Moreover, it was thought that the *will* form was relatively easy for the students to master, and that the *would* form provided a more accurate level of challenge for these students enrolled in the Advanced I level, so that more items targeting the *would* form would be appropriate.

Thus, the pilot study test, and the three Time 1 pre- and post-treatment tests were designed each to elicit eight *would* and two *will* forms. In case the students should discuss the test content with other class members, the two Time 2 pre- and post-treatment tests were designed to elicit seven *would* and three *will* forms. One delayed post-treatment test was designed to elicit seven *would* and three *will* forms, as well.

3.6 Chapter summary

In this chapter the study research questions have been introduced. The research paradigm, statement of inquiry, mixed methods rationale and design, and the split-plot framework have been explained. The development of pedagogical materials, the data collection instruments, and the data analysis instruments have been presented herein.

In Chapter 4 the study methodology will be presented in detail, including the procedures followed for using the pedagogical materials and the data collection instruments, the teacher training, and the procedures used for their analysis.

CHAPTER 4 – METHODOLOGY: THE STUDY PROCEDURES

If the tongue had not been framed for articulation,
man would still be a beast in the forest.

Ralph Waldo Emerson[11]

In the previous chapter I have described the pedagogical and data collection instruments developed for use in the study, particularly those administered during the 4-Phase treatment period. In this chapter the methodology followed in order to apply those instruments in the study will be explained. I will first explain the procedure employed in a pre-study piloting to determine the feasibility of using CM production exercises in application of AFL in an L2 EAP classroom. Next the study design will be explained, beginning with the research split-plot design, the 4-Phase treatment timeframe, and the implementation procedures followed in applying each of the instruments in this research application of AFL in an L2 classroom setting. This discussion will begin with an overview of a piloting of the main study.

4.1 Preamble: Piloting the study

4.1.1 The pilot study participants

The participants of a pilot of the study included 1teacher and her class of 16 students in an Advanced I level EAP class at the school under investigation in the main study. The students were from various language backgrounds.

4.1.2 The pilot study procedures

The objective of the pilot study was to ascertain to what extent the proposed main study procedures, the concept maps, would be feasible in an L2 classroom since they were innovative.

First, a one page essay test eliciting the contingent use modal was administered to the students, followed the administration of a 10-item FITB test. Second, the students were given a brief introduction to concept mapping, using the blackboard in their classroom. Next each of the students in turn took part in a concept mapping exercise. Each student was shown a sample CM in

[11] This quotation, besides underscoring the importance of speech, also serves as an example of the contingent use modal, *would* to describe an unreal context.

Word software on a laptop computer, in which sample sentences using *would* and *will* appeared in oval shapes. In this CM the *would* and *will* sentences appear on opposite sides of the page. This CM template appears in Appendix J. The students were instructed to manipulate the template CM in any way they wished in order to try to understand the usage of *would* and *will* visually. The student CM productions were saved in a computer file.

Next, the teacher wrapped up the exercise by explaining the grammar point to the students in a question-and-answer discussion and using the blackboard to illustrate usage. Finally, the students completed a survey to garner their perceptions of the experience.

4.1.3 Analysis of the results of the study pilot

Each essay test was analyzed and its ratio of number of correct modal forms per number of obligatory contexts was determined. These results were calculated as ratios and also as percentages. Descriptive statistics were calculated for the FITB tests. Frequency counts were made of the closed questions in the survey data, and the results of open questions recorded and categorized as being positive or negative in nature. The CM productions were analyzed to determine if a more elaborate use of concept maps might be feasible in the larger study. The decision was made to do so based on the results of the piloting.

The results suggested the following:

1. the usage of *would* and *will* presented a challenge for these learners in essay tests
2. the usage of *would* and *will* presented a lesser challenge, but a challenge nonetheless, in FITB tests
3. a high level of student interest and facility in online CM production was apparent

Based on these results, it was decided that the essay and FITB tests would be used in the main study. In addition, it was decided to use educational concept mapping software for the main study since the simpler concept maps produced in Word software were easily accessible to the learners, but would unduly limit their CM production applications. Finally, it was learned from the

pilot study that the student participants appeared to be very interested and motivated in the AFL application and in the CM production activities.

4.2 Treatment class procedures

4.2.1 Teacher training in AFL methods

The purpose of the sessions was to familiarize the teachers with the AFL approach as well as to train them in its application in an L2 classroom setting. As part of the training and prior to the treatment period, the teachers read an article in which the AFL approach was described (Colby-Kelly & Turner, 2007) and attended a presentation based on that article. Specific AFL training was then given to the 'treatment' teachers in two separate sessions led by the researcher.

In addition, the teacher training consisted of two separate sessions led by the researcher in which each treatment teacher attended. The sessions lasted approximately one hour, and each occurred during the week prior to the start of each treatment period. Right before the class where CM production exercises took place, the researcher gave each teacher ½ hour of refresher training. The study rationale and overall design was explained in the training sessions, as was the grammar feature to be used in the investigation, the contingency use modal form *would* (for denoting unreal contexts, as opposed to, for example the incorrect form **will*). In addition, the learner-centered CALL, CM and the data collection schedule were explained; the study instruments and pedagogical materials were reviewed to familiarize the teachers with the data collection tests, questionnaires and teacher and student surveys of their perceptions following interventions. Concept mapping theory was discussed and several concept maps and printouts from an online tutorial were used to illustrate it.

In preparation for the teaching application of AFL that would occur in the Phase 4 teacher-class CM production exercise, the teacher training included guided questioning and knowledge scaffolding techniques.

The AFL 10 principles were not referred to nor discussed in the training sessions, in order to avoid possible bias in advance of the teachers' completion of the Teacher questionnaire on AFL principles.

As noted in more detail in Chapter 3, the study employed a sequential, exploratory mixed methods research design (Creswell & Plano Clark, 2007). The study design and the four treatment phases are illustrated in greater detail in Figure 8. Figure 8 illustrates the CALL and CM production exercises as

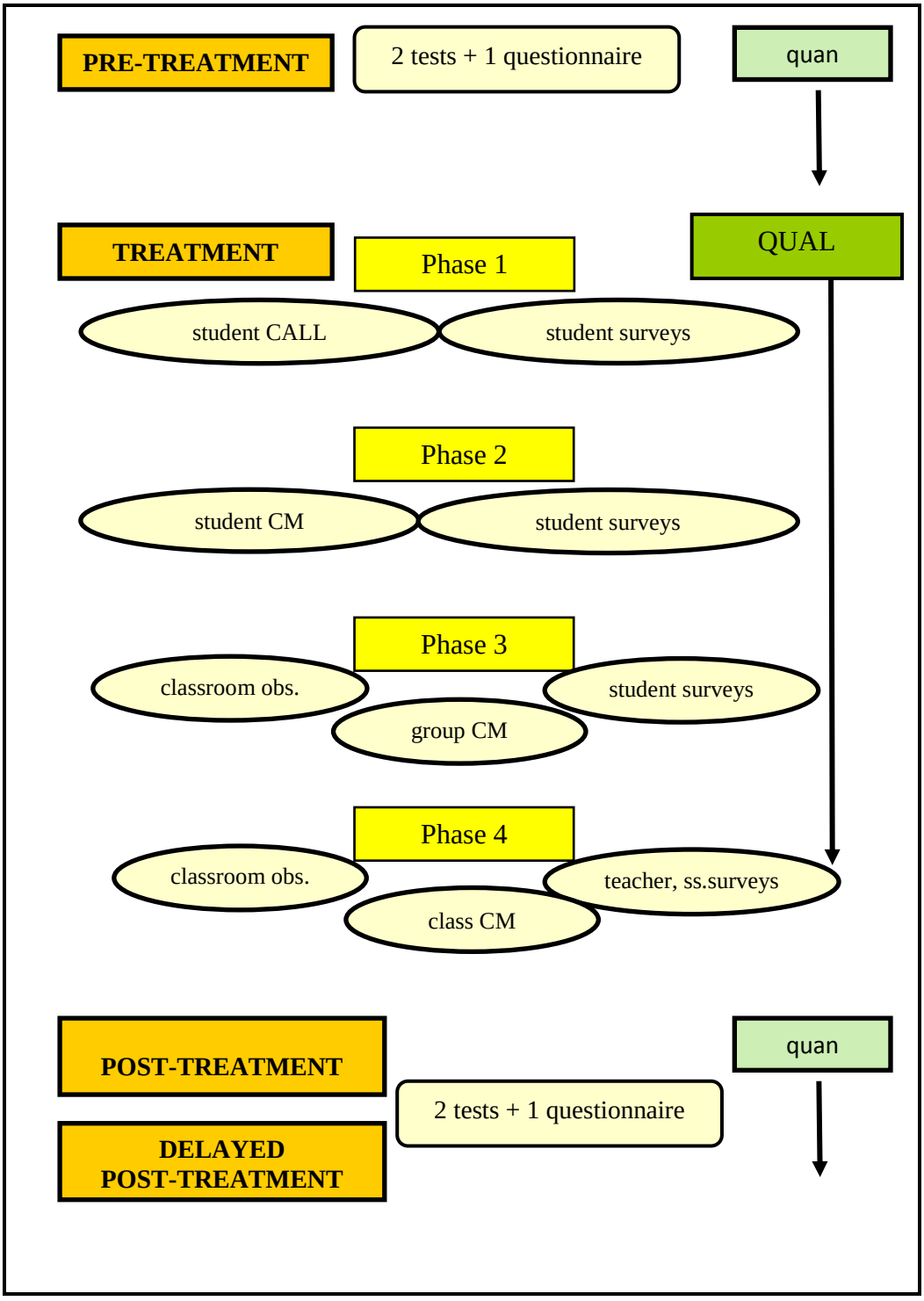


Figure 8 Sequential exploratory mixed methods research study design for treatment groups

well as the pre-, post- and delayed post-treatment data collections.

4.2.2 Phase 1: Individual student CALL

4.2.2.1 The English Mystery Module.

In Times 1 and 2 the treatment group teachers each introduced the topic of the forms *would* and *will,* to their students. Each teacher mentioned that the students had found these forms challenging in the past. Then the CALL exercises, The English Mystery Module, were introduced to the students and the students were shown printed copies of the CALL pages in order to familiarize them with the materials. In particular, the hints page, quizzes, answer grids, self-assessments, progress checklists, and guided questions, accessible using hyperlinks, were shown to the students.

In keeping with the AFL approach, the students were encouraged to self-direct their learning by using the module resources that they found worked best to help them individually in the progress of their learning.

The students were told they should not worry about making mistakes, but view mistakes as helpful indicators of 'where they were going' in their learning progress. They were encouraged to focus on the *process* of their learning rather than to be concerned about how many 'right answers' they had produced. Also they were told to focus on attempting to understand how this grammar point functioned in English since the module would serve as a preparation for subsequent exercises. (However, the subsequent concept mapping exercises were not introduced at that point in the treatment.)

Both teachers emphasized to their students that it was very important they participate in this stage of the study, since this was a necessary preparatory exercise prior to the following exercises.

The students were told that the module had been made available to them in the school's computer laboratory. At the request of both treatment group teachers, the researcher participated in this introduction and answered any technical questions that arose. The students were instructed to practice using the module in the school's computer lab over the next four days as often as they wished.

4.2.2.2 A Quizzical Mystery Quiz.

At the same time as The English Mystery Module was introduced, the students were also instructed on how to access the interactive self-assessment quiz called A Quizzical Mystery Quiz, in the school's computer lab. The way the quiz worked was explained, including its optional drop-down hints for use before or after choosing either *would* or *will* to complete each of ten sentences. The instructions from the quiz have been shown in Figure 9.

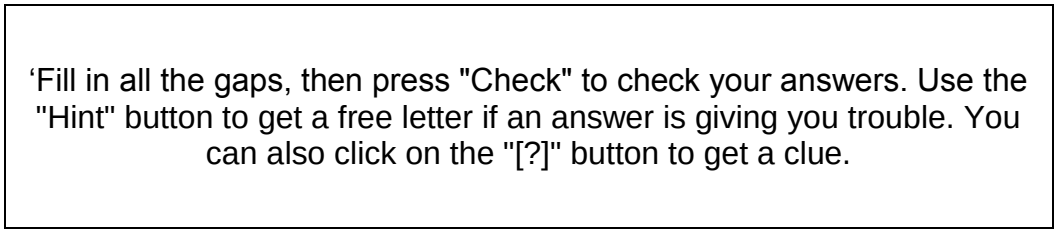


Figure 9 Instructions for interactive learning 'hints' from A Quizzical Mystery Quiz

The students were instructed to practice using the English Mystery Module and the Quizzical Mystery Quiz, in the computer lab outside of class time, over the subsequent four days. The teachers stressed the students should access and use the CALL module and interactive quiz as often as they wished in that time period, and then to print out the Module pages five and ten (which included self-assessments and metacognitive learner reflections on the grammatical concept), as well as the Quiz page for submission on the fourth day.

The students were informed that these pages were requested for submission mostly to ascertain participated in the exercises, rather than for scoring at that time. This request for submissions resonated with the school's practice of regularly assigning a considerable amount of homework for submission on an ongoing basis.

4.2.2.3 CALL data collection.

Four days after the English Mystery Module and Quizzical Mystery Quiz were assigned, the students submitted printed copies of pages five and ten of the module, and the one page of the quiz (see Appendices A and B).

4.2.2.3 Student surveys on CALL.

Following their CALL experience, the students were asked in a brief, two-question survey about their perceptions of the effectiveness of the CALL exercises in helping them learn. The surveys were submitted the same day as the three CALL printout sheets.

4.2.3 Phase 2: Individual student online CM production exercise.

Educational CM software was installed on the computers in the school's computer lab and the treatment students were instructed as to how to gain access to it. At the teachers' request, the researcher explained to the students what concept maps were, illustrating with printed pages from the software's online tutorial and also with examples of concept maps from the literature. The students were told to use the software in any way they wished, to use their imaginations to create a CM depicting their understanding of the *will* and *would* form usage in English, and to make as many maps as they wished before printing a final CM to be submitted four days later. The students were instructed that there was no right or wrong answer in the exercise, but that the focus was on the *process* of trying to understand and learn this grammar point.

Over four days, the students worked on the CM software in the computer lab and each produced one printed copy of a CM.

4.2.3.1 Individual CM production data collection.

Four days after the online CM exercise was assigned, the students submitted printouts of their individual concept maps.

4.2.3.2 Student surveys on individual CM production

Following their individual CM production experience, the students were asked in a brief, two-question survey about their perceptions of the effectiveness of the exercise in helping them learn. The surveys were submitted the same day as were their concept maps.

4.2.4 Phase 3: Co-constructed student group CM production

In this phase of the treatment, the students worked in groups to produce a group, co-constructed, paper-and-pencil CM. The previously collected individual concept maps were returned to the students who had created them

for use during the exercise. Each treatment teacher divided the class into three groups of three to four students in each. The groups were instructed to consult and discuss amongst themselves in order to create a group CM. They were also told to refer to their individually produced CM printouts, if they found that useful. Each group assigned one student to draw the group map, following the group's instructions. Sequential periods of twenty minutes each was allotted for the group exercise, which was done in a separate room away from the students' regular classroom.

4.2.4.1 Classroom observation field notes

The group CM production exercise was digitally audio recorded by the researcher, who also observed each group exercise. Field notes were taken drawing on Turner's (2009) classroom observation grid and following Tesch (1990) to describe the group members' interactions.

4.2.5 Phase 4: Co-constructed teacher-class CM production

In Phase 4 of the treatment period each treatment class as a whole produced a class concept map in concert, with their teacher guiding them. To begin, each student group presented their group CM to the teacher and the rest of the class, using their group CM productions, which had been reproduced and printed on an acetate sheet and projected on an OHP. The groups also referred to poster sized CM productions which had also been reproduced, and enlarged for display on the classroom walls for the perusal of the other students. In their presentations, the student groups explained how and why they had come to produce their CM in the manner they had done. The student group presentations were accompanied by questions and comments from the teacher and classmates.

After each student group had presented their group CM, the teacher took an unused acetate page and proceeded to draw a CM based on the suggestions of the class. The treatment teachers encouraged class discussion about target and non-target usage of *will* and *would.*

Following the AFL approach, the teachers did not provide the students with the rule of target usage, but rather attempted to lead the students towards an understanding of it through guided questioning and knowledge scaffolding

techniques, encouraging students to explore various options in discovering the 'rule' of this usage. Near the end of the lesson, the teachers provided the class with an explanation of how the contingency modal *would,* and the future aspect form *will* is used in English.

4.2.5.1 Student surveys on the treatment experience

On the last day of the treatment period, after the teacher-student CM production exercise had been completed, the *Mystery module and concept mapping exercises post-questionnaire* was administered to the students in order to survey their perceptions about the treatment experience. The item format of the questionnaire allowed for students to respond to closed questions as well as to offer comments to open questions.

4.2.5.2 Teacher surveys on the treatment experience

At the same time as the students completed the student treatment questionnaires described above, the treatment group teachers completed the *Mystery module and concept mapping exercises teacher post-questionnaire* which, as was the case with the student survey, combined closed and open questions in order to survey teacher perceptions about the treatment from a teaching perspective.

4.3 Control class procedures

During the period in which the treatment groups were undergoing the AFL treatment, the control groups followed their regular curriculum.

The regular instruction included an ad hoc mention of the use of the target modal *would* and *will* as the need arose, which was seldom, according to the control group teachers. In addition, the students had access to a writing textbook and accompanying exercise book, which the teachers were using on a trial basis (Baker, 2003; Hogue, 2003). At the time of the study the treatment and control group teachers reported they had not or rarely used those textbooks.

The school has described the course content at these students' Advanced I level in this excerpt from the school calendar:

> At this level, students focus on refining skills required to function in a university or professional setting. They take notes, discuss, answer

questions, and write summaries based on lectures and video documentaries. They also participate in various types of oral presentations and debates; present reports based on articles from magazines; and lead discussion groups. Composition skills are also stressed. (103)

Figure 10 shows the study design as it was implemented in the control groups.

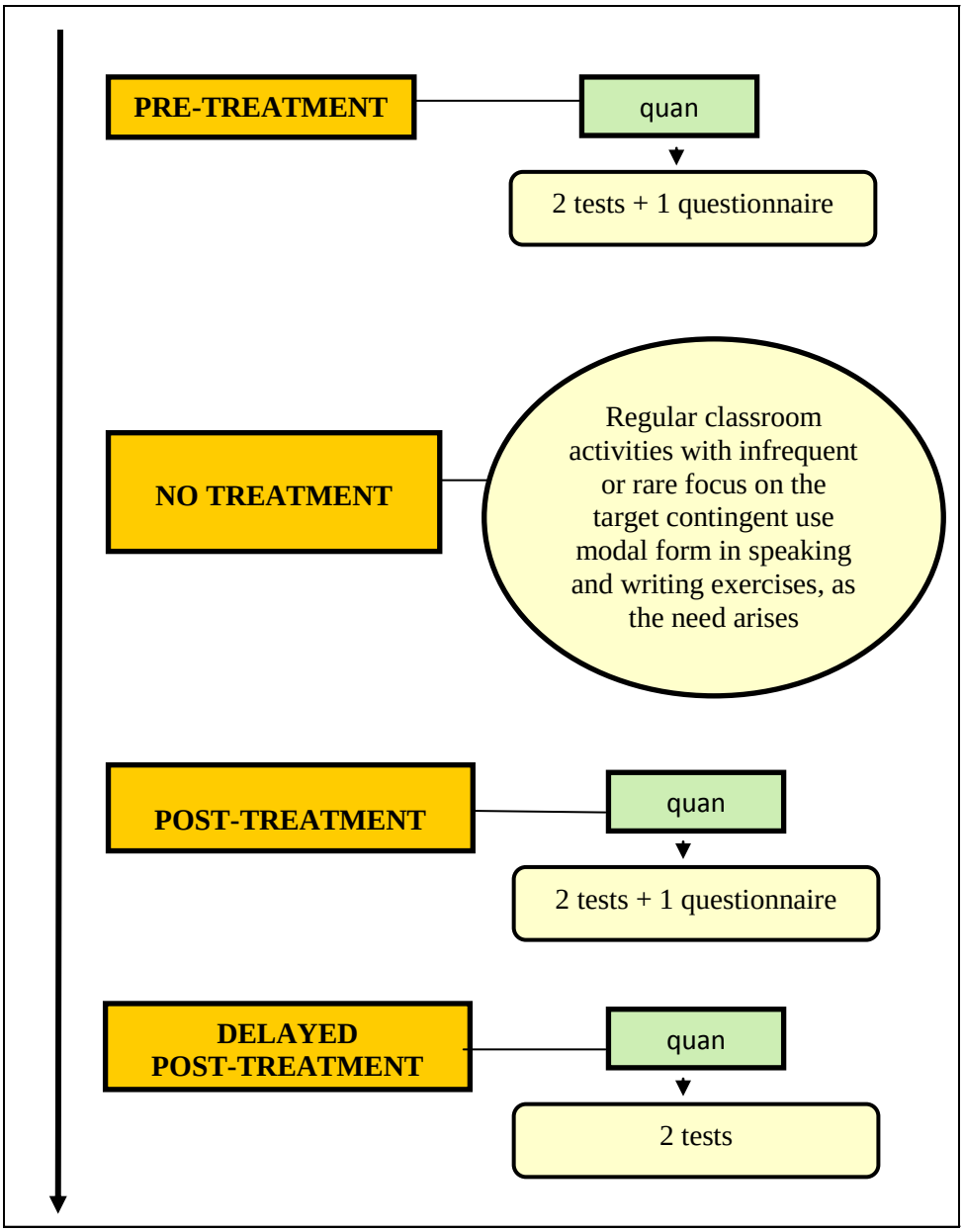


Figure 10 Sequential exploratory mixed methods research study design for control groups

4.4 Procedures of the treatment and control classes

The following pre- and post-treatment data collection occurred on the first and last days of the study in treatment Times 1 and 2, and were administered to both the treatment and the control groups.

4.4.1 Student questionnaires on the AFL 10 principles

All participant students were administered questionnaire surveys to ascertain their level of agreement with the AFL 10 principles. The Student Questionnaires were completed by the students in 10 to 15 minutes of class time.

4.4.2 Teacher questionnaires on the AFL 10 principles

All participant teachers were administered questionnaire surveys to ascertain their level of agreement with the AFL 10 principles. The Teacher Questionnaires were completed by the teachers in 10 to 15 minutes of class time, or shortly thereafter for subsequent submission to the researcher.

4.4.3 Essay tests

All participant students working individually, completed paper-and-pencil tests requiring they produce a one-page essay in class. The students were given 20 to 30 minutes in which to write the essays. Access to dictionaries or other resource materials was not allowed during test administrations.

4.4.4 Fill-in-the-blanks tests

All participant students working individually, completed paper-and-pencil FITBs tests in class. The students were given 20 to 30 minutes in which to complete the tests. Access to dictionaries or other resource materials was not allowed during test administrations.

4.4.5 Delayed post-treatment test administrations

Delayed post-treatment parallel essay and FITB tests were administered on one day, simultaneously, to all students from each of the treatment and control groups who were still enrolled in classes at the school five months following the final day of treatment Time 1, and three months following the final day of

treatment Time 2. All of these students had progressed to the school's next and final level, to Advanced II EAP.

4.5 Data analysis overview

The study data were analyzed quantitatively and qualitatively. In addition, some of the data collected from the same instruments were analyzed by mixed methods, that is, quantitatively as well as qualitatively. The qualitative measures were predominant in this study, while the quantitative measures took a secondary place as is reflected in the study design notation, quan → QUAL → quan (see Figure 10).

4.5.1 The instruments analyzed by mixed methods

Data from these instruments were analyzed both quantitatively and qualitatively:

- CALL: *The English Mystery Module*
- Individual student concept maps
- Group concept maps
- Teacher-class concept maps
- Post-treatment student surveys
- Post-treatment teacher surveys

4.5.2 The instruments analyzed by qualitative methods

Data from the following instruments were analyzed qualitatively:

- the observation field notes
- selected transcriptions from classroom observations
- teacher incidental comments

The data methods used for analysis of the study data will be explained in the following sections.

4.5.3 The instruments analyzed by quantitative methods

Data from the following instruments were analyzed quantitatively:

- the essay tests
- the FITB tests
- the teacher 10 principles questionnaire surveys
- the student 10 principles questionnaire surveys

4.6 The mixed methods analysis procedures

4.6.1 CALL: The English Mystery Module

The metacognitive self-reflection of the module (page 5) was analyzed and frequency counts of the number of completed exercises reported. The students' metacognitive comments were reported, their content analyzed, and discussed.

4.6.2 Individual student CM production.

Several concept map analysis protocols were considered for use in the study CM data analysis (Daley, 2004a; Gouveia & Valadares, 2004; Khamesan & Hammond, 2004; O'Connor et al, 2004; Takeya et al, 2004). These proved to be highly technical and with respect to the present study, would be inappropriate. This study has used concept maps in investigating the effect of AFL procedures in an L2 classroom setting, and as such the student participants were instructed to focus on learning about a point of grammar through the use of an illustrative CM to help them, not to produce a technically complex, structured diagram. For these reasons it was decided that a qualitative checklist approach to the study CM analysis protocol would be best. Figure 11 illustrates the analysis protocol instrument that was developed, the *Concept Map Binary Qualification Checklist.*

The Checklist is comprised of 10 features of concept maps, both generally present and, as noted, those found in the CM production collected in this study. The purpose of the concept map analysis is to create a 'picture' of how these students used them in order to gain insight into the effect of the concept mapping exercises on individual learning. Therefore a binary (yes/no) checklist was developed for the analysis with this in mind. It is based on the literature and also draws on the features that emerged from the CM data collected in the study, in conformity with a mixed methods approach (Creswell & Plano Clark, 2007; Tesch, 1990).

Using the qualification checklist, the student concept maps were analyzed by noting and recording which of the 10 CM features was present in each map of the TG1 and TG2 treatment groups. Frequency counts were made of the features present in each map as well as the frequency of features combined in

each group. In addition to this quantitative analysis, the CM production was analyzed for qualitative discussions.

Does the concept map include …	Y/N
1. a title	
2. an expression of a main grammatical concept	
3. the words *would* and *will* represented in a square, oval, discrete space, etc.	
4. an example of grammatical usage	
5. a grammatical explanation (that may or may not be correct)	
6. idea linking (through lines or arrows or a Venn drawing)*	
7. cross linking of ideas	
8. a linear design	
9. a non-linear design	
10. pictures to illustrate a concept visually	

Figure 11 The concept map binary qualification checklist

* A Venn drawing is a diagram of two intersecting circles side by side in which text may be inserted, such as in Figure 12 on page 128 of Chapter 5.

4.6.3 Group CM production

Group concept maps are frequently analyzed in the literature in terms of levels of cooperation or, as in the work of O'Connor et al (2004), if there is evidence of teamwork and team cognitions in group concept mapping activities. Such queries are beyond the scope of this study, which situates itself in the realm of AFL procedures by means of a group CM exercise. For this reason it was determined that it would be appropriate to use the Concept Map Binary Qualification Checklist to analyze the group concept maps data in addition to its use in analyzing individual concept maps. Another important reason for choosing to use the same checklist is for comparison purposes across the CM production exercises.

These concept maps were analyzed in the manner described above.

4.6.4 Teacher-class CM production

For all of the reasons stated above, the Concept map binary qualification checklist was used to analyze the concept maps produced in the teacher and class CM exercises. The concept maps were also analyzed in the manner described above.

4.6.5 Post-treatment student surveys

Thirteen items from the student Mystery Module and Concept Mapping Exercises Post-Questionnaire were closed questions and two items required closed, one-word answers (such as that requesting learning students identify their learning style). These responses were analyzed and reported as frequency counts.

Fourteen open-ended items of the same questionnaire were scrutinized and those results pertinent to answering the study research questions reported. Additionally, the student surveys on CALL and those pertaining to CM were collapsed with the Mystery Module and Concept Mapping Exercises Post-Questionnaire results for reporting purposes.

4.6.6 Post-treatment teacher surveys

The analysis of the Mystery Module and Concept Mapping Exercises Teacher Post-Questionnaire followed the same protocol as that described for the student questionnaire.

4.7 The qualitative analysis procedures

4.7.1 Observation field notes

The field notes taken in classroom observations were perused and data relevant to answering the research questions were documented.

4.7.2 Selected transcriptions from classroom observations

In addition to the field notes, selected student and teacher comments useful in enriching and supporting the classroom observation data were transcribed and reported. The transcription protocol used will be found in Appendix M. Drawing on previous research (Brooks, 2009: Colby-Kelly & Turner, 2007; Turner, 2006, 2009) emergent categories of the data characteristics were identified and reported.

4.7.3 Teacher incidental comments

The teacher incidental comments were perused and those relevant to answering the research questions were documented.

4.8 The quantitative analysis procedures

4.8.1 The essay tests

The essay tests were scored independently by two raters. One of the raters was the researcher and the other was a high school teacher experienced at the advanced ESL level. The scoring was done by identifying all instances requiring a contingent use modal form (obligatory instances) and by recording frequency counts of how often a grammatically correct (target) form had been produced in each essay. Thus the scores were recorded as ratios as in the following:

target forms produced : obligatory contexts

In order for the scoring to be as objective as possible, only those clear instances of obligatory contexts were included. In addition, any target modal forms (other than *would*) were accepted, such as *might*, *could*, and so on. Similarly where appropriate, alternate forms (other than *will*) were accepted, such as *going to*. In the case of differences in the initial scores between the raters, they arrived at agreed scores through discussion. The results were analyzed with frequency counts of ratio data reported as ratios and as percentages.

4.8.2 The FITB tests

The same two raters independently scored the answer keys for the FITB tests. There was no disagreement among the raters, and the keys were used to score the tests.

4.8.3 The essay and FITB tests

For these tests descriptive statistics were calculated using SPSS statistical software 14.0 to determine group means and standard deviations and for comparison across groups.

In addition, two-tailed t-tests were performed on the data to test for differences in pre- to post- and delayed post-treatment tests. Independent

samples T-tests were done to compare groups. A .05 alpha level was set for these tests at all times.

4.8.4 The teacher and student 10 principles questionnaire surveys

These questionnaire responses were calculated and reported as frequency counts across the 4-ranking Likert scale employed.

4.9 Quantitative and qualitative analyses triangulation

The study employed a sequential, exploratory mixed methods research design following Creswell & Plano Clark (2007). The quantitative and qualitative data collected were analyzed and triangulated (Tashakkori & Teddlie, 2003; Creswell & Plano Clark, 2007). Various results were discussed in concert, in triangulation of the data set as a whole.

4.10 The results interpretation

Once reported and then triangulated, the results were interpreted in conformity with the standards of the literature. The results will be reported in the next chapter.

CHAPTER 5 – PRESENTATION OF RESULTS

We have all a better guide in ourselves,
if we would attend to it,
than any other person can be.

Jane Austen[12]

5.1 Chapter overview

In this chapter I first present the results by instrument since some of the instruments were analyzed by qualitative means, some by quantitative means, and some were analyzed by means of both qualitative and quantitative or mixed methods (MM) analysis. First the instruments analyzed by MM will be presented and the results reported, followed by those of the instruments analyzed by qualitative means. This will be followed by the presentation of the results of the analysis of the instruments analyzed quantitatively.

Figure 12 illustrates the categorization of the data presented by instrument and analysis method. That is, within the converged circles will be found a list

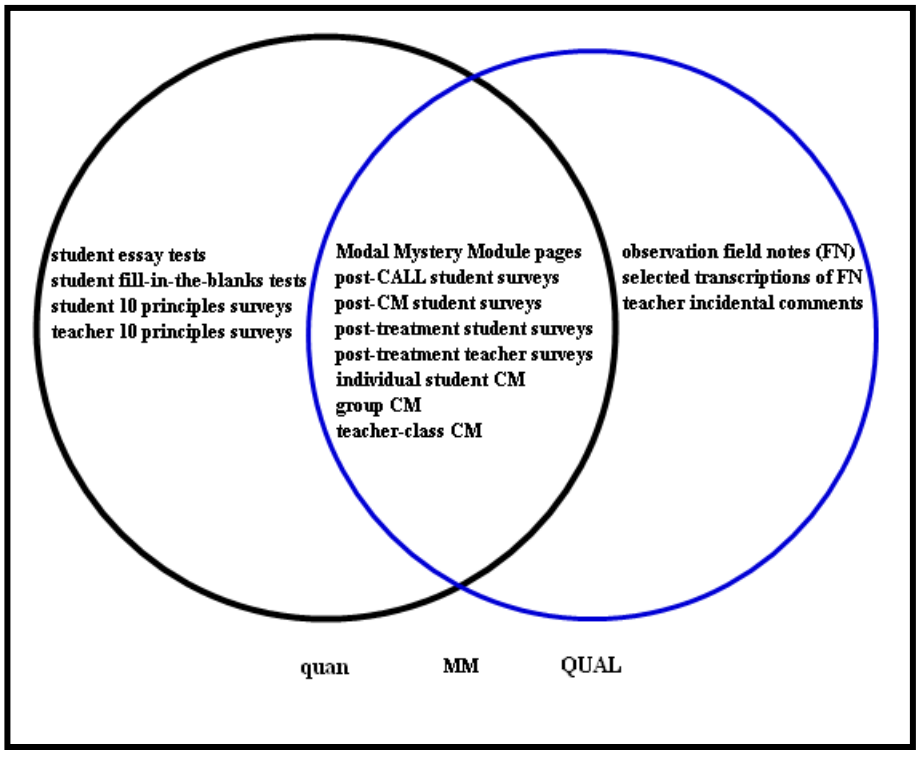


Figure 12 Separate and combined methods of instrument data analysis

[12] Jane Austen's quote here is included to show the usage of *would* in an unreal context, and as an echo of AFL's focus on learner self-reliance.

of the instruments analyzed by MM, and in the circle on the right the instruments analyzed by qualitative methods is found, and in the circle on the left will be found the instruments analyzed by quantitative methods.

Since the overriding focus of the study is qualitative, the main data reporting reflects this. Due to their being of lesser importance in the study, quantitative analyses will be reported in lesser degrees.

5.2 Results of analysis of instruments by mixed methods

The design notation for the data collection procedures of this study was quan → QUAL → quan. In reflection of this notation and to best answer the study research questions, the following data was analyzed by mixed methods. Thus, the following instruments were analyzed both quantitatively and qualitatively:

- CALL: The English Mystery Module
- Individual student concept maps
- Group concept maps
- Teacher-class concept maps
- Post-treatment student surveys
- Post-treatment teacher surveys

5.2.1 CALL: The English Mystery Module

Few students submitted pages 5, 10 and 11 (the Quizzical Mystery Quiz), presumably because submission of the pages had been presented as being on a voluntary basis. Indeed, the students had been instructed to treat the exercise as a self-directed online practice. Page 5 of the Module appears in Figure 13.

Some excerpts of the completed metacognitive self-reflection activity representative of the Module pages that were submitted by the learners, will be reported in the following section.

5.2.1.1 Qualitative results of CALL.

Of the answers to the metacognitive self-reflection activity that the students provided, many students from the Time 1 (TG1) and Time 2 (TG1) treatment classes suggested confidence in having already mastered the target form usage.

 **Complete these thoughts as they apply to you!** 
I find using *will* and *would* easy because...
I find using *will* and *would* challenging because...

Figure 13 Metacognitive student self-reflection from page 5 of the Mystery Module

Those answers follow, completing the stem '*I find using* will *and* would *easy because'*:

- TG1 – S2 *I know the rules of using both.*
- TG1 – S8 *I know a lot about it*
- TG2 – S38 *i leard it in high school* [sic]
- TG2 – S39 *I know these kind of knowledge pretty good*

In the Time 2 treatment group, TG2, several of the responses to this stem show evidence of learner metacognition. They are:

- TG2 – S31 *I listened to people how they used it*
- TG2 – S34 *it can always see the different meaning in the whole sentence.*
- TG2 – S36 *would like to, would have to, is standard, "would" also show what he/she likes and "will" is always used for the future, or what he/she want to do.*
- TG2 – S43 *of the explanation – contingent*

In completing the second stem '*I find using* will *and* would *challenging because,'* some of the reflections in both treatment groups indicated learner confusion as to the usage of these forms. Their reflections follow:

- TG1 – S7 *Some time I have diffeculty to find which is wright.* [sic]
- TG1 – S12 *I can't identify the condition*
- TG2 – S32 *I can not exactly know when I should use it*
- TG2 – S33 *I CAN'T KNOW WHEN I CAN USE IT* [capitals in the original]
- TG2 – S41 *they confuse me in some places*

Two of the reflections in response to this stem suggested some metacognition on the part of the learner. They were:

- TG1 – S9 *will and would are similar*
- TG2 – S40 *I am not native speaker*
- TG2 – S43 *sometimes the conditional context won't give you the response*

5.2.1.2 Quantitative results of CALL.

Table 2 shows the number of students who completed and submitted the self-reflection activity and those who did not.

Table 2: Frequency of student responses to metacognitive self-reflection

	Responses		
	Answered	Unanswered	Missing
I find using *will* and *would* easy because...			
TG1	2	9	(3)
TG2	3	8	(3)
I find using *will* and *would* challenging because...			
TG1	3	5	(3)
TG2	5	6	(3)

TG1 (*n*=11), TG2 *(n*=11)

5.2.2 Individual student CM production results

5.2.2.1 Qualitative and quantitative results of the Concept Map Binary Qualification Checklist analysis.

In the Time 1 treatment class (TG1), 11 students submitted 12 concept maps (1 of the students submitted 2 maps). In the Time 2 treatment class (TG2), 9 students submitted 11 maps (2 students submitted 2 maps). All of the maps were analyzed separately, with the exception of 2 of the maps submitted by S33 which were collapsed since they were clearly intended to be parts of a set (one was labeled *will* and the other *would* and they were structurally identical). First, the TG1 and TG2 CM productions were analyzed by means of the Concept Map Binary Qualification Checklist, which was introduced in the previous chapter.

It will be recalled that the Checklist, which was developed specifically for this study, was not intended to *score* the CM productions but rather to characterize or *qualify* them. Two sample concept maps produced by individual students in the TG1 and TG2 classes have been reproduced in Figures 14 and 15 to illustrate how these learners have interpreted their task relative to the Checklist. (All of the concept maps produced in the study are available from the researcher). The results of the Checklist analyses of each of the treatment classes will follow.

Frequency counts of 10-item checklist various CM features have been reported on Table 3, and have been provided here in order to contribute part of the picture of a larger, triuangulated reflection of the usefulness to these learners of this AFL application. Thus the numbers here reported do not tell the whole ‘story’; they *do* portray the kinds of CM features these learners have employed in their attempts to understand the contingent use modal. For example, clearly a greater number of CM features does not necessarily indicate a better CM; the quality of the features present would better determine worth. That said, some interpretation of this quantitative data is possible. From Table 3 it can be seen that the TG1 class included more CM features in their individual student CM production (64) than did the TG2 class (49), whose

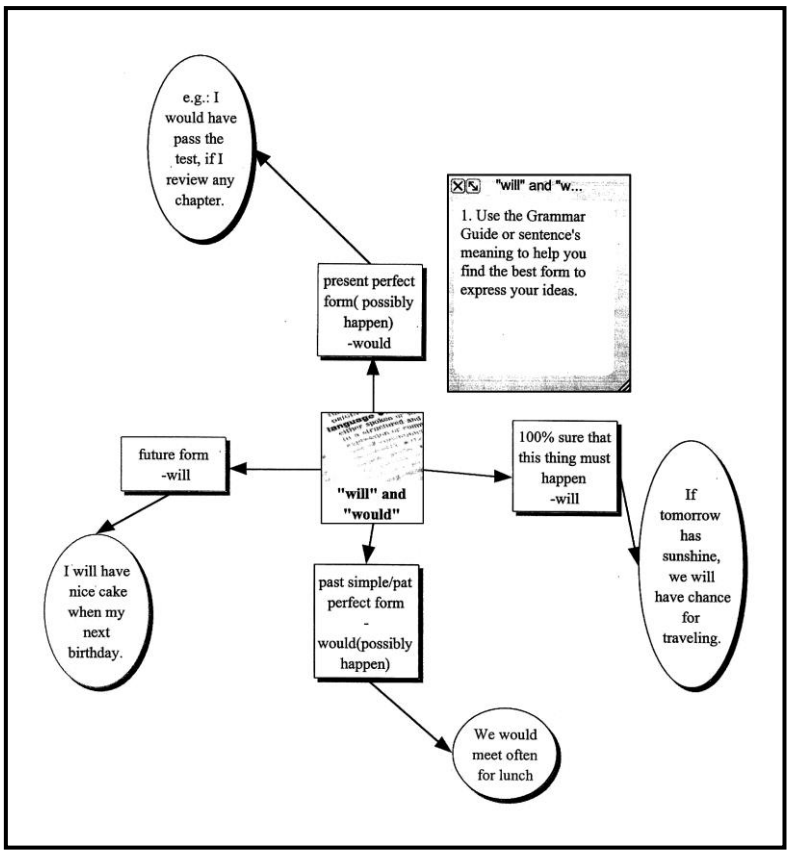


Figure 14 The concept map of an individual student (S5) from the TG1 class[13]

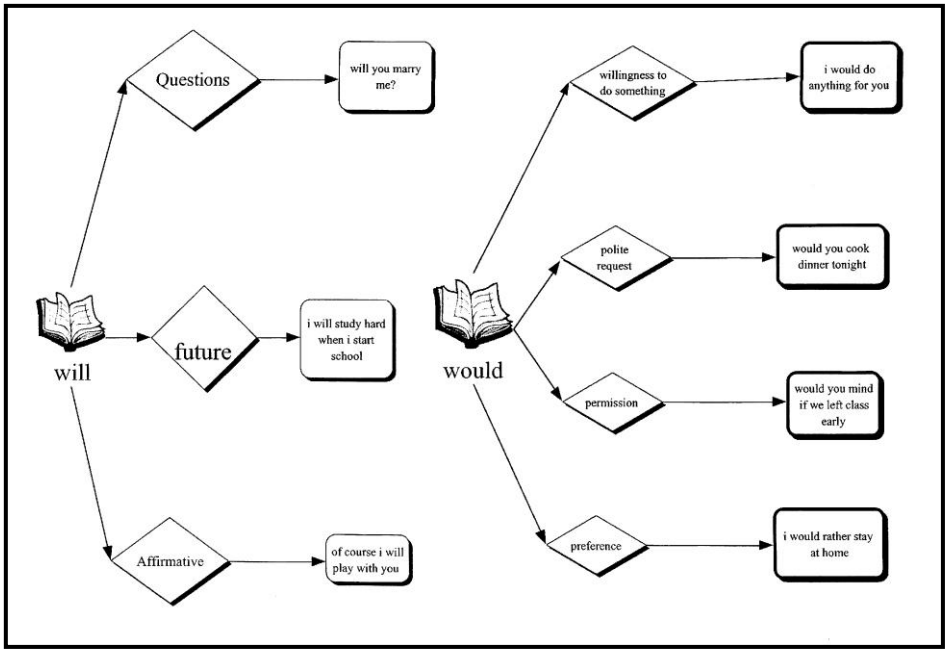


Figure 15 The concept map of an individual student (S31) from the TG2 class

[13] This concept map was made with a software icon that is purposefully blurred for effect.

Table 3: TG1 individual student CM production analysis using a concept map binary qualification checklist

Does the concept map include …	S1	S2	S3	S4	S5	S6	S7	S8	S9	S12	S12	S13	Class totals
1. a title	✓	✓	-	-	-	-	-	-	-	✓	✓	-	4
2. an expression of a main grammatical concept	✓	-	✓	✓	-	-	✓	✓	✓	-	-	✓	7
3. the words *would* and *will* represented in a square, oval, discrete space, etc.	✓	✓	-	-	-	✓	✓	✓	✓	✓	✓	✓	9
4. an example of grammatical usage	✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	-	8
5. a grammatical explanation (that may or may not be correct)	✓	✓	-	✓	-	✓	✓	✓	✓	✓	-	✓	9
6. idea linking (through lines or arrows or a Venn drawing)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	11
7. cross linking of ideas	-	-	-	-	-	-	-	-	-	-	✓	-	1
8. a linear design	-	✓	-	-	-	✓	-	-	-	-	✓	-	3
9. a non-linear design	✓	-	✓	✓	✓	-	✓	✓	✓	✓	-	✓	9
10. pictures to illustrate a concept visually	-	-	-	✓	✓	-	✓	-	-	-	-	-	3
Total CM features per student	7	6	4	5	4	5	7	6	5	5	5	5	64

(n=11) Two concept maps were submitted by S12, and both maps were analyzed.

quantitative CM analysis is portrayed in Table 4. The two most prevalent features of the TG1 class is *idea linking*, followed by including *a grammatical explanation, a non-linear design* and *an example of usage.* In the TG2 class, the most prevalent features of their CM production was the inclusion of *the words* would *and* will *represented in a square, oval, discrete space, etc.*, followed by *idea linking, an example of usage* and *a non-linear design.*

The least frequent CM attributes present in the maps of the TG1 class included *cross linking* and *pictures to illustrate a concept visually*. The least frequent CM attributes present in the maps of the TG2 class were *cross linking* and *an expression of a main grammatical concept*.

The individual student CM production in both the TG1 and TG2 classes showed evidence of metacognitive processing of the language feature, which could be seen by the content, for example in the form of explanations and examples of usage, in the maps submitted. While some were complex and some were simple, all had addressed the usage issue at hand. Some used a linear design (not always appreciated in the CM literature) which clearly portrayed the use of *would* and *will* as it was understood by the learners, others used a non-linear design to advantage.

A considerable difference in the two classes was in the inclusion of *an expression of a main grammatical concept* in much of the CM production of the TG1 group, while none of the TG2 group included that feature. Similarly, some in the TG1 class added a *title* to their CM production, and none of the TG2 class did so.

5.2.2.2 Five case studies of individual CM productions analyzed qualitatively and quantatively.

Five concept maps created by individual students from both the Time 1 and 2 treatment classes have been selected here for presentation, for the purposes of looking more closely at how these students expressed themselves using concept maps, and how the concept map production exercise may have affected their learning. Each has been chosen to serve as an example of a distinctive feature present in the concept map. Each will be discussed in triangulation across different data sources.

Table 4: TG2 individual student CM production analysis using a concept map binary qualification checklist

Does the concept map include …	S31	S32	S32	S33	S34	S38	S39	S40	S41	S44	Class totals
1. a title	-	-	-	-	-	-	-	-	-	-	0
2. an expression of a main grammatical concept	-	-	-	-	-	-	-	-	-	-	0
3. the words *would* and *will* represented in a square, oval, discrete space, etc.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
4. an example of grammatical usage	✓	✓	✓	✓	✓	✓	✓	✓	-	-	8
5. a grammatical explanation (that may or may not be correct)	-	-	-	✓	✓	-	-	✓	-	✓	4
6. idea linking (through lines or arrows or a Venn drawing)	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	9
7. cross linking of ideas	-	-	-	-	-	-	-	-	✓	✓	2
8. a linear design	✓	-	-	-	-	✓	-	✓	-	-	3
9. a non-linear design	-	✓	✓	✓	✓	-	✓	✓	✓	✓	8
10. pictures to illustrate a concept visually	✓	-	-	✓	-	✓	✓	✓	-	-	5
Total CM features per student	5	4	4	6	4	5	5	7	4	5	49

(n=9) Two differently structured concept maps were submitted by S32 and both were included in this analysis; two identically structured concept maps were submitted by S33 and they were collapsed for analysis.

5.2.2.3 Case study 1: A study in complexity.

The concept map produced by S1 of the TG1 class is quite complex, and distinctly more so than many of the student maps; it is representative of the more complex maps that were produced. The S1 map appears in Figure 16. The map incorporates an example sentence which appears to be grammatically correct using either *would* or *will*, "The player said he is injured, so he would/will not play the match." In fact, prescriptivist grammar dictates that only *would* is the correct form in this sentence, however, the form *will* is so commonly used by English speakers in sentences thus structured that the learner is, I feel, correct in assuming either would be acceptable. The fact that this learner has *noticed* this, regardless of his/her knowledge of prescriptivism, is a step in the right direction.

In addition, S1 has produced the intriguing dichotomy of usage in the example sentences, "I *would* appreciate it, it you help me." and " help me and I *will* appreciate it." These examples of usage are more sophisticated than most of the example sentences produced by the other student participants. They demonstrate that on some level this student distinguished differences in the sentences produced. The fact that they are arranged in an original conceptual format suggests that metacognition may have been involved in their production. This is in contrast to the more behaviourist, traditional approach of students being asked to list sentences in a textual document by means of typing or cutting and pasting alone (see also the CM shown in Section 5.2.2.5 for an example of rule listing which a learner has reported to have been helpful in fostering learning). This leads to the question, how did S1 perform in the other measures as possible indicators of competence? In answer, S1 performed very well in the pre-treatment essay test, with a score of 5/6 (5 target modals out of 6 obligatory contexts). There were no post- or delayed post-treatment tests administered to this student due to absences. Pre- and post-treatment FIBT scores for S1 were also a strong 8/10 and 7/10 respectively.

The next question remains, how did this learner characterize his/her learning experience? Was it perceived to have been successful using these methods? In answering the Using Concept Mapping Software and the Mystery

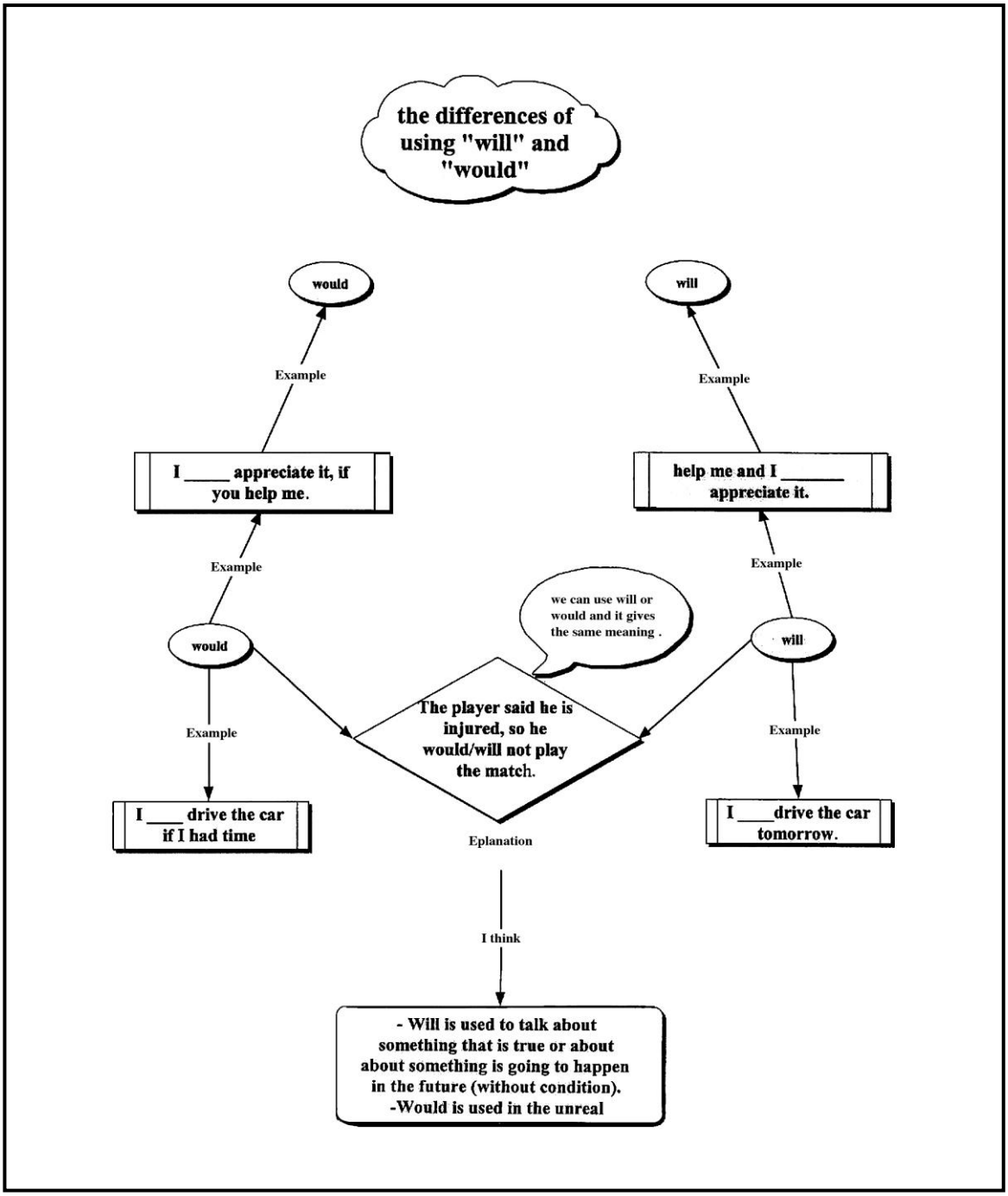


Figure 16 Individual concept map produced by S1 of the TG1 class

Module and Concept Mapping Exercises Post-Questionnaire survey questions (henceforth in this section called the survey questions) about whether the individual CM production exercise had helped the student learn, S1 answered, "*Yes, It helped me to learn more. It helped me to see the differences between "will" and "would". Also it gave me a clear idea of using them. Helped me to see what rules I'm confused with."*

When asked if the group CM production exercise had helped this student learn, S1 responded, "*Yes. To see different opinion and discussing our thouts* [thoughts] *with respect."*

When asked if the teacher-class CM production exercise had helped this student learn, S1 responded, "*Yes, helped a lot. Before, we made our work activities but we didn't know if we were right or wrong. After getting help from the teacher we know the righ* [right] *rules."*

When asked what the learner had learned in the three CM production exercises, S1 responded, "*Working in group is a better way, sharing our thoughts and asking anything I don't understand."*

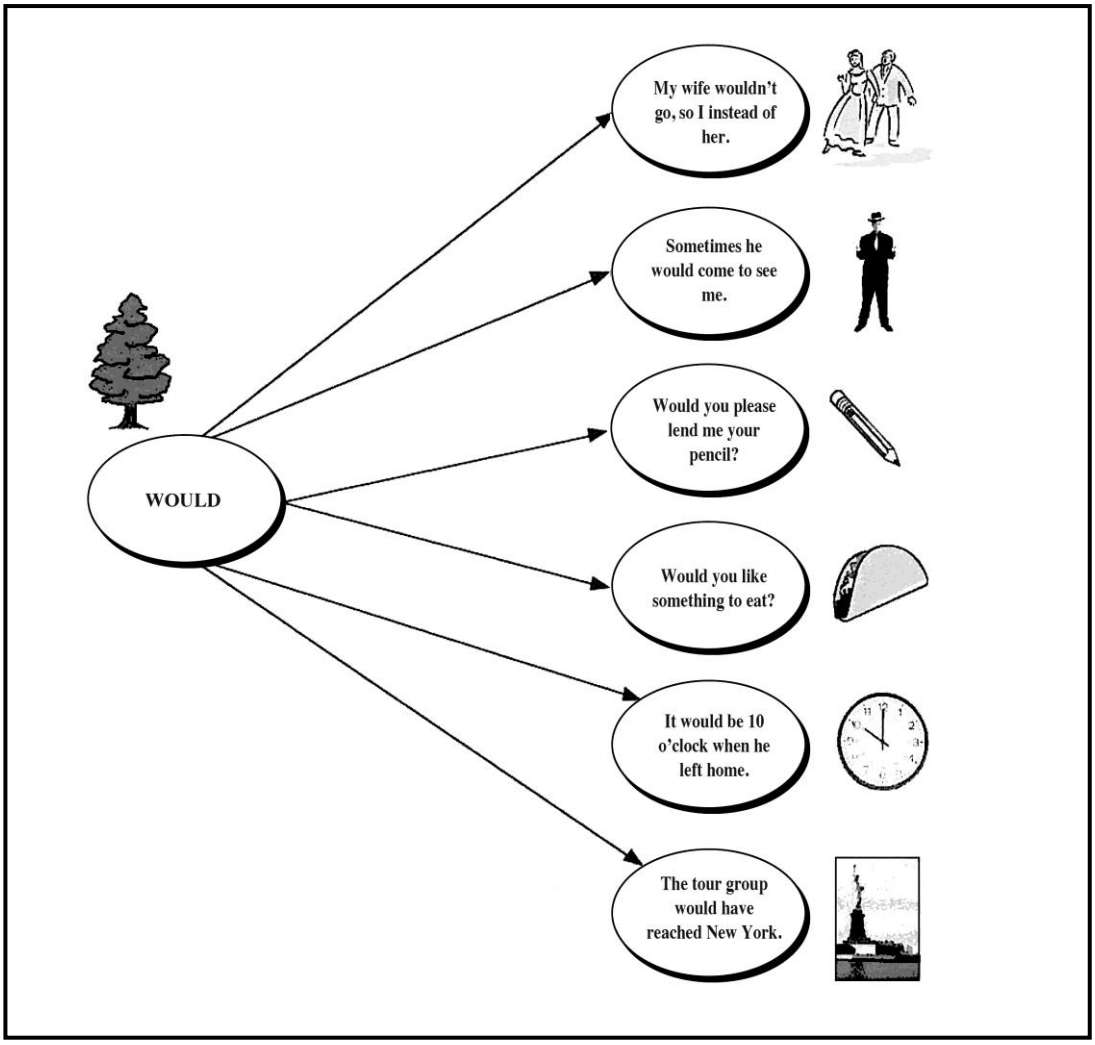
These results in triangulation, present a picture of this learner's experience using an AFL concept map application.

5.2.2.4 Case study 2: A visual learner illustrates

In the TG2 class learner S38 took a visual approach to assembling six sample sentences to show target usage of *would,* and three sample sentences to show target usage of *will*. The S38 map appears in Figure 17. S38's use of sentences as examples of usage was common amongst the student CM productions. However, S38 also chose to illustrate each sentence with a picture image, to serve as mnemonic devices. Moreover, S38 also included in her map a tree (to illustrate the word *would*) and a whale (to illustrate the word *will*) as visual-auditory reminders of the target forms. This learner clearly took pains to point these feature of the concept map to me one day when I was visiting the class and expressed enthusiasm with the CM exercise as a learning activity.

Other students from both treatment group classes also used visual images in their CM productions (as was seen in item 10 reported in Tables 3 and 4), yet this student's work was a good example of how these learners used visual means to express complex ideas.

Learner S38 completed pre-, post- and delayed post-treatment essay and FITB tests. His/her scores in the essay tests were 1/11, 1/6 and 2/7 respectively. In the FITB tests S38 scored 7/10, 10/10 and 7/10 respectively.



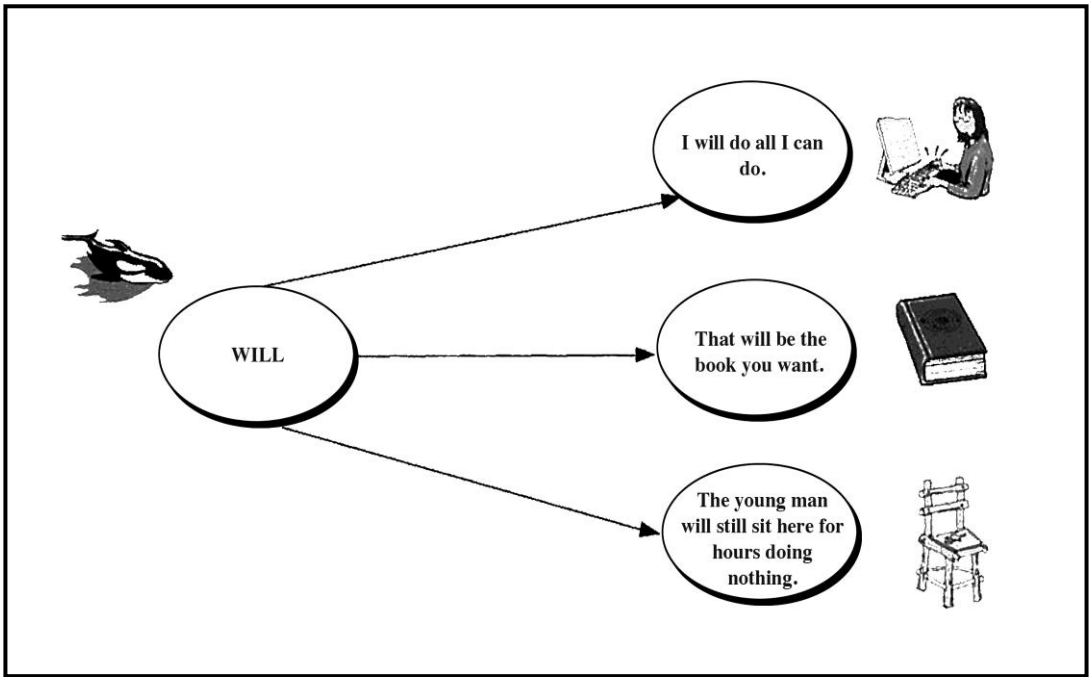


Figure 17 Individual concept map produced by S38 of the TG2 class, (divided for clarity)

This may indicate that some learning had occurred, but without evidence from the learner him/herself, assumptions of this sort cannot be made given the nature of this quasi-experimental study. So the question remains, how did this learner perceive his/her learning experience in written comments?

In answer to the survey questions about whether the individual CM exercise had helped the student learn, S38 answered:

> *Yes, it did. Before I made a concept, I was thinking about it and I read the grammar book which taught me how to use "will" and "would". it made me remember some knowledge that I forgot so when next time I meet this kind of grammar I think I wil* [sic] *do it very well. When I did it, I thought about it to improve my grammer* [sic].

When asked if the group CM production exercise had helped this student learn, S38 responded "*Yes. Team work always help each other to get better job.*" When asked if the teacher-class CM production exercise had helped this student learn, S38 responded "*Yes.".*

When asked what the learner had learned in the three CM production exercises, S38 responded "*Garmmer* [grammar]*, Teamwork."*

The point noted earlier that S38 made about consulting a grammar reference book for the individual CM production exercise leads to the next example of student CM work, which was representative of that of several students.

5.2.2.5 Case study 3: A first example of a "grammar ruler"

If grammar were a kingdom, these learners would be the grammar rulers. Thus, several student CM productions consisted primarily of applications of grammar rules which appeared to have originated in the pages of traditional grammar textbooks. This is not necessarily a bad thing. These learners may have used grammar rules in the context of concept maps to access how this usage operates in English. These learners reported learning had occurred, although in contrast to the complex concept map of S1 described above, evidence of metacognition in the concept maps consisting primarily of listed grammar rules, is far from clear.

An example of this kind of expression demonstrated in an individual student concept map production follows. In the TG1 class, S2 took a grammar

rule based approach to his/her concept map production. The concept map appears in Figure 18.

The pre- and post-treatment essay results reported for S2 were 2/7 and 5/7 respectively. The pre- and post-treatment FITB results were 8/10 and 7/10 respectively. These results suggest there were some strengths in this learner's target modal production and that learning may have occurred. S2's response to the survey questions about whether the individual CM exercise had helped the student learn was:

> *Yes, it helped me a lot. I spend some time to find a suitable map. I had to review the rules of using both "will" and "would". I spend some time to get the rules from the internet. Moreover, to make them very clear for me and for anyone who wanna take a look to my map, I included examples of each rule. I learned from my mistakes. I think that I will not fall in the same mistakes again.*

When asked if the group CM production exercise had helped this student learn, S2 responded:

> *Yes. Working as a group is much better than working alone. There a lot of discussion and arguments about the rules as well as the shape of the group map. Moreover, at the end helped us to come up with final concept map and all agreed on.*

When asked if the teacher-class CM production exercise had helped this student learn, S2 answered "*Yes, a lot. A lot of discussions about the rules and the map itself. Showing our mistakes by the teacher helped me a lot to identify."* And when asked what the learner had learned in the three CM production exercises, S2 responded "*Helping me to know the use of will & would."*

This learner expressed some operational as well as altruistic motives for including grammar rules in his/her individual CM production. Since the grammar rule based approach was quite prevalent in the CM production submitted, a second example will be cited next.

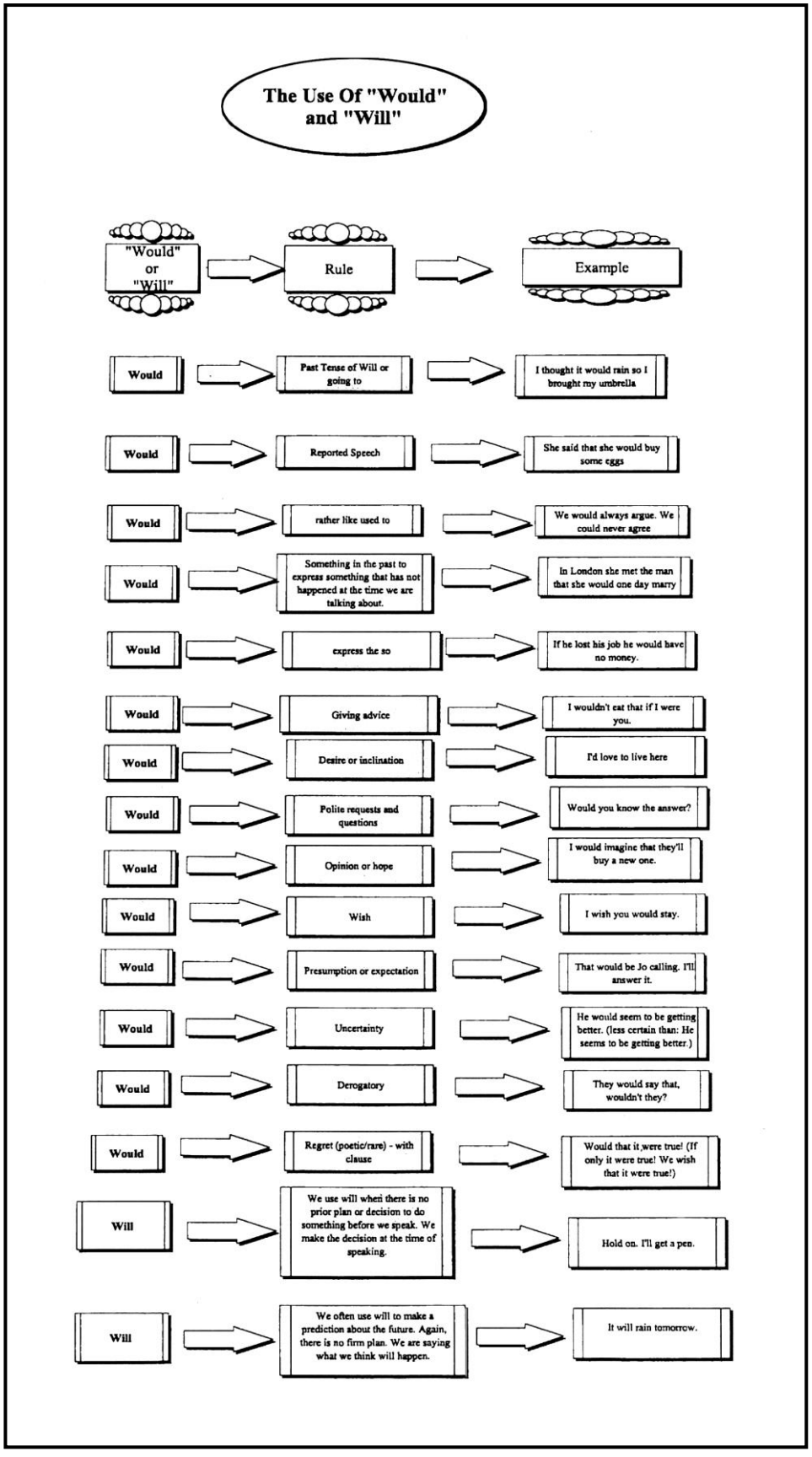


Figure 18 Individual concept map produced by S2 of the TG1 class

5.2.2.6 Case study 4: A second example of a "grammar ruler"

Again in the TG1 class, S6 used a grammar rule based approach to the CM production exercise. The concept map appears in Figure 19. The S6 concept map consists of a linear format with three rules expressing how to use *would* and three rules for expressing *will*, as well as an example of usage for each in a sample sentence.

Pre-, and delayed-post treatment essay test results for S6 were 1/1 and 4/5 respectively (no results were available for a post-treatment essay test). FITB pre-, post- and delayed post-treatment results were 5/10, 8/10 and 8/10 respectively. These results suggested there were some strengths and possible improvement in modal production over the course of the study.

In answer to the survey questions, S6 responded "*Yes, the concept map is helping me learn when to use will, when to use would. However, this map makes easier and better understanding to use will and would."* When asked if the group CM production exercise had helped this student learn, S6 answered "*Yes, more ideas."* When asked if the teacher-class CM production exercise had helped this student learn, S6 responded "*Yes."*

When S6 was surveyed about what he/she learner had learned in the three CM production exercises, S6 wrote "*I learned how to make a concept map and anys* [analyze] *some ideas became my own."*

These results demonstrate this learner believed he/she had learned as a result of the CM production exercise.

The next case study presented will discuss an instance whereby a student produced a concept map comparable to those of other students, had test results comparable to those of other students but who, in contrast to the majority of student participants, expressed the opinion that the CM production exercise had not been useful for learning. These results follow.

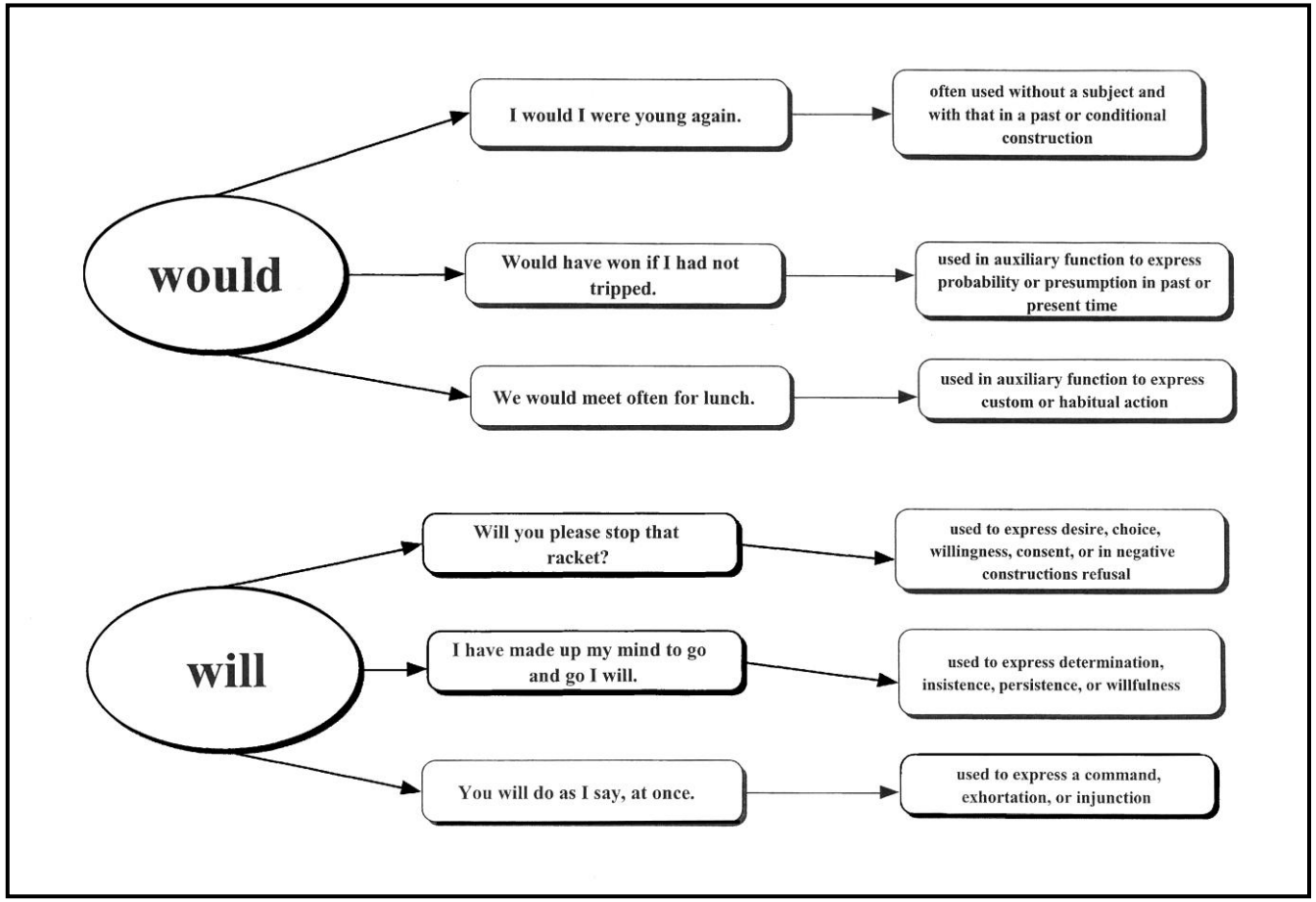


Figure 19 Individual concept map produced by S6 of the TG1 class

5.2.2.7 Case study 5: A learner who found concept maps not useful

Learner S4 from the TG1 class produced a concept map using visual elements, idea linking, and grammar explanations to present the concept of how to use *will* and *would*. The map appears in Figure 20.

Pre-, and post-treatment essay test results for S4 were 0/2 and 2/2 respectively (no results were available for a delayed post-treatment essay test). FITB pre-, and post-treatment results were 10/10 and 8/10 respectively. These results indicate the learner had a strong understanding of the subject matter either previous to the treatment, or possibly as a result of it.

In response to survey questions learner S4 reported not finding the CM production exercise to be useful in promoting learning, by responding:

> *No, It does not bring an efficient learning method. Moreover, it creates a very complex and complicated theory in my brain. In my opinion, giving formula or methods should be a priority when learning. Understanding deeply methods and exception would help us.*

When asked if the group CM production exercise had helped this student learn, S4 responded "*No."* When asked if the teacher-class CM production exercise had helped this student learn, S4 responded "*Yes. give me the formula and before I was confused with will and be going to now, I know."* When asked what the learner had learned in the three CM production exercises, S4 answered "*use of would and will."*

Case studies one through five here presented give a picture of how these learners used concept maps in their attempts at understanding the usage of *would* and *will*. In the next section, the concept maps that were co-constructed in group CM production exercises will be presented and discussed.

5.5 Group concept maps

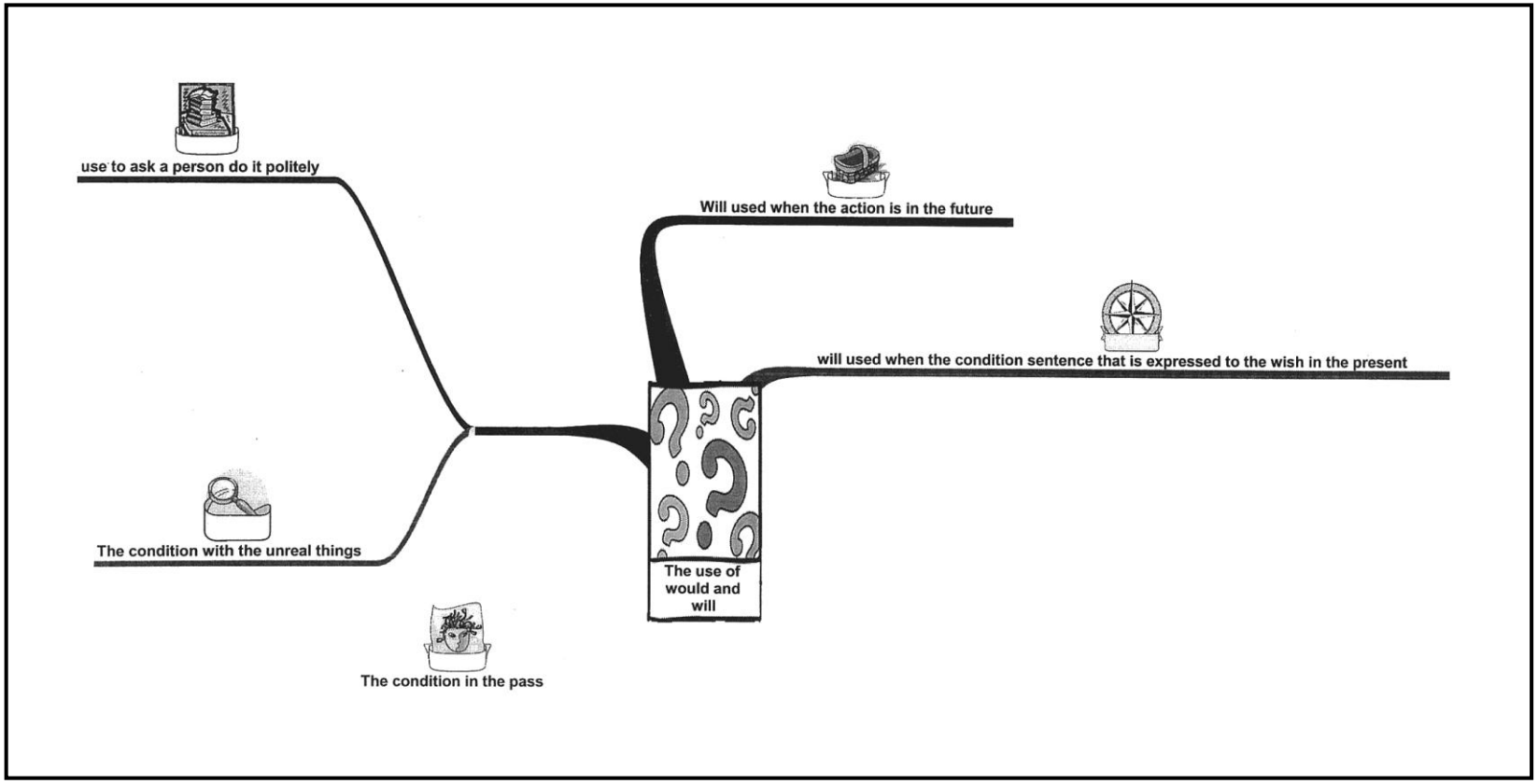


Figure 20 Individual concept map produced by S4 of the TG1 class

5.2.3 Group CM production results

Following the individual student online CM production exercise, the treatment class students worked in small groups co-constructing paper-and-pencil group concept maps. The number of group participants in each TG1 and TG2 class was the same (*n*=4, *n*=3, *n*=4).

The student group completion times in the CM production task ranged from 12 to 27 minutes. The individual group task completion times in minutes were TG1 group 1 (22), group 2 (12), and group 3 (12), and TG2 group 1 (13), group 2 (27), and group 3 (16).

5.2.3.1 Qualitative and quantitative results using the Concept Map Binary Qualification Checklist analysis.

The group CM productions were analyzed by quantitative and qualitative means using the Concept Map Binary Qualification Checklist. The analysis results are reported in Table 5. Table 5 shows there was a high level of uniformity in the Checklist features amongst the group concept maps. For example, the number of features in the TG1 class was group 1(6), group 2(5), and group 3(5), while in the TG2 class group the number of features was found to be group 1 (5), group 2 (5), and group 3 (5).

The group concept maps were more uniform in comparison with the individual CM production analyses (which were reported in Tables 3 and 4) as well. Additionally, the group concept maps were more linear than the individual concept maps. They were also more likely to include grammatical explanations, examples of grammatical usage, and linking between ideas. The group concept maps were less likely to include a pictorial image and in contrast to the individual maps of TG1 only, were less likely to include an expression of a main grammatical concept. (The student group CM productions are available from the researcher.) Two of the group concept maps have been chosen for discussion, and the data will be presented in triangulation in this section. The first group concept map was chosen for presentation because it is representative of most of the group CM productions. The other group presentation was selected by virtue of being different from

Table 5: TG1 and TG2 student group CM production analysis using a concept map binary qualification checklist

	STUDENT GROUP					
	TG1 (n=11)			TG2 (n=11)		
Does the concept map include …	1 (n=4)	2 (n=3)	3 (n=4)	1* (n=4)	2 (n=4)	3 (n=3)
1. a title	-	-	-	-	-	-
2. an expression of a main grammatical concept	-	-	✓	-	-	-
3. the words *would* and *will* represented in a square, oval, discrete space, etc.	✓	✓	✓	✓	✓	✓
4. an example of grammatical usage	✓	✓	✓	✓	-	✓
5. a grammatical explanation (that may or may not be correct)	✓	✓	✓	✓	✓	✓
6. idea linking (through lines or arrows or a Venn drawing)	✓	✓	✓	✓	✓	✓
7. cross linking of ideas	-	-	-	-	✓	-
8. a linear design	✓	✓	-	✓	-	✓
9. a non-linear design	-	-	✓	-	✓	-
10. pictures to illustrate a concept visually	✓	-	-	-	-	-
Total CM features per group	6	5	6	5	5	5

most of the maps produced by the groups. The former example will be discussed first.

5.2.3.2 A representative example of the group CM productions.

The group 1 CM production of the TG1 class has been chosen for discussion since it is representative of those of most of the other groups in both classes. The TG1 group1 concept map appears in Figure 21. The TG1 group 1 concept map exhibited six of the features included in the analysis

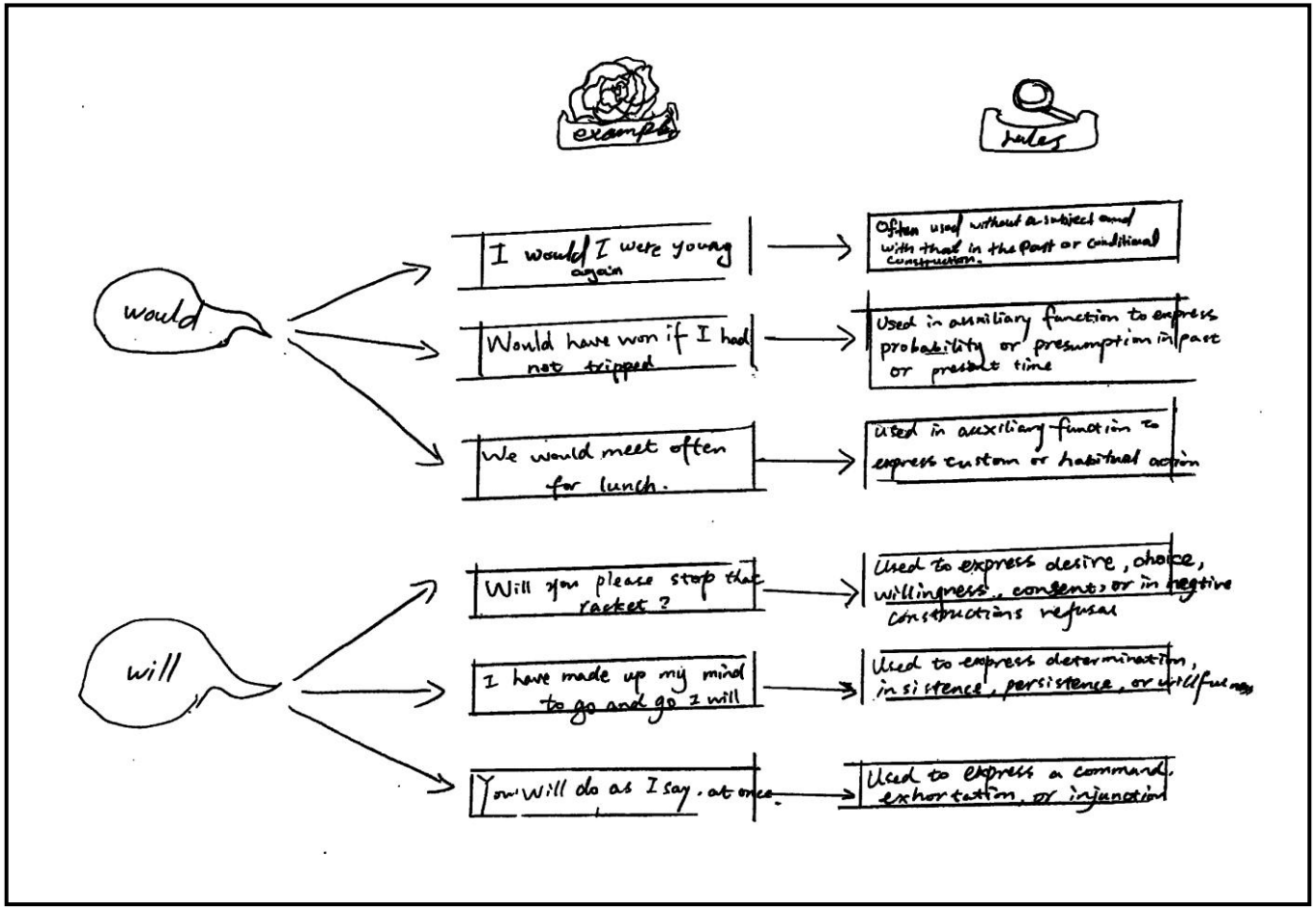


Figure 21 The group concept map produced by group 1 of the TG1 class

checklist. The design is linear, and it includes examples of usage linked to usage rules. However, this group's concept map included two pictorial images. The student who had been chosen by the group to draw the map informed me that the two images (one of a rose above the list of example sentences, and of a magnifying glass above a list of grammar rules) represented a positive and a curious image, respectively.

5.2.3.3 A 'unique' group CM production.

In some contrast, the CM production of the TG2 class group 2 was non-linear in design. It was one of the few concept maps to include idea cross-linking. The TG2 group 2 concept map appears in Figure 22. This concept

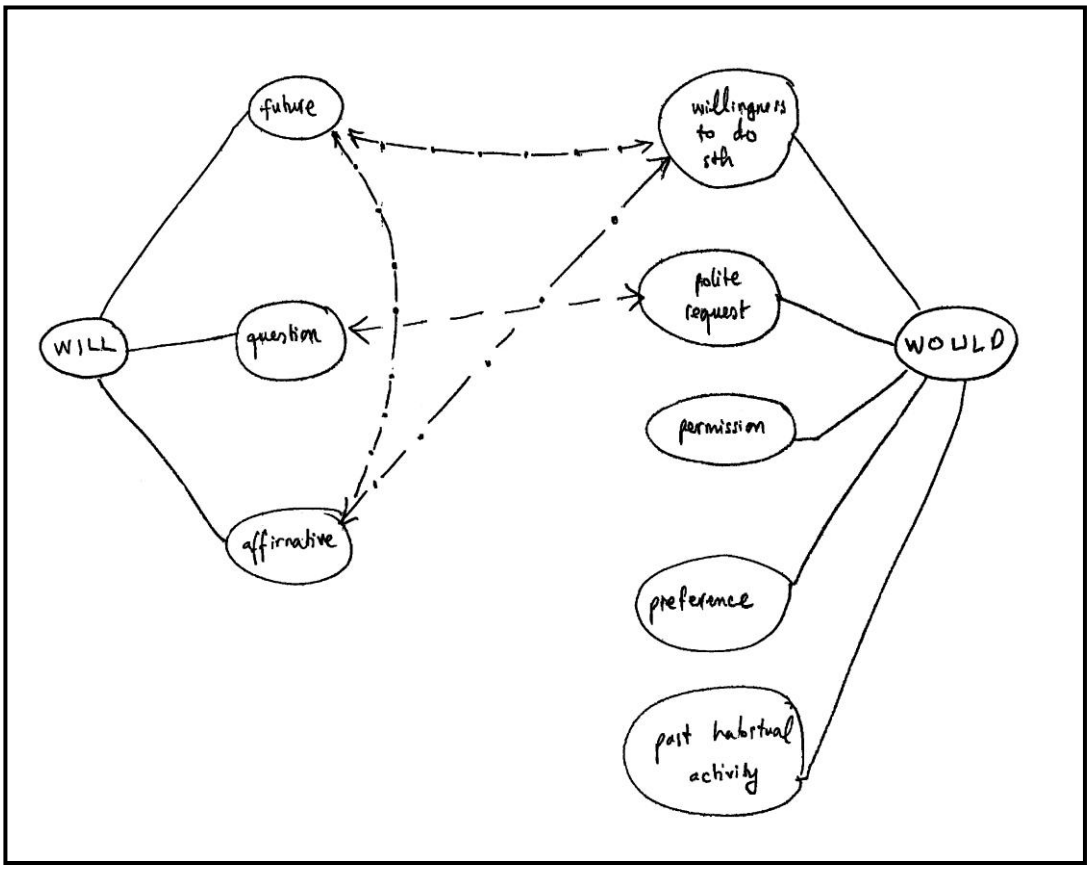


Figure 22 The group concept map produced by group 2 of the TG2 class

map is quite simple in that it includes one-word information units, as opposed to the others which include sentences to express usage examples or rules. Despite this simplicity, these students have opted to portray the interchangeable nature of this usage, for example in showing when both *will*

and *would* might be used. However, lacking examples in the form of full sentences and fuller information, the concept map is confusing in comparison with those produced by other groups.

5.2.4 Teacher-class CM production results

5.2.4.1 Qualitative and quantitative results using the Concept Map Binary Qualification Checklist analysis.

The concept maps which were co-constructed by the teachers and classes of the two treatment groups were analyzed using the Concept Map Binary Qualification Checklist. The analysis results are reported in Table 6.

As Table 6 illustrates, the concept maps of both classes were similar in that they included the Checklist features 'the words *would* and *will* represented in a square, oval, discrete space, etc.', and 'idea linking'.

Table 6: TG1 and TG2 teacher and class CM production analysis using a concept map binary qualification checklist

	GROUP	
	TG1 S(n=13)	TG2* S(n=12)
Does the concept map include …		
1. a title	-	-
2. an expression of a main grammatical concept	-	-
3. the words *would* and *will* represented in a square, oval, discrete space, etc.	✓	✓
4. an example of grammatical usage	✓	-
5. a grammatical explanation (that may or may not be correct)	✓	-
6. idea linking (through lines or arrows or a Venn drawing)	✓	✓
7. cross linking of ideas	-	✓
8. a linear design	✓	-
9. a non-linear design	-	✓
10. pictures to illustrate a concept visually	-	-
Total CM features per group	5	4

* one draft CM and one final CM produced were collapsed in this analysis

The TG1 and TG2 teacher and class concept maps are shown in Figures 23 and 24, respectively and (this is available from the researcher). From Figures 23 and 24 it can be seen that both the TG1 and TG2 class concept maps showed usage in abbreviated 'rules' of one word, such as "future" in the case of *will* The concept maps differed somewhat in that there were five features present in the TG1 class, and four in the TG2 class map. In addition, the TG1 concept map was distinguished from the TG2 concept map since it included 11 example sentences and the TG2 map had none.

Figures 23 and 24 also show that the TG1concept map appeared to be more complex than that of the TG2 class *because* of the inclusion of examples of usage in the former. In fact, the TG1 map contained only two more usage contexts (10) than did the TG2 map (8). The TG2 concept map may look simpler at first glance but in reality it is more complex in that it displays cross-linking of ideas, citing "polite requests" and "condition" as areas of usage in which either *will* or *would* may be used.

In contrast, the concept map of the TG1 class shows only separate usages for either *will* or *would*. Both of these concept maps illustrate contexts whereby *will* and *would* are used and there is some variation in usage in the two maps. (The complete CM production of individual students, student groups, and teachers and classes are available from the researcher.)

These CM production exercises may indeed have proven helpful for the students in these classes many of whom were grappling to learn the use of *will* and *would*. However, to investigate if learning did occur, it is necessary at this point to examine the evidence available from classroom observations and selected audio transcriptions of the proceedings. Accordingly, these data will be reported in the next section.

5.2.5 Post-treatment student survey results

5.2.5.1 Qualitative survey results.

Moving now to the surveys administered after the CM production exercises were completed, and as noted earlier, the student Mystery Module and Concept Mapping Exercises Post-Questionnaire was comprised of 27 items. Twelve of the items were open questions about learner perceptions to the

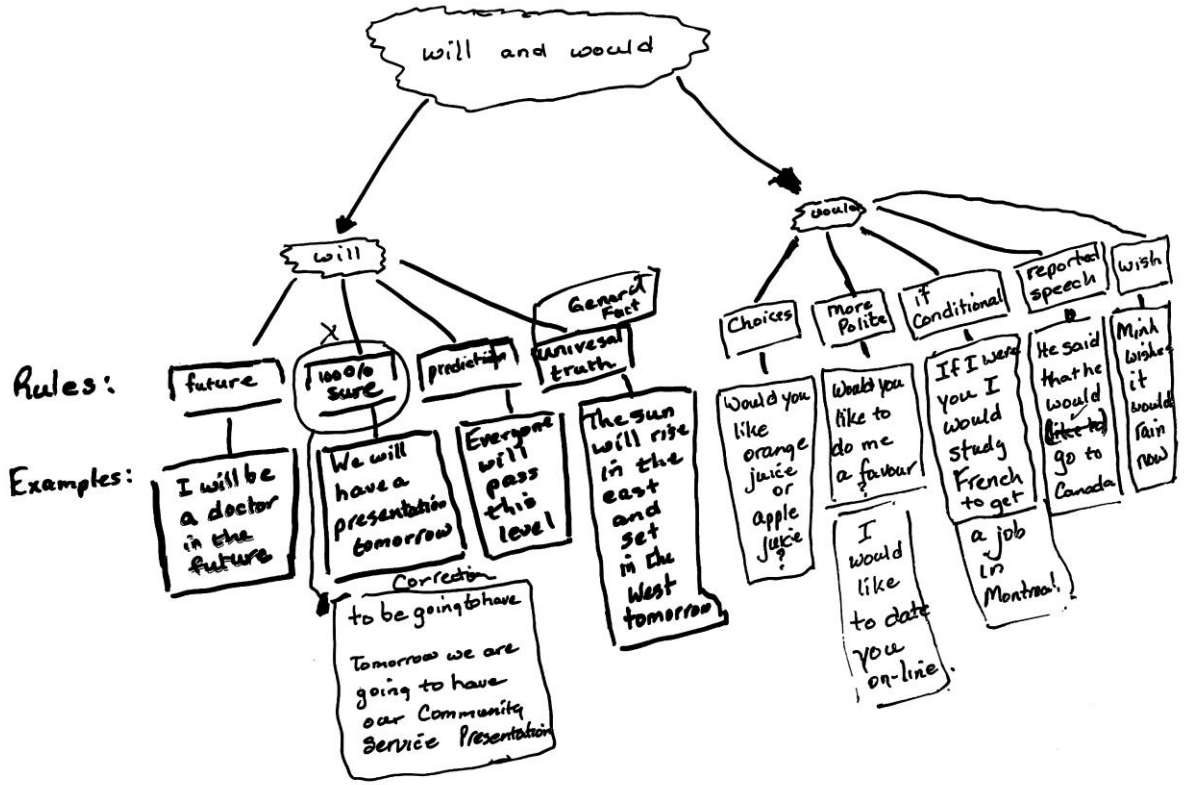


Figure 23 The teacher and class concept map produced by the TG1 class

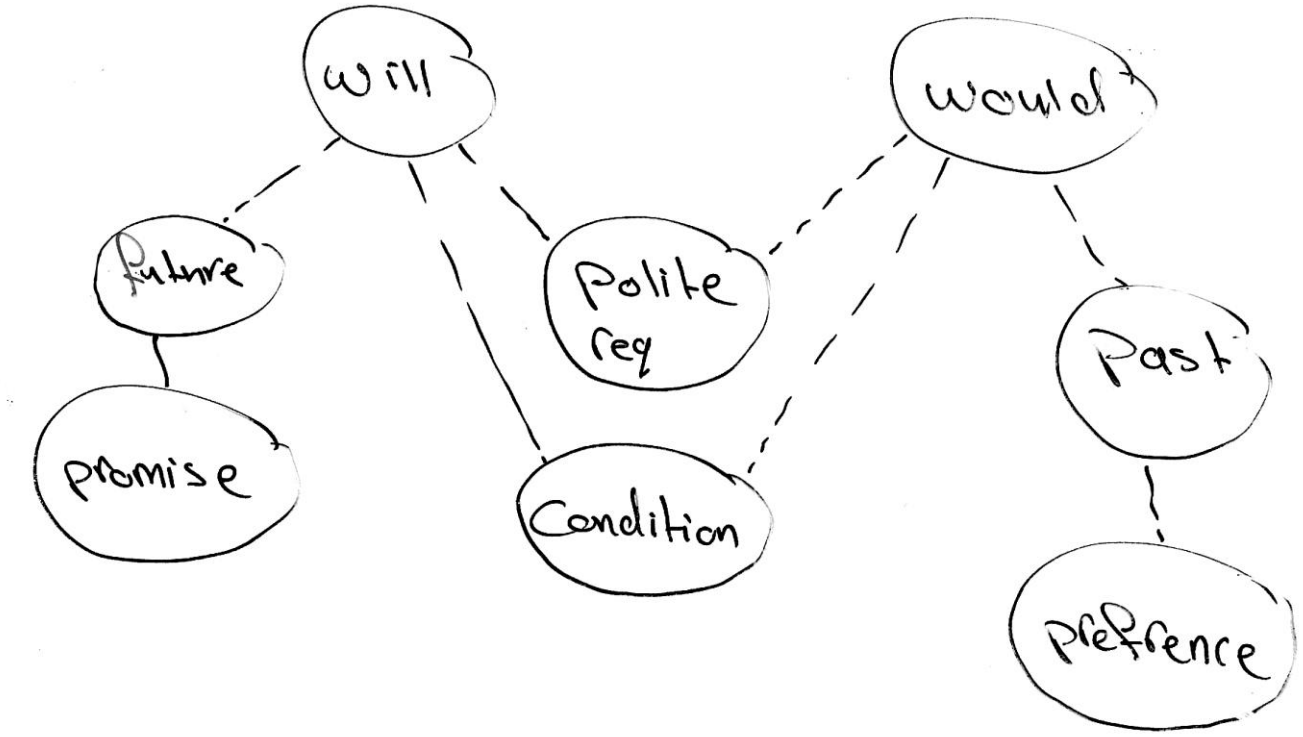


Figure 24 The teacher and class concept map produced by the TG2 class

procedures and treatment. In general these responses followed directly after the yes/no questions cited earlier and requested elaboration. The student responses are available from the researcher. As was the case with the yes/no item responses already reported, the great majority of the responses were positive. Samples of some of the responses to the questionnaire, both positive and negative will follow.

5.2.5.1.1 Did a computerized CM help you learn? In what way?

For example, positive responses to items 4 and 5, 'Did creating a computerized concept map help you learn? In what way?' from some of the students of the TG1 group were:

- *Helped me to know how to design a concept map. Also, helped me to devide [divide] the rules and attach examples to illustrated in a very easy way to learn. Now I can use it as a reference to know when and how to use* will *and* would.
- *The computerized concept map is very interesting. It gives clear information about what I want to learn.*
- *It is very clearly to see the different between* will *and* would *we can have deep memory and next time we can easily make the correct choose.*
- *Can help me to learn more about the idea of the lessons. Moreover it can divide clearly to help me understand clearly of this idea.*

Samples of positive responses to these items from the TG2 group were:

- *It made everything clear and easy to remember.*
- *Although I knew the grammar before groups and teacher can help me to understand it well.*
- *They told me information about learning.*
- *It is very clear and easy to understand the grammar.*
- *Because this kind of map help me to organize the ideas and examples.*

There was one negative responses to item 4 in the TG1 group and no comment was given to item 5 to elaborate on why the respondent felt this way. In the TG2 group there were no negative responses to these items.

5.2.5.1.2 Did a group CM help you learn? In what way?

Some positive responses to items 12 and 13, 'Did creating a group concept map help you learn? In what way?' from the TG1 group were the following:

- *To see different opinion and discussing our thoats* [thoughts] *with respect.*
- *Working as a group is much better than working alone. There a lot of discussion and arguments about the rules as well as the shape of the group map. Moreover, at the end helped us to come up with final concept map and all agreed on.*
- *According to a group concept map, I know that I should share ideas with group members and product new point for discussing.*
- *I can find some information which I didn't have in my map.*
- *It help me to lean about workin' in a group and by share or take other opinion, can gro me idea to understand it easily.*

Samples of positive responses to these items from the TG2 group were:

- *I can get ideas from my partners.*
- *It made us learn out the best way to write a complet* [complete or concept] *map.*
- *We can talk and discuss all information we have.*
- *Because each members, have different idea so we can help each other.*
- *Helped me to understand some rules.*

One student in each of the TG1 and TG2 groups responded '*no*' to item 12, but neither commented on their response in item 13.

5.2.5.1.3 Did a teacher-class CM help you learn, assess your work? In what way?

Some positive responses to items 16 ('Did creating a concept map with your teacher and class help you learn?'), 17 ('Did this teacher and class activity help you assess your own work?'), and the comments from item 18 ('In what way?') from the TG1 group were the following:

- *Before, we made our work activities but we didn't know if we were right or wrong. After getting help from the teacher we know the right rules.*
- *When I did it with my group, we shared ideas but some times we didn't know which one was right. When we did it with the teacher, she helped us identify the true.*
- *We share opinion together, and we share example together.*
- . *Give me the formula and before I was confused with* will *and* be going to *now, I know*
- *Students and teacher discuss the problem together and they create a concept map together.*

Samples of positive responses to these items from the TG2 group were:

- *Teacher gives me instruction and I learn what I should pay more attention to.*
- *Teacher can give us more ideas and explain it well.*
- *We can work together, discussion can improve our many skill, such as speaking listening and grammar*

In the TG1 group one student responded '*yes'* to item 16 and '*no'* to item 17, commenting in item 18:

- *no such a thing*

In the TG2 group several students left item 18 blank. One student responded '*yes'* to item 16, '*no'* to item 17, and left item18 blank.

5.2.5.2 Quantitative survey results.

The student Mystery Module and Concept Mapping Exercises Post-Questionnaire was comprised of 27 items. One –fifth (4) of the total (20) responses to the post-CALL survey question, 'Did this help you learn?' have been classified as *qualified yes* responses, as can be seen in Table 7. Generally speaking, these responses were in the order of "Yes, but…" answers to questions requesting a yes or no answer, and many of these answers have been reported in the previous section.

Table 7: Frequency of responses to post-CALL student survey question

	Responses			
Did this help you learn?	yes	qualified yes	no	Missing
TG1 (n=11)	7	1	1	2
TG2 (n=11)	7	3	1	0
Group totals	14	4	2	2

Thirteen of the items were yes/no questions , the responses to which are reported on Table 8. Comparisons of the TG1 and TG2 groups must first note that there were more respondents in the first group (n=13, n=10 respectively).

However, Table 8 shows the TG1 group responded positively that the treatment had been helpful to them much more frequently (135) than did the TG2 group (88). The two groups tended to answer negatively to the treatment questions slightly more in the TG1 group (33) although the TG2 group did so comparably (25). Finally, the TG1 group answered in the affirmative rather than negatively with a ratio of approximately 4:1. The affirmative: negative response ratio of the TG2 group was 3.5:1.

Table 7 shows a similar positive response trend when both treatment groups were queried on if they found the pre-CM exercises, the CALL Module, helpful in their learning. In this immediate post-CALL surveying the students responded more positively than they did to the post-treatment questionnaires reported in Table 7, both at a ratio of 7: 1.

Table 8: Frequency of responses to post-treatment student survey yes/no question items

#	Item	Treatment group			
		TG1 class (n=13)		TG2 class (n=10)	
		Y	N	Y	N
1	Did you use the Mystery Module pages for self-assessment of your work?	12	1	8	2
2	Did this self-assessment help you learn?	11	2	9	1
4	Did creating a computerized concept map help you learn?	12	1	10	0
6	Did you use the computerized concept mapping exercise for self-assessment of your work?	8	5	8	2
7	Did this self-assessment help you learn?	6	7	6	3
10	If you made drafts, did the process of re-making the map help you learn?	10	2	3	0
12	Did creating a group concept map help you learn?	12	1	8	1
14	Did other group members help you assess your own work?	9	4	4	6
16	Did creating a concept map with your teacher and class help you learn?	13	0	9	0
17	Did this teacher and class activity help you assess your own work?	12	1	7	1
19	Did self, group, or teacher-class assessments help your learning?	13	0	9	1
21	Will you use this way of learning in the future?	11	2	5	1
23	Did these activities change the way you think about how you learn?	6	7	2	7
		135	33	88	25

5.2.6 Post-treatment teacher surveys

5.2.6.1 Qualitative survey results.

Table 9 shows that both teachers refrained from an affirmative response in two cases. The TG1teacher responded in answer to item 9, 'Do you think your guided questioning helped your students learn?*'* by answering *I didn't think I was guiding them – I think they were guiding me.* The TG2 teacher qualified her answer to item 11, 'Will you use this way of teaching in the future?' with the comment, *I already do to some degree, but I haven't used concept maps for grammar points.*

The Mystery Module and Concept Mapping Exercises Teacher Post-Questionnaire also included nine open-ended items. The questionnaire responses are available from the researcher. Some samples of the teachers' responses will be given here.

Table 9: Frequency of responses to post-treatment teacher survey yes/no question items

#	item	Teacher			
		TG1		TG2	
		Y	N	Y	N
1	(If applicable) Do you think the Mystery Module helped your students learn?	✓	-	✓	-
3	(If applicable) Do you think the computerized concept map exercise helped your students learn?	✓	-	✓	-
5	(If applicable) Do you think the group concept map exercise helped your students learn?	✓	-	✓	-
7	Do you think the teacher and class creation of a concept map helped your students to learn?	✓	-	✓	-
9	Do you think your guided questioning helped your students learn?	N/A	-	✓	-
11	Will you use this way of teaching in the future?	✓	-	N/A	-
13	Did these activities change the way you *think* about student learning?	✓	-	-	✓
	Total responses	6	0	5	1

5.2.6.1.1 (1) Do you think the Mystery Module helped your students learn? (2) In what way?

In responding to these items 1 and 2, this is what the teachers said:

- *(TG1 teacher) Yes. Having to work on a computer program; completing the assignment and then being able to check the answers.*
- *(TG2 teacher) Yes. Start thinking about the grammar point.*

5.2.6.1.2 (3) Do you think the computerized concept map exercise helped your students learn? (4) In what way?

The teachers' response was:

- *(TG1 teacher) Yes. First attempt, a line was missing and they could not complete the assignment. They then returned to the program after the line was corrected. They were <u>interested</u> in knowing what the Mystery Module was all about – perseverance.* [underline in original]
- *(TG2) Yes. Helped analyze the uses of this grammar point.*

(In her comments on this item, the TG1 teacher has made reference to an early problem with the concept mapping software installation that, as she noted, was subsequently repaired.)

5.2.6.1.3 (5) Do you think the group concept map exercise helped your students learn? (6) In what way?

The teachers' response was:

- *(TG1) Yes. The class seemed to see a relationship between the two words – also the differences – and clarified a couple of points they had difficulty with. They enjoyed working together.*
- *(TG2) Yes. Confirmed some ideas and presented others.*

5.2.6.1.4 (7) Do you think the teacher and class creation of a concept map helped your students to learn? (8) In what way?

The teachers' response was:

- *(TG1) Yes absolutely. Mistakes were made and then corrected through class discussion.*

- *(TG2) Yes. Pulled all three maps together and explained some points that were not really clear…*if – will, if – would.

These responses illustrate these teachers' positive impressions of this application of AFL in this L2 setting.

5.2.6.2 Quantitative survey results.

Table 23 illustrates the results of the closed questions asked in the Mystery Module and Concept Mapping Exercises Teacher Post-Questionnaire. It can be seen in Table 23 that the two treatment group teachers were much in favour of the AFL procedures used in the study.

5.3 Qualitative results

5.3.1 Field notes of classroom observations of group CM production

All group CM production exercises were observed by the researcher and field notes were made. The observation field notes revealed group members' engagement in the task and in group discussions. The exercises were also audio taped, and selections were transcribed. The categories that emerged from these data and transcribed examples of them will be presented in the following sections. The transcription protocol that was used will be found in Appendix M.

5.3.1.1 Categorization of interactions in group CM production.

Fifteen categories of student interaction types emerged from a qualitative analysis of these data. The categories of student interaction are the following:

1. Discussions about how to draw map/making map/map set-up
2. Agreeing
3. Disagreeing
4. Negotiation of agreement and disagreement
5. Cooperative task accomplishment
6. Discussion about usage
7. Discussion about an example/s of usage
8. Correcting peer spelling
9. All talking at once
10. Rule explanation

11. Exhibiting verbal dominance
12. Exhibiting verbal submissiveness
13. Signalling the beginning of the exercise
14. Signalling the end of the exercise
15. Laughter indicative of understanding of usage

Of these categories, numbers two to eight and number ten show evidence of the AB, the interweaving of assessment, teaching (in this case co-constructed peer teaching) and learning may take place.

In addition, of these categories, two particular kinds were noted. The first nine categories had what I would define as *co-construction focussed interaction*, meaning those behaviours which have as their focal point an interactive contribution to the group experience. Negotiations amongst group members would fall into this category. The next six categories I would define as *individual focussed interaction*, or those which serve more as an expression of realization of the self. For example, in some groups one member took on a leadership 'timekeeper' role, declaring the exercise begun or ended.

The following is an example of a co-construction focussed interaction that occurred in an exchange between the group 2 students of TG2, who are identified as S31, S38, S40 and S41:

(S40) Yes, uh for *will* to express future time. Um, for example uh, "I will call you every weekend." When you, when you talk about the future, you wi-, you use uh *will*.

(S31) I'll say [?] …future.

(S41) Can you tell me uh what you base on in order to create this map?

(S31) What?

(S41) What you base on?

(S31) Uh…

(S38) I think just the different situations [?] …

(S31) You know, I take grammar book.

(S41) Grammar book?

(S31) Yeah, and I read it.

(S41) OK.

An example of an individual focussed interaction is one student's rule explanation (1) is illustrated in the following discourse of a student in TG1, group 1 (S2):

> (S2) In my concept map I have quoted the, the rule, *all* the rules of using *will* uh and *would*, OK? And uh also uh I included example for *each* rule of using uh *would* and *will*. Uh I think uh in my opinion this is, this is very easy to understand and very uh straight, uh forward. And you can use it, and you can use it as a … For me I can use it as a reference.

Both the co-constructed and individual focus interactions appear to contribute in different ways to the execution of the group exercise. While the discourse in these data in context fit quite clearly into one or the other categories, some overlap does exist between the two. For example, in some situations laughter may have a personal and internal focus while in others it may be intended as a social expression, inviting a response.

5.3.1.1.1 Group interactions and research question three.

For the purposes of this study, the co-construction focussed interactions two to seven are of most interest, since they contribute to answering the third underlying research question, which is *'Does a group concept map (CM) exercise enhance the learning of a specific language feature in a second language classroom?'* To revisit, they are:

2. Agreeing
3. Disagreeing
4. Negotiation of agreement and disagreement
5. Cooperative task accomplishment
6. Discussion about usage
7. Discussion about an example/s of usage

At this point samples of transcribed discourse will be introduced, followed by reporting of the discourse.

5.3.1.1.2 Transcriptions of TG1 student group discussions.

Transcribed learner co-construction focussed interactions taken from the group CM exercise discussions in both treatment classes have been selected to

help answer the third underlying research question. Complete transcriptions of the group CM production exercises are available from the researcher.

As noted, the focus in answering the research question at hand is on co-construction focussed interactions two to seven. Each of the following group discourse excerpts demonstrates some or all of the co-construction focussed interaction numbers two to seven. Of *most* interest among these interaction types, however, are numbers six and seven, *discussion about usage* and *discussions about an example of usage***,** respectively; they are germane to answering the research question by providing evidence of learner perceptions as to whether learning has taken place.

5.3.1.1.2.1 Student group 1.

The TG1, group 1 (*n*=4) completed the CM production task in 22 minutes. This excerpt illustrates co-construction focussed interactions five (cooperative task accomplishment) and six (discussion about usage) in a group discussion during the CM production task, in the following :

(S7) What is the rules you used for the *would* and …?

(S2) For example, past tense of *will* or *going to*, OK? We can use *would* for, the rule is for the past tense of *will* or *going to*. Here is the example, this rule, OK? We can use also it for reported speech. […]

(S7) And what is the example with *will*?

(S2) Here, here is *will*, here is the, here is the rule, the rule and the example for.

[…]

(S7) So, in mine I just write the *will* and *would* and I give example for each for *will* and *would* and just tell *will* we use it for future and *would* we can use it for choices or use it in the past sometime, so I don't have much information, just to like difference simply, simple.

(S2) So these are examples?

(S7) No, this are examples and this is rule […] So what about you?

(S6) […] "I would I were young again." So they often use without a subject

(S7) So it is like a hoping?

(S6) No, "I would I were younger" is like a thought.

5.3.1.1.2.2 Student group 2.

The TG1 group 2 (*n*=3) group took 12 minutes to complete the CM production task. This excerpt illustrates co-construction focussed interactions five (cooperative task accomplishment), six (discussion about usage), and seven (discussion about an example/s of usage). This excerpt also shows how these students have attempted to understand the use of *will* and *would* with some success and yet with some miscomprehension about the use of *would* as a past tense form and as to what should follow *so*. This is illustrated in the following:

(S1) *Will* is to talk about something that is true or about something that is going to happen in the future, and without condition, and *would* is used…

(S12) unreal.

(S1) … yeah, unreal.

(S12) Yeah, *this* one is a fact….

[…]

(S1) "If I had a million dollar I would buy my dream car."

(S12) Dream car?

(S1) Yeah.

(S12) OK, where an incident is going to be taken but not sure. Do you have example?

(S1) Yeah, […] We have uh, "John say he is tired. I think he would be absent tomorrow."

(S12) OK, yeah. "If…"

(S1) Or "He said he would be absent", no, uh. "He said he is tired. He would be absent tomorrow", yup.

(S12) "…he was tired."

(S1) "he is" or "he was" I don't know. Yeah, "…he would be absent tomorrow."

(S9) "He said he was tired so he would be absent tomorrow." So used for past time?

(S1) I don't know, do we need *so*? I think we should use *will* because I mean *so* [all talking at once]

(S12) It's a prediction.

(S1) "The global economy will be better next year." "He say he would be absent tomorrow."

5.3.1.1.2.3 Student group 3.

The TG1 group 3 (*n*=4) took 12 minutes to complete their CM production task. In this excerpt one student group member has articulated an understanding of the use of *will* and *would* in an example of co-construction focussed interaction six (discussion about usage):

> (S8) Ok, mine is very simple because me, my way to understand *would* and *will* is so restricted. So, when I use *would* is generally when it is a future possibility. For example, I am not sure about something, but I *guess* that thing will happen so this is what I call future possibility. And I took two examples. The first one is, was "I would like to return to my country" because it's something that I am guessing, but it's not something that it's you know, sure, or it's not something that it is already planned. My second example is "I would like to play computer games." … future possibility I think is the general idea behind the use of *would*. But now for the use of *will*…I think it is something which *will* happen, for example for something that will happen the next day or the next month… something that you are 100% sure that it will happen… It's not like you are guessing but you say *this will happen.*

5.3.1.1.3 Transcriptions of TG2 student group discussions.

5.3.1.1.3.1 Student group 1.

The TG2 group 1 (*n*=4) completed the CM production task in 13 minutes. The discussion that follows shows how these students interacted to co-construct meaning as they sought to understand together the usage of *will* and *would*. This excerpt illustrates co-construction focussed interactions two (agreeing), five (cooperative task accomplishment), six (discussion about usage), and seven (discussion about an example/s of usage) in the following:

> (S32) Start with *will*. I have in my map uh we can use *will* in a polite request and uh when we are sure for something…
>
> (S33) …uh about the future or about promise? That's mean you are, something will be happen in the future [talking over]
>
> (S?) Future, for the sure. For something sure that happen in the future. And for polite request. And what you have?

(S33) Promise. When you promise someone you can use *will* also.

(S34) And for quick reaction.

(S33) Quick reaction?

(S34) "I will answer the phone."

(S33, S32) Ahh.

[…]

(S33) You can use *going to*, but when you are very sure use *will*.

(S33) I have uh four point for *would*. Uh polite request, uh when you prefer something, when you discuss past habit, and pre-mission [permission]. What you have?

(S32) So you mean, it's a question, "What is your willing", not uh "What is your will"?

(S34) No, for here uh *willing* is formal and *willing* is were.

(S33) No, but I think this is right.

(S34) *Would* uh can be modal? [at 7 mins.] Modals?

(S32) Modals?

(S33) What you have?

(S34) Yeah, it's the same, show me. OK, it's the same.

(S33) Same, OK same.

(S34) Yeah, it says modals.

5.3.1.1.3.2 Student group 2.

The TG2 group 2 students (*n*=4) completed the CM production task in 27 minutes of rich discussion involving the usage of *will* and *would*. This excerpt illustrates co-construction focussed interactions five (cooperative task accomplishment), six (discussion about usage), and seven (discussion about an example/s of usage) in the following sample:

(S41) May I ask a question base on the, the way how to [?] use *will* and *would*? And I don't really understand what different between the two, *will* and *would*. But maybe for me I think *will* is, is present in, in the future, that's going to happen, and that's why predictable. And would is present in *may* be going to happen, may be unpredictable.

(S31) Yes.

(S41) And *would* is preferred to use in preferred speaking also, and it's in the formal, the formal way.

(S31) Ah, you mean *would*?

(S41) Yeah, *would*.

(S40) Also uh, for my paper I have different condition to use uh *would*. The first of uh, the first condition it's polite request. And an example for that uh, "Would you cook dinner tonight?" And uh the second things it's uh prefermence [preference] um, uh an example for that "I would rather stay at home." And um the third one it's uh permission, and an example for that "Do you mind if we leave class early?"

(S41) And is that your study uh grammar?

(S31) I'm sure there is.

(S40) Yes, actually I looked up at uh grammar books

(S41) OK.

(S40) …and uh I search a little bit from the Internet

(S41) OK.

(S40) …and I collect information and uh I put it here.

(S41) OK.

(S40) But uh I am not sure for is what I understand it's 100%...

5.3.1.1.3.3 Student group 3.

The TG2 group (*n*=3) completed the CM production exercise in 16 minutes. There was agreement on usage demonstrated in the first part of this group discussion until a friendly disagreement was observed over the usage of *would* as a past form of *will.* The excerpts that follow illustrate co-construction focussed interactions two (agreeing), three (disagreeing), four (negotiation of agreeing and disagreeing), five (cooperative task accomplishment), six (discussion about usage), and seven (discussion about an example/s of usage) in the following:

(S56)[14] 'Imagine something will happened or suppose if it's happened.' And, can, can you explain in your own world, uh word the difference between the two?

(S44) Hmm, difference I think it just uh *would* can use in, it have different way to use it. And the *would* is, can using the, uh it's uh maybe it's not sure.

(S56) Mhm. It's not sure. Mm.

(S44) Maybe it will happen or not.

(S56) Mhm.

[…]

(S43) And *would* is a past form for will.

(S44) And *would* is the past.

(S56) No.

(S44) No?

(S56) No, no…[laughs] It's not in the past. But I didn't, I didn't learn that in my life.

(S43) In your opinion what's [?] *would*?

(S56) Me, I was thinking that would it's a conditional, that it's not sure that it will, it will happen.

(S43) Yeah, it's a conditional.

(S56) Yeah, so that for me it's not in the past, it's "Would you please come with me?" that's not past form. That's not something in the past, no?

(S43) No, it's a, it's a form of *will* in the past. [?]

(S44) The same use as *should*.…uh someone talking in the past use *would*.

(S56) Can you give me an example when *would* is in the past?

[writing]

(S56) Yeah, OK, "Yesterday when your friend came, would you be happy if he brought a bottle of wine?" Is that OK?

(S43) Yeah, but it's a question.

[14] It has been reported in this study that there were 55 student participants. A female student here called S56, has not been included in the data set of the rest of the study since this student participated only in this exercise.

5.3.1.2 Summary of student group CM production observations.

These results indicate that the groups in the two classes appear to have used somewhat similar means to demonstrate an understanding of the use of *would* and *will* through their co-constructed concept maps. The evidence would suggest that these learners sometimes demonstrated an understanding of the usage of *will* and *would*, yet at other times they demonstrated inaccurate perceptions of the usage. These data suggest how difficult and challenging second language learning may be when adult learners try to decode complex and subtle L2 grammar points.

The disagreement about the use of *would* as a past tense form of *will* that was brought to light in the TG2 group 3 discussion was not resolved in the group discussion. In fact, it would remain unresolved until the teacher-class CM production exercise. In the next section, the results of the teacher-class CM production exercises will be reported and presented.

5.3.2 Field notes of teacher-class CM production exercise observations.

Field notes and audio recordings of classroom observations were analyzed qualitatively. Classroom observations revealed a high level of learner engagement and interest in class discussions about modal usage.

5.3.2.1 Categorization of interactions in teacher-class CM production.

5.3.2.1.1 Teacher interactions.

From a teaching perspective, twelve categories of teacher interactional behaviours emerged from the data. The twelve teacher interactions were the following:

1. Writing on the blackboard
2. Writing on an acetate page of an OHP
3. Rule explanation
4. Giving examples of usage
5. Recasting student responses
6. Asking learners “What do you think I want you to learn from this lesson?”
7. Guided questioning leading to a rule explanation
8. Providing negative evidence

9. Formative comprehension checking
10. Communicating to students the results of formative comprehension checks
11. Communicating to students teacher expectations and goals
12. Assigning a task with the goal of assessment and learning

The focus of interactions 6 to 12 tends more than the others towards an AFL/formative assessment student-centred approach. However any or all of the above may possibly have contributed to any classroom learning. It is noteworthy that interactions 6 and 7 have a co-constructive component in that they have required a learner response to complete them.

5.3.2.1.2 Student interactions.

In the previous sections, 15 categories of student interaction were identified and characterized as being 'co-construction focussed' or 'individual focussed.' Three of these student interaction categories were observed in the teacher-class CM production exercises. They were the following:

6. Discussion about usage
7. Discussion about an example/s of usage
10. Rule explanation

Reference to these teacher and student interactional categories will be made in reporting the transcribed discourse of the teacher-student CM production exercises in the next sections. Another interaction category that applies to both teachers *and* students that emerged from the teacher-class CM production exercise data, and will be included in this reporting. That interaction will be numbered 13 and called *discussion about learning.*

5.3.2.2 Transcriptions of teacher-class discussions.

Research question four asks *Does a teacher-class concept mapping (CM) exercise enhance the learning of a specific language feature in a second language classroom?* Samples of transcribed teacher interactions and student responses which contribute to answering this question will be reported in the following section. Complete transcriptions of the group CM production exercises are available from the researcher.

To revisit the teacher-class CM production exercise methodology, the exercises began with students from the three groups of each class presenting their group concept maps to their teacher and class, explaining the group's understanding of the usage of *will* and *would*. These group presentations were followed by a teacher and class discussion about their understanding of the target usage, in a co-constructed teacher-class concept map production exercise.

The transcription protocol that was used can be found in Appendix M. The teacher and student interactional categories that have been identified will be presented as well. When it is known, the students have been identified by their student codes (for example S3); when not known, Roman numerals have been used to differentiate between student speakers within each excerpt.

5.3.2.2.1 TG1 class.

The TG1 teacher-class CM production exercise took 1 hour, 3 minutes during part of one class period. A TG1 student's explanation demonstrating an understanding of the use of *will* and *would* illustrates co-construction focussed student interactions six (discussion about usage), seven (discussion about an example/s of usage), ten (rule explanation), and thirteen (discussion about learning), in the following excerpt:

> S3 – It said describe thing that might be happen, let's, for example like um, "I would go to the party tomorrow if I finish my homework." So, if I finish my homework I might go, but if I cannot finish it I might not be going. So it's like you're not sure what's gonna happen but you just like uh think of that's gonna happen or it's not gonna happen.

This excerpt from the TG1 teacher-class discussion, included an element of Teacher interaction seven (guided questioning leading to a rule explanation), and eight (providing negative evidence):

> TG1 teacher: OK, so for the future, um are we correct in saying "I will be a doctor in the future?"
>
> (SI) Yes.
>
> TG1 teacher: Because maybe yes, maybe no.
>
> (SI) No, because I am going, I am studying [?]
>
> [all talk together]

(SI) But this is planned OK? So you have to use I am going to…

(SII) But for the future?

TG1 teacher: For the future, OK, maybe yes, maybe no. […] Maybe when you get to university and you finish all of your courses instead of being a doctor maybe you'll want to be a teacher, alright? If you are *definite* about something, you are *going to*…

Harlen and Winter (2004) suggested AFL teachers ask students to reflect on what they thought their teacher had wanted them to learn. Drawing on this suggestion, the treatment teachers were asked to do this as well. The TG1 teacher asked this question in class, but the TG2 teacher did not. Here is a transcription of the TG1 teacher's question and some of the student responses it drew:

TG1 teacher: What do you think you have learned from this? Or what do you think I want you to learn from this?

(SI) The differences between the use of, how to use *will* and *would*?

TG1 teacher: How to use *will* and *would*?

(SII) In different ways.

(SMIII) Because, because uh a lot of people, as I heard even native speakers they don't uh practice *would* and *will* probably, no?

TG1 teacher: OK, [writes on board] and this is the native speakers. What else? Someone came up with another idea. [?]

(SIV) [?]

TG1 teacher: How to use, the idea of concept maps in university. Anything else?

(SIII) Well uh, the easiest way of [?]

(SV) Yeah, a lot of people are attract to see that map.

TG1 teacher: As a way to learn?

(SIV) Yes, because maybe I put some cont-, same contents in both of those maps but uh one will, will, uh, the information in one will be demonstrated in very good ways so it will be easiest to understand.

TG1 teacher: OK, so seeing it done in different ways has helped you?

(SIV) Yes.

TG1 teacher: OK. Anything else?

(SV) And also like for example if you write just, just write on, on board and, and other, another way you draw the map, so people more like to read the map, not just writing, so, yeah. It is attractive to people I would say.

TG1 teacher: Alright. Again we work with making it easier to learn. So some people like to see it all written out, others like to see a picture.

During the discussion that followed, the TG1 teacher summarizing the students' responses on the blackboard, which have been reproduced here in Figure 25.

When the TG1 teacher asked the students what they had learned, the student responses suggest that the *group* CM production exercise may have supported learning. The following demonstrates category 13 (discussion about learning):

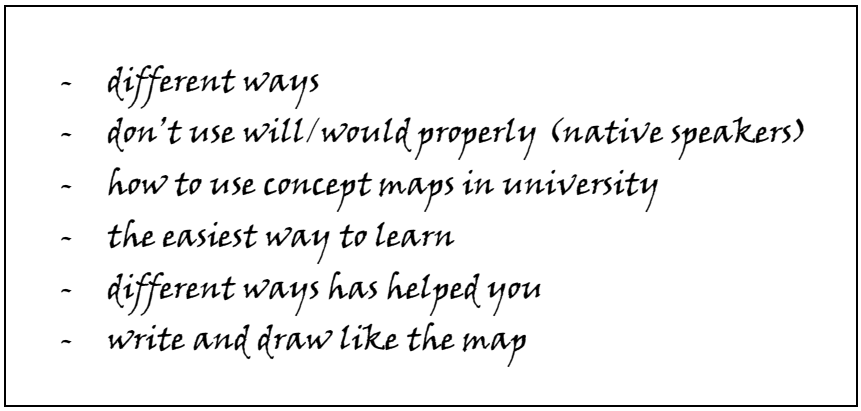


Figure 25 TG1 teacher blackboard summary notes after students were asked "What do you think I want you to learn from this?"

TG1 teacher: What have you learned?

(SI) Because they are similar, uh, revising the rules and giving some examples help us, helped us to, to remember the, the differences…

TG1 teacher: Oh, so it helped to identify…

The discourse that follows also demonstrates interaction category 13 (discussion about learning) and is presented as learner evidence of co-constructed learning:

(SII) And to work as a group.

TG1 teacher: Oh, and to work as a group… to find out your differences, in the spirit of cooperation.

(S?) …Maybe if I have a doubt about something, if I am in a group, maybe I will convince and maybe the group will convince me, so working with a group is best.

(S?) Best than alone because when I work alone first time I am confused and I don't know how to make the map. And now it's more clear as a group because I ask some question from you and from my friends, and it's better that way.

In this final comment SIII's comments show evidence of the AB since they reflect the student's need for teacher correction in order for learning to take place:

(SIII) Also we need someone to correct us because when we have different views we don't know which one is right.

In the next section excerpts from the TG2 teacher-class CM production exercise which were somewhat different from those of this class, will be presented.

5.3.2.2.2 TG2 class.

The TG2 teacher-class CM production exercise took 1 hour, 16 minutes during part of one class period. Unlike the TG1 teacher, the TG2 teacher did not ask her students "What do you think I want you to learn from this?" However, in the execution of the TG2 teacher-class CM production exercise, several AFL/formative assessment student-centred teacher interactions occurred, namely:

8. Providing negative evidence
9. Formative comprehension checking
10. Communicating to students the results of formative comprehension checks
11. Communicating to students teacher expectations and goals
12. Assigning a task with the goal of assessment and learning

and

13. discussion about learning

The following excerpt illustrates teacher interaction three (rule explanation):

TG2 teacher: So if I say "Will you come to our party" or "Would you come to our party?" …Which one is more formal? I think would is a *little* more formal - I think we use it almost equally, but *probably will* is stronger, *will* is always stronger, more certain in any way, *will* is always *certainty*, something is going to happen. What about willingness to do something?

(S) [?] "I would help you with your homework."

This excerpt illustrates teacher interaction three (rule explanation) *and* eight (providing negative evidence):

TG2 teacher: I don't think you would ever say just the sentence "I would help you with your homework" unless you had a condition, if you need help! […] I think the polite request you understand, it can be *will*, it can be *would*, OK? I think the part that I would like to talk about is the part that he [a student] mentioned, which is that *if*, he didn't really understand, or he couldn't really explain what it is. But he did say, what did you say? You said something interesting, I think you said if you imagine that something will happen or maybe will happen, supposedly will happen. Now maybe that's the part that we need to talk about a little clearer. This group talked about, […] What was your sentence? "I would help you with your homework" and how can you call that? You called it willingness. But there's a willingness with a condition, OK? […] And willingness with a condition is the form *if*.

The TG2 teacher frequently verified the students' learning progress throughout the class observed. This excerpt is an example of this, in interaction nine (formative comprehension checking):

TG2 teacher: So do you think you understand this point now? Yes or no, do you think by having done it [the group CM exercise] do you think you understand, maybe not 100% but a little bit?

(SI) Yeah.

TG2 teacher: How mu-, what percentage, do you think you understand it perfectly? 50%, 60%, are you understanding? What do you think?

(S?) [?]

TG2 teacher: It, it was still difficult. But you thought you understand, but sometimes the examples were difficult for you. OK. And in this group? Do you think you understood most of it or not? Yes, no?

(S) [?]

TG2 teacher: 80% wow [student name], good. Oh, 60, [laughs]

(S) [laughs]

TG2 teacher: OK. And you?

(S) [?]

TG2 teacher: Maybe 70%, OK.

The TG2 teacher also often exhibited teacher interaction 10 (communicating to students the results of formative comprehension checks). An example of this follows:

TG2 teacher: I'm not sure if you have grasped that clearly, but the rest, I think the rest [of it] you have more or less understood.

and

I think you understand *clearly* the request, OK, and uh the cond-, now you understand a little more the condition.

and

TG2 teacher: So I think you have understood fairly clearly.

The TG2 teacher also demonstrated teacher interaction 11 (communicating to students teacher expectations and goals), as evidenced in this excerpt:

TG2 teacher: I don't expect you to be able to use this so perfectly; I expect you to try to understand the differences.

At the end of this lesson, the TG2 teacher assigned the students a task with the goal of demonstrating their learning for formative assessment as well as to help learning, thus showing teacher interaction 12 (assigning a task with the goal of assessment and learning) as can be seen in this excerpt:

TG2 teacher: Can I ask you a little bit to try and get a sentence for me (not so much with this past, cause that's a little bit complicated), but with uh *if* and *will*, and *if* and *would* sentence? Can you try to make up a sentence to show, make you understand, in your groups? Think about it and write a sentence with *will* and with *would*. OK? Let me see whether you understand or not. Go ahead. You can write just one between all of you if you want to discuss or you can write individually.

The above shows evidence of the AB since the students' assigned task is designed to both show and help learning.

The student contribution in this exercise has been little reported up until now since it was minimal and it consisted primarily of providing examples of usage. Here are some examples of student interaction seven (discussion about an example/s of usage):

(SI) *Will* is formal and would is informal.

(SII) [?] "I would help you with your homework."

(SIII) It would cause accident if I drive too fast.

(SIV) "If you go to park, I will go with you."

(SV) "If pigs could fly, I would get married to you." [teacher and students laugh]

(S32) "If I were man, I would teach other man how to treat women". [teacher and students laugh]

Following the students' stating erroneous examples of usage, the TG2 teacher guided the students to target usage through questioning and with explanations.

The evidence here might indicate that the approach in the TG2 class was more teacher-centred than in the previous, TG1 class. However, the frequent student-centred interactions of the TG2 teacher would suggest that there was in fact a strong AFL/formative assessment component in both classes despite the teachers' somewhat varied teaching methodologies.

5.3.3 Teacher incidental comments

Prior to the study the treatment and control group teachers were asked by the researcher whether modal forms had been taught in their classes (since modal usage was part of the curricula for the level). The TG1 teacher said that prior to the study, she had in recent time been working "a lot" with her students on learning the conditional but that in her estimation, they had "not got it yet." In contrast the TG2 teacher said she had not yet taught modal usage, except for occasionally correcting students when errors in modal forms came up in student work. The control group teachers reported that they had not taught modals at the time of the study.

In classroom observation field notes and incidental comments the teachers and student participants expressed to the researcher much interest and enthusiasm with the online CM task. This was particularly evident in the TG1 group, who demonstrated a great deal of motivation and cohesion overall. In

fact, the teacher of the TG2 group mentioned that the group had been together in classes for a much shorter time than had the TG1 groups, and that she felt they would be less engaged in any task as a result.

During classroom observation of the TG1 class a student was observed to comment "We like to make the map more interesting so students can learn." This leads back again to the central topic of this research study, learning and the question of whether learning did take place in this classroom setting.

In the next section, evidence that was only analyzed by quantitative methods will be presented.

5.4 Quantitative results

In order to identify trends in these students' performance on FITBs tests and on essay task tests, and to ascertain the level of agreement with the AFL principles of the teachers and students, the following instruments were analyzed solely by quantitative means:

- the essay tests
- the FITB tests
- the teacher 10 principles questionnaire surveys
- the student 10 principles questionnaire surveys

The results of each will be presented in turn in the following sections.

5.4.1 The essay test results in the treatment and control classes

The statistics that follow have been reported as indicators of possible student progress in their learning about modal usage and in automatic production of the modal forms.

Paper-and-pencil essay production tests were administered, which required the students from the treatment and control groups at Times 1 and 2 to produce one-page essays. These tests were developed for the study in order to assess student competence in producing target contingency use modals in writing tests without the modal having been provided in the prompt (see Appendix D).

5.4.1.1 Modal production in percentages in the pre-, post- and delayed post-essay tests.

Ratio data on the total number of target modal forms produced per obligatory context, has been reported as percentages in Table 10. Table 10 shows the markedly more competent modal production of the TG1 class over all other classes in the pre-test results, with a 40% group success rate in producing target modals in obligatory contexts. (This is notwithstanding the (n=2) success rate of the two individuals to whom the delayed post-test was administered, the results of which were discounted due to the limited sample size).

Table 10: Group ratios of target modal forms produced per number of obligatory contexts in essays, expressed as percentages

	Test period		
	Pre	Post	Delayed post
Group			
Time 1			
Treatment	40 (n=11)	17 (n=10)	44 (n=2)
Control	9 (n=16)	10 (n=11)	- -
Time 2			
Treatment	7 (n=14)	20 (n=11)	34 (n=7)
Control	7 (n=11)	7 (n=8)	58 (n=6)

Nevertheless, the TG1 class results fall to 17% in the post-test. The real 'success' can be seen to occur in the TG2 class, whose results rise from 7% to 20% from the pre- to post-test. This was augmented further in the delayed post-test target production rate of 34%. Noteworthy here is the fact that the

delayed post-test was administered to exactly half the participants relative to the pre-test administration.

The control groups, however, showed less variation, as the CG1 group percentage went from 9% to 10% target production in obligatory contexts from pre- to post-test. There were no CG1 participants to take the delayed post-test. The CG2 group, similarly, changed not at all from the pre- to post-test results, with 7% target production in both. However, the CG2 group produced a somewhat surprising 58% of target modals in obligatory contexts in the delayed post-test. It is noteworthy that the 'strongest' pre-test group, the TG1 group, did not on any test occasion produce even half the number of target modals they ideally should have and their Time 2 counterparts produced only 7% on their first test.

Participant attrition resulted in a difficulty in finding students to take the delayed post-tests, particularly in the case of the TG1 and CG1 groups. Only two TG1 students and none of the CG1 students took the delayed post-tests. Thus, due to the small participant numbers, those test results were not quantitatively compared to the pre- and post-tests of Time 1.

All of these results, despite the very small number of participants, suggest that modal production presented a considerable challenge to all of these learners at the Advanced II level, as the treatment group teachers had, in fact, attested to before the study began.

5.4.2 The fill-in-the-blanks (FITB) test results

Parallel paper-and-pencil fill-in-the-blanks (FITB) tests were administered as above, to all students in the study. In these tests, *would* and *will* forms were provided to the students for them to make a choice of the correct form to use to complete each of ten sentences (see Appendix E).

5.4.2.1 Modal production in percentages in the pre-, post- and delayed post-FITB tests.

In the FITB tests, when a choice of *would* and *will* form was provided, the students across all groups performed much better than they did in the essay tests. The results were transformed into class percentages, and they appear in Table 11.

Table 11: Group percentages of correct form choices in pre-, post-, and delayed post-treatment fill-in-the-blanks tests

	Test period		
	Pre	Post	Delayed post
Group			
Time 1			
Treatment	78 (n=11)	69 (n=12)	85 (n=2)
Control	68 (n=16)	63 (n=12)	- -
Time 2			
Treatment	70 (n=14)	79 (n=12)	79 (n=7)
Control	75 (n=11)	89 (n=8)	87 (n=6)

The results of the class percentages of the FITB pre-test indicate that the treatment and control groups performed much more similarly than they did in the essay tests, within a 10% range from the lowest group scores of the CG1 group (68%) to the highest TG1 group (78%), with the other groups, TG2 (70%) and CG2 (75%), falling in between. In the post-tests the Time 1 group scores diminished from those of the pre-tests in both the TG1 (69%) and the CG1 (63%) groups. On the contrary, the Time 2 group scores rose in TG2 (79%) and in CG2 (89%). The small number of delayed post-test participants in Time 1 (n=2) preclude group comparisons.

5.4.3 Teacher questionnaire results on the AFL 10 principles

Questionnaires surveying teacher views on the ARG (2002) 10 principles of AFL were collected from the participant teachers (see Appendix K). The study had been designed with a protocol to collect one pre-treatment and one post-treatment questionnaire from two treatment and two control teachers for a total of eight surveys. However a total of six questionnaires were collected from the

teachers as noted in the following listing of the questionnaires the teachers submitted:

- The TG1 teacher submitted one post-treatment questionnaire.
- The TG2 teacher submitted one post-treatment questionnaire.
- The CG1 teacher submitted one pre- and one post-treatment questionnaire.
- The CG2 teacher submitted one pre- and one post-treatment questionnaire.

The TG1 and TG2 teachers were unable to submit a pre-treatment questionnaire. For operational reasons the TG2 teacher, it will be recalled, was also the CG2 teacher and for this reason did not submit Time 2 pre- and post-treatment questionnaires. For these reasons two out of a possible eight teacher questionnaires were not collected, three of the questionnaires were from the same person and had identical responses, and only one teacher questionnaire, (of the Time 2 control group) showed any difference in responses from the pre-treatment to the post-treatment period.

The results of the TG1 post- questionnaire, as well as the combined TG2-CG1 teacher questionnaire responses were analyzed. Thus, teacher questionnaire data could be reported for the three TG1, TG2-CG1, and CG2 teachers. The results of this analysis are reported in Table 12.

The Table 12 results show that the three teachers were in general agreement with the 10 principles, with one exception. Unlike her colleagues, the CG2 teacher disagreed with principle 10, that *formative assessment in the classroom should recognize the full range of achievements of all learners.* On no items did the teachers strongly disagree, so that ranking is not included in Table 12.

The teachers agreed with all of the other principles, in varying degrees. All three teachers (TG1, TG2-CG1, CG2) strongly agreed with principles 8 and 9. The two treatment group teachers were both in strong agreement with principles 1, 2, 4, 6, 7, and 10.

The CG2 teacher also agreed strongly with principle 2 in her post-treatment questionnaire response. The TG2-CG1 teacher alone was in strong agreement to principles 3 and 5. The TG1 and CG2 teachers were in agreement with

Table 12: Times 1 and 2 treatment and control class teacher AFL 10 principles pre- and post-treatment questionnaire responses

Formative assessment in the classroom should…	2- disagree	3- agree	4- strongly agree
1. be part of effective planning of teaching and learning		CG2	TG1, TG2 (CG1)
2. focus on how students learn		CG2: pre-treatment	TG1, TG2 (CG1), CG2: post-treatment
3. be recognised as central to classroom practice		TG1, CG2	
4. be regarded as a key professional skill for teachers		CG2: post-treatment	TG1, TG2 (CG1)
5. be sensitive and constructive given that assessment has an emotional impact		TG1, CG2	TG2 (CG1)
6. take account of the importance of learner motivation		CG2	TG1, TG2 (CG1)
7. promote commitment to learning goals and a shared understanding of the criteria by which they are assessed		CG2	TG1, TG2 (CG1)
8. give learners constructive guidance about how to improve			TG1, TG2 (CG1), CG2
9. develop learners' capacity for self-assessment so that they can become reflective and self-managing			TG1, TG2 (CG1), CG2
10. recognise the full range of achievements of all learners	CG2		TG1, TG2 (CG1)

principles 3 and 5. The CG2 teacher alone was in agreement (not strong agreement) with principles 1, 2, 4, 6 and 7. That teacher did not respond to the principle 4 item in the pre-treatment questionnaire, but all responses were obtained for all of the other items from the participant teachers.

Only one teacher response changed from the pre- to post-treatment questionnaire administration; the CG2 teacher indicated agreement in the pre-questionnaire to principle 2 (*Formative assessment in the classroom should focus on how students learn*) and then strong agreement with principle 2 in the post-treatment questionnaire.

5.4.4 Student questionnaire results on the AFL 10 principles

The students of the treatment and control groups were surveyed at pre- and post-treatment to ascertain to what extent they were in agreement with the ARG (2002) 10 principles of AFL (see Appendix L). The frequency of student responses in the four Likert-scale categories (*strongly disagree, disagree, agree, strongly agree*) and group total numbers of the rankings will be reported for all class groups in Tables 13-16.

5.4.4.1 TG1 class.

The results of the TG1 student questionnaire responses appear in Table 13 (pre- n=14, post- n=12). The TG1 students were in general agreement with the 10 principles, with the students mostly ranking the principles (pre=63, post-66) in the *agree* category. The greater part of the responses didn't change from the pre- to post-questionnaire administration, however in the *strongly disagree* category, the few students who chose it fell by half (pre=6, post=3).

5.4.4.2 TG2 class.

Table 14 shows the TG2 group responses to the questionnaires (pre- n=14, post- n=13). In the TG2 group the students displayed much more agreement with the principles than did the TG1 group. As was the case with the TG1 treatment group, most of their responses were in the *agree* category (pre=75, post=86), where there was an increase from the pre- to post-treatment period. Again, echoing the TG1 group responses, there were much fewer responses in strong agreement (pre=58, post=35) following the AFL treatment.

Table 13: Time 1 treatment class (TG1) student AFL 10 principles pre- and post-treatment questionnaire responses

	Formative assessment in the classroom should…	1- strongly disagree		2- disagree		3- agree		4- strongly agree	
		Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
1.	be part of effective planning of teaching and learning	1	0	0	1	9	8	4	3
2.	focus on how students learn	0	0	0	0	9	7	5	5
3.	be recognised as central to classroom practice	0	0	3	3	7	7	4	2
4.	be regarded as a key professional skill for teachers	1	0	5	3	3	7	4	2
5.	be sensitive and constructive given that assessment has an emotional impact	1	1	5	4	7	4	1	3
6.	take account of the importance of learner motivation	1	0	4	1	8	7	1	4
7.	promote commitment to learning goals and a shared understanding of the criteria by which they are assessed	0	0	0	2	5	7	9	2
8.	give learners constructive guidance about how to improve	0	0	0	1	5	5	9	5
9.	develop learners' capacity for self-assessment so that they can become reflective and self-managing	0	1	0	1	7	10	7	-
10.	recognise the full range of achievements of all learners	2	1	2	3	3	4	7	4
	Total responses per rank	6	3	19	19	63	66	51	30

Pre-treatment (n=14), post-treatment (n=12). Not all item responses equal 14 or 12 due to missing responses.

Table 14: Time 2 treatment class (TG2) student AFL 10 principles pre- and post-treatment questionnaire responses

Formative assessment in the classroom should…		1- strongly disagree		2- disagree		3- agree		4- strongly agree	
		Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
.	be part of effective planning of teaching and learning	0	0	0	0	8	8	6	4
2.	focus on how students learn	0	0	0	0	3	6	10	6
3.	be recognised as central to classroom practice	0	0	0	1	9	9	4	2
4.	be regarded as a key professional skill for teachers	0	0	1	0	6	11	6	0
5.	be sensitive and constructive given that assessment has an emotional impact	0	0	2	1	7	7	5	4
6.	take account of the importance of learner motivation	0	0	1	0	9	11	4	2
7.	promote commitment to learning goals and a shared understanding of the criteria by which they are assessed	0	0	0	0	10	10	4	3
8.	give learners constructive guidance about how to improve	0	0	0	0	5	5	9	8
9.	develop learners' capacity for self-assessment so that they can become reflective and self-managing	0	0	0	0	8	10	6	3
10.	recognise the full range of achievements of all learners	0	0	0	1	10	9	4	3
	Total responses per rank	0	0	4	3	75	86	58	35

Pre-treatment (n=14), post-treatment (n=13). Not all item responses equal 14 or 13 due to missing responses.

Table 15: Time 1 control class (CG1) student AFL 10 principles pre- and post-treatment questionnaire responses

Formative assessment in the classroom should…		1- strongly disagree		2- disagree		3- agree		4- strongly agree	
		Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
1.	be part of effective planning of teaching and learning	0	0	0	0	12	9	4	3
2.	focus on how students learn	0	0	1	0	6	4	9	8
3.	be recognised as central to classroom practice	0	0	2	1	13	7	0	4
4.	be regarded as a key professional skill for teachers	1	0	2	1	9	7	4	4
5.	be sensitive and constructive given that assessment has an emotional impact	1	0	4	5	10	6	1	1
6.	take account of the importance of learner motivation	0	0	1	1	10	7	5	4
7.	promote commitment to learning goals and a shared understanding of the criteria by which they are assessed	0	0	2	0	7	11	6	1
8.	give learners constructive guidance about how to improve	0	0	0	0	5	6	11	6
9.	develop learners' capacity for self-assessment so that they can become reflective and self-managing	1	0	1	0	7	6	7	6
10.	recognise the full range of achievements of all learners	1	0	0	1	13	6	2	5
	Total responses per rank	4	0	13	9	92	69	49	42

Pre-treatment (n=16), post-treatment (n=12). Not all of the pre-treatment item responses equal 16 due to missing responses.

Table 16: Time 2 control class (CG2) student AFL 10 principles pre- and post-treatment questionnaire responses

Formative assessment in the classroom should…	1- strongly disagree		2- disagree		3- agree		4- strongly agree	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
1. be part of effective planning of teaching and learning	0	0	0	1	8	2	3	5
2. focus on how students learn	0	0	2	0	7	4	2	4
3. be recognised as central to classroom practice	0	0	2	1	6	8	2	0
4. be regarded as a key professional skill for teachers	1	0	0	2	6	4	3	3
5. be sensitive and constructive given that assessment has an emotional impact	2	0	3	2	4	2	2	4
6. take account of the importance of learner motivation	0	0	1	0	5	5	4	3
7. promote commitment to learning goals and a shared understanding of the criteria by which they are assessed	0	0	1	1	7	7	2	0
8. give learners constructive guidance about how to improve	0	0	0	0	3	3	8	5
9. develop learners' capacity for self-assessment so that they can become reflective and self-managing	0	0	2	0	4	3	5	5
10. recognise the full range of achievements of all learners	0	1	2	0	8	6	1	1
Total responses per rank	3	1	13	7	58	44	32	30

Pre-treatment (n=12), post-treatment (n=8). Not all item responses equal 12 or 8 due to missing responses

5.4.4.3 CG1 class.

The results of the CG1 group questionnaire appear in Table 15 (pre- n=16, post- n=12). As in the TG1 and TG2 group responses, most responses were in agreement with the principles (pre=92, post=69), yet there were fewer responses in that category in the latter survey. The responses to the *disagree* and *strongly agree* categories were little changed (pre=13, post=9 and pre=49, post=42 respectively) in both the pre- and the post-treatment questionnaire administrations. In this group there was very little strong disagreement expressed in the pre- questionnaire, and none in the post-questionnaire (pre=4, post=0).

5.4.4.4 CG2 class.

The CG2 group response results are shown in Table 16 (pre- n=12, post- n=8). Once again, as was the case with the other groups surveyed, the CG2 group mostly agreed with the 10 principles (pre=58, post=44) although less so in the second survey. Their strong agreement ranking remained relatively unchanged over the two questionnaire administrations (pre=32, post=30). There were a few students who strongly disagreed with two of the principles, but fewer when surveyed a second time (pre=3, post=1). This was again the case with those students who disagreed (pre=13, post=7).

5.4.4.5 Overview of student responses to the AFL 10 principles.

Tables 13 to 16 present a picture of student responses to the 10 principles questionnaires in their groups and in general. For example, it can be seen from Tables 13 to 16 that the principle in most common disagreement was number 5, which states that *formative assessment in the classroom should be sensitive and constructive given that assessment has an emotional impact.*

In concert with the data in the Tables thus presented, a picture of these students' concurrence with the 10 principles, or lack of it, can be seen by collapsing the pre- and post-treatment group responses into a reporting of unqualified agreement and disagreement. Table 17 shows this binary view of the results in all of the groups surveyed.

Table 17: Frequency of class agreement and disagreement with the 10 principles of the ARG in pre- and post-treatment questionnaire responses

	AFL 10 principles	
Class	Disagreement (F)	Agreement (F)
Treatment		
TG1	47	210
TG2	7	254
Control		
CG1	26	252
CG2	24	164

From Table 14 it is clear that these groups generally agreed with the AFL 10 principles and that there was some variation in how often they did. The two treatment groups disagreed with a wide gap between them (TG1=47, TG2=7). There was a quite similar level of agreement across all of the groups. It is notable that the results of the TG2 agreement level (F=254) and that of the CG2 group (F=252) are almost identical. Similarly, disagreement in the two control groups is nearly identical as well (CG1 F=26, CG2 F=24).

5.5 Chapter summary

In this chapter the results of analysis by MM have been presented. Following that the results of the data analyzed by qualitative means alone were reported. Finally, the quantitative analysis results of the essay and FITBs tests have been shown, in their capacity to reflect trends.

In the next chapter, the results presented herein will be discussed with a view to understanding how they may help to answer the central research question at hand, which asks if there can be evidence accrued as to how L2 AFL practices may have enhanced student learning in this setting.

CHAPTER 6 – A DISCUSSION OF THE RESULTS

Therefore acknowledge your error and be attentive[15]

Christopher Marlowe

This chapter will first discuss the central research question relative to the data that have been presented. Following that, the seven secondary research questions will each be discussed in relation to these results. Then, the body of data as a whole will be triangulated in order to fully address all of the research questions.

6.1 The central research question

The primary research question of this study was, *Is there evidence that AFL practices in a second language classroom enhance the learning of a specific language feature?*

The results suggest that the AFL practices applied in this L2 setting may have enhanced the learning of the modal language feature. Strong support for this assumption has been presented in the form of learner and teacher survey reports of their perceptions that learning had occurred. Moreover, some student participants expressed a perception that, while they had not previously understood the modal feature targeted in this study, they had begun to do so during and after the treatment experience. Classroom observation field notes, audio-recordings, and audio transcriptions also support this view.

Evidence has been presented in the form of student responses to the Mystery Module and Concept Mapping Exercises Teacher Post-Questionnaire survey, in which *all* of the students in the TG1 and TG2 classes reported that they felt that they had learned from the teacher-class CM production exercise. Similarly, both of the treatment class teachers reported in their responses to the Teacher Mystery Module and Concept Mapping Exercises Post-Questionnaire that their students had learned from the treatment exercises.

Black and Wiliam (1998a) have supported the premise that for learning to take place, there must be noticing of a gap between actual and desired competence on the part of the learner or another agent such as a teacher, to

[15] Marlowe's sentiment here seems apt; in a sense he is calling for attending to and 'noticing' a gap in one's knowledge, as does AFL.

draw attention to that gap for the benefit of the learner (p. 20). This echoes the work of Sadler (1989), discussed in Chapter 2. It will be recalled that Sadler suggested learners needed three factors to improve their learning, that is, that they be able to:

1. understand the criteria and standards that were aimed for
2. be able to assess their own performance and to compare it with the standards, and
3. be able to engage in action in order to close the gap.

The evidence presented here suggests that some of these learners may have begun the process Black and Wiliam (1998a) and Sadler (1989) have alluded to. Moreover, it could be argued that the CALL and CM production exercises may have afforded these learners an occasion to begin to understand the criteria inherent in the accuracy of grammatical form required in the exercises.

In Chapter 1, I have cited Larsen-Freeman's (2009) dictum that a definition of grammar must include the elements of form, meaning and use (see Section 1.4). It is possible that in the present study, the treatments may have encouraged or allowed attention to grammatical understanding of the criteria and standards aimed for, in these students' attempts to learn how to use the target forms. Thus, these learners may have been able to or begun to assess their own performance in comparison with an 'other' in the form of a computer, peer, or teacher. This may also be in accordance with Purpura's (2004) definition of grammatical ability for assessment purposes (also cited in Section 1.4), which he deemed to be the capacity in test-taking or other language-use contexts to realize grammatical knowledge with accuracy and with appropriate meaning. In addition, the learners in the present study may have taken action to close the gap between their IL and target forms when they engaged in dynamic assessment opportunities in the CM production exercises, as demonstrated in the reporting of the results, and they may have co-constructing knowledge and, it is hoped, learned about the target forms.

Learner discourse in the co-constructed peer interactions of the group CM exercises gave the learners the possibility of comparing and contrasting their own performances with those of their student colleagues.

In fact, a strong majority of the student participants reported their perception that they had learned about the usage of the modal in question

through the group and teacher-class CM exercises. Evidence of learner metacognition observed in group and teacher and class discussions have been presented in the form of transcribed learner voices from group and class CM production activities.

In order to more carefully examine this overall interpretation of the results, the secondary research questions will be addressed and the results that have been presented will be discussed relative to each.

6.2 The secondary research questions

6.2.1 Research question 1: The CALL question

Research question 1 asked, *Is there evidence that a computer assisted language learning exercise in a second language classroom enhances the learning of a specific language feature?* The data suggest that the learning of the modal feature used in the study may have been enhanced when these learners engaged in CALL exercises.

In Chapter 3, Chapelle's (2005) view was cited, wherein she expressed the premise that multimedia computer programs may promote learner attention to the gap between their IL and target language forms, and help them advance in their learning. I have posited that similarly, the L2 AFL CALL environment used in the treatment of this study could be useful tool in helping learners notice that gap as well. I would suggest that the learner interaction with a computer as 'nonjudgmental target language producer' might have allowed the learners in the present study to be led to greater metalinguistic awareness of non-target 'errors' without compunction.

Accordingly, the CALL Module with its sectional hyperlinks was developed to give students self-learning tools. To revisit, they were:

- online practice producing IL and target forms through 'interactions' with the computer
- answer grid pages intended to promote learners' noticing of differences between their IL and target forms
- a metalinguistic component in the form of links to a 'clues' page to foster understanding of target form usage in different contexts
- self-assessment checklists to help learners track their target production progress

- checklists to help learners use metalinguistic conceptualizations to track their language learning progress.
- an interactive self-assessment quiz including a learning component in the form of access to hints to guide learners towards an understanding of modal form usage.

In summary, the CALL exercises have been designed with the objective of fostering learner engagement, reflective thinking, hypothesizing, and activating background knowledge prior to the CM exercises, and in addition, to encourage self-assessment in the learners, all of which might lead to 'noticing'. So, to return again to the research question one can ask, did this CALL experience in fact have the intended result of learning taking place? The results of this study suggest that CALL may have enhanced the learning of this specific language feature in this setting.

The learners did, in fact, submit much evidence to this effect in the form of post-treatment surveys, in which the majority of the students attested to their perceptions that the CALL exercise had helped them learn. In addition, many of the students described metacognitive self-reflections on their learning experience of doing the CALL exercises.

The majority of the students also responded positively to the post-treatment question asking whether they had found the Module self-assessments to be useful. It was found that very few of the students said the CALL Module had *not* helped them learn. Even those student who expressed reservations about the usefulness of the Module relative to their learning, answered 'yes' when asked if CALL had helped them learn, in spite of, and in some contradiction to their comments.

Both treatment teachers also supported the view that the CALL Module had helped their students learn, as evidenced in their incidental comments to the researcher and in their comments in the teachers' post-treatment surveys. The TG1 group teacher reported of the students "They were <u>interested</u> in knowing what the Mystery Module was all about," [underline in the original].

6.2.2 Research question 2: The online CM question

In Chapter 3, the work of Vygotsky (1969, 1978) was cited as having been instrumental in inspiring the idea of using concept maps in this study to help

students learn. In particular, Vygotsky's theory of learners' ZPD played a part in the decision to challenge learners to go beyond their *actual* competence and to try to reach in some measure their *potential* competence. To investigate whether using concept maps had broadened these learners' L2 learning towards their potential, the second research question asked, *Is there evidence that an online concept mapping exercise in a second language classroom enhance the learning of a specific language feature?* The data suggest that the learning of the modal feature used in the study may have been enhanced when these learners engaged in individual online concept mapping exercise.

The maps produced by these learners suggested learner interest in the activity. That is, the students displayed a range of ways of conceiving of and producing the maps, illustrating creative engagement and visualization of the modal usage at hand in the concept maps produced.

Data from analysis of the individual concept maps using the Concept Map Binary Qualification Checklist, which showed that the CM production in both treatment classes were characterized by several common features. First, most of the CMs produced displayed the *would* and *will* forms in discrete visually counterbalanced sections of the maps. In addition, all but one in the data set used idea linking to illustrate conceptual relationships. The majority of the concept maps submitted in both classes (20/22) included 'idea linking' to illustrate relationships between the usage concepts of *would* and *will* in their maps. Most of the maps (16/22) incorporated examples of usage in learner attempts at understanding modal usage. Some of the maps (13/22) included a grammatical explanation. A few of the maps (8/22) included pictures while one used colour, demonstrating some individual creativity of expression.

Five case studies were presented, illustrating different approaches the students in the two treatment classes took to the CM production exercise (stressing complexity, visual elements, or two which stressed grammar rules). The case studies reported in detail these students' individual concept maps, results of pre- and post-treatment essay and FITB tests, as well as their perceptions of their learning from their responses to the Mystery Module and Concept Mapping Exercises Post-Questionnaire survey. Four of the five case studies presented reported student perceptions that they had learned from all of the CM production exercises.

A fifth case study reported on data collected from a student who expressed a clear dislike for the CM production exercises in his/her responses to the Post-Questionnaire survey. In other responses to that survey however, this student reported having learned about the usage of *would* and *will* from the exercises.

The case study data presented in triangulation suggest that learning had taken place. Furthermore, classroom observation field notes reported that the students in the treatment classes exhibited much interest and enthusiasm with the online CM production task. This was reported to have been particularly so in the TG1 class.

The two treatment group teachers agreed that the individual student concept mapping exercise had helped students learn. One of the teachers cited the maps' use in motivating students, while the other teacher commented on the concept maps' analytical function.

Curiously, in the previously discussed CALL exercise, most of the students reported that they had participated in and learned from *both* the CALL Module and the online CM production task, however the students submitted little evidence of the CALL exercise. In contrast, nearly all of the participant students in both classes submitted concept maps. Perhaps the reason for that is that these young, pre-university learners found concept mapping more accessible to their learning styles, having grown up in a decidedly computer age. The Module, on the other hand, is much more traditional in format, though it too, is online. Of all of the student participants who were informally surveyed by the researcher as to their level of computer competence prior to undertaking the Module, only one student expressed a potential difficulty in using that medium. Subsequently, that student did participate in the CALL exercises and reported having had no difficulty in doing so.

The combined evidence suggests that concept mapping was used by these learners as a personal and cognitive exercise in attempting to conceptualize a language feature. This assumption is backed up by the evidence of the learners themselves who, in the Mystery Module and Concept Mapping Exercises Post-Questionnaire survey data, indicated their clear appreciation of the CM exercise. For example, when asked in that survey 'Did creating a computerized concept map help you learn?' the TG1 and TG2 students responded 'yes' in

strong majorities (12:1, 10:0). Similarly, most of these learners in these groups reported they would use CM production to help them learn in the future.

In summary, the evidence of this study suggests that an online concept mapping (CM) production exercise did enhance the learning of a specific language feature in this second language classroom.

6.2.3 Research question 3: The group CM question

The third research question asked, *Is there evidence that a group concept mapping exercise in a second language classroom enhances the learning of a specific language feature?* The data suggest that the learning of the modal feature used in the study may have been enhanced when these learners engaged in group concept mapping exercises.

Following analysis of the group concept maps by means of the Concept Map Binary Qualification Checklist, it was found that the group concept maps were more uniform than were the individual student produced concept maps. Furthermore, in contrast to the individual maps they were more linear in structure and more likely to include grammatical explanations, examples of grammatical usage, and linking between ideas. In addition, the group concept maps were found to be more uniform in structure and bore more resemblance to the teacher-class CM productions than to the individually-produced maps. These results point to possible effects of learner metacognition within the groups, as the learners possibly became more focused on grammatical usage and examples of that usage.

Analysis of classroom observations and audio-recorded transcriptions revealed many instances of cooperative learning amongst group members, and a high level of learner engagement in the group exercise. Following analysis, 15 categories of student interaction emerged from a qualitative analysis of these data, which were found to be either co-construction focussed interaction (meaning those behaviours which have as their focal point an interactive contribution to the group experience), or individual focussed interaction (those which serve more as an expression of realization of the self).

The fact that these categories and the co-construction or individual focussed interactions were observable, suggests a task orientation amongst these students which could be conducive to learning.

Data was presented of transcribed audio-recorded student discourse in which six kinds of co-construction focussed interactions in group discussions emerged from the data of the CM production exercise. The transcriptions suggest several outcomes these learners might have experienced. They were:

1. Some confusion regarding the usage of *will* and *would*
2. Some correct understanding of this usage
3. Some confusion and some understanding, and possibly 'noticing'

The first possible outcome may be inferred as a result of classroom observation; the second and third outcomes may only be posited, and the perhaps most intriguing possibility, of learner 'noticing' must by definition rest in the mind of the learner, obscured from the view of the researcher. Still, it is I think noteworthy that a question should arise here, did these learners begin to notice the gap between their IL constructions of *will* and *would* and those of proficient speakers? We can say that the evidence presented suggests that some of these learners *may* have begun to notice that gap. A more detailed study would be required to shed more light on that particular question.

Data in the form of student responses to the Mystery Module and Concept Mapping Exercises Post-Questionnaire supports the view that learning took place over the course of the treatment. When the students were asked 'Did creating a group concept map help you learn?' a strong majority responded 'yes' in the TG1 and TG2 classes (12:1, 8:1). Both teachers of the treatment groups agreed when surveyed, that the group CM exercise had helped their students learn.

6.2.4 Research question 4: The teacher-class CM question

The fourth research question asked, *Is there evidence that a teacher-class concept mapping exercise in a second language classroom enhances the learning of a specific language feature?* The data suggest that the learning of the modal feature used in the study may have been enhanced when these learners engaged in teacher-class concept mapping exercises.

The teacher and class concept maps produced in co-construction in the two treatment classes were analyzed using the Concept Map Binary Qualification Checklist and were found to be similar across classes. A quantitative analysis

of the Checklist analysis found the TG1 concept map had slightly more features (5) than did the TG2 map (4).

However, qualitative inspection of the concept maps showed notable differences. There was a greater number of usage contexts for *would* and *will* and examples of their usage were included in the TG1 class concept map, in contrast to the TG2 map. And the TG2 class concept map included cross-linking of contexts of *would* and *will*, in contrast to the TG1 class map.

The reported results suggest that the two classes approached the CM production task from similar vantage points, yet produced concept maps that were somewhat dissimilar in structure. The reported differences did not appear to have encumbered student learning, according to the reports of the other data sources pertaining to the teacher-class CM production exercise.

Analysis of the classroom observations and audio recordings of the teacher-class CM production exercises resulted in the categorization of 12 teacher interaction behaviours. Three student interaction behaviours previously observed in the group CM exercise were also added to the 12, bringing the total interactions observed in the teacher-class exercises to 15.

Of the 15 interactions, 2 were qualified as exhibiting a co-constructive component, since they required a response.

Moreover, of the 15 interactions, 6 were qualified as AFL/formative assessment, student-centred interactions.

The results of a review of observational and audio data suggest that these teachers took qualitatively different approaches to their task, yet similar learning results were reported by their students.

6.2.4.1 The teacher-class CM exercise teacher methodologies.

6.2.4.1.1 The TG1 teacher.

The TG1 teacher took a learner-centred approach to the exercise, commenting from her perspective, in the Mystery Module and Concept Mapping Exercises Teacher Post-Questionnaire survey, ‘I didn’t think I was guiding them – I think they were guiding me.’ The TG1 teacher was observed to allow students to elaborate on incorrect examples of the usage of *will* and *would* before stepping in to correct them, in accord with the AFL methodology of remaining neutral while allowing learners to try out various hypotheses

before teacher intervention. In addition, the TG1 teacher gave her students a full explanation of the grammar usage only near the end of the lesson, although she did give shorter explanations throughout the lesson.

Following instructions from the AFL teacher training of the study, the TG1 teacher asked her students, 'What do you think I want you to learn?' She summarized the student answers on a blackboard in six points expressing learning outcomes (see Chapter 5).

Transcribed discourse of the TG1 teacher-class CM production exercise illustrates the learner-centred focus of this lesson. Classroom observation field notes support this claim.

6.2.4.1.2 The TG2 teacher.

It was reported that the TG2 teacher-class CM production exercise did not include questioning the students, 'What do you think I want you to learn?' However, observation field notes and an audio recording of the lesson show that the TG2 teacher incorporated other elements of AFL and formative assessment in the CM production exercise. These included frequent formative comprehension checks, communicating the results of these checks to students, communicating teacher expectations and learning goals, and assigning a task which had as its goal formative assessment and learning. The observation field notes and audio recording of this lesson show that these elements appeared to engage the students in the CM production exercise.

In contrast to the TG1 teacher, the TG2 teacher took a more teacher-centred approach to teaching the lesson. The transcription evidence of the exercise illustrated that there was much less student discourse and much more teacher discourse in the TG2 class than in the TG1 class. Nevertheless, the TG2 class students seemed attentive and interested in their teacher's explanations of usage. Unlike those of the TG1 teacher, the TG2 teacher's usage explanations were frequent and detailed throughout the lesson in that class.

6.2.4.1.3 Similarities in the TG1 and TG2 teacher methodologies.

Several similarities were apparent in the TG1 and TG2 teacher-class CM production exercises. First, both classes were observed to have held jovial, good-humoured and lively debates on the subject of the correct usage of *would*

and *will*.[16] A strong teacher-student rapport was observed in several visits to both classes, and these students exhibited high motivation (not surprisingly, having as their goal entry into an English university). Both teachers incorporated the co-construction focussed interaction of 'guided questioning leading to a rule explanation.' Also, both the TG1 and TG2 teachers provided their students with negative evidence of student IL usage errors. Finally, although one teacher took a more learner-centred and the other a more teacher-centred approach, both teachers tailored their examples of usage to the students present, thus engaging their learning potential in a personal way.

6.2.5 Research question 5: The AB question

The research question 5 has asked, *Is there evidence of the* Assessment Bridge *when* Assessment for Learning *practices are employed in a second language classroom?* The data provide evidence of the AB when AFL practices were employed in this L2 classroom setting.

There was evidence of the AB in observational field notes taken in observations of the two treatment classes' group CM production and in the teacher and class CM production exercises. In the case of the group CM production exercises, 15 categories of student interaction types which had emerged from the data were presented (see Chapter 5, Section 5.3.1.1). To revisit, they were:

1. Discussions about how to draw map/making map/map set-up
2. Agreeing
3. Disagreeing
4. Negotiation of agreement and disagreement
5. Cooperative task accomplishment
6. Discussion about usage
7. Discussion about an example/s of usage
8. Correcting peer spelling
9. All talking at once
10. Rule explanation
11. Exhibiting verbal dominance

[16] The strength of the student enthusiasm that I observed was unexpected, and in my career, unprecedented relative to other observed discussions between teachers and students about grammar usage.

12. Exhibiting verbal submissiveness
13. Signalling the beginning of the exercise
14. Signalling the end of the exercise
15. Laughter indicative of understanding of usage

Of these categories, numbers 1 to 9 were further categorized as co-construction focussed interactions, which I have defined as those focussing on interactive contributions to the group learning endeavour. Of these, numbers 2 to 8 include the AB characteristics of assessment, (peer) teaching, and learning. As such, numbers 2 to 8 cited above suggesting evidence of the AB as observed in the group CM production exercises. In addition, number 10 may be cited as contributing to 'teaching' amongst peers. The following excerpt of a fuller exchange presented in Chapter 5, between students in the TG1 class, has illustrated this:

(S12) OK, where an incident is going to be taken but not sure. Do you have example?

(S1) Yeah, […] We have uh, "John say he is tired. I think he would be absent tomorrow."

(S12) OK, yeah. "If…"

(S1) Or "He said he would be absent", no, uh. "He said he is tired. He would be absent tomorrow", yup.

Evidence of the AB was presented in the teacher-class classroom observations and field notes, and in audio recordings and transcriptions of the teacher-class CM production exercises. In these data there were examples of the AB in formative teacher comprehension checks, in discussions about learning in which the students were corrected by their teachers, and when an assignment was given that included both an assessment and a learning component.

From a teaching perspective, evidence of the AB can be found in the twelve categories of teacher interactional behaviours cited, which had emerged from the data (see Section 5.3.2.1). They were:

1. Writing on the blackboard
2. Writing on an acetate page of an OHP
3. Rule explanation
4. Giving examples of usage

5. Recasting student responses
6. Asking learners "What do you think I want you to learn from this lesson?"
7. Guided questioning leading to a rule explanation
8. Providing negative evidence
9. Formative comprehension checking
10. Communicating to students the results of formative comprehension checks
11. Communicating to students teacher expectations and goals
12. Assigning a task with the goal of assessment and learning

The AFL/formative assessment student-centred approach of interactions 6 to 12 in particular indicates the presence of the AB in these classrooms. In addition, it was noted that interactions 6 and 7 have a co-constructive component in that they have required a learner response to complete them.

There was some overlap in the data presented in the student behaviours in group CM production and in teacher-class CM production exercises. Specifically, these student interaction categories which were also observed in the teacher-class CM production exercises, wherein the AB interplay can be seen in the elements of formative assessment, teaching and learning, were the following:

8. Discussion about usage
9. Discussion about an example/s of usage
11. Rule explanation

Moreover, the TG1 teacher's response to the Teacher Mystery Module and Concept Mapping Exercises Post-Questionnaire support of the premise that learning had taken place in the group as well as teacher-class CM production exercises, citing herself that student errors and group discussions about them had led the students to correct their errors. She added that this had given the students a better understanding of the grammar point in question. This is an example of the interplay of co-constructed formative assessment, teaching (and peer teaching), and possibly learning, which can be seen to be a clear example of the AB in a classroom setting.

There were also frequent instances of the AB embedded in the post-treatment student survey responses, in which students reported on informal

self-, and peer-, and teacher assessments occurring over the course of the CALL and CM production exercises, in which they reported these helped them progress in their learning.

6.2.6 Research question 6: The student AFL 10 principles question

The ARG has commented on the establishment of AFL procedures, and offered the 10 principles of AFL to help its application in classroom practice, noting that "assessment for learning requires some theoretical ideas to be put into practice if the potential benefits are to be gained. […] it is important to follow certain guiding principles which reflect the essential features of assessment for learning" (*Assessment for learning: 10 principles,* 2002). Given the widespread use of the AFL approach in many countries (as noted in Chapter 2), it was thought interesting to survey the students and teachers of this study in Quebec, Canada, to ascertain to what extent these learners agreed with the AFL 10 principles. Thus, the sixth research question of this study has asked, *Is there evidence that students in a North American second language classroom demonstrate agreement with the 10 principles of AFL proposed by the ARG?* The data provide evidence that these students demonstrated agreement with the 10 principles of AFL.

Questionnaire data from these students support the finding that these learners demonstrated agreement with the 10 principles of AFL as proposed by the ARG. The TG1 group reported mostly moderate agreement with the 10 principles. There was less strong agreement and disagreement following the AFL treatment experience. The TG2 group reported greater agreement with the 10 principles than did the TG1 group. The TG2 group agreed more often with the principles following treatment. This group differed from the other groups in that there was very little disagreement with the principles expressed, relative to the other groups.

In the control (CG1) group questionnaire responses reported, here again, there was mostly agreement with the 10 principles. There was less agreement, however, following the treatment time period. There was less strong agreement and no strong disagreement after the treatment period.

The Time 2 control (CG2) group also mostly agreed with the 10 principles and demonstrated less strong agreement and no strong disagreement following

the treatment period. There was also less agreement, and slightly less disagreement after that time as well.

There appears to have been a moderating effect on the student responses in these data since three of the groups moderated their reported views in the post-questionnaires. The two treatment groups had diminished their strong agreement considerably following treatment.

On the other hand, it is entirely possible that these learners were challenged to understand some of the concepts of the 10 principles and may have tried to moderate their responses in consequence. For example, classroom observation field notes support the view that students (despite the presence of teacher explanations) may not have fully understood the concept of self-assessment.

The principle most often disagreed with in all reported response data, was principle 5, which states that *formative assessment in the classroom should be sensitive and constructive given that assessment has an emotional impact*. One may conjecture that this view may be an artefact of some of these international students' more traditional native classroom experience abroad.

All things considered, these results support the premise that these students in a North American second language classroom demonstrated agreement with the 10 principles of AFL proposed by the ARG.

6.2.7 Research question 7: The teacher AFL 10 principles question

The seventh research question has inquired of a teacher perspective, *Is there evidence that teachers in a North American second language classroom demonstrate agreement with the 10 principles of AFL proposed by the ARG?* The data provide evidence that these teachers demonstrated agreement with the 10 principles of AFL.

The teachers' responses to the 10 principles survey demonstrated agreement with the principles much of the time and strongly agreed with them most of the time. Only one of the four teachers surveyed disagreed with any of the ideals of the 10 principles; one control (CG2) group teacher expressed disagreement with principle 10, stating *formative assessment in the classroom should recognize the full range of achievements of all learners.*

Thus, the questionnaire data from these teachers support the finding that these teachers and learners demonstrated agreement with the 10 principles of AFL as proposed by the ARG.

6.3 Triangulation of quantitative and qualitative analyses

The design and analysis of this study have reflected the researcher's belief that a mostly MM instrument analysis of much of the data was the best methodology to follow to fully investigate the research questions as impetus to the study. Thus, most of the instruments were analyzed by MM means, (including the CALL Module, the individually, group and teacher-class CM production exercises, and the Post-treatment student and teacher surveys). However, the observation field notes, incidental teacher comments and selected transcriptions were analyzed by qualitative means alone. Finally, two data sources (the FITBs, essay tests, and AFL 10 principles surveys) were analyzed quantitatively for the purposes of indicating trends.

As in the case of the proverbial elephant that several blind people would perceive differently depending on the vantage point of each, so triangulation in mixed methods affords varying viewpoints to be voiced in order to achieve a picture of the whole. For example, the quantitative results presented in this study alone suggest that the participant learners would not succeed in traditional FITBs or essay tests requiring them to produce the target modal forms either by selecting the correct form provided, or worse, by automatically producing the form in an essay writing exercise. Based on those results alone one could come to the conclusion that these students had little competence (and hence learning) relative to English modal usage. The qualitative data results, on the other hand, would suggest otherwise, that they could demonstrate a level of sophistication in examining, describing and trying out usage of various target modal forms.

In triangulation, it can be seen where the two viewpoints converge; the elusively 'unobservable' mentioned in Chapter 1 has 'appeared' in a sense. If we accept the views of Larsen-Freeman (2009), Young (1995) that learning will not necessarily be demonstrated in learner performance, as well as the strong evidence presented herein of learner and teacher attestations as to learning having taken place, we may posit that the gap between that which has

been observed learner performance (in the essay and FITBs test results) and the learning that has been attested to by the learners and their teachers (in survey reports) could both constitute correct views. Thus, the 'elephant' may have been viewed from different viewpoints, each of which may be compelling. That is, these students may have learned but may not have been able to demonstrate that learning in improved accuracy of form production. It is known that learning is an internal process, the results of which may emerge over time. A longitudinal or follow-up study could be fruitful in ascertaining if such has indeed been the case.

In the case of this study it would be prudent to say that the mixed methods design and the results it has garnered have collaborated to more effectively answer the research questions and inform the study than would either qualitative or quantitative means alone.

6.4 Interpretation of results

As noted at the beginning of this chapter, the results of the pre-, post- and delayed post-treatment essay and fill-in-the-blanks tests cannot be interpreted in a study of this design except as possible indicators of trends. In fact, as Hatch and Lazaraton (1991) have pointed out in a discussion about the quasi-experimental nature of most educational research due to the fact that the participant class members are usually pre-selected by participant educational institutions (as is the case in the present study), "When random selection is not possible, causal claims are also impossible" (85).

The results of quantitative pre-treatment tests have shown that these learners had difficulty producing correct contingency use modals in written essay and FITBs tests both in pre- and post-treatment. This suggests that they may not have learned as a result of the treatment, or that they had not learned sufficiently to be able to demonstrate learning in FITBs and essay test tasks.

However, the qualitative data are interesting in their lack of agreement with other results in the study. For example, in post-treatment student surveys the Time 1 treatment group, these learners reported having learned approximately four times as often as not learning. The Time 2 treatment group reported similarly on their perceived learning, yet at the more modest rate of an approximate three times more positive than negative responses. And as noted

earlier in this chapter, the teacher post-treatment surveys indicated strong teacher certainty that their students had learned over the course of the AFL treatment.

The variability of IL performance noted in section 2.1.5 of Chapter 2 (section 2.1.5) has implications for the present research since we know that while learning may take place under treatment conditions, this may not be reflected in learner behaviour or performance in assessment situations, as many have noted (Lightbown, 2000; Turner, in press; Turner & Upshur, 1995; Young, 1995). Moreover, Larsen-Freeman (2009) has suggested grammar teaching should help develop students' grammatical capacity within the learner, rather than aiming for formal grammatical competence (p. 527).

I would remind the reader (as noted in Chapter 1), that given the above, and while the focus of the present research is not an in-depth study of the extent to which learning may have taken place, but an investigation of the effect of applying L2 AFL from an LT research stance, the variable results of the FITBs and essay test measures, do not contribute strongly meaningful data on which assumptions of learning may be based. These results do, however, demonstrate that these forms have presented important challenges to these learners when they were tested in FITBs and essay test formats.

The results in triangulation support the premise that learning may have taken place in these teacher and class concept mapping exercises and that given the overwhelming support of the participant teachers and students, it would be worthwhile to pursue this approach to L2 AFL in further research.

Nevertheless the robust qualitative evidence supports the learner and teacher view expressed that learning may have taken place. This view is supported by classroom observation field notes and audio recordings of the observations.

One of the participant teachers (TG1) to some extent took an AFL approach to teaching the grammar point in question during the teacher and class CM production exercise, remaining neutral about her own opinions while the learners engaged in collective hypothesis testing. Observation field notes, an audio recording and the learners' surveyed perceptions suggest that technique may have been effective in supporting learning of the grammar point. On the other hand, the other teacher (TG2) took a more direct, hands-on approach to

teaching delivering a more teacher-centred classroom CM production exercise. Yet the second teacher incorporated several features of AFL, such as frequent comprehension checks and sharing of her assessments of them, and guided questioning techniques. Both appeared to engage their students and according to teacher and student perceptions, to facilitate learning.

Thus, the evidence of this study suggests that AFL practices in these classroom settings enhanced the learning of a specific language feature. It is beyond the scope of this study to postulate on whether the learning of modal usage the students and teachers reported, would at some point in the future become evident in the learners' L2 writing or speaking, in contexts where that usage was obligatory.

6.5 Summary of the findings

The results in triangulation point to the internal and personal nature of language learning and processing. An important goal of L2 teaching is that of enhancing noticing of forms in learners, in hopes of engaging learner internalization and ultimately, automatic L2 production. However as we cannot view noticing, we can only draw inferences of it from observations and especially from learner accounts. We have seen in this discussion that if 'noticing' and learning have occurred, this does not necessarily immediately result in improved learner performance.

Thus, if we take at face value the learner reports of having noticed something about the contingency use modal in the present research, then we may arrive at a logical conclusion that some development may have taken place. The best we can hope for in second language teaching is to spur on this internal process, in order that learners may, at some point manifest this learning in more target-like performance.

The fostering of learning to help students take greater responsibility for their own learning is in accord with the AFL objective of enhanced learner autonomy. Overall the results of this study suggest that the treatment group students did experience some noticing of the gap between their performance and the target modal form.

Alderson (2005) suggested learner self-assessment is important in helping learners become self-aware in relation to their language learning. Many

students in this study reported using self-assessments to help them progress in their learning. There were also many who did not see the usefulness of self-assessment procedures. Possibly the assumption that these students were all familiar with self-assessments, based on their EAP course content, was incorrect.

A different picture emerged, however, when group work occurred and informal peer-assessment comprised a large part of the interaction between group members. Classroom observation data and audio recordings revealed co-constructed evidence of learning over the sessions. These results in some measure echo those of Brooks (2009). Brooks found that students performed better in pairs engaged in co-constructed interaction than they did in individual performance testing.

Evidence of the AB was found in classroom observations of these learners in co-constructed interaction with peers and with their teachers. This was supported by audio recordings of these proceedings. The evidence of the AB reported here suggests its presence is valuable in enhancing classroom-based learning. These findings accord with a call from Pellegrino et al. (2001) in favour of moving American classrooms towards assessment intended to enhance learning. They argued that "For assessment to serve the goals of learning, there must be alignment among curriculum, instruction, and assessment" (p. 302).

In the next and final chapter, the discussion will centre on the major findings of the study, its implications for the field, recommendations for future research, limitations of the study, and its strengths and contributions.

CHAPTER 7 – CONCLUSIONS

I would fain grow old learning many things.

Plato[17]

In overview, the final chapter of the study will begin by a re-examination of the original inspiration for it and the research questions it has engendered, in relation to the theoretical frameworks cited in the study (three major theories of learning, and the AFL 10 principles) and the results presented. The overall study objective of responding to calls for application of an AFL approach in an L2 classroom setting will also be addressed.

In addition, the central research question will be discussed relative to the results presented. This will be followed by a review of some limitations and the strengths of the study. Next, the study contributions and implications for future research will be put forth, and recommendations will be made. The chapter will finish with an elucidation of the final conclusions garnered from the study as a whole.

7.1 The study inspiration revisited

At the start of this thesis, in Chapter 1, I presented the picture of students in an idealized classroom setting, and I pondered whether students might actually be brought to ask questions in class along the lines of "I understood a lot but I didn't understand the main idea," and "My comprehension is good but my expression is weak." I admitted that these learner queries sounded much like teachers' "stuff of dreams" and I pondered at how wonderful it would be if learners could become so interested in taking charge of their learning, if learning might occur in such a climate of teacher-learner cooperation. Finally, I asked if learners might ever be enticed to let go of any reluctance they might have in admitting to their limitations in a classroom setting. I noted that questions of this nature had inspired this study of an application of AFL in a second language classroom setting.

Questions that remain to be answered at this point are, have any of the above learner attributes been realized as a result of this L2-AFL application

[17] This final quote of the contingent use *would* appears herein with the purpose of lending hope to all who read it, that our collective knowledge for good works may increase with time.

with these learners, and if so, what are the implications of this study relative to AFL applications and formative assessment practices? In other words, what implications does this research have for LT and CBA here and possibly in other classroom settings? In order to consider these questions, the study results will first be examined relative to the theories of learning and AFL 10 principles that have served as the theoretical frameworks of the study.

7.2 The theoretical frameworks cited in the study and the results

7.2.1 Theories of learning

In a discussion of AFL and the ways in which the behaviourist and cognitive approaches might be best used in classroom application, James (2006) suggested that behaviourist approaches would be well suited to the development of habitual behaviours or basic skills, while cognitive approaches would work well in cases where a deep understanding of conceptual structures was a goal. I have posited that a socio-cultural approach would be exemplary in ESL classrooms, with their preponderance of group and paired interactions, leading to co-constructed learning situations. The same can be said for co-constructed learning situations in class discussions.

There was evidence of the behaviourist approach in the individual students' CM production cited in Case Studies 3 and 4 (see Sections 5.2.2.5 and 5.2.2.6) I have characterized the Case students as "grammar rulers" due to their choice of listing grammar rules in their CM productions, where many other students took a more creative approach, as shown in the results. Whether a behaviourist approach helped foster learning is beyond the scope of this research. Nonetheless, it can be said that it was the choice of the Case Study 3 and 4 learners to use a habitual repetition, behaviourist approach to the CM production exercise.

The cognitive approach to learning is concerned with learners' active engagement, construction of meaning, mental schema, and understanding of knowledge, as it may contribute to learning (James, 2006). I have cited three ways in which the cognitive approach has been incorporated in AFL, (1) in the view of the teacher-as-learning-guide, leading students towards their own learning goals, (2) with its focus on using problem-solving to advance learning, and (3) using deductive and inductive reasoning.

In addition, the self-monitoring and self-regulation of metacognition are also key factors in AFL practice, as is the scaffolding of knowledge to build mental schema.

The design of the treatment incorporated elements of the cognitive approach to learning, for example in the building up and scaffolding of knowledge intended in the sequencing of the CALL exercises, followed by the individual student, group, and finally teacher-class CM production exercises. There was evidence of the cognitive approach in the transcriptions presented of audio-recorded group and teacher-class CM production exercises, as learners voiced and discussed their reflections on grammar usage (see Sections 5.3.1.1.2, 5.3.1.1.3, and 5.3.2.2).

In fact, these transcriptions of audio-recordings also demonstrate very clearly the socio-cultural theory of learning, (also known as situated theory), which situates learning in the inescapable environment of social interaction (see Sections 5.3.1.1.2, 5.3.1.1.3, and 5.3.2.2).

We are reminded that from a socio-cultural perspective, learning cannot occur in isolation; the socio-cultural approach emphasizes group learning effects, shared learning, mediation between participants, and co-constructed learning. The data presented in Chapter 5 provide rich evidence of this approach, as well as several instances cited wherein learners fostered understanding between participants in group and in teacher-class CM production activities, which in fact suggests the exercise of Vygotsky's ZPD, wherein learners' potential rather than actual development would be stimulated by group or class participants.

In practice, the socio-cultural approach to learning reflects the characteristics of DA with its focus on mediation amongst participants, to contribute to driving learning. In fact, social interactionalists Poehner and Lantolf (2005) have called DA mediated formative assessment. DA can be said to function in concert with AFL, since both give importance to learner feedback.

A discussion of this nature relative to an L2 learning context would not be complete without mention of a particular challenge inherent in L2 teaching, learning, and particularly, assessment. As discussed in Chapter 2, I refer to the

difficulty engendered in attempting what I have called observing the unobservable, since second language learning is known to occur over time and without having necessarily manifested in increased learner accuracy, as has been noted by several (Larsen-Freeman, 2006; Lightbown, 2000; Young, 1995). In addition, Young has characterized a pattern that may occur when a learner's "intermediate stages may be further from the target than either beginning or advanced stages - a pattern commonly called U-shaped behavior" (p. 18).

In a discussion of LT theory and practice, Turner and Upshur (1995) and Turner (in press) have also argued that the complex phenomenon of learning may not be evident in learner performance. This factor clearly creates challenges in any L2 assessment setting, and may have had a direct impact on the results of the present research. That is, the results of the qualitative and quantitative analyses presented in Chapter 5 are in disaccord, given that the former strongly support the premise that learning may have taken place following the L2-AFL treatment, while the latter generally do not. On the other hand, the TG2 class *did* show marked improvement in target form production in post-treatment essays, by producing more than double the number of accurate forms in obligatory contexts following the treatment. Yet the TG1 treatment class, produced less than half the number of accurate forms in obligatory contexts following the treatment.

To reiterate what I have stated in Chapter 1, the focus of this study is not on detailed learning outcomes as might be expected in an SLA-oriented investigation. The focus of this study *is*, however, to investigate the effect of introducing an AFL methodology in an L2 classroom setting on teacher and student perceptions of learning from the perspective of LT. Thus, the main focus of the study is on the assessment methods applied and their perceived effect on learning.

7.2.2 The AFL 10 principles

The AFL 10 principles were used as the cornerstone of the new L2- AFL pedagogical materials developed for use in the study (see Sections 2.11.1 and 3.4.1.3). The 10 principles were categorized in terms of four key issues they

address, which are AFL's (a) *learner-focus*, (b) *AFL assessment objectives*, (c) *teacher and learner goal sharing* and (d) *learner motivation.*

Pedagogical materials were developed that reflected these features. Thus, the materials focused on learners and encouraged learner autonomy, took account of how students learn, incorporated constructive and sensitive feedback, and attempted to motivate the learners in their learning progress, incorporating self- and peer-assessments. Additionally, learner reflection, teacher guided questioning and knowledge scaffolding appeared in the materials.

It was reported in Chapter 2 that Leung (2007) compared AFL and DA and found similarities between them, particularly in relation to the following four of the 10 principles of AFL, 1) *Assessment for learning should be part of effective planning of teaching and learning,* 2) *Assessment for learning should focus on how students learn,* 3) *Assessment for learning should be recognised as central to classroom practice,* and 8) *Learners should receive constructive guidance about how to improve.* Leung found that the latter principle accorded with the DA practice of supporting learning through mediation.

In terms of the 10 principles, in the present research results were reported from teacher and student surveys suggesting that the participant teachers and students of the treatment as well as the control groups were in general agreement with all of the AFL 10 principles. The students were in agreement with the principles, despite having been educated in varied countries and with various pedagogical approaches (some of which, in incidental comments, the students deemed to be very traditional), prior to their coming to Canada to pursue their studies through admission to an English-language university.

7.3 A review of the study objectives

7.3.1 Answering calls for AFL application in L2 settings

Several researchers (Brookhart, 2005; Garner, 2006; Harlen & Winter, 2004; McNamara 2001a, 2001b; Pellegrino et al., 2001; Poehner and Lantolf 2005; Rea-Dickins 2004; Shohamy, 2004) have called for more research in the area of classroom assessment practices used to drive learning. Moreover,

Leung (2004) called for empirical research which would take a socially co-constructed view of formative teacher assessment.

The present study has sought to answer calls for L2-AFL from three sources in particular. First, Harlen and Winter (2004) have suggested that AFL should be adapted from general education to L2 settings. This research has attempted to put that suggestion into classroom practice. Second, the work of Colby-Kelly and Turner (2006, 2007) and Colby-Kelly (2008) called for the implementation of AFL in an L2 setting following their study of formative assessment practices used with L2 pre-university learners. Third, the present study has also responded to the words of Rea-Dickins (2004), who asserted that "assessment, with specific reference to teaching and learning in the language classroom, has remained, until recently, relatively unresearched" (p. 249).

Thus, the present study has achieved its objective of applying the AFL approach in an L2 classroom setting, in response to the impetus taken from Harlen and Winter (2004), Colby-Kelly and Turner (2006, 2007), Colby-Kelly (2008), and Rea-Dickins (2004). Specifically, it has achieved the following:

- development of L2-AFL pedagogical materials
- application of L2-AFL
- investigation of the effect of L2-AFL on learning

This brings the discussion to the other half of that inspiration, its germane and key query, the central research question, which will be examined in relation to the study findings.

7.3.2 The central research question

This exploratory study of L2-AFL application has sought to answer its primary research question, *Is there evidence that AFL practices in a second language classroom enhance the learning of a specific language feature?* There has been much evidence presented in Chapter 5 that strongly suggests that this question may be answered in the affirmative. This will now be discussed in relation to the major findings.

7.4 Major findings

In Chapter 2 it was noted that Sadler (1989) cited three conditions that he found were essential for improving learning, which were that learners:

1. understand the criteria and standards that were aimed for
2. be able to assess their own performance and to compare it with the standards
3. and be able to engage in action in order to close the gap.

The first of Sadler's conditions was not investigated in the present research. However, based on the evidence presented it can be said that these learners may have to some extent achieved the second and third conditions. This was particularly evident in view of the data presented in the form of transcriptions of audio-recorded group and teacher-class CM exercises.

Similarly, Fulcher (2010) has characterized the AFL classroom setting, asserting that "all assessment is descriptive rather than evaluative, helping learners to 'notice the gap'" (p. 1). The evidence presented herein would suggest that this may have happened as a result of the treatment used in the present research. In fact, this study has resulted in several major findings, which will now be discussed.

Robust evidence has been presented in favour of the treatment effect. Much qualitative evidence has been cited showing teacher and student perceptions that learning had taken place following the AFL treatment. Quantitative evidence of markedly improved performance on a challenging post-treatment essay test in one of the two treatment classes (TG2) supports this view. These gains were not, however, observed in the other treatment class (TG1). The focus of the study, it should be recalled, is on the assessment methods applied and their perceived effect on learning.

Strong evidence has been presented to suggest that the AFL practices applied in this L2 setting may have enhanced the learning of the modal language feature in individual student CALL exercises, in group CM exercises, and in teacher-class CM exercises. It is clear from survey data and from teacher incidental comments that this was the perception of the teacher and student participants in the two treatment classes. Moreover, transcribed audio-recorded and classroom observational evidence suggests that this AFL

application in an L2 context may have enhanced learning by fostering learners' noticing of gaps between their actual and target competency in L2 production tasks. Thus, it can be said that there is robust evidence in favour of the treatment effect.

The results of this study have demonstrated that AFL can be successfully applied in an L2 classroom setting. Additionally, I would posit that concept mapping exercises have been shown to be an appropriate means of applying AFL in terms of their usefulness in helping learners conceptualize grammatical usage. Therefore I would suggest that the evidence points to concept mapping as an effective means of helping these learners enhance their grammatical ability as defined by Purpura (2004) as "the capacity to realize grammatical knowledge accurately and meaningfully in test-taking or *other language-use contexts*" (p. 89, italics added).

Additionally, the present research has provided evidence that co-constructed schema in the form of group concept mapping and teacher and class concept mapping exercises may help learners co-construct knowledge in an L2 classroom setting. The results of the group and teacher and class concept mapping exercise presented support the premise suggested in Chapter 2 that co-constructed interaction with peers or a teacher may guide learners to a more useful conceptualization of language.

Further, these findings indicate that a dynamic assessment approach to L2 learning may benefit learners since many of these learners expressed the view that they had learned in concert with their peers in group concept mapping sessions. Field observation notes, audio recordings and transcriptions of these group CM production exercises suggest that peer assessment, constructive argumentation and discussion of grammatical usage, appeared to have contributed to learning.

In addition, as discussed in Chapter 6, evidence was found of the AB interface of assessment, teaching and learning in this L2-AFL setting.

Finally, it was also found that the participant teachers and their students in a North American context, demonstrated strong agreement with the 10 principles of AFL as proposed by the ARG in England.

7.5 Limitations to the study

One limitation to the study was that it was not possible to have a larger student participant group. Additionally, the delayed post-treatment essay and FITB tests in particular were culled from a smaller set of participants than would have been desired.

Also the small number of students still enrolled in the EAP program at the time of delayed post-test administrations was unfortunate. This was due to participant attrition. It must be said that the strong results of the delayed post-tests may have been due to the accumulative effects of exposure to the school environment and the L2 milieu over time.

Finally, the study would have benefitted from a longitudinal design, had that been possible. However that was not possible, given the administrative constraints of the participant school.

7.6 Strengths of the study

There were several strengths inherent in this study, the most substantial of which is the development of AFL pedagogical materials for an L2 classroom setting. Reliance on previous scholarship (Black & Wiliam, 1998a, 1998b; Chapelle, 2001, 2005; Colby-Kelly & Turner, 2007; Leung, 2007; Sadler, 1989; Vygotsky, 1969, 1978) in the development process, and the teacher training devised to ensure the success of the application, are other strong points of the study.

In addition, the rigour of the research design, including the use of a split-plot format with its two treatment and two control classes, is a compelling aspect of the study. Other strong points were: the innovation in choosing CM exercises as a means of interpreting AFL in an L2 classroom setting; the choice of online CM exercises to engage young adult students in exploring grammar forms; the use of CALL in preparation for the CM exercises adding to the rigour of the study; and the inclusion of exercises intended to stimulate learner metacognition.

The method that was developed for analyzing the concept maps produced in the study was innovative yet reasoned, since it employed existing methods of CM analysis and adapted them to fit this L2 classroom context.

7.7 Contributions and implications of the study

Several features of this study have contributed to the field of second language assessment. They are:

- The application of AFL procedures in an L2 classroom setting.
- The creation of L2 AFL teacher training.
- L2 AFL pedagogical materials development.
- The use of CALL and CM exercises to interpret AFL.
- The interpretation of the 10 principles of AFL to serve materials development.
- The categorization of learner and teacher behaviours that emerged from the data in this L2 AFL application.

The present study has answered calls in the literature for innovation in pedagogical assessment procedures in L2 classroom settings, and it has implications for the application of AFL in L2 classroom settings. Evidence has been presented which has demonstrated the feasibility of AFL applications for L2 teachers and learners, with implications relative to the following:

1. Teacher training in L2 classroom procedures using AFL
2. Pedagogical material use for L2 classroom-based AFL
3. Procedures for applying AFL in an L2 classroom context

In addition, the results of this study have implications for the use of individual learner concept maps to stimulate and enhance individual learning. In addition, the learner and teacher perceptions of successful learning in group and class exercises have implications for the use of co-constructed learning tasks based on AFL practices in L2 classroom settings.

7.8 Recommendations for future research

I would like to make the following research recommendations:

- AFL needs to be employed in L2 classroom settings to make better use of teacher and learner resources
- AFL should be used in L2 settings in order to increase learner autonomy, consequently resulting in possibly more effective learning

- CALL and CM production should be encouraged in L2 classrooms as effective means of engaging learner participation, and as a method of stimulating learner metacognition
- Teacher training in L2-AFL should be encouraged and included as part of pre-service teacher training
- Teachers training should include the AFL 10 principles.

7.9 A call for a research agenda relative to L2-AFL

Based on the robust evidence presented in the results of this study, I would like to put forth a call to the LT community for a research agenda to study further applications of AFL in L2 classroom settings, particularly in view of the fact that second language learning is an interactive, socially constructed phenomenon requiring action on the part of learner noticing of the gap between non-target and target forms, and in my view L2-AFL is well situated to contribute to that process.

To this I would like to add my voice to that of Turner (in press) calling for the LT community to clarify the paradigm of CBA which serves learning, as distinguished from the paradigm of more traditional LT practices.

7.10 Conclusions

In conclusion, this thesis has presented an exploratory study investigating the effects of an application of AFL practices in an L2 classroom setting, and it has also sought evidence of the AB in that setting. On a macro level, the present study attempted to (a) interpret the tenets of AFL by developing AFL procedures appropriate to an L2 classroom, (b) apply these AFL procedures in an L2 classroom setting, and (c) investigate their effect on learning.

This research has engendered a call for more application of AFL in various ways in L2 classroom settings. The conclusions support a vision of greater learner autonomy through L2-AFL application, given that evidence of learning in that setting may not be easily observable, since by its nature it resides in the minds of learners. Greater use of concept maps in L2 learning situations is also called for, especially in co-constructed peer situations.

It is my hope that this contribution to the literature will spur more applications of AFL in second language classroom settings, that through this

more learners may gain confidence in their language learning abilities, and that more teachers may discover new ways to focus their mentoring guidance in helping L2 learners uncover some of the mysteries of English, through the use of AFL procedures.

REFERENCES

Alberta Education. (2008). *Direction de l'éducation française, Guide to implementation – Grade 7 to Grade 9, Classroom Assessment*. [Brochure online]. Available at: http://education.alberta.ca/media/826689/chap8.pdf

Ali, S. & Ho, H. W. (2007). Using assessment tasks to shift focus to learning rather than evaluating students. *Journal of American Academy of Business, 10* (2) 269-274.

Ames, C. (1992). Classroom goals, structures, and student motivation. *Journal of Educational Psychology*, *84* (3) 261-271.

Alderson, C. (2005). *Diagnosing Foreign Language Proficiency.* London: Continuum.

Alderson, C. (2007). The challenge of (diagnostic) testing: Do we know what we are measuring? In J. Fox et al. (Eds.), *Language Testing Reconsidered* (pp. 21-39. Ottawa: University of Ottawa Press.

Alderson, C. (2010). "Cognitive diagnosis and Q-matrices in language assessment'': A commentary. *Language Assessment Quarterly,7,* (1) 96-103.

Anderson, J.O. & Bachor, D.G. (1998). A Canadian perspective on portfolio use in student assessment. *Assessment in Education, 5* (3) 353-327.

Anonymous. (1998). Editorial. *Assessment in Education, 5* (3), 301-308.

Assessment Reform Group. (2002). *Assessment for learning: Research-based principles of assessment for learning to guide classroom practice.* Retrieved from http://www.assessment-reform-group.org.uk

Bachman, L. (2004). *Statistical Analyses for Language Assessment.* Cambridge: Cambridge University Press.

Bachman, L. (2005). Building and supporting a case for test use. *Language Assessment Quarterly, (2),* 1, 1-34.

Bachman, L. (2006a). Generalizability, A journey into the nature of empirical research in applied linguistics. In M. Chalhoub-Deville, C. A. Chapelle, & P. Duff (Eds.), *Inference and Generalizability in Applied Linguistics, Multiple perspectives* (pp. 165-207). Philadelphia, PA: John Benjamins North America.

Bachman, L. (2006b, April). *Linking interpretation and use in educational assessments.* Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, California.

Bachman, L. (2006c, June). *A research use argument: An alternative paradigm for empirical research in applied linguistics?* Paper presented at the joint meeting of the American Association for Applied Linguistics and Association canadienne de linguistique appliquée/Canadian Association of Applied Linguists, Montreal.

Bachman, L. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In Fox et al. (Eds.) *Language Testing Reconsidered* (pp. 41-71). Ottawa: Ottawa University Press.

Bachman, L. & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford: Oxford University Press.

Baker, L. (2003). *Exercises to accompany the essentials of English, a writer's handbook.* White Plains, NY: Pearson Education, Longman.

Balliro, L. (1993). What kind of alternative? Examining alternative assessment. *TESOL Quarterly, 27,* (3), 558-561.

Bercier-Larivière, M. & Forgette-Giroux, R. (1999). L'évaluation des apprentissages scolaires: Une question de justesse. *Revue canadianne de l'éducation, 24,* (2), 169-182.

Bevan, R. M. (2007). *From black boxes to glass boxes: The application of computerized concept mapping in schools.* Unpublished doctoral dissertation, University of Cambridge, UK.

Black, P. (1993). Formative and summative assessment by teachers. *Studies in Science Education, 21,* 49-97.

Black, P. (2004). *The nature and value of formative assessment for learning.* Unpublished manuscript in preparation, King's College London.

Black, P. (2005). Editorial: Formative assessment: Views through different lenses. *The Curriculum Journal, 16,* (2), 133-134.

Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5* (1), 7-73.

Black, P. & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80,* 139-148.

Black, P. & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29,* (5), 624-637.

Black, P. & Wiliam, D. (2005a). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal, 16,* (2), 249-261.

Black, P. & Wiliam, D. (2005b). Changing teaching through formative assessment: Research and practice. In CERI (Ed.), *Formative Assessment: Improving Learning in Secondary Classrooms* (pp. 223-240). Paris: OCED.

Black, P. & Wiliam, D. (2006). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and Learning: Theory, Practice and Policy* (pp. 9- 25). London: Sage Publications.

Black, P. & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (p. 81-100). London: Sage Publications Ltd.

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

Blumenfeld, P. (1992). Classroom learning and motivation: Clarifying and expanding goal theory. *Journal of Educational Psychology, 84* (3) 272-281.

Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing, 18* (4), 393-407.

Brookhart, S. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education, 8* (2), 153-169.

Brookhart, S. (April, 2005). *Research on formative classroom assessment.* Paper presented at the American Educational Research Association, Montreal.

Brookhart, S. & Bronowicz, D. (2003). 'I don't like writing: It makes my fingers hurt': Students talk about their classroom assessments. *Assessment in Education, 10* (2), 1221-242.

Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing, 26* (3), 341-366.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing, 20* (1), 1-25.

Brown, J. D. & Rodgers, T. (2002). *Doing second language research.* Oxford: Oxford University Press.

Brown, J.D. & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, (32)* 4, 653-675.

Celce-Murcia & Larsen-Freeman (1999). *The grammar book, an ESL/EFL teachers' course.* Boston, MA: Heinle & Heinle.

Chaloub-Deville, M. (2003). Second language interaction: current perspectives and future trends. *Language Testing, 20* (4), 369-383.

Chapelle, C. (2001). *Computer applications in second language acquisition.* New York: Cambridge University Press.

Chapelle, C. (2005). Interactionist SLA theory in CALL research. In J. Egbert & G. Mikel Petrie (Eds.), *CALL research perspective* (pp. 53-64). Mahwah, NJ: Lawrence Erlbaum.

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

Cheng, L., Watanabe, Y. & Curtis, A. (Eds.) (2004b). *Washback in language testing: Research contexts and methods.* Mahwah. NJ: Lawrence Erlbaum Associates.

Colby-Kelly, C. (June, 2010). *Innovation in* Assessment for Learning *in an L2 classroom: Learner exploration of grammar through concept mapping.* Paper presentation at the Association canadienne de linguistique

appliquée/Canadian Association of Applied Linguists, Concordia University, Montreal, QC.

Colby-Kelly, C. (May, 2009a). *Assessment for learning: Investigating assessment used to serve learning in an L2 classroom*. Paper presentation at the Association canadienne de linguistique appliquée/Canadian Association of Applied Linguists, Carleton University, Ottawa.

Colby-Kelly, C. (March, 2009b). *Bridging the gap: Assessment for learning in a Quebec classroom.* Work-in-progress presentation at the Language Testing Research Colloquium, Denver, Colorado.

Colby-Kelly, C. (October, 2008). *AFL research in the L2 classroom and evidence of usefulness.* Non-refereed local presentation to the Continuing Education Language Institute of Concordia University, Montreal, Quebec.

Colby-Kelly, C. & Turner, C. (June, 2006). *At your service! Assessment at the service of learners and teachers.* Paper presented at the American Association for Applied Linguists and Association canadienne de linguistique appliqué/Canadian Association of Applied Linguists, Montreal, QC.

Colby-Kelly, C. & Turner, C. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. *Canadian Modern Language Review.*

Conlan & Bird (2004). Not yet within the mainstream: Concept mapping in a Scottish high school. In A. J. Cañas, J. D. Novak, & F. M. González (Eds.), *Concept maps: Theory, methodology, technology: Proceedings of the first internationalconference on concept mapping, vol. 2*. Pamplona, Spain.

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

Creswell, J. W. & Plano Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage Publications.

Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58,* 438-481.

Daley, B. (2004a). Using concept maps in qualitative research. In A. J. Cañas, J. D. Novak, & F. M. González (Eds.), *Concept maps: Theory, methodology, technology: Proceedings of the first international conference on concept mapping, vol. 2.* Pamplona, Spain.

Daley, B. (2004b). Using concept maps with adult students in higher education. In A. J. Cañas, J. D. Novak, & F. M. González (Eds.), *Concept maps: Theory, methodology, technology: Proceedings of the first international conference on concept mapping, vol. 2.* Pamplona, Spain.

Davison, (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly, 4,* (1), 37-68.

Dunn, W. E. & Lantolf, J. P. (1998). Vygotsky's zone of proximal development and Krashen's *i + 1*: Incommensurable constructs; incommensurable theories. *Language Learning, 48*, 3, 411-442.

Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist, 41* (10), 1040-1048.

Elder, C., Barkhuizen, G., Knoch, U., & von Randow, J. (2007). Evaluating rater responses to an online training program for L2 writing assessment. *Language Testing, 24* (1), 37-64.

Flores, M. (2005). Teachers' views on recent curriculum changes: Tensions and challenges. *The Curriculum Journal, 16,* (3), 401-413.

Fox, J. (2008). Alternative assessment. In N. Hornberger and E. Shohamy (Eds.), *Encyclopedia of language and education 2nd Edition, Volume 7: Language testing and assessment* (97-109). New York: Springer.

Fox, J. (2009). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes, 8,* 26-42.

Fox, J. & Hartwick, P. (in press). Taking a diagnostic turn: Reinventing the portfolio in EAP classrooms. In D. Tsagari & I. Csepes (Eds.), *Classroom-based language assessment.* Frankfurt: Peter.

Fulcher, G. (2007). *Language Testing and Assessment, an Advanced Resource Book.* London: Routledge.

Fulcher, G. (2010). *An introduction to Assessment for Learning.* Retrieved: May 19, 2010 from http://languagetesting.info/features/afl/formative.html

Garner, J. (Ed.). (2006). *Assessment and Learning.* London: Sage Publications.

Genesee, F. & Upshur, J. (1996). *Classroom-based evaluation in second language education.* Cambridge: Cambridge University Press.

Gipps, C. V. (1994). *Beyond Testing: Towards a theory of educational assessment.* London: Falmer Press.

Glaser, R. (1990). Toward new models for assessment. *International Journal of Educational Research, 14* (5), 475-483.

Gouveia, V. & Valadares, J. (2004). Concept maps and the didactic role of assessment. In A. J. Cañas, J. D. Novak, & F. M. González (Eds.), *Concept*

maps: Theory, methodology, technology: Proceedings of the first international conference on concept mapping, vol. 2. Pamplona, Spain.

Half-baked. (2004). Hot Potatoes. Version 6.2 Half-baked Software, inc. http://web.uvic.ca/hrd/halfbaked/

Hamp-Lyons, L. (1999). Implications of the "examination culture" for (English language) education in Hong Kong. In V. Crew, V. Berry & J. Hung (Eds.), *Exploring diversity in the language curriculum.* (pp. 133-140). Hnog Kong: Hong Kong Institute of Education.

Hamp-Lyons, L. (2006). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins & C. Davison (Eds.), *The international handbook of English language teaching.* (Vol. 1) (pp. 487-506.) Norwell, MA: Springer.

Harlen, W. (1996). Editorial. *The Curriculum Journal 7*, 129.

Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal, 16,* (2), 207-223.

Harlen, W. (2006). The role of assessment in developing motivation for learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 61-80). London: Sage Publications Ltd.

Harlen, W. & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education, 4,* (3), 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.

Hatch, E. & Lazaraton, A. (1991). *The Research Manual, Design and Statistics for Applied Linguistics*. Boston, MA: Heinle & Heinle.

Herring, S. C. (1996). *Computer-mediated communication: Linguistic, social and cross-cultural perspectives.* Amsterdam: Benjamins.

Hirai, A. & Koizumi, R. (2009). Development of a practical speaking test with a positive impact on learning using a story retelling technique. *Language Assessment Quarterly, 6,* 151-167.

Hodgen, J. & Marshall, B. (2005). Assessment for learning in English and mathematics: a comparison. *The Curriculum Journal*, *16* (2), 153-176.

Hogue, A. (2003). *The essentials of English, a writer's handbook.* White Plains, NY: Pearson Education, Longman.

Huerta-Macías, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal,* (5), 8-11.

Hutchinson, C. & Hayward, L. (2005). The journey so far: Assessment for learning in Scotland. *The Curriculum Journal, 16* (2), 225-248.

Inspiration 8 [Computer software]. (2007). Oregon, USA: Inspiration Software, Inc.

James. M. (2000). Measured lives: The rise of assessment as the engine of change in English schools. *The Curriculum Journal, 11* (3), 343-364.

James, M. (2005). Editorial. *The Curriculum Journal, 16* (1), 3-6`.

James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 47-60). London: Sage Publications Ltd.

James, M. & Gipps, C. (1998). Broadening the basis of assessment to prevent the narrowing of learning. *The Curriculum Journal, 9,* (3), 285-297.

James, M., Pollard, A., Rees, G. & Taylor, C. (2005). Researching learning outcomes: Building confidence in our conclusions. *The Curriculum Journal, 16*, (1), 109-122.

Jang, E. (2009). Demystifying a Q-matrix for making diagnostic inferences about L2 reading skills. *Language Assessment Quarterly, 6,* (3), 210- 238.

Jang, E. (2010). Demystifying a Q-matrix for making diagnostic inferences about L2 reading skills, the author responds. *Language Assessment Quarterly,7*, (1), 116-117.

Jonassen, D. et al (2008). *Meaningful Learning with Technology*. Upper Saddle River, NJ: Pearson Education Inc.

Kachru, Y. (2005). Teaching and learning of world Englishes. In Eli Hinkel (Ed.) *Handbook of Research in Second Language Teaching and Learning.* Mahwah, N.J.: Lawrence Erlbaum Associates.

Khamesan, A. & Hammond, N. (2004). Taxonomy of analysis levels of learning effectiveness in collaborative concept mapping. In A. J. Cañas, J. D. Novak, & F. M. González (Eds.), *Concept maps: Theory, methodology, technology: Proceedings of the first international conference on concept mapping, vol. 2*. Pamplona, Spain.

Krakow, E. (2005). Metacognition: Helping students understand how they learn. *Schoolscapes, Ministère de l'Education du Gouvernement du Québec*. Quebec, QC: MEQ.

Lantolf, J. P. & Aljaafreh, A. (1995). Second language learning in the zone of proximal development: A revolutionary experience. *International Journal of Educational Research, 23*, 619-632.

Larsen-Freeman, D. (2009). Teaching and testing grammar. In M. Long & C. Doughty (Eds.), *The Handbook of Language Teaching*. (pp. 518-542). Boston: Blackwell Publishing.

Lee, Y.-W. & Sawaki, Y. (2009a). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly, 6,* 172-189.

Lee, Y.-W. & Sawaki, Y. (2009b). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly, 6,* 239-263.

Lee, Y.-W. & Sawaki, Y. (2009c). Cognitive diagnosis and Q-matrices in language assessment. *Language Assessment Quarterly, 6,* 169-171.

Lee, Y.-W. & Sawaki, Y. (2010). Cognitive diagnosis and Q-matrices in language assessment: The authors respond. *Language Assessment Quarterly, 7,* 108-112.

Legendre, M-F. (2001). Favoriser l'émergence de changements en matière d'évaluation. *Vie pédagogique, 120,* 15-19.

Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly, 1* (1), 19-41.

Leung, C. (2005). Classroom teacher assessment of second language development: Construct as practice. In E. Hinkel (Ed.) *Handbook of Research in Second Language Teaching and Learning.* New Jersey: Erlbaum Associates.

Leung, C. (2007). Dynamic assessment: Assessment *for* and *as* teaching? *Language Assessment Quarterly, 4* (3), 257-278.

Leung, C. & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing, 21* (3), 336-359.

Leung, C. & Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly, 40,* (1), 211-234.

Leung, C. & Rea-Dickins, P. (2007). Teacher assessment as policy instrument: Contradictions and capacities. *Language Assessment Quarterly, 4,* (1), 6-36.

Lightbown, P. (2000). Anniversary article, classroom SLA research and second language teaching. *Applied Linguistics, 21* (4), 431-462.

Linn, R., Baker, E., & Dunbar, S. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20,* (8), 15-21.

Luckin, R. & du Boulay, B. (1999). Ecolab: The development and evaluation of a Vygotskian design framework. *International Journal of Artificial Intelligence in Education, 10*, 198-220.

Lynch, B. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18* (4), 351-371.

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education, 10* (2), 209-220.

McNamara, T. (1996). *Measuring Second Language Performance.* London: Longman.

McNamara, T. (2001a). Editorial: Rethinking alternative assessment. *Language Testing, 18* (4), 329-332.

McNamara, T. (2001b). Language assessment as social practice: Challenges for research. *Language Testing, 18* (4), 333-349.

Messick, S. (1989). Validity. In R. L. Linn (Ed.). *Educational Measurement*. American Council on Education and the National Council on Educational Measurement.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13,* 241-256.

Ministère de l'Éducation du Québec. (2003). *Policy on the Evaluation of Learning, Abridged Version.* Québec, Québec : Gouvernement du Québec.

Mislevy, R., Steinberg, L. S. & Almond, R.G. (2002). Design and analysis in task-based language assessment. *Language Testing, 19* (4), 477-496.

Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice, 22,* 4, 13-25.

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist, 22* (2), 155-175.

Norris, J., Brown, J.D., Hudson, T. & Yoshioka, J. (1998). *Designing Second Language Performance Assessments.* Honolulu, HI: University of Hawai'i.

Novak & Cañas, 2008. The theory underlying concept maps and how to construct and use them. (Technical Report IHMC Cmap Tools No. 2006-01 Rev 01-2008) Florida: Florida Institute for Human and Machine Cognition.

Nyikos, M. & Hashimoto, R. (1997). Constructivist theory applied to collaborative learning in teacher education: In search of ZPD. *The Modern Language Journal, 81*, iv, 506-517.

O'Connor, D., Johnson, T. & Khalil, M. (2004). Measuring team cognition: Concept mapping elicitation as a means of constructing team shared mental models in an applied setting. In A. J. Cañas, J. D. Novak, & F. M. González (Eds.), *Concept maps: Theory, methodology, technology: Proceedings of the first international conference on concept mapping, vol. 2*. Pamplona, Spain.

Oscarson, M. (2004). Alternative forms of assessment and student participation. *Language Testing Update, 36*, 114-115.

Pellegrino, J. Chudowsky, N. & Glaser, R. (2001). *Knowing what Students Know: The Science and Design of Educational Assessment.* Washington, D.C.: National Academy Press.

Poehner, M. E. & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research 9*, (3), 233-265.

Purpura, J. (2004). *Assessing Grammar*. Cambridge: Cambridge University Press.

Qualifications and Curriculum Authority. (2001). Formative teacher assessment. Retrieved April 4, 2007, from http://www.qca.uk/ca/5-4/afl/definitions/asp

Quirk, & Greenbaum (1973). *A university grammar of English.* Essex: Longman.

Read, J. (2008). Identifying academic language needs through diagnostic assessment. *Journal of English for Academic Purposes, 7*, 180-190.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18* (4), 429-462.

Rea-Dickins, P. (2004). Editorial: Understanding teachers as agents of assessment. *Language Testing, 21* (3), 249-258.

Rea-Dickins, P. (2006a). Currents and eddies in the discourse of assessment: A learning-focused interpretation. *International Journal of Applied Linguistics,16* (2), 163-188.

Rea-Dickins, P. (2006b, June). *Teachers-as-raters, teachers-as-facilitators of language development: Is there a tug of war?* Paper presented at the joint meeting of the AAAL and ACLA/CAAL, Montreal.

Rea-Dickins, P. & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing, 17* (2), 215- 243.

Richards, J., Platt, J. & Platt, H. (1992). *Longman dictionary of language teaching & applied linguistics*. Essex, UK: Pearson Education Ltd.

Robinson, P. (1995). Attention, memory and the 'noticing' hypothesis. *Language Learning, 45*, 285-331.

Ross, J., Hogaboam-Gray, C. & Rolheiser, A. (2002). Influences on student cognitions about evaluation. *Assessment in Education, 9* (1), 81-95.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119-144.

Sawaki, Y., Kim, H-J., & Gentile, C. (2009). Q-matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessments. *Language Assessment Quarterly, 6,* (3), 190-209.

Sawaki, Y., Kim, H-J., & Gentile, C. (2009). Q-matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessments, the authors respond. *Language Assessment Quarterly, 7,* (1), 113-115.

Scriven, (1967). OR CITE BLACK & WILIAM (2003) FOR THIS. SEE PG. 2 OF LIT. REVIEW.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11* (2), 129-158.

Shakleton, J. (2005). 'Really useful ideas': Building users' powers of persuasion. A response to the discussion of the Learning Outcomes Theme Group from a practicioner in the post-compulsory sector. *The Curriculum Journal, 16,* (1), 101-103.

Shepard, L. (1991). Psychometricians' beliefs about learning. *Educational Researcher, 20,* (7), 2-16.

Shohamy, E. (2004). Democratic assessment as an alternative. *Language Testing, 18* (4), 373-391.

Solomon, P. (2002). *The assessment bridge. Positive ways to link tests to learning, standards, and curriculum improvement.* Thousand Oaks, CA: Corwin Press Inc.

Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing 18,* 4, 463-481.

Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn from test scores. *Language Testing, 18* (3), 275-302.

Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16* (3), 371-391.

Takeya, M. & Arroyo, A. (2004). A performance scoring method based on quantitative comparison of concept maps by a teacher and a student. In A. J. Cañas, J. D. Novak, & F. M. González (Eds.), *Concept maps: Theory, methodology, technology: Proceedings of the first international conference on concept mapping, vol. 2*. Pamplona, Spain.

Tashakkori, A. & Teddlie, C. (2003). *Handbook of Mixed Methods in Social & Behavioural Research.* Thousand Oaks, CA: Sage Publications.

Teasdale, A. & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing, 17* (2), 163- 184.

Tesch, R. (1990). *Qualitative research: Analysis types and software tools.* New York: Falmer Press.

Triggs, P., Weeden, P., Winter, J. & Broadfoot, P. (2000) The LEARN project: Phase 2: Developing assessment in the primary school project report. (University of Bristol Graduate School of Education CLIO Centre for Assessment Studies). *Report for the Qualifications and Curriculum Authority.* Available at: http://qca.uk/ages3-14/afl/292.html (September 2005).

Tunstall, P. & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22* (4), 389- 404.

Turner, C. E. (2006). Professionalism and high-stakes tests: Teachers' perspectives when dealing with educational change introduced through

provincial exams. *TESL Canada Journal/Revue TESL du Canada, 23,* (2), 54-76.

Turner, C. E. (2009). Examining washback in second language education contexts: A high stakes provincial exam and the teacher factor in classroom practice in Quebec secondary schools. *International Journal of Pedagogies and Learning*, (5), 103-123.

Turner, C. E. (in press). Classroom Assessment. In G. Fulcher & F. Davidson (Eds.), *Routledge Handbook of Language Testing*. (pp. 1-31). London: Routledge.

Turner, C. E., Tan, M., & Deschambault, E. (2010, June). *Development and field-testing of the writing task for Quebec's Secondary Five provincial ESL exit exams: Using the washback effect to create an assessment bridge.* Poster session presented at the annual conference of the Canadian Association of Applied Linguistics/Association canadienne de linguistique appliquée. Montreal: QC.

Turner, C. E. & Tan, M. (2010). [Examining teacher mediation between classroom assessment activity and preparing students for upcoming external exams: The assessment bridge and the potential for synchronized assessment.] Unpublished raw data.

Turner, C. & Upshur, J. (1996). Developing rating scales for the assessment of second language performance. In G. Wigglesworth & C. Elder (Eds.), *Australian Review of Applied Linguistics: Series S, No. 13. The Language Testing Cycle: From Inception to Washback.* (pp. 55-79). Melbourne: ARAL.

Turner, C. & White, J. (March, 2004). *What does it take to get a high score on a paired task? Examining discourse, scores and the construct of L2 proficiency?* Paper presented at the Language Testing Research Colloquium, Temecula, CA.

Van de Watering, G. & Van der Rijt (2006). Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. *Educational Research Review, 1*, 133-147.

Vygotsky, L. (1962). *Thought and language.* Cambridge, MA: MIT Press.

Vygotsky, L. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Weeden, P. & Winter, J. (1999). The LEARN project: Learners' expectations of assessment for learning nationally. (University of Bristol Graduate School of Education CLIO Centre for Assessment Studies). *Report for the Qualifications and Curriculum Authority.* Available at: http://qca.uk/ages3-14/afl/292.html (September 2005).

Western and Northern Canadian Protocol for Collaboration in Education. (2006). *Rethinking classroom assessment with purpose in mind.* [Brochure online]. Available at: www.wncp.ca

Wiliam, D. (2007). Five "key strategies" for effective formative assessment. *National Council of Teachers of Mathematics assessment research brief.* Reston, VA: NCTM.

Wiliam, D., Lee, C., Harrison, C. & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11* (1), 50-65.

Wolf, A. (1998). Portfolio assessment as national policy: The National Council for Vocational Qualifications and its quest for a pedagogical revolution. *Assessment in Education, 5* (3) 413-433.

Young, E. (2005). Assessment for learning: Embedding and extending. *Smarter Scotland Scottish Executive*. Available at: http://www.ltscotland.org.uk/assess/for/indez.asp (January, 2007).

Young, R. (2002). Discourse approaches to oral language assessment. *Annual Review of* Applied Linguistics, 22, 243-262.

APPENDICES

Appendix A – A Quizzical LearnModal Quiz

A Quizzical LearnModal Mystery

Gap-fill exercise

Fill in all the gaps, then press "Check" to check your answers. Use the "Hint" button to get a free letter if an answer is giving you trouble. You can also click on the "[?]" button to get a clue.

Joe [ ] **[?]** like to see that car in his parking space.

It [ ] **[?]** be too bad if rainshowers were predicted for this afternoon.

Marie has agreed she [ ] **[?]** be the class president starting next September.

How long do you think your funds [ ] **[?]** last?

If we run out of money, what [ ] **[?]** you suggest we do?

A lottery win [ ] **[?]** certainly help with tuition costs.

[ ] **[?]** you be so kind as to help me with my homework?

You [ ] **[?]** help me, or else!

Next Thursday, Ali [ ] **[?]** drop by the student affairs office to get the key.

How [ ] **[?]** you happen to know that information?

Check Hint

Appendix B - A Quizzical LearnModal Quiz Hints

1. What does Joe mean by this? Is he talking about his own car?
2. Have rain showers been predicted?
3. Is it certain that Marie is going to be the class president?
4. Do you have funds now?
5. Are we out of money NOW?
6. Have we already won the lottery?
7. You might say this to the Queen of England.
8. The speaker sounds very certain!
9. Are we sure of this? Why?
10. Think about 'happen' used here - what does it signify?

Appendix C - Agenda for Treatment Teacher Training

C. COLBY-KELLY PhD STUDY JUNE-JULY, 2009 DATA COLLECTION TEACHER TRAINING AGENDA

📖 **Study inspiration**
- No blame for errors approach
- Effective bio-learning - ensuring learning has been effective before moving on to the next level (what good teachers and ants have in common)
- Assessment for Learning (AFL)

📖 **Grammar feature in question**
- Contingent use modal 'would'
- When to use 'would' and when to use 'will'

📖 **Pre-treatment tests (2)**
- Fill-in-the blanks
- One-page essay

📖 **Format of study**
- 'Treatment' and 'control' groups
- 2 sessions + 2 teachers each session
- Revised calendar

📖 **Computer-Assisted Language Learning (CALL) module**
- English Mystery pages in computer lab for use after classes in 10-pg. module
- Link to learning quiz
- Concept mapping software for student online CM production

📖 **In-class group exercise**
- Paper and pencil group concept mapping

📖 **In-class teacher-class exercise**
- Blackboard class concept mapping

📖 **'Homework' assignments**
- As per calendar dates

📖 **Post-treatment tests (2)**
- Fill-in-the blanks
- One-page essay

📖 **Delayed post-treatment tests (2)**
- Fill-in-the blanks
- One-page essay

Appendix D - Essay Test Formats

Piloting test:
Imagine it is a very windy day and a lottery ticket has fallen at your feet on the sidewalk. You have no way of knowing who the ticket belongs to. Later in verifying the numbers on the ticket you discover you have just won $10,000.00. Using the space below describe your plans for your winnings. (Use the space on the back of the page if you need it.)

Session 1 Essay Pre-test:
Imagine you have just been offered a two-week all-expenses paid trip to the destination of your choice at the end of this school year. Use the space provided to say what you plan to do with this offer.

Session 1 Essay Parallel Pre-test:
Imagine you have just been offered a two-week all-expenses paid trip to the North American destination of your choice *by train* at the end of this school year. Use the space provided to say what you plan to do with this offer.

Session 1 Essay Post-test:
Suppose you are about to enter university and you find out that among other students, a company has chosen *you* to have all of your living and tuition expenses paid during your studies. Use the space provided to say what you plan to do with this offer.

Session 2 Essay Pre-test:
Imagine you have just won a brand-new car with gasoline expenses paid for the next three months. Use the space provided to say what you plan to do with this offer.

Session 2 Essay Parallel Pre-test:
N/A

Session 2 Essay Post-test:
Instructions: Imagine a radio station has chosen your name in a draw to receive $5,000.00 of free merchandise of your choice from any "Better Buy" electronic store located in Montreal. Use the space provided to say what you plan to do with this offer.

Delayed Essay Post-test:
Instructions: Imagine you have entered a newspaper contest and you have just received a call to tell you that *you* have been named the winner! In the space provided please say what you intend to do with the $5,000.00 prize money you have won.

Appendix E - Fill-in-the-blanks Test Formats

Modal Choices

Choose and write the correct form (*would/will*) in the following sentences:

Piloting test:

1. I _____ love to get a new iPod for my birthday.
2. Pierre _____ like to go to Mexico on the March break.
3. If Suzanne were here today, she _____ have a laugh.
4. I know I _____ be going to the cottage this weekend.
5. If I were you, I _____ not mention that to Georges.
6. If they were really poor, they _____ not be able to afford that car.
7. If you work very hard, you _____ become a doctor.
8. We _____ love to go skiing more often.
9. Where _____ we be without music in our lives!
10. Tomorrow I will get a new computer. I _____ like to have some help setting it up.

Session 1 pre-test TG1 and CG1:

1. My parents _____ love to come visit me more often.
2. If you work hard, you _____ succeed.
3. Were I you, I _____ not tell Marie about that!
4. I know I _____ be going to the Comedy Festival this summer.
5. On Friday I'm getting a new notebook. I _____ sure like to have some help with it.
6. If he were really weak, he _____ not be able to lift that piano.
7. You _____ love to get new earbuds, right?
8. Juan _____ like to go to see his family in Spain before September.
9. If your teacher saw that, she _____ laugh!
10. How _____ we live without friendships!

Session 1 pre-parallel test TG1:

1. You _____ love to get a newer cell phone, right?
2. I know I _____ be going to see the fireworks festival this summer.
3. If your mother saw that, she _____ laugh!
4. How _____ we live without pizza!
5. On Friday I'm getting a new cell phone. I _____ sure like to have some help setting it up.
6. If he were really not hungry, he _____ not be able to eat that pizza.
7. My parents _____ rather visit in the summer than in the winter.
8. If you try really hard, you _____ succeed.
9. If I were you, I _____ not tell Pierre that!
10. Ali _____ like to go to see his grandfather before September.

Session 1 post-test TG1 and CG1:

1. On the weekend Mela and I are going to Sara's cottage in Granby. We _____ sure like to get there in time for lunch.
2. Milad _____ like to go to his country before the end of the summer.
3. If you save your money, you _____ be successful.
4. Did you say you _____ love to get some new clothes?
5. I _____ not tell your boyfriend/girlfriend that!
6. My sister _____ prefer to visit me around the time of my birthday.
7. How _____ we do anything without computers!
8. If Wei saw that, he _____ not believe it!
9. I know I _____ be going out of the city sometime this summer.
10. If she were really angry, she _____ not be smiling.

Session 2 pre-test TG2 and CG2:

1. Did you say your friend _____ love to come, too?
2. The next time we have a holiday, they said they _____ prefer to come for dinner rather than lunch.
3. I know I _____ be going on a picnic sometime this summer.
4. Tomorrow after school Marie and I are going to the University Lost-and-Found Centre. We _____ sure like to find the sweater she lost.
5. When _____ you graduate from University?
6. Mohamed _____ like to go to his country to be at his brother's wedding at the end of the summer.
7. If your mother said that, you _____ not believe it!
8. If you invest your money wisely, you _____ have some success.
9. If that comedian really were funny, we _____ all be laughing by now.
10. If I were you I _____ not speak to my friend like that!

Session 2 post-test TG2 and CG2:

1. The next time you visit, I _____ make you a cup of my special coffee.
2. Wei _____ like to go to the Fireworks Festival in Montreal before the summer is over.
3. If your father did that, you _____ be proud of him!
4. When _____ you finish classes at this school?
5. Did you say that you _____ like to come, too?
6. The next time, the class decided they _____ prefer to have the lunch party in the classroom.
7. I know I _____ need to buy an umbrella and a raincoat if this keeps up.
8. We _____ all be sorry if that were to happen.
9. Anya and I are going to the Jazz Festival. We _____ like to hear some American jazz played on the saxophone.
10. If I were you, I _____ not talk on the cell phone so much!

Delayed post-test for all groups:

1. When _____ you complete your studies?
2. Did you tell me you _____ like to see that new film with me?
3. If my sister explained the idea, my mother _____ understand it better.
4. I know I _____ need to buy a laptop when I start the first year of university.
5. Another time, we all decided we _____ rather read the book than see the movie.
6. Moe and I want to go to the bookstore after class. We _____ like to buy all of our textbooks for next term at once.
7. The next time we go out together, I _____ 'text' you when I'm almost ready to leave, so you won't have to wait for me.
8. Joey _____ like to go to the market to get some fresh vegetables and some tofu.
9. If I were you, I _____ not worry about that!
10. I _____ try to cure people if I were a doctor.

Appendix F - Post-CALL Student Survey Questions

Post-CALL survey Questions:

1. Did this help you learn?
2. In what way

Appendix G - Post-CM Production Student Survey

Name: _____

Date: _____

Using Concept Mapping Software

---

1. Did making a concept map help you learn?

_____

_____

_____

_____

2. In what way?

_____

_____

_____

Appendix H - Teacher Mystery Module and Concept Mapping Exercises Post-Questionnaire

MYSTERY MODULE AND CONCEPT MAPPING EXERCISES TEACHER POST-QUESTIONNAIRE

TEACHER: _____

3. **(If applicable) Do you think the Mystery Module helped your students learn? _____**

4. **In what way? _____**
   _____
   _____

5. **(If applicable) Do you think the computerized concept map exercise helped your students learn? _____**

6. **In what way? _____**
   _____
   _____

7. **(If applicable) Do you think the group concept map exercise helped your students learn? _____**

8. **In what way? _____**
   _____
   _____

9. **Do you think the teacher and class creation of a concept map helped your students to learn? _____**

10. **In what way? Can you give an example of this? _____**
    _____
    _____

11. **Do you think your guided questioning helped your students learn?**

    _____

12. **In what way? Can you give an example/s of this? _____**
    _____
    _____

13. Will you use this way of teaching in the future? _____

14. Why or why not? _____
_____
_____

15. Did these activities change the way you *think* about student learning?

16. In what way? _____
_____
_____

17. What kinds of learning styles do you think these learners have? (analytical, visual, auditory, etc.) _____
_____
_____

18. **What impact do you think these learning styles might have on student learning in these exercises? _____**
_____
_____

19. What did you think the students learned in these exercises? _____
_____
_____
_____

THANK YOU VERY MUCH
FOR YOUR PARTICIPATION!

Appendix I - Student Mystery Module and Concept Mapping Exercises Post-Questionnaire

MYSTERY MODULE AND CONCEPT MAPPING EXERCISES

POST-QUESTIONNAIRE

---

20. Did you use the Mystery Module pages for self-assessment of your work? _____

21. Did this self-assessment help you learn? _____

22. In what way? _____
_____
_____

23. Did creating a computerized concept map help you learn? _____

24. In what way? _____
_____
_____

25. Did you use the computerized concept mapping exercise for self-assessment of your work? _____

26. Did this self-assessment help you learn? _____

27. In what way? _____
_____
_____

28. How many concept map drafts did you make before you made a final map? _____

29. If you made drafts, did the process of re-making the map help you learn? _____

30. In what way? _____
_____
_____

31. Did creating a group concept map help you learn? _____

32. In what way? _____
_____
_____

33. Did other group members help you assess your own work? _____

34. In what way? Can you give an example of this? _____

_____

_____

35. Did creating a concept map with your teacher and class help you learn?

_____

36. Did this teacher and class activity help you assess your own work?

_____

37. In what way? Can you give an example of this? _____

_____

_____

38. Did self, group, or teacher-class assessments help your learning?

_____

39. In what way? _____

_____

_____

40. Will you use this way of learning in the future? _____

41. Why or why not? _____

_____

_____

42. Did these activities change the way you *think* about how you learn? __

.

43. In what way? _____

_____

_____

44. What kind of learner are you? (analytical, visual, auditory, etc.) _____

45. **What impact do you think your learning style had on your learning in these activities?** _____

_____

46. What did you learn in these exercises? _____

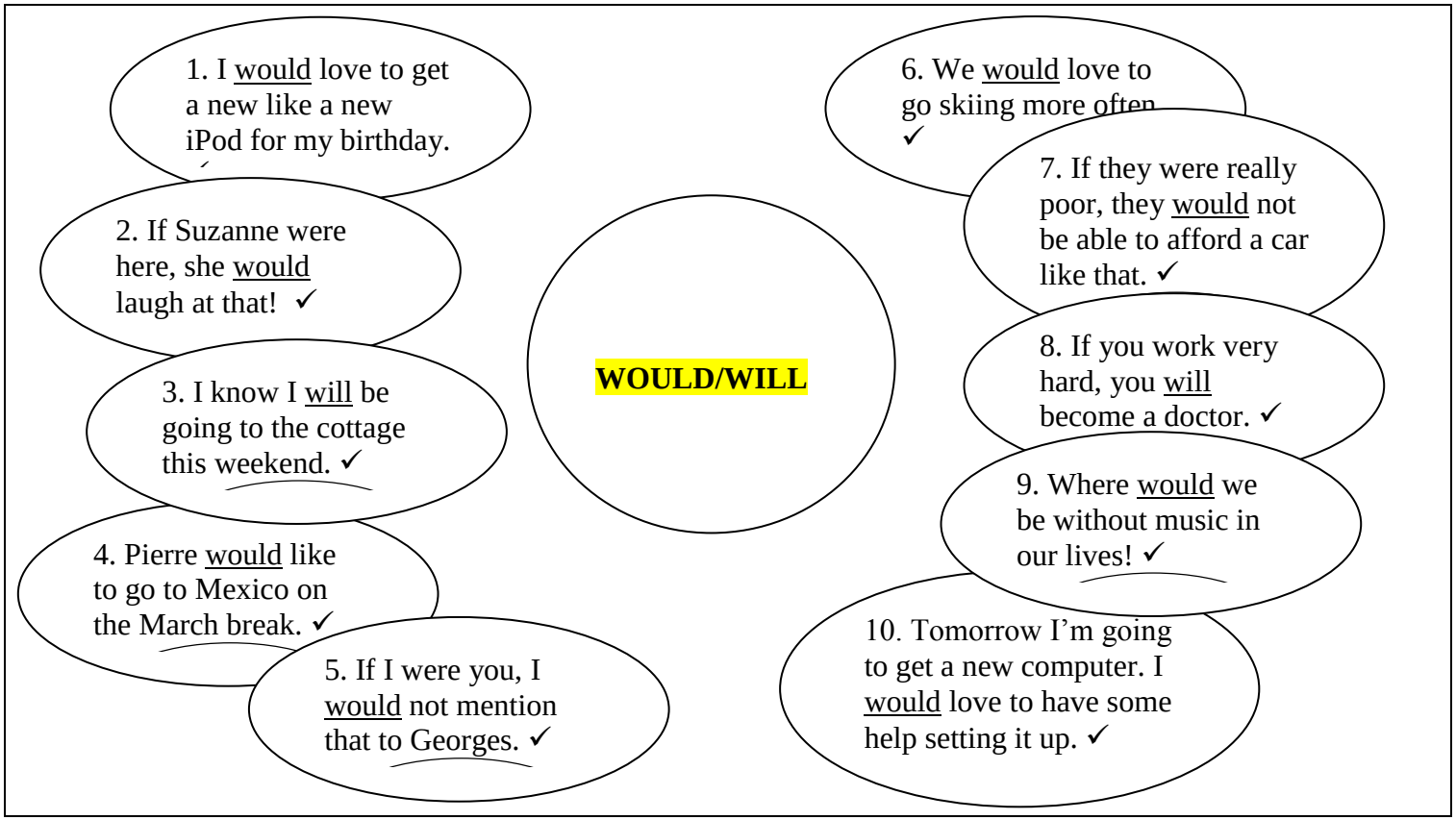
_____

_____

THANK YOU FOR YOUR PARTICIPATION!

Appendix J - Pilot Study Concept Map Template

C. Colby-Kelly PhD Pilot Project Concept Map Activity



Appendix K - Teacher 10 Principles Questionnaire

(Reduced for formatting purposes; in the original the instructions and questionnaire were on one page)

TEACHER QUESTIONNAIRE

Context: Formative assessment refers to all assessments that are used during a course to show learning progress and to help students attain the goal of learning *while* they are learning; they usually don't "count" in students' final marks.
Instructions: Using the scale given, please indicate how much you agree with the statements in the table below.
Scale: 1 = strongly disagree 2 = disagree 3 = agree 4 = strongly agree

FORMATIVE ASSESSMENT IN THE CLASSROOM SHOULD…	1 STRONGLY DISAGREE	2 DISAGREE	3 AGREE	4 STRONGLY AGREE
1. be part of effective planning of teaching and learning.				
2. focus on how students learn.				
3. be recognised as central to classroom practice.				
4. be regarded as a key professional skill for teachers.				
5. be sensitive and constructive given that assessment has an emotional impact.				
6. take account of the importance of learner motivation.				
7. promote commitment to learning goals and a shared understanding of the criteria by which they are assessed.				
8. give learners constructive guidance about how to improve.				
9. develop learners' capacity for self -assessment so that they can become reflective and self-managing.				
10. recognise the full range of achievements of all learners.				

Appendix L - Student 10 Principles Questionnaire

(Reduced for formatting purposes; in the original the instructions and questionnaire were on one page)

STUDENT QUESTIONNAIRE

Context: Formative assessment refers to all assessments that are used during a course to show your learning progress and to help you attain the goal of learning while you are learning; these assessments usually don't "count" in your final mark.

Instructions: Using the scale given, place an 'X' to indicate how much you agree with the statements in the table below.

Scale: 1 = strongly disagree 2 = disagree 3 = agree 4 = strongly agree

FORMATIVE ASSESSMENT IN THE CLASSROOM SHOULD…	1 STRONGLY DISAGREE	2 DISAGREE	3 AGREE	4 STRONGLY AGREE
1. be part of effective planning of teaching and learning.				
2. focus on how students learn.				
3. be recognised as central to classroom practice.				
4. be regarded as a key professional skill for teachers.				
5. be sensitive and constructive given that assessment has an emotional impact.				
6. take account of the importance of learner motivation.				
7. promote commitment to learning goals and a shared understanding of the criteria by which they are assessed.				
8. give learners constructive guidance about how to improve.				
9. develop learners' capacity for self -assessment so that they can become reflective and self-managing.				
10. recognise the full range of achievements of all learners.				

Appendix M – Transcription Protocol

…	unfinished utterances
[…]	omitted discourse
[?]	unclear utterances
[ ]	words inside brackets explain context or actions taken
(SM)	male student
(SF)	female student
“ ”	used to indicate sentences used as examples
Italics	to show speaker emphasis, to show words discussed as forms

In addition, words have been spelled as they sounded and so pronunciation errors affecting meaning were reported and the target word furnished, for example ‘pre-duct [predict]’. The sentences were transcribed as they were heard, including grammatical errors, word repetitions, and half-words uttered. Filled pauses were transcribed as ‘um’, ‘mhm’, ‘uh’, and ‘ah’. In a discourse sequence the student speakers were numbered with Roman numerals to indicate different unidentified speakers. Where possible, student speakers were identified by their student participant codes.