Getting real in the language classroom: Developing Japanese students' communicative competence with authentic materials

Alexander Gilmore

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

September 2007

ABSTRACT

The research described in this thesis reports on a 10-month quantitative/qualitative classroom-based study, carried out at a Japanese university, investigating the potential of authentic materials to develop learners' communicative competence. It was hypothesised that the 'richer' input provided by authentic materials, combined with appropriate awareness-raising activities, would be better able to develop a range of communicative competencies in learners (linguistic, pragmalinguistic, sociopragmatic, strategic and discourse competences).

Ninety-two 2nd year English major students, of similar proficiency levels, were assigned to either a control or experimental group for the period of the trial. The control group received input from two textbooks commonly used in Japanese universities, while the experimental group received input from authentic materials (films, documentaries, 'reality shows', TV comedies, web-based sources, home-produced video of native speakers, songs, novels and newspaper articles), designed to allow students to 'notice' features of the discourse which could help them develop some aspect of their communicative competence. The hypothesis was tested with a batch of eight pre/post-course measures, designed to tap into different aspects of learners' communicative competence or language skills: a) Listening; b) Pronunciation; c) 'C'-Test; d) Grammar; e) Vocabulary; f) Discourse completion task (DCT); g) IELTS oral interview; h) Student-student role-play. These were supported with qualitative results from learners' diaries, case-study interviews with subjects from both groups and transcripts of classroom interaction.

Univariate analysis of the pre/post-course tests, using ANCOVA, indicated statistically significant differences between the two treatment groups, with the experimental group out-performing the control group in five of the eight communicative competence measures. The qualitative results of the trial helped to account for these differences in performance, suggesting that the authentic materials, and their associated tasks, allowed learners to notice a wider range of discourse features than those generally available in textbook input. They also indicated a clear preference in the experimental group for authentic materials over textbooks, suggesting that learners found them more interesting, varied and challenging, and better able to meet their perceived future language needs. Finally, the qualitative results demonstrated that, for learners, social goals often override instructional goals in the classroom, suggesting that classroom-based research benefits from both an emic and etic perspective in order to fully account for results.

ACKNOWLEDGEMENTS

My sincere thanks to the many people who helped with this research, and without whom this thesis would never have been completed. In particular, I would like to thank my supervisor, Professor Ronald Carter, for his patience, encouragement and insights into language and pedagogy. I am also deeply indebted to my second supervisor, Professor Zoltán Dörnyei, for the many hours he willingly gave up to help with the methodological aspects of my investigation ~ I couldn't have hoped for better mentors. Many thanks also to my friend, Professor Masanori Toyota, for his insights and advice on Japanese language and culture.

Thanks are also due to Dr. Gila Schauer (Lancaster University) who very kindly allowed me to use her multimedia DCT in this investigation, as well as the staff and students of Kansai Gaidai University, Japan, who participated in this research. In particular, I would like to thank Tim Micklas, Danielle Talerico, Mark Seimelink, Richard Cleveland, Chris Burrows, Jerry Gordon, David Althaus, Gareth Jones, Steve Müller, Francoise Mercier, Richard Harrison, Neil Klein and David Stormer for the time they gave up to interview students and rate the oral interviews and DCT responses. My appreciation also goes to all the students of classes 1, 2, 3 and 4 in 2004/5 who flung themselves enthusiastically into the lessons and classroom research. Finally, an extra special thank you goes to my wife, Michiyo, for the hours she sacrificed translating the Japanese interaction, and for all of her support and many compromises over the years it has taken to complete this work.

Dedicated to the memory of my mother, Wendy Patricia McDonald, whose love, like a distant star, continues to shine down on me long after she has gone.

PART I

INTRODUCTION

Background to the study

When I look back on my own language learning experiences, and consider what elements have remained in my memory over the passage of time, I realise that the majority of experiences that have lasted are associated with authentic materials.

One example, from my Spanish classes in Mexico over 10 years ago, is a song called 'Unicorn' by the Cuban songwriter Silvio Rodriguez (1982):

Unicornio

Mi unicornio azul ayer se me perdió.
Pastando lo dejé y desapareció.
Cualquier información bien la voy a pagar.
Las flores que dejó
no me han querido hablar.

Mi unicornio azul ayer se me perdió,
no sé si se me fue, no sé si extravió, y yo no tengo
más que un unicornio azul
Si alguien sabe de él, le ruego información.
Cien mil o un millón, yo pagaré.
Mi unicornio azul se me ha perdido ayer,
se fue.

Mi unicornio y yo hicimos amistad,
un poco con amor, un poco con verdad.
Con su cuerno de añil pescaba una canción,
saberla compartir era su vocación.

Mi unicornio azul ayer se me perdió,
y puede parecer acaso una obsesión,
pero no tengo más que un unicornio azul
y aunque tuviera dos yo sólo quiero aquel.
Cualquier información, la pagaré.
Mi unicornio azul se me ha perdido ayer,
se fue.

Unicorn

My blue unicorn, I lost him yesterday.
I left him grazing and he disappeared.
For any information, I'll pay well.
The flowers that he left behind
don't want to talk to me.

My blue unicorn, I lost him yesterday,
I don't know if he left me, or if he got lost,
and I only have one blue unicorn.
If anyone's got any news, I beg to hear it.
A hundred thousand or a million, I'll pay.
My blue unicorn, I lost him yesterday,
he went away.

My unicorn and I became friends,
a little bit with love, a little bit with truth.
With his indigo horn, he fished for a song
knowing how to share it was his calling.

My blue unicorn, I lost him yesterday,
and it may seem perhaps like an obsession
but I only have one blue unicorn
and even if I had two I'd only want him.
For any information, I'll pay,
My blue unicorn, I lost him yesterday,
he went away.

The song, with its very beautiful but melancholic melody, immediately caught our attention and it quickly became apparent from the first read-through that its central theme

was ‘loss’. What encouraged us to engage with the text from this point on was the discovery of what exactly the ‘blue unicorn’ represented. Obviously it was someone, or something, very special: unicorns are hard enough to find, but blue ones are unique indeed. We struggled in class to arrive at an answer to this puzzle and, in the process, had to deal with the rich vocabulary and grammatical forms used by the songwriter to express himself, such as reflexive verbs, future tenses and irregular past forms. Our teacher finally explained to us that the blue unicorn represented the son of Silvio’s friend, who died in El Salvador, fighting for the guerrillas during the civil war. The powerful lyrics, attractive melody and content, relating to real people and real events, combined to make this learning experience highly memorable, and our desire to understand the meaning of the text encouraged us to engage with it and deal with the lexicogrammatical obstacles in our way. Because the words were embedded in a song, I effectively learnt them as one, long fixed expression. Later, I was able to deconstruct this as my Spanish improved, and use elements of it, such as ‘se fue’, in my own conversations.

In Japan, one of my clearest memories from the classroom is of reading ‘Yuki-onna’ (Snow Woman), a fairytale about two wood-cutters, Mosaku and Minokichi, who become trapped in a forest hut during a terrible snow storm. In the night, a pale and beautiful snow spirit enters the hut and her icy breath leaves the older Mosaku a frost-covered corpse. Minokichi, with his youth and handsome looks, is spared from death and promises never to mention what he has seen to anyone else in return for his life. He escapes from the forest and, on the way back to his hometown, meets an attractive young lady who he falls in love with, marries, and has a family. He never speaks of his terrifying experience again until, one night, while watching his wife sew, her face brings

back memories of the past. He reveals his secret to her and, to his horror, his wife turns into the snow woman before his eyes:

Figure 1. Depiction of Yuki-onna

「お前の顔を見ていると、昔のことを思い出すんだよ。」
"When I look at your face, it reminds me of something that happened a long time ago."
「昔のことって、昔、何かあったんですか。」
"A long time ago, you say, what happened?"
「あれは18才の時だった。お前によく似た女を見たんだよ。」
"I was 18 years old, I saw a woman who looked just like you."
おゆきは顔を上げずに仕事を続けながら言いました。
"Oyuki, still working, asked without looking up.
「どこでその方をご覧になったんですか。」
"Where did you see this woman?"
巳之吉はあの約束を思い出しました。
At that moment, Minokichi remembered his promise.
「いやいや、これはお前にも言えない。」
"No no, I can't tell even you that."
「どうして、どうして言えないんですか。私たち夫婦の間に秘密などないではありませんか。今までだって、何でもお互いに打ち明けて、助け合って来たんじゃありませんか。さあ、お話しになって。」
"Why, why can't you tell me? There are no secrets between a husband and wife. Up to now, we've always told each other everything and helped each other. Go on tell me!"
あまりにも強く言われたので、巳之吉はとうとうあの晩のことを全部話してしまいました。あの晩の寒さ、恐さ、それにあの女のことも。
Since she insisted so strongly, Minokichi finally told her everything about that night. The cold, the fear and finally about the woman too.
「あの女は人間じゃなかった。あれは一体。。。。」
"The woman wasn't human. I wonder what she was……."
と、そこまで話して、巳之吉は驚きのあまり、口がきけなくなりました。
Before he could finish the sentence, Minokichi froze and couldn't utter another word.
「約束を破ったな。」
"You broke our promise!"
おゆきの姿がいつの間にか、あの恐ろしい雪女の姿に変わっていたのです。
Oyuki turned into the terrifying Snow-woman.
「あの時、誰かに話したら殺すと言っておいたのに、どうして話したのだ。お前を殺さなければならない。。。。。でも、あそこに寝ている子供たちを見ると、今、お前を殺すことはできない。。。。子供を大切にしてください。子供にやさしくしてください。もしやさしくしなかったら、私はすぐにわかる。。。。。」
"That night, I told you if you ever told anyone about me, I'd kill you. Why did you tell me? Now I've got to kill you……..But when I see our children sleeping over there, now I can't kill you…..Please look after our children. Please be kind to them. If you're not, I'll know immediately……."
話しているうちにおゆきの声は小さくなりました。そしておゆきは白い霧になって、窓から静かに出て行ってしまいました。その後、村の人たちは二度とおゆきの姿を見なかったそうです。
While Oyuki was talking, her voice became quieter and quieter. Then Oyuki turned into white mist and quietly disappeared through the window. After that, the people of the village never saw Oyuki again.

The translation of the scene into English doesn't do justice to the story because much of the terror of the transformation from wife to malign spirit is achieved through the way Yuki-onna speaks. Before the promise is broken, she uses the polite, gender-specific language expected of an obedient wife: '*doko de sono kata wo goran ni nattan desu ka*' (Where did you see this woman?); '*watashi tachi fuufu no aida ni himitsu nado nai dewa arimasen ka'* (There are no secrets between a husband and wife, are there?). This changes to cruder, more masculine language, marked by the verb endings (*yabutta*, rather than *yaburimashita*), sentence final particles[1] (*yabutta na*, rather than *yabutta ne*), and lexical adaptions (the use of *omae* instead of *anatta* for 'you'). These grammatical and lexical signals, combined with paralinguistic changes in the delivery of the lines, have a truly spine-chilling effect. The text also highlights for the learner some of the complex ways gender and register are marked in Japanese.

It is hardly surprising that these are the materials that have stayed in my memory when so many others have been lost. They connected with my emotions and imagination, and in the process created a desire to engage with and understand the texts ~ and a huge amount of language learning was stimulated as I attempted to interpret the messages the writers were conveying.

But there are other benefits associated with authentic materials other than their potential to motivate students. They provide learners with a richer source of L2 input to work with, which has two advantages: a) It is more likely to meet the varying interlanguage needs of individual students within the class and; b) It is more likely to develop a range of communicative competencies in learners (see chapters 1 & 2).

[1] The sentence final particles *ne* and *na* have the same function as tag questions in English.

The tightly structured, lexicogrammatical syllabuses we have been working with in language classrooms for decades may create the comforting impression of order and comprehensiveness, but in reality they often only present a tiny fraction of the full richness of the language, through the mouths of cardboard cut-out characters discussing topics as dull as ditchwater. After over a decade of boring my learners (and myself) to death with textbooks imposed on us by higher authorities, I have reached a point where I would like to explore other possibilities: What would happen if we abandoned textbooks in favour of exclusively authentic materials in the classroom? How would learners' communicative competence develop with authentic input instead of textbooks and how would they respond to this (often) more challenging material? This investigation aims to explore these questions.

Organisation of the thesis

The thesis is divided into two parts, with part I containing all the main chapters and part II the appendices and bibliography.

Chapters 1, 2 & 3 review the literature associated with the three principle concerns of this investigation: communicative competence, authenticity and research methodology.

Chapters 4, 5, 6, 7 & 8 describe the classroom-based mixed methods research carried out over a 10-month period at Kansai Gaidai University in Japan. Chapter 4 outlines the setting, participants and testing instruments used in the main trial. Chapter 5 summarises the quantitative results from inferential statistical analysis of the pre- post-course tests of communicative competence. Chapter 6 describes the qualitative results from the learners' diaries and case study interviews, while chapter 7 outlines a sequence of lessons from the

experimental input, which illustrate how the authentic materials might have led to greater levels of overall communicative competence in the experimental group. Finally, chapter 8 outlines the conclusions of the study, and discusses its strengths, limitations and research or pedagogical implications.

Part II includes the appendices, bibliography and a DVD containing sample video extracts from the IELTS oral interviews, student-student role-plays and classroom interaction.

CHAPTER 1. COMMUNICATIVE COMPETENCE[2]

1.1 Historical roots of the communicative competence model

It is no accident that models of communicative competence developed in the 70s and 80s in parallel with calls for greater authenticity in the literature. Both stemmed from the increasing influence of sociolinguistics on the profession and the growing realisation that language cannot be separated from the context in which it is produced. Labov (1966) and Hymes (1972), for example, both argued for a move away from abstract, ideal notions of native speaker competence towards a focus on actual, contextualised performance. Hymes strongly attacked Chomsky's notion of an 'ideal speaker-listener' and his trivialisation of performance:

> It is, if I may say so, rather a Garden of Eden view. Human life seems divided between grammatical competence, an ideal innately-derived sort of power, and performance, an exigency rather like the eating of the apple, thrusting the perfect speaker-hearer out into a fallen world. Of this world, where meaning may be won by the sweat of the brow, and communication is achieved in labor…little is said. The controlling image is of an abstract, isolated individual, almost an unmotivated cognitive mechanism, not, except incidentally, a person in a social world. (Hymes, 1972: 272)

Instead, he proposed a broader definition of competence that accounted for a speaker's knowledge of the language itself as well as his/her ability to *use* the language in a social context. Hymes was vague as to what competence actually meant beyond a concern for what language is possible, feasible and appropriate in a given situation but he sparked a

[2] A modified version of chapters 1 & 2 was originally published in *Language Teaching* 40.2: 97 – 118, April 2007, © Cambridge University Press.

frenzy of interest in pragmatics and a model of communicative competence slowly began to take shape, first with Clyne (1979) and Schmidt & Richards (1980) and then, most importantly, Canale & Swain (1980) and Canale (1983). This resulted in a framework composed of four areas of knowledge: grammatical competence, sociolinguistic competence, discourse competence and strategic competence. This model was further refined by Leech (1983) who divided sociolinguistic competence into two separate components, pragmalinguistics and sociopragmatics, and Celce-Murcia, Dörnyei & Thurrell (1995) who re-named grammatical competence 'linguistic competence' in recognition of the fact that this area includes a speaker's lexical and phonological knowledge as well as grammatical knowledge.

1.2 Components of the communicative competence model

Today, there seems to be general agreement in the literature on a model of communicative competence consisting of five components:

a) **Linguistic competence**: This refers to a speaker's lexical, morphological, orthographical, syntactical and phonological knowledge of the language. In other words, how to build up morphemes into words and words into clauses and sentences, how to spell them in the written form or pronounce them in speech. It only deals with the literal meaning (or locutionary force) of utterances. This is the type of knowledge that has traditionally been the staple diet of ELT classrooms and it is important to note that it is not rejected in the current model of communicative competence, but rather assumes a lesser role, seen as only one aspect of language proficiency.

b) **Pragmalinguistic Competence**: This refers to a speaker's ability to understand or convey communicative intent appropriately in a given context based on a knowledge of phrases typically used by native speakers to express speech acts such as apologies, requests, refusals and so on. This kind of competence therefore describes a speaker's ability to interpret the illocutionary force, or conversational implicature (Grice 1975), of utterances, for example, understanding that 'Could you open the door?' is a request rather than a question about ability to complete an action.

c) **Sociopragmatic Competence**: This refers to a speaker's knowledge of what is socially or culturally appropriate in a particular speech community. This might include an appreciation of politeness and social conventions, taboo topics and non-verbal factors such as kinesics and proxemics. For example, the knowledge that, in Japan, business cards should be exchanged at the beginning of an initial meeting, handed to the recipient with both hands and treated with great reverence is a kind of sociopragmatic competence.

d) **Strategic Competence**: This refers to a speaker's ability to exploit verbal or non-verbal communication strategies when communication problems arise, compensating for deficiencies in other competences. These include four common types:

    i) *Avoidance or reduction strategies* such as topic avoidance or message abandonment to try to keep conversation inside areas where the speaker feels in control;

ii) *Compensatory strategies* such as circumlocution or mime when a word is not known;

iii) *Stalling strategies* such as using hesitation devices or repetition to hold the turn in conversation while a message is formulated;

iv) *Interactional strategies* such as asking for repetition or clarification where the speaker makes use of the linguistic resources of other interlocutors to maintain conversation.

e) **Discourse Competence**: This refers to a speaker's ability to produce unified, cohesive and coherent spoken or written discourse of different genres (Halliday & Hasan 1989). In writing this might include the knowledge of the correct layout for a letter or how to use anaphoric reference in a text. In speaking it would include how to develop a conversation naturally through 'topic shading' where a sub-topic from preceding talk is taken up and expanded into the main topic (Crow 1983; Bublitz 1988). It could also include knowledge of different generic structures such as narratives, gossip or jokes (Eggins & Slade 1997) or discourse intonation (Brazil, Coulthard & Johns 1980).

1.3 Alternative views of communicative competence

Things aren't quite as simple as this, however. Different writers have, confusingly, created their own terminology for essential the same components, for example pragmalinguistic competence is called 'sociolinguistic competence' by van Ek (1986), 'illocutionary competence' by Bachman (1990) and 'actional competence' by Celce-

Murcia et al. (1995). This inability to settle on terms is perhaps indicative of the state of flux this area of research is currently in but it hinders comparison and is rather unhelpful.

Some researchers have added extra components or merged existing ones in their models. van Ek's 'framework for comprehensive foreign language learning objectives' (1986: 33) includes a sixth component, 'social competence', which he defines as 'the will and the skill to interact with others, involving motivation, attitude, self-confidence, empathy and the ability to handle social situations' (ibid: 65). Personally, I believe these areas of a speaker's ability to be qualitatively different from those indicated above and outside of a model of communicative competence. They relate too closely to an individual's personality rather than a body of knowledge that can be used to shape a syllabus. That is not to say that these factors don't play an important role in language learning, of course. Cortazzi & Jin (1999) suggest adding 'intercultural competence' to the four areas proposed by Canale & Swain (1980) and Canale (1983):

> […] intercultural competence is seen in social effectiveness (i.e., the ability to achieve instrumental and social goals) and appropriateness (i.e., suitable communication in a given situation in a particular culture) (Martin, 1993). (Cortazzi & Jin 1999: 198)

However, it seems to me that these areas of competence are already included in the model proposed in section 1.2: 'social effectiveness' requires the deployment of all five types of competence and 'appropriateness' corresponds to pragmalinguistic and sociopragmatic competence. The idea of intercultural competence is important though and we shall return to this later.

With respect to merging different areas of the model, Hall (1999: 137) uses the term 'interactional competence' which appears to combine pragmalinguistic, sociopragmatic and discourse competences into one component. This, I also regard as unhelpful since I believe these components to be distinct enough from each other to warrant separate names. In addition, it further confuses what is already a complex area through the introduction of yet another term.

1.4 Roles and relative importance of different components in language learning

What we should be looking for now is a general consensus in the field on the types of communicative competence which exist and appropriate terminology to describe them so that we can move forward in our understanding of how the various components interact with each other and their relative importance in language learning.

These remain areas where very little research has been done to date. Bachman & Palmer (1982) developed a battery of tests to measure (using their terminology) 'grammatical competence' (morphology and syntax), 'pragmatic competence' (vocabulary, cohesion and organization), and 'sociolinguistic competence' (sensitivity to register, naturalness and cultural references). They found that grammatical and pragmatic competences were closely associated with each other while sociolinguistic competence was distinct. However, the components they included within pragmatic competence are more commonly associated with linguistic or discourse competence so their results are difficult to interpret within the framework proposed here. Allen et al. (1988) found no significant differences between grammatical competence (morphology and syntax), discourse competence (cohesion and coherence) and sociolinguistic competence

(sensitivity to register) in their study using factor analysis of test scores. On the other hand, Schmidt's (1983) three-year longitudinal study of the development of communicative competence in Wes, a Japanese artist living in Hawaii, found that his discourse and pragmatic competence developed significantly while his grammatical knowledge changed very little, suggesting that these components *are* distinct from each other. More recent investigations have focused principally on linguistic versus pragmalinguistic competence and confirm the distinctness of these two components (for example, Bardovi-Harlig & Dörnyei 1998; Kasper 2001a/b; Salsbury & Bardovi-Harlig 2001).

Turning to discourse competence, Hatch (1978) and Day (1986) note, interestingly, that most researchers into both first and second language acquisition assume that linguistic competence precedes discourse competence but that the opposite might in actual fact be true:

> One learns how to do conversation, one learns how to interact verbally, and out of this interaction syntactic structures are developed. (Hatch 1978: 404)
>
> In interactions, the discourse frames, the scripts for interactions, develop: the language appropriate to the interaction builds on this development; and the language, in turn, refines the frame. (Day 1986: 6)

With respect to the relative importance of different areas of competence, Celce-Murcia et al. (1995) see discourse competence as playing a central role:

> Discourse competence concerns the selection, sequencing, and arrangement of words, structures, sentences and utterances to achieve a unified spoken or written text. This is where the bottom-up lexico-grammatical

microlevel intersects with the top-down signals of the macrolevel of communicative intent and sociocultural context to express attitudes and messages, and to create texts. (ibid: 13)

Bachman (1990: 103), on the other hand, expands the notion of strategic competence and sees it as mediating between the communicative goal in a given situation and the language resources available to a speaker. It assesses how best the speaker's competencies can be exploited to achieve the communicative goal, retrieves the relevant items and plans the execution of the message. It then assesses how well the goal has been achieved. This seems remarkably similar to the 'core role' assigned to discourse competence by Celce-Murcia et al. (1995).

In summary, although there is some general agreement as to the types of competence native speakers have at their disposal to achieve communicative goals, there is little consensus over how distinct they are from each other, how they interact with each other or their relative importance in successful communication. Byram (1997: 10), following van Ek (1986), sees the components of communicative competence as different aspects of the same concept. We can focus on one particular component of the model but it can never be completely understood in isolation: 'At any one point, one aspect will be central but others, and their relationship to that aspect, will also be in view.'

In terms of language learning, however, this state of flux in the theoretical aspects of the communicative competence model need not concern us overly. All of the components identified in the model above are clearly important to learners if they want to communicate successfully in the L2. Using this model as a starting point, it is clear that current ELT materials are skewed in favour of linguistic competence at the expense of the other four types of competence and I believe that a solution to this imbalance is well

overdue. Celce-Murcia, Dörnyei & Thurrell's (1995) pedagogically motivated model of communicative competence, in particular, provides a useful starting point for teachers, material writers and language testers who wish to address this issue in the classroom.

1.5 Beyond the communicative competence model

Communicative competence, as it has been outlined so far, has recently been criticised by a number of writers because it models itself on educated native speakers and takes their communicative competence as the ultimate goal of foreign language learning. This is problematic for a number of reasons. Firstly, there is the difficulty of defining native speaker norms 'in a time of large-scale migrations, cross-national and cross-cultural encounters, and increasing linguistic and pragmatic differences among speakers of the same language' (Kramsch 1998b: 16). Even if we are able to agree on what constitutes native speaker competence, many question how appropriate this model is to learners, both because it sets the impossible target of becoming like a native speaker, something which could potentially de-motivate learners and which devalues the social identity and competences they have developed within their own culture (Byram 1997), and because the communicative needs of non-native speakers (NNSs) are very different from native speakers (NSs) existing in a particular speech community and vary according to the social context in which they wish to operate (Saville-Troike 1989). In place of this native speaker communicative competence, Byram & Fleming (1998: 12) propose a model based around intercultural communicative competence (ICC):

> Instead of the assumption that learners should model themselves on 'the native speaker', it is becoming apparent to teachers and their learners that successful cross-cultural communication depends on the

acquisition of abilities to understand different modes of thinking and living, as they are embodied in the language to be learnt, and to reconcile or mediate between different modes present in any specific interaction. This is not the 'communicative competence' on which people using the same language in the same, or closely related, cultures rely; it is an 'intercultural communicative competence' which has some common ground with communicative competence, but which also has many unique characteristics.

Thus, rather than expecting learners to abandon their own social identities and communicative competencies in an attempt to replicate some native speaker ideal, ICC emphasises the knowledge and skills needed to understand people from other, unfamiliar cultures and to mediate between the foreign culture and the learner's own culture in a way that leads to successful communication (see also Lustig & Koester 1993). This is a variety of communicative competence that many native speakers, particularly those with limited experience of 'otherness', tend to lack. As Byram & Fleming (ibid) suggest, learners may still want to acquire many of the aspects of native speaker communicative competence but with the goal of mediating between disparate cultures rather than complete integration into a particular community.

1.6 The representation of communicative competence in language learning materials

It has long been recognised that the language presented to students in textbooks is a poor representation of the real thing, 'far away from that real, informal kind of English which is used very much more than any other during a normal speaking lifetime' (Crystal & Davy 1975: 2) and although more recently much has been done to redress the balance, there remain numerous gaps. Research into different areas of communicative competence through discourse or conversational analysis, pragmatics and sociolinguistics has

exploded and, with our deepening understanding of how people make meaning through language, it has become clear that it is time for a fundamental change in the way we design our syllabuses:

> […] awareness of discourse and a willingness to take on board what a language-as-discourse view implies can only make us better and more efficient syllabus designers, task designers, dialogue-writers, materials adaptors and evaluators of everything we do and handle in the classroom. Above all, the approach we have advocated enables us to be more faithful to what language *is* and what people use it for. The moment one starts to think of language as discourse, the entire landscape changes, usually, for ever. (McCarthy & Carter 1994: 201)

What follows, is a review of some of the relevant research that supports the need for the paradigm shift, alluded to above. It is far from comprehensive but serves to illustrate how inadequate many current language textbooks are in developing learners' overall communicative competence.

1.6.1 *Linguistic competence*

This area of communicative competence, as is well known, has historically dominated foreign language teaching but the linguistic knowledge imparted to learners was largely based on intuitions gleaned from examination of the written form and sentence-based, classical notions of grammar. With the introduction of audio recording technology and, subsequently, the development of procedures to transcribe and analyse authentic spoken language (through discourse, conversation & corpus analysis), much of the focus in applied linguistics has shifted to speech in recent years. It is not surprising, therefore, that

the majority of work in this area of competence focuses on the lack of adequate models for spoken grammar in textbooks.

Holmes (1988) provides data on the relative frequencies of lexical items expressing doubt or certainty in written and spoken corpora and, surveying four well-known ESL textbooks, finds that the more common modal lexical items are often under-represented in comparison to modal verbs (see also McCarthy 1991: 84). This could potentially have serious consequences for learners because of the important pragmatic function of this group of words. Altman (1990), using a ranking test of 7 common modal auxiliaries, found that low-intermediate learners were unable to accurately assess the relative strengths of 'should' and 'had better', judging the former to be much stronger than the latter. This he blames on a bias in textbooks towards linguistic, rather than sociolinguistic, rules. Tannen (1989) examines speakers' use of repetition in conversation and finds it to be a ubiquitous feature. She explains its presence not in terms of some kind of real-time performance limitation, but rather as an important affective tool for creating rapport between people. McCarthy (1991) agrees with this view and, in addition, illustrates how reiteration, or reworking, of previously mentioned lexical items (*relexicalisation*), allows for coherent topic development in conversation. This has important implications for the teaching of vocabulary because it assumes that learners need to be 'armed' with a wide variety of hyponyms and synonyms to converse naturally in English, 'using a range of vocabulary that is perhaps wider than the coursebook or materials have allowed for' (ibid: 68). As McCarthy goes on to point out, other languages may not rely on relexicalisation in the same way as English does to develop discourse so learners need to be made aware of this feature. Williams finds, in her 1990 study, that

native speakers of American English and Singaporean English both prefer an invariant SVO order in Yes/No questions when talking casually to close friends or family members. She sees this as a production strategy employed by both groups to avoid semantically redundant syntax and urges teachers and researchers to refer back to authentic data when making judgments on learners' performance, rather than relying on prescriptive notions. Powell's (1992) analysis of spontaneous conversation from the London-Lund corpus finds high frequencies of evaluative, vague, intense or expressive language in informal contexts. This meets the interactional and affective needs of speakers in informal contexts and contrasts sharply with the 'safe, clean, harmonious, benevolent, undisturbed, and PG-rated' world presented to learners in textbooks (Wajnryb 1996: 1). Channell (1994), in her book 'Vague Language', provides the most comprehensive description of linguistic vagueness so far undertaken, arguing that it is a key element in the communicative competence of native speakers and, therefore, has important pedagogical implications. McCarthy & Carter (1994) focus on the evaluative role of idioms in natural language and, as a result, their high occurrence in specific types of discourse (problem-solution or narrative genres) and predictable parts of the discourse. As the authors claim, however, textbooks rarely deal with this language in a systematic way and idioms are often regarded as 'something to tag onto the higher levels or terminal stages of language courses' or, alternatively, 'left to the twilight world of (in publishers' parlance) 'supplementary materials'' (ibid: 109). McCarthy & Carter (1995) present early results on distinctions between spoken and written grammar found in CANCODE (Cambridge and Nottingham Corpus of Discourse in English), a spoken corpus of around 5 million words collected between 1995 and 2000. They show how standard grammars

fail to account for pervasive features in spoken discourse such as ellipsis or 'slots' at the beginnings and ends of clauses ('heads' and 'tails') for speaker orientation/evaluation and stress the importance of an interactive interpretation on verb-form choices in real data. Hughes & McCarthy (1998) argue that sentence-based grammars are inadequate to explain speaker/writer choices at the discourse level. They show, for example, how *it*, *this* and *that*, which are normally not taught together in language pedagogy, frequently operate as alternatives in real discourse. Whereas *it* signals continued, ongoing topics, *this* marks new or significant topics and *that* has a distancing or marginalising function (see also McCarthy & Carter 1994: 91). The discourse grammar approach that they recommend has important implications for the classroom because it relies on learners being presented with longer stretches of text in order to interpret grammar choices made. Wray (2000) (but see also Willis 1990, Lewis 1993, Aijmer 1996) focuses on the importance of *formulaic sequences* (also referred to by others as *idioms*, *collocations* and *sentence frames*) in language learning, stating that even proficient non-native learners have difficulties distinguishing what is natural from what is grammatically possible but non-idiomatic. She blames this on the lack of natural language models in the classroom (despite their common occurrence in television and film) and on the problems teachers have selecting the right formulaic sequences to present, concluding that 'It seems difficult to match in the classroom the 'real world' experience of language.' (ibid: 468). Perhaps this difficulty can most easily be overcome by presenting learners with carefully selected authentic language to work with in the classroom; at least until we understand more about the processes involved in sounding idiomatic in English. Basturkmen (2001) illustrates how learners are often misled by descriptions of questioning found in ELT materials and

argues for authentic texts to be used with higher-level learners to give more realistic models. Shortall (2003) reports that the emphasis in textbooks on adjectival comparatives and superlatives underestimates the importance of the 'noun + *more*' construction for comparing, as illustrated by frequency data from the British National Corpus. Carter & McCarthy (2003) illustrate, with spoken corpus data from CANCODE, how *E*-language (the 'external' language of real-world communication) consistently differs from *I*-language (the language of introspection or Chomsky's ideal speaker-listener). In spoken language, question tags, relative clauses and subject-verb concord often fail to conform to prescriptive descriptions. Their frequency data also highlight the pervasiveness of words such as *like*, the morpheme *–ish*, and response tokens such as *right*, which all play an important affective role in discourse but are rarely taught in the language classroom.

These inadequacies in the way that language is presented to learners in textbooks are not only confined to English: similar results have also been found in French by Walz (cited in Herschensohn 1988), and O'Conner Di Vito (1991). The most comprehensive description of variation in authentic spoken and written English grammar to date is Carter & McCarthy's (2006) 'Cambridge Grammar of English'. This will prove useful to teachers wishing to assess the extent to which the grammar patterns in their classroom input conform to authentic, native speaker norms.

1.6.2 *Pragmalinguistic competence*

There is a substantial body of work available now which points to the lack of appropriate pragmatic models in textbooks (see, for example, Kasper 2001a: 1). This is generally blamed on the fact that material writers have relied on intuitions about language rather

than empirical data and have focused on imparting lexicogrammatical knowledge at the expense of pragmatics.

Pearson (1986) (cited in Salsbury & Bardovi-Harlig 2001) notes that agreement/disagreement speech acts are frequently given equal emphasis in language textbooks, perhaps painting a misleading picture for learners since native speakers are more likely to agree with each other than disagree and frequently employ face-saving strategies when they do disagree. Williams (1988) compared the language used for meetings in authentic business interactions with the language taught for meetings in 30 business English textbooks. She found almost no correspondence between the two, with only 5.2% of the 135 exponents presented in the classroom materials actually occurring in the genuine meetings. She criticises material writers for relying on introspection rather than empirical research when selecting which exponents to present in the classroom. Bardovi-Harlig et al. (1991) surveyed conversational closings in 20 ESL textbooks and found that, despite claims of naturalness or authenticity, the models presented were often only partially complete, with the pre-closing or closing moves missing. They criticise the lack of pragmatic information available to learners in textbook materials. Boxer & Pickering (1995) assess the presentation of complaint speech acts in 7 EFL textbooks, finding that all deal with direct (Ds) rather than indirect complaints (ICs) (in Ds, the addressee is seen as being responsible for the perceived offence whereas in ICs they are not). This is despite the fact that, in normal conversation, ICs are much more common and play an important affective and discoursal role. They give an addressee the opportunity to show rapport by commiserating with the speaker's complaint and open up the subject of 'what's wrong with X' to further topical development. The authors also

criticise the lack of contextualisation in the textbooks examined, without which it is impossible for learners to know in what situations, and with whom, the target language is appropriate. They recommend that material writers rely on spontaneous authentic interaction rather than intuition when creating textbooks in order to better reflect the sociopragmatic norms of a culture. Bouton (1996) provides a useful overview of Nessa Wolfson's work on invitation speech acts in the 1980s in which she identified three types: *unambiguous invitations* which are direct and specify a time, place or activity; *ambiguous invitations* in which the invitation is co-constructed through negotiation by the participants; and *non-negotiable non-invitations*, along the lines of 'We must get together some time', which seem to function as positive politeness strategies rather than actual invitations. Bouton compares the distribution of these 3 types of invitation in naturally occurring language (from Wolfson's data) with 'Say It Naturally' (Wall 1987), which, he believes, provides 'one of the better presentations of this speech act (ibid: 16). The results are dramatically different:

	Wolfson data	**Wall examples**
Unambiguous invitations	26%	80%
Ambiguous invitations	41%	8%
Non-negotiable non-invitations	33%	0%

(Bouton 1996: 17)

The representation of invitations in the textbook clearly gives learners a distorted picture of reality, one that is likely to have serious repercussions on their pragmatic competence. Ambiguous invitations are used in situations where the relationship between speakers is still 'under negotiation'– arguably the most typical scenario to be encountered by NNSs attempting to make friends in a new environment. Learners are also likely to misinterpret non-negotiable non-invitations as genuine if they have never seen them in

the classroom, leading to disappointment or frustration when the offer is not realised. Bouton calls for authors to incorporate far more pragmatic information into their materials, using the wealth of data now available in the research literature. Wajnryb (1996) examines two popular EFL textbooks for the pragmatic features of distance, power or face threatening acts (FTAs) between speakers – factors that effect what kind of language is appropriate in a given situation. She finds 67% of exchanges in the textbooks are between speakers where there is high social distance and this means that the language used tends to be explicit and textually coded because of the lack of shared knowledge between interlocutors. As a consequence, learners may be deprived of examples of the more implicit language used in low social distance discourse, affecting their ability to interpret implicature (see, for example, Bouton 1990/99). Wajnryb reports that, in terms of power, 89.5% of interactions are symmetrical in the textbooks and this limits the examples of negotiation in the scripts, since negotiation is more typical of asymmetrical relationships. Finally, she notes the very low incidence of FTAs in the textbooks and, even when they do occur, the learning opportunity for 'facework' they provide is rarely exploited. Bardovi-Harlig & Dörnyei (1998) compared the ability of ESL/EFL students to recognise grammatical and pragmatic violations in 20 videotaped scenarios with one of three conditions: with grammatical mistakes; with pragmatic mistakes; with no mistakes. They asked subjects to identify whether or not the scenarios contained mistakes and, if they did, how serious they were. While the ESL learners (studying English in the United States) rated the pragmatic mistakes as more serious than the grammatical ones, exactly the opposite pattern was found with the EFL learners (studying in Hungary and Italy).

The authors explain this greater pragmatic awareness in the ESL learners as stemming from the quality of their experience with the L2:

> It seems likely, then, that the pragmatic awareness of the ESL learners may have come from the friction of their daily interactions: the pressure not only of making themselves understood but also of establishing and maintaining smooth relationships with NSs in the host environment. (ibid: 253)

They suggest that EFL students' pragmatic awareness could be improved by increasing the amount of pragmatic input in the classroom and by placing a greater emphasis on this area of communicative competence.

1.6.3 *Sociopragmatic competence*

Much of the work relating to sociopragmatics highlights the wide variation in behavioural norms around the world and the risk of miscommunication, negative stereotyping and conflict when different cultural groups interact with each other. The consequences of sociopragmatic failure are likely to be much more serious than any kind of linguistic problem since, whereas linguistic difficulties are normally recognised as such, sociopragmatic problems due to cultural differences seem to be much harder for participants to identify:

> […] when a listener does not react to a cue or is unaware of its function, interpretations may differ and misunderstanding may occur. It is important to note that when this happens and when a difference in interpretation is brought to a participant's attention, it tends to be seen in attitudinal terms. A speaker is said to be unfriendly, impertinent, rude, uncooperative, or to fail to understand. (Gumperz 1982a: 131)

Scollon & Scollon (1981) describe how Athabaskans, an indigenous North American group, expect the person who initiates a conversation to immediately relinquish the floor to the other speaker, in contrast to the normal conversational rules in English, which allow the initiator to keep the speaking turn (see Schegloff 1968, Schegloff 1972, Sacks, Schegloff & Jefferson 1974). This can result in negative stereotyping within both groups, with the Athabaskans perceiving the English speakers as dominating and the English speakers perceiving the Athabaskans as aloof (see also Clark 1998/9 for more on Native American discourse strategies). Neu (1990) demonstrates how differences in the non-verbal competence of Japanese and Saudi students has significant effects on their overall assessment in oral interviews, suggesting that training in NVC deserves to have a greater role in language learning. Gumperz (1982a) reports that newly employed Indian or Pakistani ladies serving food in the canteen of a British airport were perceived by their supervisor as surly or unhelpful because of their manner and use of falling intonation on questions such as "Gravy?". This was interpreted by NS customers as a statement, "This is gravy", rather than an offer, "Would you like gravy?" and therefore caused offence. Chick (1985, 1989) analysed intercultural encounters in educational and workplace settings between native speakers of English and Zulu in South Africa. He found that mismatches in the frames of reference, back-channeling, turn-taking and politeness behaviour in conversation contributed to the negative perceptions each group often had of the other and encouraged prejudice and stereotyping (see also Bilbow 1997 for similar negative stereotyping resulting from differing interaction styles between ethnic Chinese and 'expatriates' in workplace contexts in Hong Kong). Cultural differences in interaction patterns have also been shown to disadvantage minority groups in other

educational contexts (see Sato 1981, Brice Heath 1983, Johnson 1995). In a wide-ranging survey, Argyle (1988) illustrates cultural variations in non-verbal communication, covering areas such as facial expressions, gesturing, gaze, spatial behaviour and touch. He claims that cultural differences in NVC are, 'a major source of friction, misunderstanding, and annoyance between cultural and national groups' (ibid: 49). Cook-Gumperz & Gumperz (1997) analysed the transcripts from interviews of British and Pakistani workers applying for a further education program in the United Kingdom. They found that the NNS were often unable to recognise the interviewer's questions as cues for them to begin narratives, recounting past accomplishments, which would allow them to demonstrate their suitability for the course. As a result of differing expectations of how the discourse of interviews unfolds, NNS missed opportunities to convey an 'institutional self' in the fashion expected by the interviewers. This is likely to seriously affect their chances of success. Bailey (1997) examined the divergent communicative practises of Korean retailers and their African American customers during service encounters in Los Angeles liquor stores. He found that while the Korean storekeepers preferred a less involved, transactional style in the encounters, their African American customers tended towards a more personal, interactional style of communication (see also Aston 1988 for evidence of a preference for a more interactional style in service encounters in Italy). Bailey argues that both groups are attempting to show politeness in their meetings but whereas the Koreans tend to emphasise negative face, the African Americans emphasise positive face (see Brown & Levinson 1987). These difference styles of interaction often lead to conflict between the two ethnic groups:

The seeming avoidance of involvement on the part of immigrant Koreans is frequently seen by African Americans as the disdain and arrogance of racism. The relative stress on interpersonal involvement among African Americans in service encounters is typically perceived by immigrant Korean retailers as a sign of selfishness, interpersonal imposition, or poor breeding (ibid: 353)

Stubbe (1998) reports on quantitative and qualitative differences in the use of verbal supportive feedback between cultures, focussing principally on the analysis of Maori and Pakeha[3] informal conversation, taken from the Wellington Corpus of Spoken New Zealand English. She found that Maori informants in her sample used approximately a third less verbal feedback than Pakehas, often punctuated by quite long pauses, supporting previous work which suggests that Maori speakers place greater emphasis on non-verbal signals than on verbalisation. She believes that the different styles of communication can cause irritation, confusion or complete misunderstanding, leading to negative attitudes between the two groups which can have serious consequences, particularly for the group with less power in New Zealand society, the Maoris (see also Lehtonen & Sajavaara 1985 on lower rates of back-channeling and interruption in Finnish; Clancy 1982, Lo Castro 1987, White 1989 & Yamada 1997 for evidence of higher rates of back-channeling in Japanese).

Gilmore (1998) contrasts Japanese and English styles of interaction in casual conversation, finding differences in turn-taking, back-channeling, pausing, gaze and gesturing behaviour between the two groups. He discusses the detrimental effects that these differences might have for Japanese learners attempting to participate in English

[3] New Zealanders of British or European extraction.

conversation and suggests that awareness raising and practice of the target culture's interaction style could be of great benefit in the classroom.

Boxer (2002) provides an excellent review of recent research in cross-cultural pragmatics in social, educational and workplace contexts. She concludes that, in a world where the opportunities for people from distinct speech communities to interact are increasing, cross-cultural understanding will become increasingly important.

Many researchers point out that in instances of cross-cultural miscommunication, such as those illustrated above, it is usually the minority group members who have the most to lose, denied access to valued resources or positions by gatekeepers from the dominant group (see, for example, Loveday 1982, Gumperz 1990, Chick 1996, Schiffrin 1996). From the perspective of foreign language learners, who will often find themselves as minorities, being able to understand and adapt towards the sociopragmatic norms of the target culture will therefore be crucial if they hope to gain control of the situations they find themselves in.

But can sociopragmatics be taught? Although many sociolinguists stress the value of teaching cultural differences in order to avoid miscommunication, the evidence in the literature to demonstrate its effectiveness is rather limited. What there is, however, does suggest that when learners are exposed to input illustrating cultural variation in behavioural norms, in conjunction with awareness raising and practice activities, their sociopragmatic competence can be improved.

Collett (1971), in his well-known study, found that it was possible to train Englishmen in the non-verbal behaviour of Arabs. After only 10-15 minutes of specific training on common features of Arab interaction and a single practice run, an experimental group

were found to approximate Arab NVC much more closely than a control group, who had received no training, during a 5 minute interview with Arabic volunteers. Statistically significant differences were found between the two groups in features such as interpersonal distance and levels of eye contact or touching. The different behaviour of the two groups of Englishmen was also found to affect the Arab subjects' attitude to the encounters, with nine out of ten of them stating that they would prefer to: a) share a flat with; b) be friends with; or c) trust Englishmen from the experimental group. These results therefore supported the hypothesis that Arabs prefer Englishmen who behave more like themselves, a finding which is likely to be duplicated across all cultures.

Argyle (1988: 264) describes how training in NVC is a crucial part of social skills training (SST). Professionals such as teachers, doctors and custom officers can be taught to recognise the meaning behind facial expressions or tones of voice and politicians can be taught to give more convincing presentations through vocal and gestural emphasis. Although not directly related to cross-cultural communication, it does suggest that people can learn to understand NVC better and can adapt their own body language to become more effective communicators through awareness-raising and practice activities.

Poyatos (1992) proposes a model for the teaching of NVC (paralanguage and kinesics) in addition to, what he terms, the standard 'verbal language' offered by most textbooks. His methodology essentially involves integrating non-verbal information, organised into categories such as emblems, adaptors and externalizers, with a more traditional, lexicogrammatical syllabus. The NVC would be demonstrated by the teacher or through audiovisual materials and the students given practice of the target behaviour before

finally being assessed on their ability to use it. Poyatos does not, however, provide any evidence that his methodology actually improves learners' NVC skills.

Hurley (1992) discusses the potential of film and television from the target culture to raise students' awareness of differences in pragmatics, prosody and NVC. He suggests that asking students to role-play identical scenes, both before and after viewing may help them to assess how native-like their behaviour is but stresses that the effectiveness of techniques such as these has not been proven with empirically-based classroom research yet (see also Swaffar & Vlatten 1997).

Kellerman (1991) claims that knowledge of the kinesic behaviour of the target language community is a vital component in the development of learners' communicative competence. She argues for exposing learners to video or film in the class, rather than just audio-taped material, but remains uncertain as to whether body language needs to be taught explicitly or systematically.

Scollon (1999) examines the potential of television sitcoms for raising awareness of cultural differences in patterns of conversational openings. He analysed programmes from Hong Kong, Korea, Japan and America and found that they demonstrated variations (in the normally subconscious rules) on how to enter a house/room and how to greet appropriately in each culture. Scollon argues that sitcoms can be effectively employed in the classroom to contrast and compare these differences and, by making learners conscious of them, improves their chances of developing sociopragmatic competence (see also Lebra 1987 on the potential of Japanese and American soap operas to raise awareness of quantitative and qualitative differences in the use of vocalization).

Even if further empirical research into the teaching of a target culture's sociopragmatics shows that it can be effective, it is not something that should be imposed on learners (see, for example, Loveday 1982, Littlewood 1983, Hurley 1992). They might, quite naturally, resist attempts to be re-socialised as Siegal (1996) found with white women in Japan studying Japanese who avoided accepted behavioural norms which they considered demeaning or silly, such as speaking in a high pitch on the telephone or using humble forms of the language. The most sensible approach would seem to be to raise students' awareness of important cultural differences, to illustrate what can happen if they are flouted and then to let learners decide for themselves how far they wish to adapt their behaviour:

> […] sociopragmatic rules and conventions of the L2 should not be prescriptively taught – i.e. that the learner should not be required to conform to them – but that he should as far as possible develop a 'metapragmatic awareness' which enables him to avoid patterns detrimental to personal relationships, while maintaining the maximum freedom in expressing his own personality. (Aston 1988: 36)

A further issue raised by Chick (1996) is that sociopragmatic rules vary to such a degree, depending on the context and participants involved, that it would be difficult to select what body of knowledge it would be appropriate to include in a language course. This focus on the variability of language in its social context, depending on constraints in *field*, *tenor* and *mode* (Halliday 1978), is a common theme running through all the research into communicative competence. The solution, in terms of developing sociopragmatic competence, would seem to be to concentrate on giving learners the tools

they need to interpret different behaviour they encounter, to analyse more objectively and critically the reasons for miscommunication through an ethnographic approach:

> […] to help students inhabit a more uncertain and critical world in which their own assumptions and stereotypes are challenged by the constant questioning of the sources of evidence presented to them. As linguists, their jobs and interests are likely to involve them in multilingual worlds where their whole social being must interact with others. They need the cultural tools for making sense of new intercultural contacts and experiences rather than positivistic facts about other countries, structures and systems which are, despite the textbooks' attempts to freeze-dry them and turn them into fresh-looking digestible items of information, constantly in a process of contestation and change. (Barro, Jordan & Roberts 1998: 97)

1.6.4 *Strategic competence*

In language learning, strategic competence is generally understood to refer to the use of verbal or non-verbal communication strategies (CSs) by learners to compensate for L2 deficiencies or other language-related problems. Although it has been assigned a broader definition by some researchers as 'the overall ability of a speaker to enhance the effectiveness of communication' (Canale 1983; Tarone & Yule 1987), this discussion will limit itself to the narrower definition.

Since the 1970s, when research onto CSs first began, a variety of taxonomies and terms have been proposed to describe them (see Dörnyei & Scott 1997 for a review). Most of these classifications were derived from an analysis of the surface structures used by native or non-native speakers when communication breakdown occurs, such as circumlocution or appeals for help, but these have been criticised more recently as having dubious validity from a psycholinguistic standpoint (for example, Bialystok 1990; Kellerman 1991). More recent taxonomies have attempted to link CS use to Levelt's

(1989; 1993; 1995) cognitive model of speech production (see Poulisse 1993; Dörnyei & Kormos 1998) which, ultimately, should lead to a deeper understanding of the processes at work, merging insights from linguistics and psycholinguistics into a single coherent and scientifically valid model.

From the perspective of language pedagogy, what is of more immediate concern than the ongoing theoretical debate is the practical implications of communicative strategies for learners. As Dörnyei & Kormos (1998: 350) point out, alluding to work by Gass & Varonis 1991, a significant amount of L2 communication is problematic:

> […] even a brief analysis of any spontaneous piece of L2 oral discourse will reveal that L2 speakers tend to spend a great deal of time and effort negotiating meaning and struggling to cope with the various problems they encounter during the course of communication.

Despite this, language textbooks tend to portray conversation as smooth and problem-free (Wajnryb 1996; Carter 1998) and CSs are rarely represented in input or made the focus of explicit teaching (although they do appear more in supplementary materials, for example, Nolasco & Arthur 1987; Ellis & Sinclair 1989; Dörnyei & Thurrell 1992). Given that a 'core group of specific strategies' (Bialystok 1990: 61) emerge from the various taxonomies proposed in the literature, it would seem worthwhile investigating the teachability of these CSs and their usefulness to learners.

But can communicative strategies be taught? Some researchers (for example, Bialystok 1990; Kellerman 1991) argue that since FL learners are already competent users of CSs in their L1, these skills should be readily transferable into the L2 and therefore it is language, rather than strategies, that needs to be taught. Others, see a place for

awareness-raising tasks, the teaching of L2 realisations of CSs, and practice activities (Brown 1979; Canale 1983; Faerch & Kasper 1983; Willems 1987; Tarone & Yule 1989; Rost & Ross 1991; Dörnyei 1995; Gilmore 2005). Certainly, there are strong pedagogical justifications for this. Since CS use is normally a subconscious process, an explicit focus on them might allow learners to notice examples when they occur in input and promote their acquisition (Schmidt 1990). They are composed of a core group of words and structures (Tarone & Yule 1989) which are easy to learn and extremely versatile and are likely to give, particularly lower-level learners, greater confidence in using the L2. Finally, practice manipulating the L2 forms in the classroom might encourage learners to use them spontaneously and fluently in real-time interaction later.

Ultimately, this kind of debate will only be settled through empirical, classroom-based research but that has been limited to date and much of what *has* been done is often narrow in scope or lacking in experimental rigour. Wildner-Bassett (1986) found that an explicit focus on 'fillers' as stalling strategies improved the quantity and quality of learners' use of this type of CS in post-experimental tests. Dörnyei (1995) trained Hungarian high school students, over a six-week period, in the use of three CSs: topic avoidance/replacement, circumlocution and fillers/hesitation devices. Subjects were assessed orally by talking on an abstract topic, describing a cartoon strip, and defining or describing a series of Hungarian words in English. He found that the treatment group improved with respect to the quality of circumlocutions and the quantity of fillers used irrespective of the learners' proficiency level and suggests that strategy training be incorporated early on in language programs. Cohen, Weaver & Li (1996) examined whether instruction in a broad range of language use/learning strategies could improve

the speaking proficiency of French or Norwegian language students at the University of Minnesota. The treatment group received 10 weeks of explicit training in strategy use and were assessed on their speaking ability in three tasks: a self-description, a story re-telling and a city description. The experimental group outperformed the control group (but only on the third task) and the authors conclude that 'explicitly describing, discussing and reinforcing strategies in the classroom can have a direct payoff on student outcomes' (ibid: 29). However, since their speaking test consisted solely of taped monologues, it was not able to capture the negotiation of meaning characteristic of real interaction where CSs are most important. Nakatani (2005) investigated to what extent oral communicative strategies could be taught to Japanese females at a private college in Japan, and whether explicit instruction led to improvements in their communicative ability. The experimental group received 12 weeks of training in CSs and their oral proficiency was tested pre- and post-treatment with a role-play, which was videotaped and rated by native speakers and also transcribed for quantitative analysis. She found that the strategy training significantly improved the oral test scores in the treatment group whereas the control students showed no significant gains. Furthermore, her analysis of the transcribed role-play data revealed how these gains in oral proficiency had been achieved. The experimental group were found to make longer utterances (a measure of increased fluency), used more achievement strategies (which are seen as active attempts to negotiate meaning, by using CSs such as clarification or time-gaining fillers), and fewer reduction strategies (which are seen as attempts to avoid negotiation by using CSs such as message abandonment).

In summary, the classroom-based research into CSs, largely supports the notion that they *can* be taught, and that teaching them has a positive effect on learners' strategic

competence. This may help to build students' confidence to use the L2 and to stay engaged in conversation even when communication problems occur, as they inevitably will. By continuing to attempt to negotiate meaning, rather than abandoning the interaction altogether, learners are likely to expose themselves to greater quantities of comprehensible input and acquire more language:

> […] there could be considerable gains in teaching learners how to compensate for insufficient linguistic resources by using the totality of their communicative resources creatively and appropriately. (Faerch & Kasper 1983: 55/56)

This is likely to be particularly important at lower proficiency levels where communication failure is more common and, since Dörnyei (1995) has shown that CSs can be successfully taught to even pre-intermediate students, it would seem sensible to incorporate them early into the curriculum. At present, however, strategic competence is rarely the focus of much attention in textbooks and their largely contrived dialogues tend to portray a world of problem-free interaction, resulting in input that has few examples of CSs for students to learn from. Authentic interactions involving NNSs are clearly the best source to examine for natural examples of CSs and, carefully selected, these could provide the input needed to develop strategic competence.

1.6.5 *Discourse competence*

Historically, FLT has principally been concerned with static, sentence-level descriptions of language and has paid scant attention to the social context in which it is produced. This resulted in such teaching practices as the Grammar-Translation Method where students were offered isolated sentences of dubious authenticity to learn from (although, to be fair,

literature was also considered important in this particular methodology), such as Henry Sweet's favourite example, 'The philosopher pulled the lower jaw of the hen' (Howatt 1984: 145). Discourse analysis brought with it an awareness of the higher order patterns in text and an appreciation of the dynamic and interactive nature of language (McCarthy & Carter 1994), out of which the notion of discourse competence emerged. This ability to produce unified, cohesive and coherent spoken or written texts is a critical part of learners' overall communicative competence.

Textbooks have, without doubt, been more effective in improving learners' discourse competence in the written mode than the spoken. Introducing authentic examples of written genres appears to be less problematic for material writers, and larger patterns in the discourse of, for example, narrative and exposition texts or personal and business letters are commonly highlighted. For this reason, the discussion here will mainly focus on spoken discourse, in particular three areas which are noticeable by their absence in textbooks: conversational management, spoken genres and discourse intonation.

(a) Managing conversation

Conversation, in particular casual conversation, is of critical importance to us as social beings. We spend a large part of our lives using language to make conversation and it plays a significant role in defining who we are, something that goes widely unrecognised:

> […] despite its sometimes aimless appearance and apparently trivial content, casual conversation is, in fact, a highly structured, functionally motivated, semantic activity. Motivated by interpersonal needs continually to establish who we are, how we relate to others and what we think of how the world is, casual conversation is a critical linguistic site for the negotiation of such important dimensions of our social

identity as gender, generational location, sexuality, social class membership, ethnicity, and subcultural and group affiliations […] (it) is concerned with the joint construction of social reality. (Eggins & Slade 1997: 6)

Although Communicative Language Teaching methodology has placed a lot more emphasis on conversation and negotiating meaning than earlier methodologies, the general trend has been to 'practice talking' rather than giving learners specific advice to help them manage conversation in the L2. While time spent talking in the classroom is likely to benefit students' fluency**,** it ignores the wealth of knowledge from discourse and conversational analysis about how we 'do talk' that has potential pedagogic applications.

In speech, the ability to produce coherent and cohesive discourse depends on a number of factors such as taking turns appropriately, maintaining topical continuity or developing new topics, making appropriate listener responses and opening or closing down the talk effectively (see Wardaugh 1985 for a useful overview). All of this takes place, 'under the communicative stress […] of real time processing' (Stubbs 1983: 36).

The systematicity of turn-taking in English was first pointed out by Sacks, Schegloff & Jefferson (1974), who identified some general features spanning all contexts and genres:

a) Typically, only one party talks at a time.
b) Occurrences of more than one participant speaking at any time are common but brief.
c) Transitions from one turn to the next with no (or slight) gap and no (or slight) overlap, make up the majority.
d) Current speakers may select the next speaker or participants may self-select.
e) Repair mechanisms exist for dealing with breakdowns in the turn-taking system.

f) Turn changes generally occur at transition relevance places (TRPs); points in a conversation when syntactic, semantic, visual and prosodic clues come together to indicate that a speaker is willing to relinquish the floor.

Houtkoop & Mazeland (1985: 597), however, pointed out some inadequacies in this model in terms of its ability to describe what they term *discourse units*, points in conversation when turn-taking is suspended and larger projects such as jokes, narratives or descriptions are produced:

> […] it appears that not all turn-taking behaviour can be explained in terms of Sacks et al.'s model. There are stretches of talk where the completion points of turn-contructional units do not present themselves as transition relevance places ie. as opportunities for turn-transfer […] A speaker who is producing such a larger project not only has the right to take a turn which is constructed out of more syntactical units, but also has the right to take as many turns as necessary to finish the project.

Slade (1997) refers to these characteristics of conversation with the catchier terms *chat* and *chunks*, where chat segments are managed locally, turn by turn, in the manner described by Sacks et al., while chunks, which have a global or macro structure, are 'triggered' by themes arising from the chat.

What conversation is 'about' at any point in its co-construction by participants is referred to as *topic*. Although problematic to define or identify precisely, conversationalists pay close attention to the developing topic, even if they are not consciously aware of doing so:

> That a speaker can, without any difficulty, sum up the topic of a conversation in a single sentence, suggests that native speakers have an intuitive awareness of topic. (Gardner 1987: 132)

This is important to maintain the coherence of a conversation, in accordance with Grice's (1975) co-operative principle:

> Speakers do not cast their utterances at random into the stream of conversation. Rather they link them deliberately with the succession of contributions from all participants and place them at certain points rather than at others. Above all they ensure that their utterances form a contribution to the topic at hand and so can be understood by other participants as 'topically coherent'. (Bublitz 1988: 26)

Crow (1983) suggests that, there are four ways for speakers to institute topic shifts:

a) *Coherent shifts* where a sub-topic in the preceding talk is taken up as the main topic (topic shading).
b) *Renewal* where there is a shift back to earlier topics in the conversation, often marked explicitly by discourse markers such as 'Anyway, going back to…'
c) *Non-coherent shifts* where topic connection is not maintained.
d) *Inserts* where abrupt shifts do not succeed in gaining the topical floor.

Not surprisingly, he finds that coherent shifts are 'the structurally preferred mechanism for topic change within conversation' (ibid: 142).

Bublitz (1988: 16/17) sees topic management as one of the principle methods for speakers to exert control over conversation and move it in directions that suit their own needs or goals:

> […] the participants agree on the discourse topic and the topical actions connected with it either explicitly or, as is the rule, implicitly […] it is normally the privilege of the primary speaker to develop discourse topic, and, if not to initiate it, then certainly to perform topical actions (such as changing the topic, closing

the topic, digressing from the topic etc.). The secondary speaker can violate these principles in many ways, both subtle and obvious, in order to manipulate topics, topical actions and topic talk thus steering the conversation into a direction which does not accord with the direction previously determined by his interlocutor's contributions.

Speakers frequently use *reactive tokens* (RTs) to help weave their turns together and 'grease the wheels of conversation' (Tottie 1991: 255), as they co-construct discourse. Clancy et al. (1996: 355) define a reactive token as:

A short utterance produced by an interlocutor who is playing a listener's role during the other interlocutor's speakership. That is Reactive Tokens will normally not disrupt the primary speaker's speakership, and do not themselves claim the floor.

RTs can fulfil several, simultaneous roles in discourse, taking on a 'supportive function' to signal agreement or understanding and a 'regulative function' to encourage the primary speaker to continue with the turn (see Gardner 1998 for a discussion of *yeah*, *Mm hm* & *Mm*). Clancy et al. (1996) categorise RTs into five types:

a) *Back-channels* are non-lexical, vocalic forms, serving as continuers or claims of interest/understanding.
b) *Reactive expressions* are short, non-floor taking, lexical words or phrases produced by the non-primary speaker.
c) *Collaborative finishes* are completions of the primary speaker's utterances by the non-primary speaker.
d) *Repetitions* are echoes of parts of the primary speaker's utterances by the non-primary speaker.

e) *Resumptive openers* are back-channels used by the non-primary speaker and immediately followed by a full turn, registering the prior turn before taking the floor.

Reactive tokens can be viewed from a different perspective as a type of *discourse marker* (see, for example, Schiffrin 1987; Fraser 1993; McCarthy & Carter 1994; Eggins & Slade 1997). This important class of words (e.g. *oh*, *right*, *well*) and phrases (e.g. *I know*, *I mean*, *for a start*) serve a critical role in maintaining the coherence and cohesion of text by organising, monitoring and managing the discourse. Carter & McCarthy (2006) give a useful overview of the forms and functions of discourse markers, noting the important affective role they play as well as providing a means to exercise control in conversation.

Visual signals also play a key role in managing conversation in English, as Duncan & Niederehe (1974) pointed out in their important paper, 'On signalling that it's your turn to speak'. They list four behavioural cues that tend to occur at the beginning of a speaker's turn:

a) A shift away in head direction.
b) Audible inhalation.
c) Initiation of gesticulation.
d) Paralinguistic over-loudness.

Learners can potentially benefit from being made aware of, and practising, these features of conversational management, particularly when the L1 conversational patterns differ markedly from those of the target language. The evidence that different cultures *do* manage conversation in divergent ways is growing, although at present it is limited to

only a few ethnic groups. Sakamoto & Naotsuka (1982) use the metaphors of tennis and bowling to describe the different styles of communication in English and Japanese, respectively. English conversation is seen as being livelier, with the floor up for grabs by anyone quick enough to get it and participants working together to keep the same ball (topic) going. Japanese conversation, on the other hand, is seen as a gentler affair, with each participant taking a turn in a regulated manner and with no obligation to continue with the same topic as the previous floor-holder. The author, an American married to a Japanese man, found herself experiencing problems as a result of these conflicting styles of interaction:

> No wonder everyone looked startled when I took part in Japanese conversations. I paid no attention to whose turn it was, and kept snatching the ball halfway down the alley and throwing it back to the bowler. Of course the conversation died. I was playing the wrong game. (ibid: 83)

Clancy (1982) and Clancy et al. (1996) comment on the highly affect-laden, fragmented style of Japanese conversation, which contains pauses at the end of most tone units to allow for the involvement of all participants and the verbal encouragement (aizuchi) of the floor-holder:

> […] a speaker tends to communicate information gradually, unintrusively, allowing plenty of opportunity for the listener to assimilate the input, ask questions if necessary, and indicate comprehension. The listener, in turn, is concerned that the speaker feels supported, understood, and appreciated, and times feedback so that at the conclusion of each separate unit of production, the speaker is reassured of the listener's continuing co-operation and participation in the interaction. (Clancy 1982: 76)

The high levels of back-channeling in Japanese have also been noted by many other researchers (see, for example Lo Castro 1987; Yamada 1997; Gilmore 1998; Iwasaki & Horie 1998). Lo Castro (ibid) points out that, since it is considered rude to interrupt a speaker or to ask for clarification or repetition directly, auditors may express their wishes indirectly by ceasing to back-channel.

Ulijn & Xiangling (1995) investigated turn-taking behaviour during intercultural business negotiations involving Chinese, Finnish and Dutch participants. They found that Chinese interrupted more, and in a more marked way, than the other two nationalities and suggest that, although this might be perceived as impolite behaviour by others, it is probably meant to indicate an eagerness to do business by the Chinese participants.

Strauss & Kawanishi (1996) compared the conversational patterns of Japanese, Koreans and Americans, talking on the same topic: their experiences during the 'Northridge Earthquake' in Los Angeles, 1994. The authors noted marked differences between the groups, with the Japanese discourse having much longer stretches of simultaneous talk with a higher frequency of assessment tokens (RTs) than the other two nationalities. Iwasaki & Horie (1998) used the same research design as Strauss & Kawanishi to compare Japanese and Thai conversational behaviour. They found that the Japanese preferred to construct the floor cooperatively, without creating conflict while the Thais were more self-assertive and made more moves to control topic development. They explain these different styles of interaction in terms of a desire to display mutual dependence in Japanese culture, in contrast to a desire for independence in Thai culture.

Gilmore (1998) investigated the characteristics of dinner-table conversation in Japanese and English and found a wide range of differences in the interaction styles of the two

groups. In the Japanese conversation, gender differences were marked, with the males taking the majority of turns and performing topical actions while the females took on a supportive, listener role with a much higher rate of back-channeling than was found either in the Japanese males or the English data. Pausing inside turns was much more common in the Japanese dinner chat but direct gaze and gesturing were much lower, as might be expected in a 'non-contact culture' (Argyle & Cook 1976).

(b) Generic structures

The ability to recognise or produce unified texts, which is an important part of the definition of discourse competence, suggests that proficient language users have an awareness of written and spoken genres. These larger patterns in discourse have been recognised by a number of researchers, although terminology differs: Schank & Abelson (1977) call them *knowledge structures* or *scripts* and Hatch, Flashner & Hunt (1986), *discourse frames*. Bakhtin (1986: 52) defines spoken genres as:

> […] the typical form of the utterance associated with a particular sphere of communication […] which have therefore developed into 'relatively stable types' in terms of thematic content, style and compositional structure.

They occur across a wide range of contexts such as telephone calls, service encounters, classroom language, language-in-action and casual conversation (McCarthy 1991) and Rost (2002: 126) suggests that learners are exposed to a wide variety of them 'in order to 'develop a feel' for the range of spoken language.' One example that has been the focus of particular attention and is of great potential value to learners is that of oral narratives.

Eggins & Slade (1997) remark on the high frequency of storytelling in their casual conversation data, something that is not surprising given its important role in relationship-building:

> It provides conversationalists with a resource for assessing and confirming affiliations with others […] in stories, values, attitudes and ways of seeing the world are created and represented. (ibid: 229)

Stories normally emerge naturally from chat segments of conversation and when participants recognize that a story has begun, normal turn-taking is suspended while the floor-holder takes the listeners through the various obligatory or optional stages expected: an *abstract*, *orientation*, *complication*, *evaluation*, *resolution* and *coda* (Labov 1972). This, of course, places considerable responsibility on the teller to make their contribution relevant and interesting and competent speakers have a wide variety of techniques at their disposal to do this, such as the use of elaboration, exaggeration or idiomatic language, direct speech, shifts to historic present or progressive forms and the employment of 'story voices' (the assignment of distinctive accents, rhythms and pitches to different characters) (Schiffrin 1981; Polanyi 1982; Longacre 1983; Polanyi 1985; Yule 1995). Rintell (1990) reports that learners have considerable difficulties recognising and producing these characteristics of oral narratives.

Proficiency in written and spoken genres is important for learners for a number of reasons. Firstly, it allows them to employ effective, top-down, reading or listening strategies, predicting what is likely to come next and focusing their attention on critical points in the discourse (Brown & Yule 1983; Rost 2002). For example, in a presentation, knowledge of the typical patterns of this genre would encourage particular attention on

the introduction where the speaker is likely to outline the entire talk. Secondly, it allows learners to frame their own discourse in expected ways in the L2, easing the task of comprehension for the target group and giving NNSs control over 'genres of power' which can profoundly affect their acceptance and success within the target community (Clyne 1981; Alexander 1990; Hyland 2002):

> The teaching of key genres is seen as a means of helping learners gain access to ways of communicating that have accrued cultural capital in particular communities […] The central notion here is that students stand more chance of success in a transparent curriculum which makes the genres of power visible and attainable through explicit induction. (Hyland 2002: 125)

(c) **Discourse intonation**

Just as Hughes & McCarthy (1998) argue that sentence-based grammars are inadequate to explain speaker/writer choices at the discourse level, traditional descriptions of intonation based on attitude or grammar fail to account for the turn-by-turn prosodic decisions speakers make in authentic interaction to structure, highlight and background information or to indicate convergence or divergence. Once authentic discourse becomes the focus of a syllabus, the critical role that intonation plays in making meaning needs to be accounted for:

> In general, the further one's interests move towards some notion of *communicative competence* and away from the lesser ability to produce and understand grammatical sentences, the greater the pressure one feels to take proper account of how intonation contributes to the communicative value of an act of speech. One also begins to realise more and more that engagement with intonation is not merely a cosmetic exercise, concerned with the removal of residual, and comparatively unimportant, marks of foreignness in the

otherwise competent utterances of an advanced learner, but that in fact it leads one to a consideration of some quite fundamental aspects of the communicative process. (Brazil, Coulthard & Johns 1980: xiii)

Although our understanding of discourse intonation is incomplete, systematic descriptions do exist for the English language (see for example, Brazil, Coulthard & Johns 1980; Brazil 1985; Bradford 1988; McCarthy 1991; Brazil 1995; Clennell 1997; Rost 2002) and have the potential to help learners develop their discourse competence.

The first characteristic worthy of focus is the way proficient speakers package up their utterances into digestible *intonation units*, or *tone groups*. These reflect the real-time planning constraints on speakers as they formulate their ideas moment by moment and probably also coincide with the receiver's short-term memory processing limitations (Chafe 1992, cited in Rost 2002). Rost (1990) asserts that NNSs have often internalised inappropriate 'metrical templates' (schema for the prosodic organisation of the target language) and one reason for this may be that spoken language is normally represented with the punctuation of the written form in textbooks. As Brazil (1995) points out, representing speech organised into intonation units is likely to give the listener a better sense of the original discourse and may help learners to develop appropriate target language prosodic patterns. Compare, for example, two representations of the same oral narrative (see Brazil 1995: 100 for the original version):

Example 1:

She'd been standing in the car park and it was freezing cold. And she asked her to take her round to her daughter's so she agreed to take her round. What else could she do? She couldn't leave her standing in this car park.

Example 2:

//she'd been STANding in the CAR park//

//and it was FREEZing COLD//

//and she asked her to TAKE her round to her DAUGHTer's//

//so she aGREED to take her round//

//WHAT else could she DO//

//she COULDn't leave her STANDing//

//in this CAR park//

The second model provides valuable information for the learner on how to parcel up the discourse naturally, where to pause and which words to emphasise. It may also improve their conversation management skills by helping them to recognise transition relevance places (TRPs), to co-construct discourse by weaving together intonation units and to use reactive tokens (RTs) in appropriate places. Lastly, it may help to highlight other features of natural speech such as the tendency to chain together utterances with simple discourse markers like *and*, *but*, or *so* (Wardaugh 1985) by positioning these lexical items at the front of each tone group where they are more likely to be noticed as salient.

What information is chosen for emphasis, or *prominence*, marks another level of discourse choice available to the speaker. Often, it is words carrying the informational load (*content words*), which are highlighted. This can be seen when the prominent words from the narrative above are extracted without losing the sense of the original: standing/car park/ freezing cold/ take daughter's/ agreed/ what do/ couldn't standing/ car park. This is useful for learners because, in terms of reception, it can provide the basis for successful listening strategies, focusing on stressed words. In terms of production, it can ease the strain on NSs of comprehension of NNS talk by ensuring that the content words

stand out in the flow of language. However, speakers may also choose to highlight non-preferred words when the interaction demands it, so that, for the example narrative above the following three deliveries of the opening utterance are all possible in other contexts:

a) //SHE'd been standing in the CAR park//

b) //she HAD been standing in the CAR park//

c) //she'd been standing IN the CAR park//

Learners need to develop sensitivity to the discourse choices available to them for assigning prominence in intonation units and the meanings associated with different patterns of emphasis. This is likely to be particularly difficult for learners from cultures where prosody does not have a highlighting function, such as Japan where syntactical changes are employed in place of intonation to create the same effect.

Clennell (1997: 122) points out the potentially serious consequences for learners of inappropriate prosodic choices. For example, a student in a library who says: //excuse ME// you have GIVEn me// the WRONG form//, rather than the preferred: //exCUSE me// you've given me the wrong FORM//, is likely to cause offence. As we saw with other kinds of sociopragmatic failure, when sub-conscious cultural norms are flouted in this way, they often lead to negative stereotyping, both parties considering the other rude (see, for example Gumperz 1982). Learners are unlikely to discover the source of their communication problems in genuine encounters such as the one above, so awareness raising of the use of prominence to mark topic salience would seem to be best tackled in the classroom.

Tone choice represents yet another prosodic decision which speakers need to make in genuine interactions. Within an intonation unit, the last prominent syllable, known as the *tonic syllable* (Halliday 1985), is where the pitch movement begins and, in English, this has communicative significance. It is generally agreed that there are five significant tones in English: the rise, fall-rise, fall, rise-fall and level. However, the fall-rise and the fall, termed the *referring tones* and *proclaiming tones* by Brazil (1985), seem to occur most frequently and play an important role in maintaining the coherence of the on-going discourse so it is these two tones that have the greatest pedagogic significance (Bradford 1988). The fall-rise is used by speakers to convey information that is considered to be common ground with the hearer – something they already know or have some experience of. The fall, in contrast, is used to tell the hearer something the speaker believes is new information. Rost (2002: 34) claims that speakers generally maintain a pattern of two intonation units with referring tones, followed by one with a proclaiming tone, thus situating new information in a context where it can be readily understood by the listener. Learners, therefore, have to continually monitor the 'state of play' (McCarthy 1991: 109) of the discourse and mark information as new or given depending on the specific context and participants. Bradford (1988: 16) provides useful practice activities to encourage students to begin making these prosodic choices:

In this example the same words are used as responses in two different contexts. First listen, and then practise making the responses in the two different ways.

A: There's a very good fish restaurant where we could have dinner tonight.

B: // ↘↗ I had FISH // ↘ for LUNCH //

A: We won't have time to eat later. So I hope you've had something already.

B: // ↘ I had FISH // ↗ for LUNCH //

Now go on. Use the same words to make suitable responses in the two different contexts.

Unfortunately, this kind of discourse competence has been widely ignored by material writers and publishers.

The final principle intonation choice available to speakers as they communicate is *key*, which refers to a noticeable change in pitch level on a prominent syllable in the flow of discourse. Brazil (1985) sees speakers as having three choices for key: *mid*, *high* or *low*. The mid key comes in the middle of a speaker's pitch range and if this is used, simply marks continuance of the ongoing discourse. High key is used for utterances which the speaker considers to be contrary to the hearer's expectations (contrast stress). Low key is used for reiteration or asides and thus has a 'backgrounding' effect. McCarthy (1991) points out that key also plays an important role in topic management, with high key used to begin new topics and low key to end them (see also the work on *paratones* by Brown & Yule 1983b).

For students to learn how to manage conversation effectively in the target language, they need to have realistic models of proficient users doing the same thing, as Brown & Yule (1983: 52) pointed out over twenty years ago. In terms of conversation management, the kind of talk requiring the most work by participants, and therefore also providing the best model to develop this aspect of discourse competence, is casual conversation but this is largely ignored by textbooks, perhaps because it is seen as unstructured and, as a result, unteachable (Eggins & Slade 1997: 315). Language teaching materials tend to concentrate on monologues or dialogues where turn-taking is

structured and predictable, with some kind of transactional goal. More interactional, non goal-oriented language, used to develop relationships, is much less common and it is hardly surprising, therefore, to find that learners experience more difficulties with this kind of talk. Belton (1988) found that advanced Italian NNSs of English displayed 'virtually native speaker competence' on transactional tasks but 'striking dissimilarities' with NS talk on interactional tasks and blames this on the predominantly transactional input and tasks of EFL materials. Authentic recordings of casual conversation are the most likely source of useful models to illustrate how proficient speakers effectively manage discourse and build relationships, employing a range of strategies such as recognising transition relevance places (TRPs) where they can appropriately make a bid for the floor (Sacks, Schegloff & Jefferson 1974), employing 'topic shading' to ensure that their turns are coherent with preceding talk (Crow 1983; Bublitz 1988), making subtle topical moves which move the conversation in a direction to suit their own goals, using reactive tokens to empathise (Clancy et al. 1996) and discourse markers to signal how their turns relate to the ongoing conversation (Schiffrin 1987; Carter & McCarthy 2006). Once learners are aware of these strategies, they can practice using them in their own conversations, even recording and transcribing their own discourse and comparing it with NS samples – effectively becoming 'mini conversational analysts' themselves, something recommended by a number of researchers (Brown & Yule 1983; Willis & Willis 1996; Celce-Murcia & Olshtain 2000; Schegloff et al. 2002; Wong 2002). The process of transcribing speech is a critical step for exploitation of spoken discourse in the classroom because it allows us to 'freeze' the interaction and highlight salient features for the learners that would otherwise be lost in the normal, transient flow of communication.

With respect to spoken genres in textbooks, a number of problems exist, the first of which relates to the *range* of genres illustrated. In a principled approach, we would expect to see the relative importance and frequency of generic types (for a specific target context) reflected fairly in classroom input but this is often not the case. Eggins & Slade (1997), for example, identified five common generic types in their casual conversation data: storytelling (narratives, anecdotes, exemplums and recounts) (43.4%), observation/comment (19.75%), opinion (16.8%), gossip (13.8%) and joke-telling (6.3%). They claim that, despite the important role these structures play in establishing peoples' identities, they are largely unrepresented in language teaching materials (although there are exceptions such as Viney & Viney 1996; Gairns & Redman 2002).

A second concern is with the *accuracy* of spoken genres represented in textbooks since many researchers, such as Yule (1995: 185), have reported that there 'continues to be a substantial mismatch between what tends to be presented to learners as classroom experiences of the target language and the actual use of that language as discourse outside the classroom.' Myers Scotton & Bernsten (1988) compared direction-giving in natural conversations with textbook dialogues and found that authentic interactions were much more complicated than the standard, three-step, model presented to students (request for directions – direction-giving – thanks). They typically included other elements such as: a) an opening sequence which could be a filler, a pause, a repetition of the question, an interjection or a comment such as 'It's really far'; b) a pre-closing where the direction-giver provides a kind of *coda* (an evaluative comment which brings the conversation back to the present) such as 'It's way, way on the other side of campus from here'; c) orientation checkers, parenthetical comments and confirmation checkers interspersed

throughout the exchange; d) non-fluencies, particularly in the opening sequence (see also Psathas & Kozloff 1976 for more on the discourse structure of directions). The authors point out that this more complicated generic structure in the natural discourse places considerable interactional demands on the direction-seeker to 'edit out' essential from non-essential information and to respond to confirmation and orientation checkers. They suggest that learners be given authentic interactions in the classroom with awareness-raising tasks to highlight the discourse structure of direction-giving. Wong (2002) (but see also Wong 1984) examined model telephone dialogues in eight ESL textbooks and assessed their faithfulness to the canonical sequencing identified by the conversational analyst, Emanuel Schegloff, in American English (see, for example, Schegloff 1993a). The opening segment is typically composed of four parts: a) a *summons-answer sequence*, where the telephone rings and the receiver answers, typically with a 'hello', which provides the caller with a voice sample for recognition purposes; b) an *identification-recognition sequence*, where the caller identifies him/herself with a voice sample such as 'hi' or by name, depending on the relationship with the receiver; c) a *greeting sequence*; an adjacency pair, often 'hi' or 'hello', and d) a *how-are-you (HAY) sequence*, where the caller normally produces the first 'How are you?' inquiry (to which the receiver can reply with a neutral response, such as 'fine', that closes down the topic, or a plus/minus response, such as 'great' or 'terrible', that invites further topical moves), followed by a second 'How are you?' from the receiver. Wong found that none of the textbook telephone dialogues she examined contained all four canonical sequences and concludes, 'As routine, simplistic, or ritualistic as telephone openings appear to be, it is striking that they were not designed in a more authentic fashion by textbook writers.'

(ibid: 53/4). The lack of realistic models in course books means that learners are unlikely to get a feel for the typical patterning of this genre, particularly how to enter and exit the talk naturally. This is exactly the kind of information that can instill a greater sense of control over TL interactions and engender confidence. Gilmore (2004) compared seven textbook service encounters with their equivalent authentic interactions and found considerable differences across a range of discourse features: length and turn-taking patterns, lexical density and the frequency of false starts, repetition, pausing, terminal overlap, latching, hesitation devices and back-channels. Similarly to Myers Scotton & Bernsten (1988), the authentic samples were found to have a more complicated structure than the regular A-B-A-B question-answer patterning displayed in the textbooks. Instead, the smooth flow of the discourse was frequently disrupted by the 'information giver' seeking clarification or further information from the 'information receiver'. Thus, in authentic service encounters, learners may have considerably more interactional demands placed on them than they are given to expect by classroom models.

The final concern with respect to the presentation of spoken genres in textbooks is that, even when the model dialogues *are* accurate, material writers typically do not attempt to highlight key components of the generic structure. This contrasts notably with written genres where larger patterns, such as the introduction-main body-conclusion structure of discursive essays, are often pointed out. Presumably, noticing generic patterns in the spoken mode can be just as beneficial for learners' discourse competence as it appears to be in the written mode and, although little empirical research has been done so far on this question, a number of writers advocate awareness-raising activities. Interest has mainly focussed on oral narratives to date (see, for example, Slade 1986; Rintell 1990; Yule

1995; Corbett 1999; Jones 2001) but Hawkins (1985) (cited in Celce-Murcia, Dörnyei & Thurrell 1995) demonstrated that learners were able to complain more effectively after a focus on the generic structure of complaint scripts.

1.6.6 *Implications for materials design*

What emerges from this review of some of the literature comparing authentic and textbook discourse is that our deepening understanding of language has profound implications for syllabus design:

> With a more accurate picture of natural discourse, we are in a better position to evaluate the descriptions upon which we base our teaching, the teaching materials, what goes on in the classroom, and the end products of our teaching, whether in the form of spoken or written output. (McCarthy 1991: 12)

The contrived materials of traditional textbooks have often presented learners with a meagre, and frequently distorted, sample of the target language to work with and have failed to meet many of their communicative needs (Schiffrin 1996). Authentic materials, particularly audio-visual ones, offer a much richer source of input for learners and have the potential to be exploited in different ways and on different levels to develop learners' communicative competence.

A further point that becomes clear from the discussion above is how context-sensitive language is. Since the discourse created in any situation is so dependent on the unique set of characteristics (the place, participants, topic and mode) prevailing at the moment it is produced, how can we begin to help learners cope with all the variety and uncertainty they are likely to face during communication in the L2? The first step is to present

language solidly contextualised and to sensitise learners to the ways in which the discourse reflects its context. The kinds of contexts selected for inclusion will often mirror those most likely to be encountered by learners in their future lives (Zuengler & Bent 1991) (although literature can still play an important role: Dissanayake & Nichter 1987; Brown 1990), and the focus of tasks will need to take into account the differences between the learners' culture and the target culture. For example, learners from low-contact cultures such as Japan (who tend to touch and look at each other less: Argyle & Cook 1976), who wish to integrate into high contact cultures are likely to need more help adapting their non-verbal communication. Similarly, those from low-context cultures such as Norway (who rely predominantly on verbal means to communicate meaning) will need more help in interpreting subtle contextual clues when integrating into high-context cultures (Hall 1989; Christopher 2004). This suggests that each classroom is quite unique in terms of its students' needs – internationally marketed textbooks are unlikely to meet these needs adequately.

CHAPTER 2. AUTHENTIC MATERIALS AND AUTHENTICITY IN FOREIGN LANGUAGE LEARNING

2.1 Historical overview

The use of authentic materials in foreign language learning has a long history. Henry Sweet, for example, who taught and wrote at the end of the nineteenth century and is regarded as one of the first linguists, made regular use of authentic texts in his books and was well aware of their potential advantages over contrived materials:

> The great advantage of natural, idiomatic texts over artificial 'methods' or 'series' is that they do justice to every feature of the language […] The artificial systems, on the other hand, tend to cause incessant repetition of certain grammatical constructions, certain elements of the vocabulary, certain combinations of words to the almost total exclusion of others which are equally, or perhaps even more, essential.
> (Sweet 1899: 177)

During the twentieth century, however, prevailing linguistic theories of the time spawned a multitude of methods such as the 'New Method' and the 'Audiolingual Method' (Richards and Rodgers 1986) which all imposed carefully structured (and therefore contrived) materials and prescribed behaviors on teachers and learners, leading to what Howatt (1984: 267) refers to as a 'cult of materials', where 'The authority of the approach resided in the materials themselves, not in the lessons given by the teacher using them'.

Large-scale trials in the 1960s, comparing the merits of different methods in the classroom, not surprisingly, proved inconclusive (see, for example, Smith 1970) since researchers were seriously underestimating the role of teachers and learners in the

learning process and the profession grew disillusioned with the search for a 'perfect method' (Howatt 1984; Alderson & Beretta 1992).

The issue of authenticity reappeared in the 1970's as the debate between Chomsky (1965) and Hymes (1972) led to a realisation that communicative competence involved much more than knowledge of language structures and contextualised communication began to take precedence over form. This culminated in the approach which, at least in EFL circles, still holds sway today – Communicative Language Teaching – and paved the way for the reintroduction of authentic texts which were valued for the ideas they were communicating rather than the linguistic forms they illustrated. However, despite appeals for greater authenticity in language learning going back at least 30 years (O'Neill & Scott 1974; Crystal & Davy 1975; Schmidt & Richards 1980; Morrow 1981), movements in this direction have been slow.

The debate over the role of authenticity, as well as what it means to be authentic, has become increasingly sophisticated and complex over the years and now embraces research from a wide variety of fields including discourse and conversational analysis, pragmatics, cross-cultural studies, sociolinguistics, ethnology, second language acquisition, cognitive and social psychology, learner autonomy, information and communication technology (ICT), motivation research and materials development. Unfortunately, many researchers limit their reading to their own particular area of specialization and, although this is understandable given the sheer volume of publications within each field, it can mean that insights from one area don't necessarily receive attention from others. With a concept such as authenticity, which touches on so many areas, it is important to attempt to bridge these divides and consolidate what we now

know so that sensible decisions can be made in terms of the role that authenticity should have in foreign language learning in the future. In this chapter, I attempt to do this although, given the scale of the undertaking, some areas of discussion are necessarily superficial.

2.2 Defining authenticity

There is a considerable range of meanings associated with authenticity, and therefore it is little surprise if the term remains ambiguous in most teachers' minds. What is more, it is impossible to engage in a meaningful debate over the pros and cons of authenticity until we agree on what we are talking about. At least eight possible inter-related meanings emerge from the literature:

a) Authenticity relates to the language produced by native speakers for native speakers in a particular language community (Porter & Roberts 1981; Little, Devitt & Singleton 1989).
b) Authenticity relates to the language produced by a real speaker/writer for a real audience, conveying a real message (Morrow 1977; Porter & Roberts 1981; Swaffar 1985; Nunan 1988/9; Benson & Voller 1997).
c) Authenticity relates to the qualities bestowed on a text by the receiver, in that it is not seen as something inherent in a text itself, but is imparted on it by the reader/listener (Widdowson 1978/9; Breen 1985).
d) Authenticity relates to the interaction between students and teachers and is a 'personal process of engagement' (van Lier 1996: 128).

e) Authenticity relates to the types of task chosen (Breen 1985; Bachman 1991; van Lier 1996; Benson & Voller 1997; Lewkowicz 2000; Guariento & Morley 2001).

f) Authenticity relates to the social situation of the classroom (Breen 1985; Arnold 1991; Lee 1995; Guariento & Morley 2001; Rost 2002).

g) Authenticity relates to assessment (Bachman 1991; Bachman & Palmer 1996; Lewkowicz 2000).

h) Authenticity relates to culture, and the ability to behave or think like a target language group in order to be recognized and validated by them (Kramsch 1998a).

From these brief outlines we can see that the concept of authenticity can be situated in either the text itself, in the participants, in the social or cultural situation and purposes of the communicative act, or some combination of these. Reviewing the multitude of meanings associated with authenticity above, it is clear that it has become a very slippery concept to identify as our understanding of language and learning has deepened. This raises the question, should we abandon the term on the grounds that it is too elusive to be useful? My own preference would be to limit the concept to objectifiable criteria since, once we start including subjective notions such as learner authentication, any discourse can be called authentic and the term becomes meaningless. To this end, I define authenticity in the same way as Morrow (1977: 13): 'An *authentic text* is a stretch of real language, produced by a real speaker or writer for a real audience and designed to convey a real message of some sort.' Using these criteria, it is possible to say whether a text is authentic or not (within these terms) by referring to the source of the discourse and the context of its production. The concept also has validity since, as Porter & Roberts (1981:

37) point out (referring specifically to listening texts but equally valid for all texts), native speakers are usually able to identify authentic text 'with little hesitation and considerable accuracy'. Furthermore, by defining authenticity in this way, we are able to begin identifying the surface features of authentic discourse and evaluating to what extent contrived materials or learner output resemble it (see, for example, Trickey 1988; Bachman & Palmer 1996; Gilmore 2004).

How far does this more specific definition of authenticity take us? Not a great distance. Even if we limit our description to real language from a real speaker/writer for a real audience with a real message, this still encompasses a huge amount of language variety. Graded teacher-talk in the classroom, motherese, international business negotiations between non-native speakers and scripted television soap operas would all be classified as authentic. But all these types of authentic input can be expected to have very different surface discourse features and some will serve as better input to stimulate language acquisition in our learners than others. Authenticity doesn't necessarily mean 'good', just as contrivance doesn't necessarily mean 'bad' (Widdowson 1979; Clarke 1989; Cook 2001; Widdowson 2003). As Cook (1997) points out, terms such as 'authentic', 'genuine', 'real' or 'natural' and their opposites 'fake', 'unreal' or 'contrived' are emotionally loaded and indicate approval or disapproval whilst remaining ill-defined. I would argue that, from the classroom teacher's perspective, rather than chasing our tails in pointless debate over authenticity versus contrivance, we should focus instead on *learning aims* or as Hutchinson & Waters (1987: 159) call it, 'fitness to the learning purpose'.

The key issue then becomes 'What are we trying to *achieve* with classroom materials?' A logical response to this would be that the goal is to produce learners who are able to communicate effectively in the target language of a particular speech community, that is to say, learners who are *communicatively competent*. To reach this goal, I would suggest that teachers are entitled to use any means at their disposal, regardless of the provenance of the materials or tasks and their relative authenticity or contrivance.

2.3 The English-as-a-world-language debate

The spread of English around the world and its success as the primary medium of global communication has considerably complicated the issue of teaching the language and the concept of authenticity in the process. With its expansion across the globe, English has naturally diversified into a proliferation of forms, varying in pronunciation, intonation, grammar, vocabulary, spelling and conventions of use, as it has been adapted to suit new surroundings so that 'it becomes ever more difficult to characterize in ways that support the fiction of a simple, single language' (Strevens 1980: 79). An estimated 1 billion people are learning English as a foreign language (Graddol 1997) and by 2010 it is predicted that there will be 50% more speakers of English as a foreign language than native speakers (Crystal 1997). Currently, it is believed that something like a staggering 80% of English used worldwide does not involve native speakers at all (Crystal ibid; Prodromou 1997b). All of this has led to 'doubts and anxieties among professionals and the general public alike' (Strevens ibid: 78) as the concepts of 'native speaker' and 'standard English' become ever more difficult to pin down (Crystal 2003; Carter & McCarthy 2003).

2.3.1 *What is a native speaker?*

Most of us probably imagine a prototypical American or Englishmen when we think of a native speaker but this model quickly begins to disintegrate under closer inspection (Davies 1995). Even assuming that the term native speaker can be defined precisely as those speech communities in Kachru's (1985) 'inner circle', the rapid development of 'non-native' varieties and the use of English as an International Language has called into question their ownership of the tongue (Alptekin & Alptekin 1984; Strevens 1987; Bowers 1992; Widdowson 1994; Nelson 1995; Graddol 1997; Seidlhofer 1999; Jenkins 2000; Modiano 2001; House 2004). Graddol (ibid:10) criticises Kachru's (1985) 'inner, outer and expanding circles' model because 'it locates the 'native speakers' and native-speaking countries at the centre of the global use of English, and, by implication, the sources of models of correctness'. This view is increasingly challenged 'by the growing assertiveness of countries adopting English as a second language that English is now *their* language, through which they can express their own values and identities, create their own intellectual property and export goods and services to other countries' (ibid: 3).

The use of authentic language in the classroom has often been challenged because it is typically seen as the discourse produced by those in Kachru's inner circle (Widdowson 1994). However, when the definition of 'native speaker' expands to include all proficient speakers of English, of whatever variety, this argument ceases to be valid. The question then becomes: Whose authentic English should we use as our model, if any, or is some form of contrived lingua/cultura franca more appropriate in the classroom? There are

cases to be made for either choice although the pedagogical consequences are quite different.

2.3.2 *Is a lingua/cultura franca model more appropriate in the classroom?*

The concept of a 'lingua franca' is not something that can be readily codified but for the purposes of ELT it is most likely to mean a reduced form of English, incorporating what textbook writers perceive to be the most relevant features of the language for communication between non-native speakers in international contexts. This may include a pronunciation syllabus which only models the core phonological distinctions necessary for intelligibility, as proposed by Jenkins (2000). It will also tend to be a more standard, formal variety of the language devoid, as far as possible, of its cultural associations and set in 'cosmopolitan' contexts like international airports and hotels (Strevens 1980; Brown 1990; Prodromou 1996b). This has several potential advantages for the learner. Firstly, it maximises their chances of learning a variety of English which can be understood by a wide range of nationalities and can be put to immediate, practical use in what we have seen is the most likely scenario: one non-native speaker talking to another non-native speaker. Secondly, it avoids culturally loaded language, which is often difficult to understand once removed from its context of use, and may, in any case, be perceived as 'utterly boring' by learners (Prodromou 1996b: 88). Prodromou (1997b) illustrates this point in a simple experiment. He compared the ability of students to complete two gap fill exercises with vocabulary items, one using made-up sentences taken from a traditional dictionary, the other real examples taken from a corpus-based dictionary. He found, not surprisingly, that learners had considerably more difficulty

completing the real examples than the more self-contained, contrived ones. Furthermore, 76% of the teachers polled believed the made-up samples were more appropriate for the classroom. In this sense, contrived language would appear to be better suited to the learning process (see Widdowson 2003, Ch. 8 & 9 for a detailed discussion of this issue). Thirdly, by avoiding 'inner circle' varieties of English in textbooks, the balance of power shifts from native speaker to non-native speaker teachers (Seidlhofer 1999); something many are keen to see after the accusations of linguistic imperialism put forward by the likes of Phillipson (1992a/b) and Pennycook (1994).

Many researchers see problems with using some type of lingua franca as the model for language teaching, however. Firstly, this approach, generally though not necessarily, relies on the textbook writer's intuitions about language and these are notoriously unreliable:

> […] rules of speaking and, more generally, norms of interaction are not only culture specific, they are also largely unconscious. What this means is that native speakers, although perfectly competent in using and interpreting the patterns of speech behavior which prevail in there own communities, are, with the exception of a few explicitly taught formulas, unaware of the patterned nature of their own speech behavior. (Wolfson 1986a: 693)

This means that writers often run the risk of presenting a distorted view of the language to learners (Sinclair 1991; Biber, Conrad & Reppen 1994). O'Connor di Vito (1991: 384) points out that students naturally assume, unless otherwise indicated, that the language presented to them in course books is 'equally generalizable, equally important communicatively, and equally productive in the target language' so any distortions in the

materials will have serious knock-on effects for learners' use of the target language. By limiting ourselves to authentic samples of discourse, researchers argue that we are less likely to fall into this trap. A further problem with the lingua franca model is its emphasis on more formal varieties of English. This limits students' exposure to the more evaluative, interactional features of the language which tend to be associated with informal, spoken English (Brown & Yule 1983; Richards 1990; Carter & McCarthy 1996; McCarthy & Carter 1997) and may therefore affect their ability to 'be friendly' in the L2.

Cultura franca?

A third issue, related to the topic of 'cultura franca', is to what extent it is possible or advisable to separate a language from its cultural associations. Pulverness (1999: 6) points out that many modern ELT textbooks try to side-step the issue of culture altogether by presenting their target language in 'international contexts' outside the domain of any particular country but these attempts are doomed to failure for a number of reasons. Firstly, the materials generally consist of contrived dialogues written by native speaker authors who, despite feigning to represent other nationalities, cannot possibly dissociate themselves from their own cultures sufficiently to do the job justice and reflect the lexicogrammatical, topical or interactional choices natural for people from different cultures (Dissanayake & Nichter 1987; Brown 1990; Alptekin 1993). Even if textbook writers could realistically portray international encounters, they are still not culture-less; for example, Japanese and Saudi businessmen at a meeting in New York carry their own cultural expectations to the table. It would seem, then, that culture-free language is an

impossible goal (see, for example, Valdes 1986; Byram 1991, 1997; Kramsch 1993; Nelson 1995) but, if this is the case, what choices are available to material writers? Cortazzi & Jin (1999) suggest that there are three types of English language textbook on the market: those that teach the students' own culture (C1); those that teach the target culture (C2); and those that teach a wide variety of other cultures that are neither source nor target cultures (C3, 4, 5...). There are potential advantages and disadvantages for all three of these options, which are worth examining in more detail.

Teaching the target language through the learners' own culture may help to reinforce their national identity in a world increasingly dominated by western paradigms:

> Being at the receiving end of a virtually one-way flow of information from Anglo-American centres, the host country runs the risk of having its own culture totally submerged, and thus imposes restrictions in educational and cultural domains to protect its way of life. (Alptekin & Alptekin 1984: 15)

However, this view has been challenged more recently for being rather patronizing, underestimating, as it does, the non-native speakers' ability to take from the language materials only what they consider useful, and to appropriate English for their own needs, or in Kramsch & Sullivan's (1996: 210) words, 'the unique privilege of the NNS to poach on the so-called authentic territory of others, and make the language their own.' (see also Byram 1991; Bisong 1995; Siegal 1996; Seidlhofer 1999; Gray 2000; Carter & McCarthy 2003). The desire to impose restrictions on cultural input from abroad is, in any case, more likely to emanate from political institutions within the country seeking to maintain control over the population (see McVeigh, 2002 for a discussion of the Japanese context). Materials based on the C1 do, however, allow learners to practise explaining about their

country in English (Cortazzi & Jin 1999) and, because they start from familiar content, provide greater support, allowing for more top-down processing (Richards 1990) which may be particularly beneficial at lower levels of proficiency. Furthermore, in Widdowson's (2003) opinion at least, C1 language input better suits the social reality of the classroom because it is real for the learners and therefore more effective in activating the learning process.

The disadvantages with these kinds of materials are that they fail to exploit the language learner's natural curiosity in other cultures and, in the absence of information to the contrary, students are likely to assume that other cultures operate in the same way as their own (Byram 1991: 18). Also, although the intention may be to reinforce the learners' national identity, paradoxically, they may be prevented from doing this because they have nothing to compare their culture with (Cortazzi & Jin 1999); true understanding of our own culture can only come from seeing how other societies operate. Finally, restricting the cultural input to the C1 limits the marketability of textbooks, rendering them less cost effective for publishers (Alptekin 1993).

Which target language culture?

Materials that teach the C2 (the target culture of a speech community where English is used as a first language) are the traditional fare of the ELT industry and, although historically they may have included as much contrived as authentic discourse, are the obvious place to exploit authentic texts. For many languages, such as Japanese or Danish, it would seem natural to introduce the target culture and language concurrently since the destinations of the learners and the communities they will need to operate in are more

predictable. As we have already seen, however, the situation with English is much more complicated because of the wide variety of cultures which call the language their own. Decisions over whose culture to represent in language teaching materials are likely to vary from place to place. Prodromou (1992), in his survey of Greek students' attitudes to English-speaking cultures, found a marked preference for the British over the American model which he accounts for in terms of the historical tensions between Greece and the United States, but this is likely to be the reverse in Japan where students tend to have a far greater affiliation with America. There is, of course, no reason why a wide variety of English-speaking cultures cannot be represented in language textbooks and this might be more fitting to its international status, while at the same time rendering publications more marketable worldwide. In my own opinion, it is essential to include the target culture (or rather cultures) within language teaching materials in order to serve the broader educational goal of developing learners' intercultural communicative competence (Byram & Fleming 1998). In modern urban societies, characterised by their social and cultural heterogeneity (Schiffrin 1996: 313), successful communication depends on much more than a superficial command of a target language, it also requires an ability to see the world from different perspectives:

> What is at issue here is a modification of monocultural awareness. From being ethnocentric and aware only of cultural phenomena as seen from their existing viewpoint, learners are to acquire an intercultural awareness which recognises that such phenomena can be seen from a different perspective, from within a different culture and ethnic identity. (Byram 1991: 19)

Authentic materials, such as television sitcoms (Scollon 1999) are uniquely placed to bring about this shift in awareness and to heighten learners' understanding of both their own and the target culture. This kind of approach sees learners as comparative ethnographers (Byram 1991; Cortazzi & Jin 1999; Pulverness 1999), forced to re-examine their own culture-specific schemata by comparison with other patterns of behaviour.

The risk with introducing the target culture(s) into the classroom is that we disenfranchise learners who then 'switch off, retreat into their inner world, to defend their own integrity.' (Prodromou 1988: 80). It can also disadvantage NNS teachers, undermining their confidence (Seidlhofer 1999). Materials such as these therefore, obviously, need to be selected carefully, with the specific needs of the learners in mind, and handled intelligently, allowing students to move from the familiar to the unfamiliar in a way that keeps them engaged in the learning process. They also need to provide teachers with sufficient support to confidently deal with the syllabus.

Cortazzi & Jin's third and last type of textbook are those that teach a wide variety of other cultures that are neither source nor target cultures. The advantage of these kinds of materials is that they meet the needs of the increasing number of learners who want to use English as an International Language to speak to other non-native speakers around the world. Similarly to the arguments made above, they can also be exploited to develop students' inter-cultural competence by exposing learners to unfamiliar behavioural patterns or instances of cross-cultural miscommunication but *only when the discourse is authentic, NNS-NNS interaction* (see, for example, Firth 1990; Newman 1996). Contrived dialogues written by native speakers of English are unlikely to capture the true flavour of

NNS-NNS interactions so we should be wary of textbooks that embrace internationalism only superficially in an attempt to make themselves more marketable.

One disadvantage of materials such as these is that non-native speakers of English are often unable to express their thoughts as precisely in the L2 as they can in their mother tongue. We therefore run the risk of providing learners with 'dumbed down' models of English which, although perhaps meeting their transactional needs, fail to illustrate the true expressive potential of the language. Carter & McCarthy (1996), in a series of articles debating authenticity with Luke Prodromou, argue that we should never hold back information about the language because it disempowers learners (see also Phillipson 1992a/b; Sinclair 1997).

Only a small number of researchers have bothered to ask the learners *themselves* what they think about these issues. One of the few who has is Timmis (2002), who received responses to his questionnaire on teacher and student attitudes to 'native' vs. 'standard' English from respondents in 14 different countries. He found a continued preference for native-speaker models in his sampling, concluding that:

> There is still some desire among students to conform to native-speaker norms, and this desire is not necessarily restricted to those students who use, or anticipate using English primarily with native speakers. (ibid: 248)

2.4 Authenticity & Motivation

Claims that authentic materials are a motivating force for learners are widespread through the literature (Cross 1984; Deutsh 1984; Hill 1984; Wipf 1984; Swaffar 1985; Freeman & Holden 1986; Keinbaum, Russell & Welty 1986; Little, Devitt & Singleton 1989;

Morrison 1989; Bacon & Finnemann 1990; González 1990; King 1990; Little & Singleton 1991; McGarry 1995; Peacock 1997). This opinion appears to be mirrored in the language teaching population at large, since authenticity is frequently used as a selling point in the marketing strategies of publishers. Various justifications have been put forward to support these claims, the most common being that authentic materials are inherently more interesting than contrived ones because of their intent to communicate a message rather than highlight target language (although contrived materials aren't only produced to focus on form) (Swaffar 1985; Freeman & Holden 1986; Hutchinson & Waters 1987; Little, Devitt & Singleton 1989; King 1990; Little & Singleton 1991). This position is rejected by others, however, who argue that the difficulties associated with authentic texts (because of the vocabulary used or the cultural knowledge presumed), de-motivate learners (Williams 1983; Freeman & Holden 1986; Prodromou 1996a; Widdowson 1996, 1998, 2003). Cross (1984) suggests that showing learners that they can cope with authentic materials is, in itself, intrinsically motivating which introduces the idea of motivation as the result, rather than the cause, of achievement (Ellis 1985; Little et al. 1989; Skehan 1989). Some attribute the motivating nature of authentic materials to the fact that they can be selected to meet students' specific needs, unlike textbooks which cater to an international audience (Morrison 1989; McGarry 1995; Mishan 2005), but this would appear to be an argument for more selection, adaptation or supplementation of course books rather than the exclusive use of genuine texts. Finally, some see the fact that students perceive them as 'real' as being the motivating force (Hill 1984; Peacock 1997). The fact is, however, that researchers and teachers are largely unaware of learners' true motivations for learning a language (Oxford & Shearin 1994) and empirical research in

support of any of the claims outlined above is scarce (González 1990; Peacock 1997). This is not altogether surprising given the problems associated with establishing a causal link between authenticity and motivation. The first difficulty relates to the definitional ambiguities surrounding the term 'authenticity' in the literature (see section 2.2) since, before we can make any claims about the effects of authentic materials, we need to ensure that we are all talking about the same thing. Most researchers use the term to refer to cultural artifacts like books, newspapers & magazines, radio & TV broadcasts, web sites, advertising, music and so on but this kind of discourse, which is often more considered, or even scripted, typically has very different surface features from that produced in spontaneous conversation between native speakers. Produced by talented communicators to entertain a wide audience, it is also often much more interesting than the humdrum discourse of everyday life (Porter Ladousse 1999):

> Most conversations are appallingly boring. It is the *participation* in conversations which makes us such avid talkers, the 'need to know' or 'the need to tell' or 'the need to be friendly'. You can listen to hours and hours of recorded conversation without finding anything that interests you from the point of view of what the speakers are talking about or what they are saying about it. After all, their conversation was not intended for the overhearer. It was intended for them as participants. (Brown & Yule 1983: 82)

Some researchers (for example Swaffar 1985) classify any text with a true communicative objective as authentic, which could include much of that written specifically for language learners, so we obviously need to be very careful when we compare the results from different trials. The second problem is that the success of any particular set of authentic materials in motivating a specific group of learners will depend

on how appropriate they are for the subjects in question, how they are exploited in the class (the tasks) and how effectively the teacher is able to mediate between the materials and the students, amongst other variables (Kienbaum et al. 1986; Omaggio 1986; Rings 1986; Rogers & Medley 1988; González 1990). Where the effects of authentic materials are compared with those from a control group using a 'standard textbook', the results will depend as much on the quality of the control text chosen as the experimental materials. Since many modern course books contain a lot of authentic texts anyway, researchers may end up comparing like with like. These influencing factors are seldom mentioned in research reports and are, in any case often very difficult to judge objectively, all of which poses a serious threat to the internal validity of this kind of classroom investigation (Brown 1988). A further consideration is that the learners' location and goals are likely to affect their attitudes towards authentic materials. Those with integrative motivation (Gardner & Lambert 1959), typically second language learners, are more likely to react positively to authentic materials than those with instrumental motivation, typically foreign language learners, (Dörnyei 1990; Oxford & Shearin 1994; Mishan 2005), although this is not always the case; medical students, for example, studying ESP with no desire to integrate into a native-speaking community, may respond more enthusiastically to authentic medical texts than contrived textbook material. Another issue that may influence the research results is the learners' familiarity with authentic materials prior to the study. González (1990) and Peacock (1997) both detected a time effect in their research with students' motivation increasing as they became more familiar with using authentic materials. The length of time over which motivation is measured may therefore be important. Lastly, there is the problem of how to accurately measure learners'

motivation in classroom-based studies. Most empirical research of this type has relied on student self-report data, which runs the risk of being contaminated by the 'approval motive' where 'the respondent works out what the 'good' or 'right' answer is, and gives it.' (Skehan 1989: 62).

In summary then, it is clear that there are many dangers inherent in this kind of research (Duff 2005). This does not mean, of course, that we should give up on our attempts to establish a link between motivation and authenticity; after all, a consensus amongst researchers on this issue could have major implications for materials design. However, meaningful results will depend on carefully conceived experimental designs that attempt to account for all of the variables outlined above. To my knowledge, only three empirical studies have so far been conducted into the effects of authentic materials on motivation (Keinbaum, Russell & Welty 1986; Gonzales 1990; Peacock 1997). Keinbaum et al. hypothesised that a communicative methodology used in conjunction with authentic materials could increase students' motivation towards studying German, French and Spanish as a foreign language. 29 American college students received either the control or experimental treatment over a period of 30 weeks and, although no statistically significant differences were found between groups at the end of the trial in terms of language performance, they report that their qualitative data indicated that students were well motivated by the use of authentic materials. Unfortunately, they do not establish whether this was as a result of the materials or the methodology used in the experimental group. Kienbaum and associates used an attitude survey to try and quantify differences in motivation between the control and experimental groups but only 3 items out of 23 on the questionnaire actually focused on the method or materials employed so their results are

far from convincing. Gonzalés (1990) investigated whether exposure to authentic materials (but only as textbook supplements) would have any effect on Spanish-language learners' attitude, motivation and culture/language achievement. 43 students at an American college, assigned to either control or experimental groups, received the treatment over a period of 10 weeks but no statistically significant differences in either 'levels of satisfaction' (ibid: 105/6) or achievement were found. Unfortunately, the learners' feelings towards the use of authentic materials were only measured by one item on a self-report Foreign Language Attitude Questionnaire. Some of the qualitative data in the study from student feedback and instructors' logs did indicate a positive reaction towards the authentic supplements but to what extent this is due to the materials themselves and not just a desire to do something other than the assigned textbook is impossible to determine. Peacock (1997) provides the most convincing empirical results on authenticity and motivation available to date. He used a more sophisticated model of motivation – interest in and enthusiasm for the materials used in class; persistence with the learning task, as indicated by levels of attention or action for extended periods of time; and levels of concentration or enjoyment (Crookes & Schmidt 1991: 498-502) – to investigate the effects of authentic materials on beginner-level, English language university students in South Korea over a period of 20 days. He found highly significant ($p < 0.001$) increases in both on-task behaviour and overall class motivation when students were using authentic materials, as judged by an external observer. Student self-reported motivation also increased significantly with the authentic input ($p < 0.05$) but only after day 8 of the study, which Peacock attributes to a period of adjustment to the experimental materials. However, although students found authentic materials more

motivating than contrived ones, they also found them *less interesting*, suggesting that interest and attention to task or persistence with learning tasks are 'separate components of classroom motivation' (ibid: 152).

In summary, despite the widespread belief in the motivating potential of authentic materials, very little empirical support for the claim currently exists.

2.5 Text difficulty & task design

Widdowson (1978, 1983, 1996, 1998, 2003) has argued consistently that learners are unable to authenticate real language since the classroom cannot provide the contextual conditions for them to do so. Instead, he sees simplified texts that gradually approximate authentic ones as more pedagogically appropriate. In Widdowson (1998: 710), he gives the following example from The Guardian newspaper (30/11/95) to illustrate his point:

IT TAKES BOTTLE TO CROSS THE CHANNEL

Bibbing tipplers who booze-cruise across the Channel in search of revelry and wassail could be in for a rough ride. Itchy-footed quaffers and pre-Christmas holiday-makers are being warned not to travel to France, widespread disruption continues despite the lifting of the blockade on trapped British lorry drivers.

This does, without doubt, show the potential dangers of introducing authentic texts into the classroom: the high lexical density, idiomatic language, low frequency vocabulary used for satirical effect, and opaque cultural references all combine to make it 'pragmatically inert' (ibid: 710) for most learners. However, Widdowson chooses a particularly extreme example to make his case and many researchers disagree with his

point of view, believing that all levels of learner can cope with authentic material if the texts and tasks are carefully selected.

Rating a text's difficulty is not an exact science and is, to some extent, dependent on the learning context in which it is used. Anderson & Lynch (1988: 81), for example, point out that low frequency words are generally assumed to be difficult but whether they are or not depends on how common the lexis is in the target community (the word '*stalker*', for example, despite only a handful of hits on the British National Corpus, is widely understood in Japan), the context in which the word occurs, the learners' knowledge of the topic and whether there are any cognates in the L2 (see also Wallace 1992: 76). Similarly, rating text difficulty on grammatical criteria is not straightforward either, since it will be influenced by the degree of similarity between the L1 and L2 grammatical systems. In addition, SLA research has shown that just because a grammatical point, such as 3rd person 's', is easy to analyse doesn't necessarily mean that it is easy to learn (Nunan 1988, 1989). However, it has long been recognized (see Sweet 1899) that authentic texts are naturally graded and some general guidelines can be offered. Brown & Yule (1983) mention a range of factors affecting text difficulty:

a) Different spoken genres can be represented on a cline of increasing inherent difficulty (description < description/instruction < storytelling < opinion-expressing), depending on whether they represent static, dynamic or abstract concepts.
b) The number of elements in a text and how easily they can be distinguished from one other, so that a short narrative with a single character and a few main events

will be easier to comprehend than a long one involving more characters and events.

c) The delivery speed and accents used in spoken texts.

d) The content (grammar, vocabulary, discourse structure and presumed background knowledge in a text).

e) The visual support offered in conjunction with listening texts (video images, realia or transcripts).

Anderson & Lynch (1988) report on a range of other factors that have been shown in experimental research to affect listening comprehension (although mainly with young native-speakers), such as the way in which information is organized, topic familiarity, and degree of explicitness. Bygate (1987: 16) points out that spoken text is generally syntactically simpler than written text because of the performance pressures speakers operate under. Rather than producing complex sentence structures, they tend to employ 'parataxis' to string simple clauses together with coordinating conjunctions (discourse markers might be more appropriate terminology), leading to less dense text with a lower lexical density (Ure 1971; Stubbs 1986), which can ease the task of comprehension. Text length is mentioned by Nunan (1989) as a further factor affecting difficulty because it can lead to reader/listener fatigue, but, as Anderson & Lynch (1988: 85) citing Wallace (1983) note, there is not necessarily a one-to-one relationship since, 'the longer someone speaks on a topic the more chance there is of understanding the point of what he is trying to say'.

A second way to control for difficulty in authentic materials, which has become increasingly important since the 1980s and the emergence of the 'strong' version of the

communicative approach (Howatt 1984: 279), is to vary the task rather than the text (Prabhu 1987; Nunan 1989; Willis 1996). This approach allows for only partial understanding of texts by learners on the basis that even native speakers typically operate with less than total comprehension (Willis 1996; Guariento & Morley 2001; Widdowson 2002):

> Even native speakers do not impose a standard of total comprehension on themselves, and tolerate vagueness. For example, on the BBC weather forecasts for shipping, millions of listeners may hear that a wind is 'backing south-easterly'. To a layman, 'backing' will mean 'moving' and he is quite content with that, though aware that there is probably a finer distinction contained in the term. His comprehension is partial, but sufficient for his needs, and in proportion to his knowledge. (Porter & Roberts 1981: 42)

From this perspective, authentic materials are seen as both encouraging a tolerance of partial comprehension and enhancing learners' inferencing skills (Morrison 1989; Brown 1990; Duff & Maley 1990; McRae 1996; Guariento & Morley 2001). Many writers have demonstrated how it is possible to adapt authentic texts to different levels of learner by varying the tasks associated with them (Windeatt 1981; Wipf 1984; Swaffar 1985; Nunan 1988, 1989; Morrison 1989; Little & Singleton 1991; Devitt 1997). They do not, however, provide any empirical evidence that this approach is more effective than adapting the texts themselves.

2.5.1 *Text modification, comprehensibility and SLA*

Nation (2001: 232) believes that incidental learning of lexis through guessing from context should be the most important method of vocabulary acquisition for learners (as it

is for NSs). In order to do this effectively, he estimates that learners need to understand 95% to 98% of running words in a text (or one unknown word in every two to five lines). Ensuring this optimal ratio suggests that text modification could have an important role to play. However, studies investigating the effects on language acquisition of modifying input have produced mixed results which suggest that, if there are benefits, they may vary with factors such as learner proficiency, mode (spoken or written), type of modification (linguistic, syntactic, articulation rate, pauses etc.), approach taken (simplification or elaboration), text characteristics (rhetorical style, lexical density etc.), topic familiarity and so on. In addition, comparisons between studies are frustrated by differences in the method of assessment (multiple choice questions, recall, self-assessment, dictation, cloze tests, etc.) and the time of assessment (during or after exposure to the text) (Leow 1993; Yano, Long & Ross 1994; Young 1999). Yano et al. (ibid) summarize the results of fifteen studies into the effects of simplified and elaborated input on non-native speaker comprehension, concluding that text modification tends to have a positive effect. They note, however, that many of these trials do not adequately distinguish between simplifying and elaborative changes and often generalize from small samples. In their own study, they therefore sought to determine the relative effectiveness of these two approaches on learners' reading comprehension in Japanese college students. They found that both types of text modification improved learner comprehension compared to the unmodified NS versions and conclude that text elaboration is 'a viable alternative to simplification' (ibid: 214). Although this result may seem to disfavour the use of authentic texts, it is important to remember that elaboration is likely to occur in the classroom anyway, even when it is not explicitly designed into the materials. Teachers

naturally clarify, rephrase, and make connections explicit to mediate between the materials and learners, and learners also negotiate meaning between themselves in order to comprehend input (Hammond & Gibbons 2005).

Other researchers have tried to simplify spoken texts by altering the delivery rate or by inserting pauses into the discourse, again with mixed results. Griffiths (1990) observed that above average speech rates led to a significant reduction in comprehension (as did Conrad 1989) but slower than average rates had no significant effects (see also Blau 1990; Derwing & Munro 2001). Blau (ibid) and Derwing (2006) both noted improvements in learners' comprehension when pauses were inserted at sentence, clause or phrase boundaries or after key lexical items respectively. However, Derwing (1990) found that increased total pause time had an inhibiting effect on learner comprehension. These results do not appear, therefore, to favour contrived over authentic listening texts as long as the authentic recordings are selected carefully to filter out above average articulation rates. Pauses, even if they are found to be beneficial, can easily be introduced mechanically in the class by the teacher. However, much more research is needed in this area before we can come to any reliable conclusions. How, for example, does varying the lexical density affect comprehension and can learners cope with higher articulation rates in authentic speech (rather than writing) which, as we have seen, tends to be more 'spread out' (Bygate 1987: 16)? Does slowing articulation rates or inserting pauses benefit different proficiency levels to different degrees? What difference does inclusion of visual support through the use of video make to learner comprehension?

Writers who dispute the benefits of text simplification often do so on the grounds that: a) it makes the task of reading more difficult by reducing the number of linguistic and

extralinguistic cues (Grellet 1981; Johnson 1982; Clarke 1989; Willis & Willis 1996); b) it can cause unnaturalness at the discourse level (McCarthy 1991); and c) it can prevent learners from looking beyond the most obvious meanings of words and from acquiring the ability to interpret representational as well as referential language (Swaffar 1985; Vincent 1986; McRae 1996). In terms of empirical evidence against text modification, the evidence is rather limited, however. Allen et al. (1988) found that high school foreign language students coped well with authentic texts compared to modified texts, even though the teachers involved in the trial had judged them to be too difficult for the learners. Young (1999), investigating reading comprehension in Spanish language students, noted a tendency for better recall scores on authentic, as opposed to simplified, versions of texts and concludes that simplification is not necessarily more effective.

Leow (1993) disputes the results of studies such as those mentioned above on the basis that they assume a causal link between comprehension and language acquisition. The rationale is that, by simplifying input, it becomes more comprehensible and this, in turn, eases the cognitive demands on learners and allows them to pay more attention to forms in the input that are not part of their current interlanguage system and, thereby, acquire more language (Krashen 1982, 1985, 1989; Long 1985; McLaughlin 1987). Leow, instead, looked at learners' *intake* (elements of the input that are noticed by the learner, and become available for acquisition) of selected linguistic items from authentic and simplified texts and found that, although the simplified versions were significantly more comprehensible, they did not facilitate greater levels of intake. He concludes:

> Consequently the findings of this study appear to provide empirical support for proponents of unedited authentic written materials in the classroom. If we consider the rather small increase in intake in this study

by learners exposed to the simplified passage and the amount of time, effort, and expertise needed to modify texts for the classroom, it can strongly be argued that the use of authentic texts provides a more practical alternative to simplified texts. (ibid: 344)

We will, therefore, need more empirical evidence before we can make any strong claims about the relationship between authentic or modified input and language acquisition. Leow's work is particularly interesting though, because it grounds itself firmly in SLA theory. He hypothesizes that it is probably the learners' own internal language system that determines what is taken in so that 'external manipulation of the input may not only be haphazard but also inadequate to address what may appropriately facilitate learners' intake' (ibid: 342). This concurs with constructivist theories from developmental psychology that see learning as a process of actively selecting out the data necessary for personal development from the overwhelming range of stimuli we are constantly exposed to:

In contrast to more traditional views which see learning as the accumulation of facts or the development of skills, the main underlying assumption of constructivism is that individuals are actively involved right from birth in constructing *personal meaning*, that is their own personal understanding, from their experiences. In other words, everyone makes their own sense of the world and the experiences that surround them. (Williams & Burden 1997: 21)

Nunan (1996) uses the metaphors of building a physical structure or growing a garden to describe these different views of learning. The traditional view sees language acquisition as a linear, step-by-step process, like laying bricks in a wall, where we can only move on to building the next level once the previous one has 'solidified'. This is the

model that the PPP methodology in language teaching aims to serve, presenting learners with 'graded' linguistic items to digest one at a time, but as Skehan (1996) says, it has now largely been discredited in the fields of linguistics and psychology. The garden metaphor, on the other hand, sees language learning as a more organic process where many things are learned imperfectly all the time (Nunan ibid: 370). A text-driven approach to learning (Mishan 2005) is more in tune with this model of language acquisition. Providing learners with 'rich input' from (authentic) texts, allows them to take different things from the lesson to suit their own particular developing interlanguage systems. As Allwright (1984), Slimani (1992) and Bygate, Skehan & Swain (2001) point out, this is what learners do anyway, even when we force them to march lock-step in the classroom: 'Learners are perfectly capable of reinterpreting tasks, in such a way that the carefully identified pedagogic goals are rendered irrelevant as a learner invests a task with personal meaning' (Bygate et al. ibid: 7).

Another concept emerging from SLA studies that is having an increasing impact on materials selection and task design is *noticing* (Schmidt 1990; Batstone 1996; Skehan 1998). Schmidt & Frota (1986) and Schmidt (ibid) challenge Krashen's (1985) view that language acquisition can proceed without any attention to form, claiming that a degree of awareness is important before items can be incorporated into the developing interlanguage system, or as Ellis (1995: 89) puts it, 'no noticing, no acquisition'. Intake does not necessarily become part of the developing IL system but it is seen as making it as far as the learner's short/medium-term memory, from where it can interact with, and reshape, information stored in long-term memory in a process that Piaget termed 'assimilation' and 'accommodation' (see Williams & Burden 1997: 23). Schmidt (ibid)

sees six influences operating on noticing (which Skehan 1998 incorporates into his information processing model): a) Frequency of forms in the input; b) Perceptual salience of forms in the input (how much they stand out); c) Explicit instruction; d) Individual differences in processing abilities; e) Readiness to notice; and f) Task demands. Schmidt & Frota (ibid), expanding on an idea first put forward by Krashen, propose a second process that can enhance the acquisition of intake, which they term 'noticing the gap'. This means learners seeing a difference between their current competence and the information available to them as intake. Ellis (1995: 89) incorporates both of these processes into his 'weak-interface' model of L2 acquisition.

What impact do these models of information processing and language acquisition have on the authenticity debate? Authentic material is likely to expose learners to a wider variety of grammatical and lexical features but with less frequency than contrived input specifically designed to highlight particular target language. Ellis (1999), in his summary of studies looking at the effects of 'enriched input', concludes that it can help learners acquire new forms so this may favour contrivance *if* we are able to accurately predict when learners are ready to notice something. On the other hand, it could be argued that exposing learners to a wider variety of language increases the likelihood that there is something in the input that they are predisposed to acquire, which would favour authenticity. A second difference relates to what exactly learners are able to notice in the input they are exposed to in the classroom. As we saw in Chapter 2, authentic discourse is typically very different from the language presented to learners in textbooks and this will inevitably impact on the way their IL develops: learners can't notice things that aren't made available to them in the input. Recently, a number of authors have exploited the

concept of noticing with authentic materials to raise learners' awareness of features not normally brought to their attention in textbooks. For example, Hall (1999) and Basturkmen (2001) both highlight typical features of interactive speech and Jones (2001) focuses on the linguistic realizations of oral narratives.

In terms of designing tasks to use with authentic materials, we will want to ensure that we do not overload learners' language processing systems by asking them to analyse input for meaning and form simultaneously. This is typically done by allowing them to focus on meaning first before shifting attention to language forms (Batstone 1996; Willis 1996; Willis 2003). Mariani (1997: 4) sees the whole issue of text difficulty and task design from the very practical standpoint of providing *challenge* and *support* in the classroom. He argues that all pedagogic activities can be described along two dimensions in terms of the level of challenge and support they provide, and that different combinations of these two factors have different learning consequences:

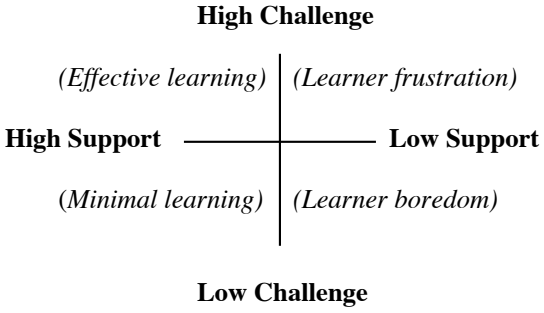


Figure 2.1 The effects of challenge & support in the classroom

The most effective classrooms are seen as those where learners have both high challenge and high support, a view which is consistent with both Bruner's (1983) model of 'scaffolding' and Vygotsky's (1978) notion of learning only taking place when learners are working inside their *zone of proximal development* (ZPD). This is where the

challenge of a task is just beyond the learner's level of competence so that it can only be achieved with support.

Hammond & Gibbons (2005) see scaffolding as operating at both macro and micro levels in the classroom: at the 'designed-in level', careful planning, selection and sequencing of materials and tasks ensures that learning opportunities are created where students can operate within their ZPD while at the 'interactional level', teachers and learners engage with each other contingently to jointly construct meaning from those opportunities (see also Swain, Brooks & Tocalli-Beller 2002).

These constructivist and interactionist views of learning to some extent push any distinctions between authentic and contrived discourse to the periphery since, as long as materials and tasks allow learners to operate within their ZPD, it could be argued that their origin is irrelevant. However, we might speculate that authentic materials are often superior because they provide rich input that is more likely to cater to the different stages of development and individual differences that exist within any classroom population.

Skehan (1998) summarizes research which suggests that task design can have different effects on the accuracy, complexity or fluency of learners' output. In the future, then, we can expect task design to be more in tune with information processing models from second language acquisition research.

2.6 Conclusion

Although much of the research reviewed above points to the inadequacies of current language textbooks and often makes specific recommendations on ways to improve them, change has been slow to take place. Indeed, Tomlinson et al. (2001) identify a growing resurgence of grammar-based syllabuses by major British publishers of ELT courses

(although these are not necessarily incompatible with authentic texts). Where change has occurred, it generally takes the form of 'bolt-on activities' added to a more traditional, structural syllabus (see, for example, the Headway series) and an evolution into a 'multi-syllabus', rather than a complete break with the past (Yalden 1987; McDonough & Shaw 1993). There are a number of possible reasons for this rather conservative approach:

a) With all the wild pendulum swings our profession has been subjected to over the last fifty years or so, there is an understandable reluctance to embrace yet another fashionable trend.

b) The division of applied linguists and language practitioners into two distinct, and at times hostile, bodies (for a discussion of this issue, see Strevens 1980; van Lier 1984; Allwright & Bailey 1991; Shaw 1996; Hopkins & Nettle 1994; Cook 1998; Judd 1999; Lightbown 2000; Clemente 2001; Thornbury 2001a,b; Widdowson 2003) leads to what Clarke (1994) calls a 'dysfunctional discourse'. Poor communication between researchers and teachers means that potentially useful findings from research often 'linger in journals' (Bouton 1996) instead of making it into the classroom.

c) Publishers are reluctant to take risks with innovative materials or to change the status quo, given the enormous costs involved in developing global textbooks (Tomlinson 1998/2001). As Thornbury (1999: 15) says, 'Form is safe. It sells books'.

d) There are practical difficulties that discourage teachers or institutions from abandoning textbooks in favour of authentic materials, even when this is seen as desirable. Finding appropriate authentic texts and designing tasks for them can, in

itself, be an extremely time-consuming process (Crystal & Davy 1975; Kienbaum et al. 1986; Kuo 1993; Bell & Gower 1998; Hughes & McCarthy 1998) but to be able to exploit authentic materials to their maximum potential also requires a familiarity with the kind of research literature reviewed in section 3. Few teachers have either the access to these studies, or the time (inclination?) to read them (Judd 1999) and, even if they did, the sheer volume of work available would make it difficult to identify areas with the greatest pedagogic significance. Admittedly, teacher friendly resource books are quickly spawned from new ideas arising in the literature (for example, the 'Resource Books for Teachers' series from Oxford University Press) and these help to bridge the gap between theory and practice.

e) Teaching learners and testing their progress becomes considerably more complicated once a discrete-point syllabus is abandoned. As Skehan (1998: 94 remarks, the 3P's approach 'lends itself very neatly to accountability, since it generates clear and tangible goals, precise syllabuses, and a comfortingly itemizable basis for the evaluation of effectiveness'.

Woodward (1996) notes a growing dissatisfaction with current practices within the language teaching profession and suggests that there are signs of an imminent paradigm shift, although, as yet, there is little in the way of consensus as to what exactly we should shift *to*. One possibility is a text-driven approach (Tomlinson 2001; Mishan 2005) which, rather than starting from a predetermined list of lexicogrammatical items to be taught, focuses on teachers (or students themselves) selecting and exploiting authentic materials appropriate to their own particular contexts and needs, using a task-based methodology (Prabhu 1987; Nunan 1989; Bygate, Skehan & Swain 2001; Willis 1996). The syllabus is

arrived at retrospectively, from what is made available for noticing in the input, and in this sense it is more in tune with constructivist theories of language acquisition. Although the text-driven approach would address many of the criticisms cited in this paper, it lacks any real control over the language learning goals since the curriculum is randomly shaped by whatever features happen to occur in the texts selected. Willis (2003: 223), however, attempts to systematise this approach through his notion of the *pedagogic corpus*. He suggests that the texts chosen for inclusion in a syllabus are analysed for coverage of key lexical items (based, for example, on corpus frequency lists) so that words that don't arise naturally can be included in supplementary materials.

A second possible way forward is along the lines of van Ek's (1986) 'framework for comprehensive foreign language learning objectives' and Celce-Murcia, Dörnyei & Thurrell's (1997) 'principled communicative approach'. This would use current models of communicative competence to structure the syllabus, an approach that often favours authentic materials because of their ability to illustrate a broader range of competencies, but does not discount contrivance. Noticing features in the input would continue to be crucial in this kind of approach, but rather than limiting ourselves to predominantly lexicogrammatical items, the focus would broaden to encompass all aspects of communicative competence – this is the approach adopted in the experimental design reported here. The fundamental question facing us, then, is: *What should we get learners to notice in the target language?* With an ever-expanding number of features vying for inclusion, but no more class time to teach them, curriculum design is destined to become increasingly complicated and solutions are more likely to be found at the local level rather than through globally published textbooks.

CHAPTER 3. RESEARCH METHODOLOGY

The broad aims of the investigation described in this thesis stem from insights emerging from discourse and conversation analysis over the last few decades. It is based on a belief that current classroom materials could benefit from being informed by a 'language as discourse' approach, which 'enables us to be more faithful to what language is and what people use it for.' (McCarthy & Carter 1994: 201). As discussed in chapters 1 & 2, it would seem that the best way to remain true to what language is and how it is used by real people in real contexts would be to exploit authentic materials in the classroom and to highlight relevant features of the language from these texts. But this is all theory – the recommendations of applied linguists hunched over their computers in cramped university offices. Researchers, who rarely have contact with real classrooms and are, understandably, stronger on theory than practical applications. There is clearly a need then, to test these theories out in genuine classrooms. But how can we design a research framework that can provide answers to the questions we are asking?

A survey of the literature available on research methodology reveals a startling number of pit-falls, problems, compromises and contradictions, which leave the reader with the distinct impression that he or she is entering an area difficult to emerge from professionally intact, rather like Indiana Jones going into the cave in Raiders of the Lost Ark. But, of course, we do not really have much choice if we want to test our theories out. The best approach would seem to be to engage in research with our eyes wide open to potential difficulties, to operate a 'best practice' policy which pre-empts as many potential criticisms as possible through careful design of the research framework. The discussion below outlines the issues relevant to an investigation of the effects of authentic

materials on learning and makes some recommendations as to the best way to approach this type of research.

3.1 The Quantitative, Qualitative Debate

Traditionally, research (in the 'human sciences' at least) has been composed of two approaches: the first approach, which is quantitative and aligns itself with the 'pure sciences' is known as psychometry and is defined by the Concise Oxford Dictionary as 'the science of measuring mental capacities and processes'. The second, which is qualitative, is usually termed ethnography and 'involves the study of the culture/characteristics of a group in real-world rather than laboratory settings. The researcher makes no attempt to isolate or manipulate the phenomena under investigation, and insights and generalisations emerge from close contact with the data rather than from a theory of language learning and use.' (Nunan 1992: 55)

Largely as a result of the historical dominance of the pure sciences, psychometry, with its emphasis on measuring processes, has often been valued more than ethnographic approaches, as Konrad Lorenz recognised in 1971:

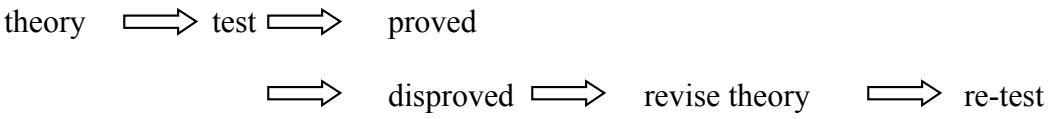
> We are living in a time in which it has become fashionable to assess the exactitude, and with it the value, of any scientific result by the extent to which quantitative methods have taken part in producing it.

A further perceived advantage of psychometry is its focus on theory, which as van Lier (1984: 112) notes, is held in higher esteem than more practical, 'hands-on' approaches:

The terms ‘theoretical’ and ‘applied’ carry connotations of class distinction, of an academic pecking order, where ‘applied’ attaches to a more blue-collar variety of research worker, and ‘theoretical’ is identified with the academic upper crust. This can lead to educational researchers not deeming it necessary to go into classrooms themselves, or leaving the more ‘menial’ tasks of data collection, transcription, coding, and quantification to junior colleagues or secretaries.

Recently, however, there has been a noticeable shift in attitude (in educational research at least) towards a greater appreciation of the value of both approaches. There has also been a growing realisation that making black-and-white distinctions between quantitative and qualitative approaches is an oversimplification (Nunan, 1996) since ‘researchers in no way follow the principles of a supposed paradigm without simultaneously assuming methods and values of the alternative paradigm’ (Reichardt and Cook 1979, cited in Chaudron 1986: 709).

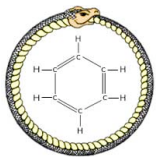
I agree with this view that the distinction represents a false dichotomy: proponents of the quantitative approach would like to imagine that they preserve the objectivity of their investigations by developing theories first, then testing them out under controlled conditions to prove or disprove them. They therefore see the process of discovery as proceeding in the following way:



In this sense, it can be regarded as theory *testing* whereas ethnography is more concerned with theory *forming*:

Ethnography places great store on the collection and interpretation of data, and in marked contrast with the experimental method, questions and hypotheses often emerge during the course of the investigation, rather than beforehand. This is anathema to the proponents of experimental approaches to research. (Nunan 1992: 56)

But where do the psychometrician's theories come from in the first instance? Often, from hunches or intuitive guesses after prolonged exposure to data: in other words, in much the same way as theories are formed in ethnographic approaches. This applies just as much to the 'pure sciences' as to the 'human sciences'. In the 1800's, Freidrich Kekule, a German chemist, was puzzled by the structure of benzene, a hydrocarbon extracted from oil. We now know that benzene is a 6-carbon hexagonal ring, but at the time of Kekule's flash of inspiration, this was not realised. One day, while dozing in front of his fire, Kekule dreamt of snakes chasing one another. Suddenly, one of the snakes twisted round and took hold of its own tail and he awoke with an answer to the puzzle. Kekule described his dream in his diary:

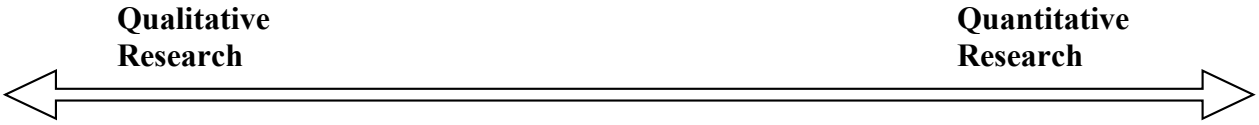


[….] I was sitting writing on my textbook, but the work did not progress; my thoughts were elsewhere. I turned my chair to the fire and dozed. Again the atoms were jumbling before my eyes. This time the smaller groups kept modestly in the background. My mental eye, rendered more acute by the repeated visions of the kind, could now distinguish larger structures of manifold conformation; long rows sometimes more closely fitted together all twining and twisting in snake-like motion. But look! What was that? One of the snakes had seized hold of its own tail, and the form whirled mockingly before my eyes. As if by a flash of lightning I awoke....

(Retrieved 22/9/07. http://p-i-a.com/Magazine/Issue19/Physics_19.htm)

So although both the psychometric and ethnographic approaches can be equally intuitive, they differ in the way they seek to validate their theories. Quantitative approaches attempt to control all the variables in an experiment, to prove a causative relationship between two things by keeping everything else constant. They then attempt to convince the reader of the validity of their theories through measurement of the changes and statistical manipulation. Qualitative approaches, on the other hand, prefer to allow the theories to reveal themselves naturally from, often, intense and prolonged contact in the field – Shirley Brice Heath (1983), for example, spent almost a decade living, working and playing with the families of two communities, Roadville and Trackton, in the Piedmont Carolinas before reaching her conclusions. The reader is persuaded of the validity of the writer's theories by sheer weight of detail and exemplification.

It would seem more productive then, to see purely quantitative or qualitative approaches as being the two extremes of the same continuum, each seeking to reach the truth but through different means (Onwuegbuzie & Leech 2005). The characteristics of each are represented below:



1. Naturalistic (non-interventionist)	1. Experimental (interventionist)
2. Theory forming	2. Theory testing
3. Large amounts of data collected	3. Limited data collected
4. No pre-conceived ideas ('open technique')	4. Ideas pre-formed ('closed technique')
5. Small amounts of total data included	5. Usually, all data collected is included in final account
6. Difficult to analyse independently (low internal reliability)[4]	6. High internal reliability
7. Difficult to replicate (low external reliability)[5]	7. High external reliability
8. Subjective evidence	8. Claims to be objective
9. Contexts resemble those the researcher wants to generalise to (higher external validity)[6]	9. Contexts unlike those generalised to (lower external validity)
10. Intervening variables mean causal relationships cannot be ascribed (lower internal validity)[7]	10. Variables carefully controlled (higher internal validity)
11. Focuses on the social context of learning	11. Blind to social context

(Mehan, 1979; Beretta, 1986; van Lier, 1988; Nunan, 1991; Alderson & Beretta, 1992; Johnson, 1995)

Figure 3.1 Characteristics of qualitative & quantitative approaches to research

From figure 3.1 it can be seen that neither a qualitative nor a quantitative approach is without its problems when it comes to classroom-based research. When internal validity is strengthened in a quantitative approach (by controlling variables), external validity is weakened and the results become less generalisable. But if we attempt to replicate the

[4] Reliability: The consistency of the results obtained from a piece of research.
Internal reliability: The consistency of the data collection, analysis & interpretation in the research.
[5] External reliability: The extent to which a piece of research can be reproduced by an independent researcher.
[6] Validity: The extent to which a piece of research investigates what it purports to investigate.
External validity: The extent to which the results are generalisable to the larger population.
[7] Internal validity: The extent to which differences found can be ascribed to the experimental treatment.
(Nunan, 1992: 14)

classroom context more faithfully, the large number of intervening and uncontrolled variables between those we are interested in means that the internal validity is threatened (Beretta 1986).

To stick dogmatically to one or the other approach is likely to lead to weaker results, particularly with respect to educational research where it is impossible to control all the variables. For example, when exploring the effects of authentic materials on learners' developing communicative competence, it would be possible to design a laboratory experiment where an experimental group received authentic input over a given time and a control group received contrived input. Both groups could then be tested to discover how their communicative competence had changed, or how enjoyable they had found the exercises, but this approach could be criticised for being quite unlike genuine classrooms, thereby lacking external validity, or for being blinkered by pre-conceived ideas. Equally, it would be possible to design an ethnographic study, following the implementation of a syllabus using authentic materials with a particular class, but this could be criticised for its lack of a comparison control group or its subjective interpretation of the data. A far more sensible solution to experimental design would therefore seem to be 'mixed methods' approach, which seeks to measure as much as possible, from as many perspectives as possible, a point made by a number of researchers:

> As for the practical matter of figuring out the best way to test some particular hypothesis, the best way to test it is to test it in every relevant way possible: questionnaires, interviews, large-scale electronic searches, analysis of natural behaviour. Yes, it's an investment, but it is bound to be a worthwhile one; if the results of the tests are not all consistent, you still learn something; namely, that the tests are not all testing the same thing. (Green 1995: 14)

Yet despite exhortations for a more eclectic approach, classroom-based research rarely takes this form (Nunan 1992: 52). When a mixed methods approach has been taken, it inevitably leads to a better understanding of the situation than could have been achieved otherwise (see for example Schmidt & Frota 1986; Lett & Shaw 1986; Lynch 1992) so it is difficult to understand why it has not been adopted more extensively, unless it relates to the 'investment' it entails, as alluded to by Green (1995) above.

So what might a mixed methods approach involve when investigating the effects of authentic materials on learners' developing communicative competence? The first thing to note is that a true experimental approach is impossible with classroom-based research. Subjects are not normally randomised before they are grouped, which weakens the statistical generalisations that can be made to the wider population. Where a control group and experimental group are compared, it is difficult for the researcher to ensure that all the variables, except the ones being investigated, are kept constant since each classroom is composed of individuals who together create a unique environment. In my own teaching at university in Japan, I am often required to teach the same material to more than one class and rarely do they unfold in identical ways. However, by careful control of the variables (learner age/nationality/language ability, learning context, class size, teacher, etc.) and by making explicit those factors which are less well controlled, we might feel that it is worthwhile carrying out a study of this nature. In these situations, the investigation would be termed *quasi-experimental*.

When investigating the effects of authentic materials on learners, a quasi-experimental approach would seem valuable because we might expect to see quantifiable differences between an experimental group receiving authentic input and a control group receiving

standard, textbook input. But at the same time as setting up a study to test a pre-conceived hypothesis, there is no reason why a more ethnographic approach cannot be adopted too. This would involve careful observation of both experimental and control classes, using a range of ethnographic tools: video/audio recording, transcription of classroom interaction, learner diaries, case studies, interviews, and so on. It would also involve resisting the temptation to come to premature conclusions, instead allowing the data to speak for itself, for hypotheses to emerge gradually after intense contact with the data. The mixed methods approach thus requires the researcher to have a 'psychometric head' and an 'ethnographic head' which are quite distinct and never really come together until the post-study stage when the results are contrasted and reconciled.

Once the research context has been examined from a variety of perspectives, the results need to be related to each other in a process known, alternatively, as *triangulation* (Denzin 1970), *the principle of convergence* (Labov 1972), *progressive focusing* (Parlett & Hamilton 1972) or *interactive methodology* (Cicourel n.d.) (cited in van Lier 1984: 120). This allows researchers to enhance the credibility of their work. I shall use Denzin's terminology, triangulation, for the remainder of this discussion. Denzin (1970) describes four types of triangulation:

a) Data triangulation where different sources of data (teachers, students, researchers etc.) contribute to the investigation.
b) Theory triangulation where different theories are applied to a study.
c) Researcher triangulation where more than one researcher contributes to the investigation.

d) Methods triangulation where more than one method is used to collect data (test scores, diaries, questionnaires, classroom observation, etc.).

While it might not be possible for all of these to be incorporated into a single investigation, multiple triangulation (van Lier 1988), where feasible, should be used to enhance the study's credibility: 'the greater the triangulation, the greater the confidence in the observed findings' (Denzin 1970: 472). In this study, some elements of all four types of triangulation were incorporated:

a) Data triangulation: with learner, teacher and researcher participation (although both the teacher and researcher roles were largely borne solely by the author).
b) Theory triangulation: with quantitative and qualitative approaches considered, as well as a number of relevant theories on second language acquisition.
c) Researcher triangulation: with native English-speaking teachers conducting blind evaluations of students' performances in the discourse completion task, oral interview and student-student role-play.
d) Methods triangulation: with a wide variety of methods exploited to collect data, including pre/post-course tests designed to measure different aspects of communicative competence (listening, pronunciation, 'C' test, grammar, vocabulary, discourse completion task, oral interview & ss-ss role play), learner diaries, an initial Personal Learning History questionnaire, case studies, likert scales and video/audio recording & transcription of student discourse.

However, triangulation is only one of the justifications given for adopting a mixed methods approach in the literature. Bryman (2006: 105-107) gives a comprehensive list of possible reasons, based on a review of 232 social science articles:

a) *Triangulation*: quantitative and qualitative research is combined to triangulate findings and give greater validity.

b) *Offset*: combining quantitative and qualitative research allows researchers to offset the weaknesses and draw on the strengths of both.

c) *Completeness*: combining quantitative and qualitative research allows researchers to give a more comprehensive account of events.

d) *Process*: qualitative research gives a better sense of process.

e) *Different research questions*: quantitative and qualitative research allows researchers to answer different questions.

f) *Explanation*: findings from one approach are used to explain findings from the other.

g) *Unexpected results*: surprising results from one approach can be understood by employing the other approach.

h) *Instrument development*: qualitative research is employed to develop better research instruments such as questionnaires and scale items.

i) *Sampling*: one approach is used to facilitate the sampling of respondents or cases.

j) *Credibility*: combining quantitative and qualitative research enhances the integrity of findings.

k) *Context*: qualitative research provides contextual understanding while quantitative research provides generalizable, externally valid findings.

l) *Illustration*: qualitative data helps to illustrate quantitative findings.

m) *Utility*: combining quantitative and qualitative research improves the usefulness of findings to practitioners.

n) *Confirm and discover*: qualitative data is used to generate hypotheses, which are then tested using quantitative approaches.

o) *Diversity of views*: researchers' and participants' differing views can both be incorporated by combining quantitative and qualitative research.

p) *Enhancement*: the findings from one approach can augment or enhance the findings of the other approach.

Many of these justifications will apply equally in any individual example of mixed methods research ~ as is the case in this investigation.

3.2 Classroom-Based Research

'*the classroom is the crucible – the place where teachers and learners come together and learning, we hope, happens*.' (Gaies 1980, cited in Allwright & Bailey 1991: 18)

The analogy of the classroom as a crucible is, undoubtedly, an apt one; in a crucible metals or ores are combined and heated to give a product. But its final form will depend on a large number of variables: the metals used and the impurities they contain, the temperature, the gases present during smelting and so on. Similarly in the classroom, the final product (learning) is affected by a large number of interacting variables but where the analogy breaks down is that the results of the smelting process can be accurately predicted because they deal with inorganic compounds. When it comes to studying human behaviour, there are no such certainties. So what is it about classrooms that makes them such difficult places to investigate? Below, I outline some of the complicating

factors, which need to be taken into consideration when planning and executing classroom-based research.

3.2.1 The multi-faceted nature of the classroom

Classrooms are complex places. There are so many variables at play: the context in which learning takes place; the participants (teachers and learners); their previous experiences, beliefs about, and reasons for, learning; the materials used to learn with; the methods employed to exploit those materials and the resulting interaction in the classroom. All of these factors intertwine to make each classroom unique: any one set of materials can be interpreted and exploited in an infinite number of ways by the individuals who make up a class:

> […] although a syllabus may be written down for (teachers), it inevitably becomes shaped by them into something personal which reflects their own belief systems, their thoughts and feelings about both the content of their lessons and their learners, and their view of the world in general. In addition to this, the curriculum that they actually deliver becomes itself interpreted in different ways by their learners, so that the whole learning experience becomes a shared enterprise. (Williams & Burden 1997: 21)

This has been explored in classroom research by Allwright (1984) and more recently Slimani (1992) who distributed an 'Uptake Recall Chart' to learners at the end of each lesson she observed, asking them to recall, in as much detail as possible, the events of the preceding lesson with particular reference to grammar, words and expressions, pronunciation and spelling. Slimani's results support a constructivist model of language learning and she concludes that, 'learners' uptake is strongly idiosyncratic' (ibid: 206).

So, not only is each classroom event unique in itself, shaped by the participants and materials from which it is created, but the way that each event is perceived and exploited by individuals for learning will also be unique. Although it will forever be impossible to give a full and complete account of what exactly has taken place in a class, an ethnographic element to classroom-based research at least recognises and attempts to face these issues.

Historically, the response from researchers to the complexity of the learning context has been one of two options. The first is to ignore classroom interaction altogether, to regard it as a 'black box' (Long 1983) between input and output:

> It seems to me that a great deal of research in our field is conducted in contexts where classroom noise either is unheard or is considered irrelevant and therefore removed from the equation before the numbers are added up and their significance determined. This lack of contact with the reality of the classroom has driven a wedge between researcher and practitioner which threatens to become a gulf unless steps are taken to bridge it. (Nunan 1996: 42).

Nunan (1991) reports on a survey he conducted into the methods used in fifty classroom-oriented studies. He found that only thirty per cent were carried out in genuine language classrooms.

The second option is to limit classroom investigations to those events that are easiest to measure:

> […] the overall picture we have of classroom language learning from research so far is already distorted by this bias towards the visible. We only know about what we have looked at, and what we have looked at

over the last two decades consists largely of whatever has been easiest to observe. (Allwright & Bailey 1991)

But of course any research which attempts to gloss over these complexities is unlikely to get anywhere near the truth and so the only option open to the responsible researcher would seem to be an attempt to recognise these different factors, to document them all as far as possible and to try to reconcile them at the end of the investigation:

At some point all these factors [setting, content, interaction, participants, method] must be taken into account, for all are relevant, many are related, and as yet we know little about their potential contribution to L2 language development. In the classroom they all come together and produce the undefinable quality, the dynamics of classroom work. It is clear that, unless we are to oversimplify dangerously what goes on in classrooms, we must look at it from different angles, describe accurately and painstakingly, relate without generalizing too soon, and above all not lose track of the global view, the multifaceted nature of classroom work. (van Lier 1988: 8)

Essentially, what is being recommended here is an integrated, mixed methods approach exploiting both quantitative and qualitative dimensions, as alluded to earlier, which brings us to a further problem with classroom-based studies.

3.2.2 Mixed methods approaches are time-consuming

Whilst psychometric testing, which focuses on the *products* of learning, is generally quite straightforward to implement in the classroom, ethnographic investigations, looking at the *process* of learning, are usually much more time consuming. They involve techniques such as observations of classes, audio and video recording, transcription of classroom

interaction, diary-keeping, case studies and interviews. All of these activities can generate a huge amount of data, which then has to then be analysed and interpreted. With limited resources available, the researcher has to make critical decisions about what to focus on and in how much depth, a delicate balance to maintain:

> Comprehensiveness must be seen in conjunction with exhaustiveness, or, in other words, breadth must be related to depth. The purpose of the exercise will determine whether breadth or depth is emphasized. One will always be stressed to some extent at the expense of the other. I can spread my jam very thinly over a huge slice of bread, or heap it onto a tiny morsel. (van Lier 1984: 127)

The research reported on here aims to investigate the effects of authentic materials on learning which, in order to be given a real chance of success, needs to be carried out over a long time period – with university students, a one-year trial is the most obvious choice. Furthermore, in order to be able to claim any link between the experimental materials and their effects on learning, it is necessary to have a control group to compare with. All of these requirements, combined with the demands of a mixed methods approach, mean that the investigation will inevitably lie closer to the 'comprehensive' end of van Lier's continuum. However, if the process of data collection has been thorough, there is no reason why the material cannot be 're-visited' at a later date by other researchers who might wish to look in more depth at a specific issue, or to validate the study's conclusions for themselves.

It should be clear by this stage that the designing of effective quantitative/qualitative, classroom-based research requires considerable thought and preparation. This point is also made by Alderson & Beretta (1992: 98), commenting on Lynch's REST project

(Reading English for Science and Technology), one of only a handful of investigations to have employed an integrated approach at that time:

> […] a quasi-experimental, control type of design is very difficult to realise in practice [...] At least it is clear that very careful attention needs to be paid to the detail of the design, the practicality of plans for administration of instruments and the gathering of data, long before any programme can be evaluated.

With this in mind, it is clearly important to trial the chosen instruments before the main investigation begins.

3.2.3 Transcription of classroom interaction

The production of high quality transcriptions of classroom interaction involves a considerable investment of time and energy. van Lier (1988: 241) estimates about 20 hours of work to produce a transcript for one lesson and Johnson (1995) concurs with this:

> Transcribing native-speaker dyads normally takes about five times the length of the interaction [...] It takes even longer to transcribe the speech of pairs or groups of non-native speakers interacting in their second language. Transcriptions of classroom interaction, where there are large numbers of speakers whose voices and accents may be similar, where voices often overlap, and where some speakers will be heard more clearly than others, can be very time-consuming indeed. (In our experience, one hour of language classroom data can take up to twenty hours to transcribe accurately).

So is it really worth all the effort? Many researchers believe it is:

> […] verbatim transcripts, which display all the hesitations, false starts, pauses and overlaps of natural speech, are extremely valuable records of interaction. Transcripts show us, in ways that coded data and frequency counts often mask, how classroom interaction develops, as a dynamic phenomenon. (Allwright & Bailey 1991)

van Lier (ibid) lists a number of reasons why transcriptions are so valuable:

a) Recording and transcription of classroom interaction acts as an estrangement device. Classrooms are such familiar environments for us all, it is very difficult to evaluate what we see objectively but through transcription we are able to become one step removed from the discourse and to see it with fresh eyes:

   > […] to frame the questions and answer them we must grope towards our invisible knowledge and bring it into sight. Only in this way can we see the classroom with an outsider's eye but an insider's knowledge, by seeing it as if it were the behaviour of people from an alien culture. Then by an act of imagination we can both understand better what happens and conceive of alternative possibilities. (Barnes 1975: 12-13)

b) To understand classroom interaction requires 'intensive immersion in the data'; a necessary part of the transcription process. Interesting phenomena often only reveal themselves after detailed investigation of the data.

c) Transcription of classroom interaction allows other researchers to analyse the primary data for themselves and confirm or refute claims made by the original authors. This can help to strengthen the internal validity and reliability of the study.

d) Transcribed data from one study can be compared with that from other studies, which leads to 'cumulative research'. (Classroom-based investigations are often criticised for not being explicit enough).

e) It permits evaluation of the entire interaction in the context in which it occurred.

f) Small amounts of transcribed data can lead to impressive insights, as Harvey Sacks demonstrated in his paper 'On the analysability of stories by children' in 1972, using just two utterances. For a quasi-experimental study stretching over a long period of time (such as the one reported on here), it is clearly impossible to transcribe and analyse more than a tiny percentage of the interaction, even if it has all been recorded on audio or videotape. The best that can be hoped for is to produce a fair representation of the discourse by careful selection of extracts from various stages of the course.

Once classroom interaction has been transcribed, it can be analysed in different ways. How this is done depends on which of three related research traditions informs the process: discourse analysis, conversation analysis or interaction analysis. Nunan (1992: 160) summarises the essential differences between these three approaches:

	Discourse analysis	**Conversation analysis**	**Interaction analysis**
Method of generating data:	Invented Elicited Naturalistic	Naturalistic	Elicited Naturalistic
Mode:	Spoken / written	Spoken	Spoken
Type of analysis:	Categorical	Interpretive	Interpretive
Units of analysis:	Linguistic	Non-linguistic	Linguistic & non-linguistic

Table 3.1 Summary of differences between discourse, conversation & interaction analysis

Since the transcribed data from my research originates from (naturalistic) classroom interaction, it is presumably open to interpretation from any of the three types of analysis

mentioned above. Discourse analysis tends towards a psychometric approach, categorising and quantifying various discourse features in the interaction while conversation and interaction analysis are more ethnographic in nature. As such, each has its strengths and weaknesses:

> DA [discourse analysis] theorists can accuse CA [conversation analysts] of being inexplicit, or worse, plain muddled, about the theories and conceptual categories they are employing in analysis […] CA practitioners can retort that DA theorists are so busy with premature formalization that they pay scant attention to the nature of the data. The main strength of the DA approach is that it promises to integrate linguistic findings about intrasentential organization with discourse structure; while the strength of the CA position is that the procedures employed have already proved themselves capable of yielding by far the most substantial insights that have yet been gained into the organization of conversation. (Levison 1983: 287)

The research literature is full of examples of transcripts being exploited in quantitative and qualitative ways to help illuminate features of interaction. Lennon (1990) investigated the developing fluency of four subjects during a period of study in Britain by examining transcripts and quantifying: a) Words per minute (including repetition); b) Words per minute (excluding repetition); c) Repetitions; d) Self-corrections; e) Filled pauses; f) Repetitions and self-corrections as a percentage of total speech; g) Unfilled pause time as a percentage of total speech; h) Filled pause time as a percentage of total speech; i) Mean length of speech 'runs' between pauses (in words); j) Percentage of T-Units followed by a pause; k) Percentage of total pause time at all T-Unit boundaries; and l) Mean pause time at T-Unit boundaries. Lennon found that all the subjects paused less, spoke faster (words per minute) and had longer speech runs between pauses at the end of their stay.

Schmidt's (1983) well-known case study of Wes, a Japanese artist living in the USA, used a series of transcribed taped conversations and monologues to investigate Wes's developing discourse and strategic competence. Schmidt's approach is more ethnographic in nature but he demonstrates quite convincingly Wes's competence in these areas. On the subject of Wes's discourse competence, he comments:

> The good-natured' teasing type of humour of this passage [...] is typical of Wes's conversations, as is his skill in listening to what people say and picking up topics for further development. Wes is not a passive conversationalist but nominates topics frequently. Moreover, the topics he nominates are almost always relevant to previous topics. I have never observed any instances of conversation coming to a halt because Wes has raised a topic (or commented on a topic already on the floor) in a way that indicated he had not understood what the previous speaker had said or made an unfathomable connection to a new topic. (Schmidt, 1983: 160)

To substantiate this claim, he offers the following transcript as an example:

M:	I would like eggs benedict (to waitress) / that's the specialty (to Wes)
Waitress:	how about you?
Wes:	here eggs benedict is good?
M:}	yeah
G:}	it's the specialty
Wes:	yeah? / OK / I have it (waitress leaves)
M:	you never ate before?
Wes:	no, I ate before / but not this hotel
M:	it's very good over here
Wes:	but only just English muffin / turkey / ham and egg / right?
G:	right
Wes:	so how different? / how special?

M: because it's very good here / maybe it's the hollandaise / I don't know

G: maybe it's just the atmosphere

Wes: yeah / I think so / eggs benedict is eggs benedict / just your imagination is different / so / this restaurant is belong to hotel?

G: no / not exactly

(Schmidt 1983: 159-60)

Schmidt also illustrates effectively Wes's developing strategic competence through his ability to repair communication breakdown (which is quite common due to his limited linguistic competence) through transcribed data:

Wes: Doug / you have dream after your life?

NS: Whaddya mean?

Wes: OK / everybody have some dream / what doing / what you want / after your life / you have it?

NS: you mean after I die?

Wes: no no / means next couple of years or long time / OK / before I have big dream / I move to States / now I have it / this kind you have it?

NS: security I suppose / not necessarily financial / although that looms large at the present time

(ibid: 165)

It would seem sensible, therefore, for researchers to take an eclectic approach to the analysis of transcribed data and to exploit it in any way that illuminates the processes at work during interaction. However, it must be remembered that transcribing interaction is not an objective process:

The acts of observing, transcribing, and any form of labelling or coding, being of necessity selective, involve interpretation. The interpretation therefore does not start when observation and descriptive work are completed, it pervades the entire activity. (van Lier, 1984:126)

Because of this, it is important for researchers to be as explicit as possible about the whole process.

3.3 Adopting a methodological 'frame of reference' for an investigation

A discourse analysis of classrooms is basically about language, and the nature of research will be largely determined by the researcher's views about the nature of language-in-use. It is important, for the relevance and clarity of any study, to be as explicit as possible about these views, which the researcher carries with him/her as basic assumptions. Together with views concerning the proper conduct of scientific activity, here referred to as research methodological principles, they can be formulated as a methodological frame of reference [...] The frame of reference may enable 'consumers' of the 'product' of the research to see why the researcher chose to focus on certain aspects of the observed data more than on others [...] and why the chosen phenomena are described in the way that they are described. (van Lier 1984: 119/20)

So far in this chapter, I have tried to follow van Lier's very sensible advice and to be as explicit as possible about the methodological principles underlying my research. I now turn to the assumptions underlying the work and provide an explanation for the choice of constructs used to explore learners' developing communicative competence with different types of input.

The central belief underlying the research is that traditional language courses based on a structural syllabus tend to provide learners with an impoverished input. Through excessive control of lexico-grammatical items and contrivance, many of the features of

natural language are lost and I propose that this inhibits learners' L2 development. Authentic materials (particularly audio-visual ones), whilst encumbered with different sorts of disadvantages, provide richer input with greater contextualisation, visual and acoustic information, and samples of real language used by native speakers with all of their hesitations, repetitions, false starts, interruptions and misunderstandings; their fights, gossip, jokes, flattery and deceit. As well as often being more interesting, it is suggested that these types of materials are better able to develop students' communicative competence.

The next issue to resolve is how to operationalise this hypothesis. Clearly, the way we choose to measure learners' development of communicative competence will affect our results and whether we consider the treatment successful or not. Cohen (1997), for example, participated in a semester-long accelerated Japanese course in which his performance according to the classroom syllabus was very successful (he was placed top in his class). However, pragmatically he felt that he had made little progress: something that was not revealed in the class because the teacher's structural, rote-learning approach was insensitive to this particular feature of language.

In this study, I have chosen to measure learners' L2 development with reference to Celce-Murcia, Dörnyei & Thurrell's (1995) model of communicative competence (based on Canale & Swain 1980), which is made up of five components: linguistic, pragmalinguistic, sociopragmatic, strategic and discourse competences. This model provides a useful framework for assessing the learners' progress because it considers a broad range of 'knowledge areas', necessary for natural and effective communication. Most teachers would probably agree that to have students with competence in all five

areas would be a desirable outcome for a language curriculum. It is also useful because there is evidence of learners developing competence in these areas in variable ways, depending on their learning context. Of course, there is a certain amount of inter-connectedness amongst the different components in the communicative competence model, and improvements in one area are likely to have knock-on effects on others (see chapter 1).

Given the established link between the input or context of learning and a variable development in different types of communicative competence (although, to my knowledge, research has so far only focused on pragmatic and linguistic competence), we might expect measures such as these to be sensitive to post-treatment differences between the control group and experimental group of this study.

The five components of communicative competence, in addition to learners' language skills, were measured both pre- and post-course using the following instruments:

a) **Linguistic competence:** Pronunciation test; Grammar test; Vocabulary test; C-test
b) **Strategic competence**: Oral interview; Student-student role-play
c) **Pragmatic competence (pragmalinguistic + sociopragmatic):** Discourse completion task; Oral interview; Student-student role-play
d) **Discourse competence:** Oral interview; Student-student role-play
e) **Language skills:** Listening test; Oral interview; Student-student role-play; C-Test.

Vocabulary, grammar and pronunciation tests are all relatively uncontroversial so I do not intend to discuss them in any detail here, but it is worth spending a little time looking at the C-test and discourse completion tasks (DCTs). The C-test is a similar to a traditional cloze test except that it involves deletion of the second half of every second word, starting and ending with an intact sentence (see Klein-Braley & Raatz 1984). Dörnyei & Katona (1992) compared the C-test with a standard cloze test and found it to be a superior measure of general language proficiency, particularly with more homogeneous groups (as Japanese university classes tend to be). It was found to be a reliable and valid instrument, and avoids some of the criticisms leveled at cloze tests, such as their difficulty to score objectively and the fact that the choices available to the test designer (choice of text, deletion rate, starting point for deletion & scoring method) have considerable effects on the scores (Dörnyei & Katona, ibid).

Discourse completion tasks are widely used in pragmatics research as measures of pragmalinguistic competence. Learners have a situation described to them (such as bumping into an old lady in a supermarket) and have to say or write down how they would respond. Elicitation devices have included cue cards, role-play scenarios or video clips, and responses can be written (open-ended/fill-in-the-blanks forms, multiple-choice, ranking exercises) or spoken (recorded and transcribed). Both spoken and written forms of the DCT have advantages and disadvantages:

> The advantage of written surveys is that they can be administered efficiently, and a large amount of data can be amassed quickly. The disadvantage is that because it is more tiring to write than to speak, responses are likely to be shorter, and may be less carefully considered. Oral interviews, with responses tape-recorded, take longer to administer, and require an investment of time and effort to transcribe, but the

results are more likely to contain richer responses – ones which consider the question in greater depth, and from multiple perspectives. (Green 1995: 13)

Obviously, since DCTs are testing a learner's ability to make socioculturally appropriate responses, it is important that the context is crystal clear ~ variables such as age, social distance, power, and imposition will all affect the appropriateness of replies. In addition, measures need to be taken to ensure that each context is interpreted similarly by respondents from the cultures being investigated (in this case, Anglo-Americans and Japanese). As Bardovi-Harlig (1999: 242) points out, 'conveying just the right amount of information to respondents is not always easy'. Too little contextual detail and respondents use their imaginations to fill in the missing information, effectively answering different questions, too much and their linguistic resources may be stretched or they may become bored. She recommends the use of film for eliciting responses because 'it is one way to increase the likelihood that everyone is responding to the same scenario. It increases the richness of the scenario while avoiding the pitfalls of taxing the linguistic ability or the patience of the respondents.' (ibid: 243)

Hudson, Detmer & Brown (1995) advise researchers to design DCTs in which the respondents play themselves, in familiar contexts which avoid reference to family members. It would also seem sensible to assess a variety of speech acts in order to evaluate learners' sociocultural competence fairly; Kasper & Schmidt (1996: 154) report that ILP studies to date have found no speech communities which lack speech acts for requesting, suggesting, inviting, refusing, apologising, complaining, complimenting and thanking so perhaps these could act as a basis for DCT design. One of the earliest attempts to design a rating scale for sociocultural competence was by Cohen & Olshtain

in 1981. They concluded that their efforts had, 'produced at best a crude measure of such competence.' (ibid: 130). It is clear then that the validity and reliability of these instruments is less well established than tests for grammatical or lexical proficiency. However, since they are the only measures available to tap this area of competence, they should not be ignored.

One of the problems of having a comparative study with two groups receiving different treatments is the difficulty of devising tests which are equally fair to both programs. Beretta (1986) addresses this issue in his paper 'Program-Fair Language Teaching Evaluation' where he outlines five strategies available to the researcher to investigate the effects of different programs: a) Standardized tests; b) Specific tests for each program; c) Program-specific and program-neutral measures; d) Common/unique objectives identified for each program; and e) Appeal to consensus. As usual, there are advantages and disadvantages to each of these approaches and these are summarised below:

Strategy	Advantages	Disadvantages
Standardised test	Known characteristics, may be impartial.	May be insensitive to the features of a particular program.
Specific tests	Focus on test validity.	Direct comparisons between groups often impossible.
Program-specific & program-neutral tests	More integrated approach.	Complicated to administer & interpret.
Common/unique objectives	More integrated approach.	Goals for each program may not be amenable to measurement.
Appeal to consensus	May be impartial.	Prevailing consensus on language pedagogy may change.

Table 3.2 Advantages & disadvantages of different strategies for evaluating programs (Beretta 1986)

In this investigation, the strategy adopted is essentially 'appeal to consensus'; what is being tested is external to the claims of either program and is based on a generally accepted model of communicative competence (see Frolich, Spada & Allen 1985 for a similar type of experimental design). Within this framework, specific tests were selected to test different aspects of communicative competence. These avoid the criticism often leveled at specific tests (i.e. that direct comparisons cannot be made) because in this study, the aims, and therefore the tests, for both groups are identical (to develop the communicative competence of the learners). It should therefore be possible to compare the development of proficiency:

a) Between learners in the control and experimental groups over time;

b) For each individual learner over time.

As mentioned earlier, this investigation attempts to take an integrated approach, which involves the use of both quantitative and qualitative methodologies. In taking an ethnographic stance, we want to gather as much extra information about the process of learning in the control and experimental groups as possible. In this qualitative approach, we hope to see insights emerge gradually from the data. An important part of this process is to gather information from all the participants in the study; researchers, teachers and learners. By accumulating different points of view about the same event, we can hopefully arrive at a fairer appraisal of what really took place. This is referred to as taking an *emic* or *etic* perspective: an emic approach uses concepts or frameworks derived from within the culture while an etic approach uses theories or belief systems coming from outside the culture (Allwright & Bailey 1991). This is important because the way a particular event is perceived by the participants can vary enormously, particularly

when they come from markedly different cultures (as they do in this study). For example, in this quote taken from Allwright & Bailey (1991: 53), Sun-yu, a Taiwanese student, describes the difficulties she faced in adapting to graduate-level education at an American university:

> When I first came here, I couldn't believe how much Americans talked in class. In Taiwan, students never speak in class unless the teacher calls on them. At first, I was afraid to talk in class because I thought I might ask a question that I should know the answer, or I might say something that was already said. I was afraid that what was interesting for me might not be interesting to the rest of the students. I kept waiting for my teachers to call on me, but they never did. Then I realized that this way of talking was what teachers expected, and so I would have to get used to it. I think I have talked more in classes here than all my years of schooling in Taiwan.

A Western researcher, observing Sun-yu in class, might have jumped to the conclusion that she was rather shy or had no opinions of her own but of course this would have been quite wrong. So by incorporating both emic and etic perspectives in our research, we can achieve a more balanced view. To this end, the following additional instruments were used:

a) An initial 'Personal Learning History' questionnaire to elicit as much background information as possible about each participant;
b) Learners' diaries;
c) Case studies of three learners (strong, average and weak) from both the control and experimental groups.

Diary studies are defined by Bailey (1990: 215) as 'a first-person account of a language learning or teaching experience, documented through regular, candid entries in a personal

journal and then analyzed for recurring patterns or salient events'. Most diary studies investigating second language acquisition have been done by applied linguists or language teachers, investigating their own learning (for example Rivers 1983; Schmidt & Frota 1986; Bailey 1983; Schumann & Schumann 1977) but they are being used increasingly with learners too these days. Although, as Nunan (1992: 118) points out, diaries are 'important introspective tools in language research', collecting useful data from learners is far from easy. Halbach (2000) reports that in his study, learner diaries were often of poor quality, being short and incomplete. In addition, the rate of return was low with less than fifty per cent handing in their diaries (see also Carroll 1994 for similar compliance rates). Richards & Lockhart (1994: 18) comment that 'the discipline of diary keeping was a burden on time and energies of participants'. These problems can be eased slightly by incorporating the diary writing into the lesson, leaving ten or fifteen minutes free at the end for learners and teacher to record their thoughts about what has happened while events are still fresh in their minds, as recommended by Bailey (1990) and Porter, Goldstein, Leatherman & Conrad (1990). Bailey (1983; 1990) and Bailey & Ochsner (1983) recommend a five-step procedure for implementing diary studies research (which was also followed in this investigation):

a) Diaries begin with a full account of the diarists' personal learning history;
b) Diarists should be encouraged to be as candid as possible in their diary entries, despite the possibility of embarrassment;
c) This initial database is then revised for public consumption;
d) Patterns and significant events in the diary entries are identified;
e) The data is interpreted and discussed.

Despite the practical difficulties of implementing diary studies in the classroom, it was a line of inquiry felt worth pursuing in this study because of its potential for tapping into affective and personal factors which influence learning (van Lier 1988). From previous experience of teaching in Japan, I have noticed that Japanese learners are often much more comfortable expressing their feelings privately, in written form and therefore expected this method of enquiry to yield some interesting data. Diary studies should be particularly valuable for illuminating the learners' reactions to authentic and textbook materials and the extent to which they find them interesting, motivating or useful.

Case studies are designed to focus in on one or a handful of learners and to look in detail at their development, or as van Lier (1984) puts it, 'spreading your jam on a tiny morsel'. In this way, they provide a useful contrast to the more product-oriented aspects of the research. Nunan (1992: 88) has this to say about them:

> Despite possible problems of validity and reliability, the case study has great potential as a research method in applied linguistics, and has already established itself in the area of second language acquisition.

They are particularly valuable in this research as supplements to the diary studies, to investigate in more detail the attitudes and impressions expressed in diary entries. To this end, three learners from the control and experimental groups were selected and interviewed on four occasions during the study. In an effort to be representative of the whole group, strong, average and weak students were chosen. There are a number of ways to structure interviews, ranging from completely controlled to completely open:

An unstructured interview is guided by the responses of the interviewee rather than the agenda of the researcher. The researcher exercises little or no control, and the direction of the interview is relatively unpredictable. In a semi-structured interview, the interviewer has a general idea of where he or she wants the interview to go, and what should come out of it, but does not enter the interview with a list of pre-determined questions. Topics and issues rather than questions determine the course of the interview. In the most formal type, the structured interview, the agenda is totally predetermined by the researcher, who works through a list of set questions in a predetermined order. (Nunan 1992: 149)

The unstructured interview then, tends to be more emic in nature, focusing on the learners' own perceptions of events while the structured interview is more etic, revolving as it does around the researcher's agenda. Done badly, both structured and unstructured interviews can yield next to no insights; the unstructured one because the conversation can easily veer off into areas that have no bearing on the investigation and the structured one because the interviewer's questions may restrict the extent to which the interviewee can express their true feelings. It is also worth bearing in mind that the interviewer needs to be sensitive and friendly in order to encourage open and honest responses. van Lier (1989: 499) shows what can happen if they are not in this painful example of an oral proficiency interview:

Interviewer:	Where is your mother? What does your mother do?
Subject:	She's dead.
I:	Ah – she's dead. Very good.
I:	What's your father's name?
S:	[no response]
I:	What does your father do?
	Where does he work? Where does your father work?

Come on girl, talk! Talk! Don't be afraid. Where does your father work?

S: [no response]

I: What do you do at home? Do you help your mother? What does your mother do?

S: [no response]

I: (into microphone) Doesn't talk.

Nunan (1992) notes that the semi-structured interview seems to have found favour with many researchers because it gives interviewees a certain level of control over the direction the interview takes and also gives the interviewer more flexibility. This was the approach taken in this study, with learners' diaries operating as the 'jumping off points' into discussion.

3.4 Conclusion: Avoiding the pitfalls in classroom-based research

At the beginning of this chapter and through the detailed discussion that followed, I have tried to emphasise the difficulties and complexities of classroom-based research. I have also suggested that researchers should enter the process with their eyes wide open to the potential pitfalls. Below is a summary of criticism leveled at previous classroom-based research from the literature with suggested remedies:

Criticism from the literature	Suggested solution
a) Quantitative studies are blind to the social context of learning and results are not generalisable to real classrooms while qualitative studies are overly subjective and lack internal reliability & validity. (Mehan 1979; Beretta 1986; van Lier 1988; Nunan 1991; Alderson & Beretta 1992; Johnson 1995)	a) Use an integrated (mixed methods) approach, which includes psychometric and ethnographic instruments. Collect as much information as possible about all aspects of the investigation and attempt to interpret the results in a way that satisfies the varied data through 'triangulation'.
b) Academics often hide behind complex presentations & gloss over problems encountered in their research. (Dingwall 1984; Hughes 2002)	b) Avoid excessive jargon in reports and give a comprehensive account of the investigation 'warts and all'.
c) Research designs are flawed, researchers have applied inappropriate statistical measures, critical data is not reported or is hidden away. (Ritchie & Bhatia 1996)	c) Design research carefully, including control & experimental groups & ensuring that statistical measures are appropriate. Give a full account of the procedures adopted.
d) Theories about pedagogical matters are rarely based on classroom research. (Long 1980; Chaudron 1988)	d) Carry out classroom-based research to resolve pedagogical issues.
e) Explicit methodological frames of reference not given. (van Lier 1984)	e) Make methodological frames of reference clear in the write-up.
f) Impossible to be both comprehensive & exhaustive. (van Lier 1984)	f) Choose a compromise between breadth & depth, which best satisfies the research questions posed.
g) A lack of detailed accounts of classroom processes. (Chaudron 1986; Nunan 1991)	g) Give as much detail as possible about classroom events.
h) A distorted picture of classroom processes because of a bias towards the visible. (Allwright & Bailey 1991; Johnson 1995)	h) Invest the time & energy necessary to develop classroom-based research, which gives a more comprehensive description of the classroom.
i) Research periods are too short in longitudinal studies, sample sizes are too small, comparison groups / teachers are not equivalent. (Alderson & Beretta 1992; Ritchie & Bhatia 1996)	i) Ensure that time-length for the study is sufficient for changes to be observable, ensure that student numbers are large enough to produce statistically significant results, ensure control & experimental groups / teachers are as similar as possible, give a full account of the research process.

Table 3.3 Summary of problems with classroom-based research & suggested solutions

CHAPTER 4 MAIN STUDY

4.1 Research design

The main study was conducted over a ten-month period, from April 2004 to January 2005, at Kansai Gaidai University (Kansai Institute of Foreign Languages), in Hirakata, Japan. Participants were enrolled on a (compulsory) Communicative English Course, which focused principally on developing learners' listening and speaking skills. Classes were held twice a week in quiet, well-lit classrooms for all groups, each lesson lasting 90 minutes (a total of 82.5 hours of input over the course of the investigation), and were all taught by one teacher (the author). Quantitative data on the students was collected pre- and post course with a batch of eight different tests, while qualitative data was collected at regular intervals during the trial using a variety of techniques: learner diaries, case study interviews and transcripts of classroom interaction. All training and testing took place during scheduled classes, except for the IELTS oral interviews and case study interviews, which were arranged outside of class time by appointment.

4.2 Participants

A total of 92, 2nd year English-major students from 4 separate classes took part in the trial. The classes, identified by the numbers 1, 2, 3 and 4, represented the students with the highest proficiency in the university, streamed by TOEFL score. Classes 1 and 3 were assigned to the experimental treatment and classes 2 and 4 the control treatment. The TOEFL ranges within each class are shown in table 4.1.

Class	TOEFL Range
1	567-520
2	520-503
3	520-503
4	503-493

Table 4.1 TOEFL ranges within each class

Both the experimental and control groups consisted of 46 students each, at the commencement of the study. However, students at Kansai Gaidai University have the opportunity to study abroad from the end of the first term, on the International Exchange Program, and this meant that the mortality rate during the study was high (almost 33%). Significantly, it tended to be the most able and motivated students who dropped out of the trial, since acceptance on the exchange program involves passing a proficiency test and enrollment in extra lessons. Any students who did not participate in the full 10-month trial were automatically excluded, leaving a total of 62 students (31 in the experimental group and 31 in the control group) for the final analysis.

4.2.1 Student profiles

A total of 87 students out of 92 (94.6%) completed the 'Personal Learning History' questionnaire (see appendix I) at the beginning of the study and the data collected was used to build up a more comprehensive picture of the participants. The results of this are shown below in table 4.2.

Gender	
Female	60
Male	27
Age (years)	
Range	19-22
Mean	19.2
Nationality	
Japanese	86
Korean	1
Years studying English	
Range	7-15
Mean	7.9
English level (TOEFL)	
Range	493-567
Mean	514.3
English level (TOEIC)	
Range	570-860
Mean	702.4
Official English input per week (hours)	
Range	5-16
Mean	9.3
Types of classes taken	Communicative English (listening & speaking skills); English II (reading & writing skills); Practical English (TOEIC preparation); Business English; Current English; Area Studies; English Literature; Phonetics
Enjoy studying English?	
Yes	82
Sometimes	3
No	2
Preferred classroom activities? (Ranking)	
Speaking	19.4%
Listening	18.6%
Pronunciation	15.6%
Reading	13%
Vocabulary	12.4%
Writing	11%
Grammar	10%
Most preferred classroom activities?	
Speaking with other students	48
Watching movies	12
Listening activities	7
Games	5
Other responses (reading activities, presentations, tests, writing essays, speaking with teacher, learning vocabulary, cultural information, no response)	15
Least preferred classroom activities?	
Presentations	28
Working alone	16
No response	16
Grammar activities	6
Tests	5
Speaking with other students	6
Other responses	10

(passive lessons, difficult lessons, pronunciation activities, easy material, 'no topic chat', working with students who speak in Japanese)	
What changes would you like to make in the class?	
No response	39
More speaking activities	25
Frequent change of partners	6
No Japanese in class	4
Other responses (smaller classes, more movies, pleasant learning environment, etc.)	13
Have you ever travelled abroad?	
Yes	57
No	30
Have you ever lived abroad? (> 4 months)	
No	78
Yes (USA, Canada, New Zealand, Hungary, Finland)	9
Do you speak English outside of the classroom?	
No	53
Yes	23
Sometimes	11
Future plans?	
Unspecified work using English	21
Live abroad	20
Unsure	16
English teacher	13
Airport work	11
Other (tour guide, interpreter, patent attorney, TV director, service industry)	6

Table 4.2 Summary of responses to 'Personal Learning History' questionnaire

The questionnaire indicates a female to male ratio of 2:1 in the study, with an average age of 19.2 years. The vast majority of students were Japanese, with approximately 8 years of English study behind them. The TOEFL and TOEIC scores were consistent with each other and indicated that the students' level would be rated upper-intermediate to advanced on Oxford University Press's level guide:

Course ---------- Student level	TOEIC	TOEFL	STEP/ Eiken	UCLES exams
False beginner	250	380/85	4	N/A
High beginner	310	400/100	3	KET
Pre-Intermediate				
Intermediate	380	425/115	Pre-2	PET
Upper-Intermediate	520	475/150	2	N/A
Advanced	700	545/210	Pre-1	FCE
Advanced+	800	575/210	1	CAE
	950+	630/265+	N/A	CPE

Table 4.3 Oxford University Press' level guide (O.U.P. ELT catalogue 2004)

These results are somewhat misleading, however, since they rate students on the basis of 'paper tests' only. Japanese learners typically perform much better in tests than in real communicative contexts (largely as a result of the Japanese education system in junior & high-school) and are often unable to cope with listening and speaking tasks at the assumed level in internationally produced textbooks.

Students typically had around 9 hours of official English input per week, only 3 hours of which relate to the experimental/control treatment. Furthermore, around 40% of students had some informal English input each week from native speakers outside of the classroom. These moderating variables were impossible to control for in the trial and their effect on the development of students' communicative competence over the period of the trial is not known.

The majority of respondents reportedly enjoyed studying English (as might be expected for English-major students at university), and most envisaged a need for English in their future lives, so motivation amongst participants was higher than might typically be found in Japanese English language programs. Students demonstrated a marked preference for listening and speaking activities in the classroom, perhaps reflecting awareness on their part that these were the skills in most need of development. Since both experimental and control groups in this trial received predominantly listening and speaking input in the Communicative English classes, this also led to the expectation of high levels of motivation.

The least preferred class activity was presentations, reflecting Japanese students' high levels of anxiety in language classrooms and their typical reluctance to use the L2 in public, where their English can be judged by peers. Other least preferred activities seem to be a reaction against earlier language learning experiences in junior and high school (passive lessons, individual study, grammar activities, tests, etc.). These findings are consistent with those found in the pilot study.

4.3 Testing instruments

As outlined in chapter 3, the approach taken in the classroom-based research reported on here was both quantitative and qualitative in nature. The concern to present as comprehensive a picture as possible of classroom events meant that a variety of different psychometric and ethnographic measures were employed.

4.3.1 Quantitative measures

The quantitative measures used in the trial were designed to show the effects of the experimental and control treatments on learners' overall communicative competence. As was discussed in chapter 1, the current model of communicative competence envisages the existence of five, inter-related components: linguistic competence, strategic competence, pragmalinguistic competence, sociopragmatic competence and discourse competence.

A batch of eight different tests was used to tap into these different types of competence, as well as to assess learners' listening, speaking and reading skills. The same tests were used for both pre- and post-course evaluation, making the assumption that the time between tests (36 weeks) would be sufficient to counter any practice effects. This is supported by at least one student in comments from his diary at the end of the treatment:

SN: I completely forgot that we had to take some tests again. They were exactly the same texts we had at the beginning of this course, but I could not remember answers! Gee, how annoying!

The content of the communicative competence tests is outlined below:

a) **Listening test**: The listening test used was an IELTS practice test, taken from *Passport to IELTS* (Hopkins & Nettle 1995: 130-132). It is composed of four separate dialogues, each centred on an Australian female studying at a British university:

   i) Talking about a university competition.
   ii) Listening to a presentation on university security.
   iii) Asking for help operating a computer in the university's computer room.
   iv) Talking about a year spent travelling around the world.

Four different voices (and a range of accents) are heard in the test: Julie, Julie's friend, a police officer and a student in the computer room. Sections i), iii) and iv) are dialogues between a male and female which eases the problem of identifying the interlocutors. Section ii) is a monologue. Test responses include picture selection, gap-fill with a word or phrase, true/false and open questions, with a total of thirty-three items.

b) **Pronunciation test**: The receptive pronunciation test used was taken from *Speaking Clearly* (Rogerson & Gilbert 1990: 2-6). Nine sections are included in the test, covering syllable stress, weak forms, individual sound recognition, rhythm, word recognition and catenation, sentence stress and intonation, with a total of sixty items.

c) **'C'-Test**: The 'C'-Test was adapted from texts taken from four different levels of the Headway Series (Oxford University Press) (see Appendix II). As discussed in chapter 3, the 'C'-test has been shown to be a reliable indicator of general L2 proficiency and is similar to a traditional cloze test except that it involves deletion of the second half of every second word, starting and ending with an intact sentence (Klein-Braley & Raatz 1984). Dörnyei & Katona (1992) compared the 'C'-Test with a standard cloze test and found it to be a superior measure of general language proficiency, particularly with more homogeneous groups (as was the case with the participants in this trial).

The texts selected for use in the 'C'-Test consisted of four paragraphs, of increasing difficulty, taken from the Headway series of language textbooks, with a total of one hundred and sixteen gaps to be completed by students:

- Extract 1: ‘The traditional English breakfast’, Headway Elementary (Soars & Soars 1993: 66)
- Extract 2: ‘Billionaire Milton Petrie’, Headway Pre-Intermediate (Soars & Soars 2000: 50)
- Extract 3: ‘My Aunt Emily’, Headway Intermediate (Soars & Soars 1996: 23)
- Extract 4: ‘The dangers of modern life’, Headway Upper-Intermediate (Soars & Soars 1987: 49).

d) **Grammar test**: The grammar test used was taken from *English Grammar In Use: Intermediate Level*, 2nd Edition (Murphy 1994: 301-309). It consists of fifteen sections and a total of one hundred and twenty one multiple-choice items. A wide range of grammatical structures which should be familiar to a proficient user of English are covered in the test: present, past, present perfect & future tenses; modal auxiliaries; conditionals; passives; reported speech; questions & auxiliary verbs; -ing & infinitive constructions; articles & nouns; pronouns & determiners; relative clauses; adjectives & adverbs; conjunctions; prepositions.

e) **Vocabulary test**: The (receptive) vocabulary test used was Schmitt’s Vocabulary Levels Test (Version 1) (Schmitt 2000) which aims to test learners’ receptive knowledge of lexis from different frequency ranges. It consists of four sections, each with thirty items, making a total of one hundred and twenty words tested:

- Section 1: 2,000 word level.
- Section 2: 3,000 word level.
- Section 3: 5,000 word level.
- Section 4: 10,000 word level.

According to Schmitt, Schmitt & Clapham (2001), the Vocabulary Levels Test has good internal consistency, with a Cronbach alpha coefficient reported of over .9.

f) **Discourse Completion Task**: The DCT used was a Multimedia Elicitation Task (MET) developed by Gila Schauer, now at Lancaster University (see Schauer 2005). It consists of audio-visual prompts requiring students to use request speech acts with, either same-status, or higher-status individuals. While looking at a computer screen, students listen to an audio recording giving details of the context for the request. After the prompt, "You say…" students record their answers on to tape. There are sixteen scenarios in total with eight different requests to either same status or higher status individuals:

i) Asking someone to open a window.

ii) Asking someone for directions.

iii) Asking someone to fill in a questionnaire.

iv) Asking someone to lend books or articles.

v) Asking someone to move aside.

vi) Asking someone to meet in the holidays.

vii) Asking someone to meet on a different day.

viii) Asking someone to explain a concept.

In the pilot study, it was found that the students' poor listening comprehension often prevented them understanding the scenarios in the MET. In these instances, they tended to move on without responding to the prompt at all, meaning that the test became more one of listening comprehension than pragmatic competence. On the other hand, when participants were given English transcripts of the audio prompts to support their understanding of the context, they tended to incorporate elements of the prompt into their answers, perhaps distorting measurement of their true pragmatic ability. In the main trial, students were therefore given only a Japanese translation of the audio prompts as a support (see appendix III).

Students' responses to the 16 different scenarios in the MET were tape-recorded, transcribed and blind-rated for pragmatic appropriateness by five NS teachers, using the rating guidelines shown in appendix IV.

The DCT's internal consistency (the degree to which individual items in the test 'hang together') was investigated by calculating Cronbach's alpha coefficient. This was found to be .68 for the scale as a whole, which is just below the value of .7 considered the cut-off point for reliability. Analysis of the corrected item-total correlations indicated that scenarios 1, 2, 9 & 12 had particularly low values, below .3, which suggest that they were measuring something different from the scale as a whole. These were therefore removed from the final analysis, giving an improved Cronbach alpha value of .8, which is considered reliable.

g) **Oral interview**: The oral interview used was based on a 1998 -2000 version of the IELTS speaking test. Interviews were conducted by one of eleven different NS teachers at Kansai Gaidai University (from America, Britain, Canada or New

Zealand) and recorded on both audio and videotape. The interview consisted of 5 phases and lasted a total of 11-15 minutes:

Phase 1: Introduction (1-2 minutes)
- Exchange greetings.
- Candidate's identity checked.
- Candidate settled.
- Candidate asked basic personal questions about their life, hometown, etc.

Phase 2: Extended discourse (3-4 minutes)
- Candidate encouraged to speak at length on one of the following topics: marriage rituals, city versus country life, festivals, leisure interests.

Phase 3: Elicitation (3-4 minutes)
- Candidate asked to elicit information from the interviewer through a role-play activity: asking about a wedding (pre-course); asking about an English language school (post-course).

Phase 4: Speculation & attitudes (3-4 minutes)
- Candidate encouraged to speak about academic and vocational interests and talk about future plans.

Phase 5: Conclusion (1 minute)
- Interview brought to a close.

The video interviews for both experimental and control groups were then blind-rated on 5 criteria (phonology, body language, fluency, context-appropriate vocabulary & interactional competence) by three or four trained NS volunteers, using the descriptors shown in appendix V.

h) **Student-student role-play**: After the pilot study and pre-course testing, it became clear that the oral interview with a NS teacher was not giving students the opportunity to display their speaking skills in the best light. Many students appeared anxious meeting the NS teachers for the first time, which, to some extent, inhibited them in the interviews. Furthermore, the IELTS oral test only gave students control over the conversation for 3-4 minutes, in the elicitation

section (Phase 3) and this meant that they had limited opportunities to display their discourse competence (making topical moves, back-channeling and so on). It was therefore decided to add a second speaking test in the form of a student-student role-play.

Participants read a role-play card (see appendix VI) and were then given a few minutes preparation time, before performing the role-play in pairs. The role-plays were video recorded and rated on two criteria, a) conversational behavior and b) conversational management (see appendix VII), by the author.

Eight post-course role-plays, representing a total of sixteen low and high-level students, were blind-rated by a second (trained) NS teacher and checked for inter-rater reliability, using Pearson product-moment correlation coefficient. There was a positive correlation between the mean scores on both components of the rating criteria. For conversational behaviour, the Pearson correlation coefficient was slightly lower [r=.70, n=16, p=.003], with a coefficient of determination (r squared) value of .49 indicating that the two variables shared 49% of their variance. Given the small sample size and the fact that an r value of .7 is considered large (Pallant 2005: 126), this scale was considered reasonably reliable. For conversational management, the Pearson correlation coefficient was higher [r=.85, n=16, p<.0005], with a coefficient of determination (r squared) value of .72 indicating that the two variables shared 72% of their variance. This scale was also, therefore, considered reliable.

Trialing of the communicative competence measures

This section describes, in some detail, an assessment of the quantitative measures used in the pre- and post-course testing of students' communicative competence from the pilot study. Those readers willing to accept that the tests were, indeed, tapping into different kinds of competence may prefer to move on to the summary of the quantitative tests on page 158.

The tests selected for the investigation were initially trialed with 10 university student volunteers. The results of a proficient non-native Japanese speaker of English (TOEIC 845, 3 years living in the UK) and an American native speaker were also included in order to assess the sensitivity of the tests to different proficiency levels and these are shown below:

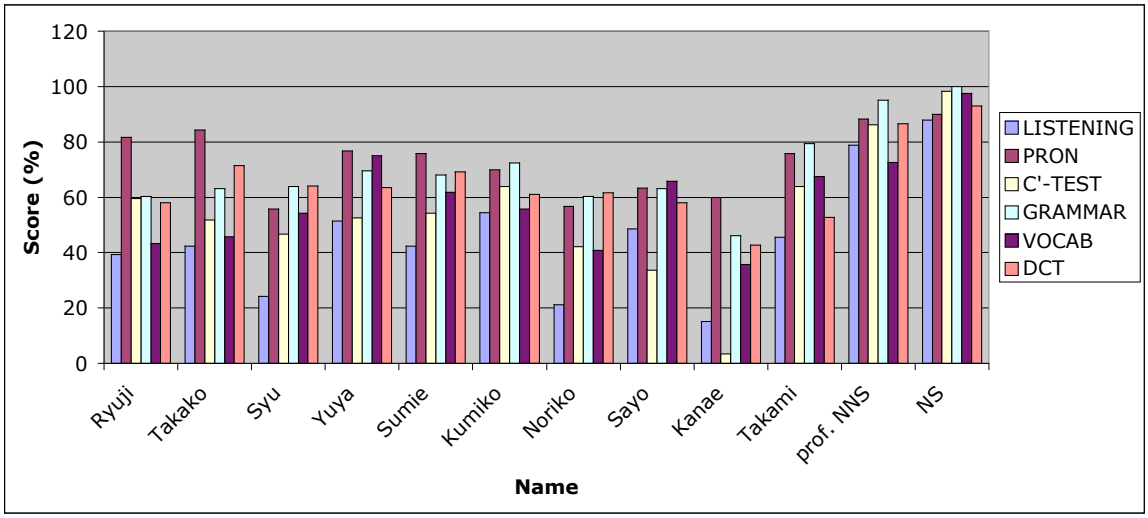


Figure 4.1 Results from the pilot study trialing of communicative competence measures

As the results in figure 4.1 illustrate, the tests produced a range of scores, which appear to effectively distinguish between different proficiency levels within the participants. As

expected, the native speaker consistently achieved the highest scores, with the proficient non-native speaker also out-performing the students on most of the measures (one student scored higher on the vocabulary test).

The scores from the DCT appeared to be successful in identifying the participants likely to have the highest levels of pragmatic competence, with the NS and proficient NNS (both with experience of living in English-speaking countries) significantly out-performing the students (with no experience of living outside Japan). These results are consistent with Bardovi-Harlig & Dörnyei (1998) who found that ESL learners, with experience of living in an English speaking community, had greater pragmatic awareness than EFL students, whose experience with the L2 was predominantly classroom-based (see chapter 1).

Five of the volunteer students from the pilot study also took the IELTS oral interview and their band scores, as judged by a NS interviewer, were in line with their mean scores on the other measures of communicative competence, with Takami emerging as the strongest member of the group and Kanae the weakest:

Name	IELTS score[8]	Mean score on other tests (%)
Takami	6	64.1
Sumie	(High) 5	61.9
Takako	5	59.8
Sayo	5	55.4
Kanae	4	33.9

Table 4.4 Proficiency ratings on IELTS oral interview compared with mean scores on other measures for 5 volunteer students

[8] IELTS band descriptors run from 1 to 9:
9 = Fluent, situationally appropriate & fully acceptable speech.
5 = Broadly able to communicate meaning on most general topics though errors in structure & vocabulary may interfere with communication.
1 = Essentially unable to speak English.

Takami's and Kanae's oral interviews were also transcribed and compared with a NS performing the same task in order to provide further evidence of the reliability of the different measures (see appendix VIII). A brief read-through of the transcripts indicates a clear qualitative difference in proficiency between the three participants with Kanae's conversation displaying a number of elements which reduce its overall coherence: widespread use of Japanese to compensate for lexical deficiencies, poor pronunciation with 'katakana-ised' English[9], frequent communication breakdown due to comprehension difficulties and so on. Discourse analysis of eight individual features of the three conversations also provides quantitative support for this impressionistic evidence:

a) Number of interviewer's questions mis-/not understood.

b) Number of clarification or confirmation strategies employed.

c) Number of instances of Japanese used as a compensatory strategy.

d) Percentage of self-corrections.

e) Mean number of (English) words per turn.

f) Percentage and range of discourse markers employed.

g) Percentage and range of reactive tokens employed.

h) Speech rate.

Figure 4.2 shows a range of discourse features indicative of a low proficiency in English:

[9] Japanese sounds normally consist of a consonant paired with a vowel: *ka ki ku ke ko, ma mi mu me mo*, etc. Students with low proficiency or limited exposure to NS English therefore find it difficult to pronounce consonant-final words and tend to add a vowel sound on to the end, producing 'katakana-ised' English such as *and-o*, *but-o* or *because-u*.

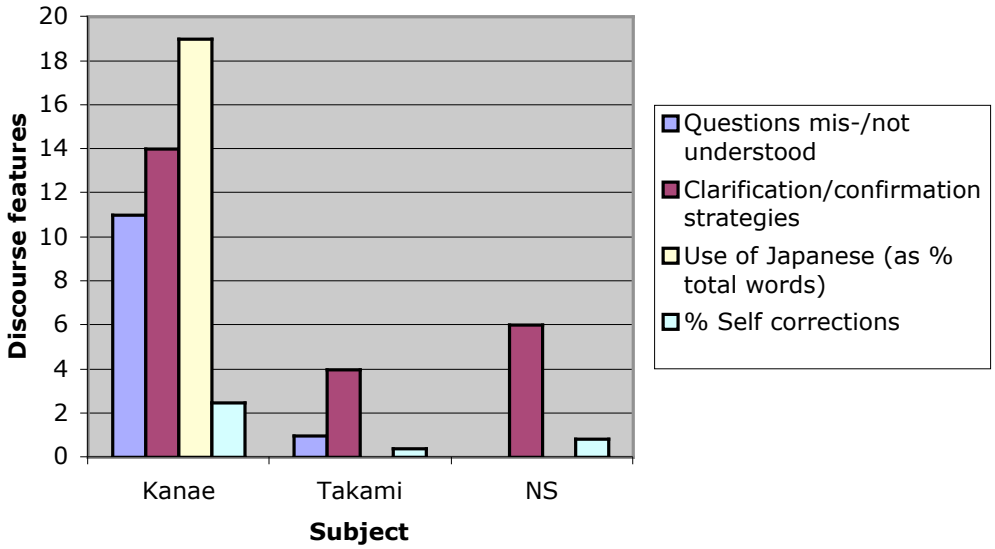


Figure 4.2 Discourse features illustrating low proficiency in the target language

Because of Kanae's poor listening skills, she frequently misunderstood, or failed to comprehend altogether, the interlocutor's questions and, as a result, both she and the interviewer were forced to use a higher number of clarification or confirmation strategies during the interview:

40 I: What do you think your life would be like if you lived in the countryside?
41 K: Oh I want to go abroad but ah *e to*
42 I: If you were to live in the country [K: Living no] rather than in a city how do you think your life
43 would be different?
44 K: Oh (incomp. Japanese)
45 I: Yeah so city life country life [K: Mm] what do you think?
46 K: Ah *tokai to inaka to?* [I: Mm hm] *wa dou chigaimaska?* [I: Mm hm] *e to* I live in Osaka nearby
47 convenience store [I: Ah ha] *yappari e to* there is convenient for live [I: Ah ha] near the station
48 near the *suupaa* [I: Ah ha] *e* car many cars many [I: Ah ha] but countryside isu many forest
49 [I: Mm hm] ando natural natural there [I: Mm hm] mm is good pointo [I: Mm hm] mm I liku
50 countryside er summer ah during summer vacation [I: Mm] I went to Nagano [I: Ah] mm summer
51 vacation and winter vacation [I: Ah ha] I went to Nagano it isu ah I go I like to snowboarding

Surprisingly, clarification or confirmation strategies were employed more often by the native speaker than Takami, but whereas both Japanese subjects use them to compensate

for comprehension problems, the NS participant uses them to confirm questions and narrow their scope before answering:

7 I: Great thanks very much [C: Mm hm] alright so erm can we talk a little bit about your hometown?
8 C: My hometown?
9 I: Yeah
10 C: Er which one? I've got two
11 I: You've got two oh ok
12 C: Erm the place I was born or?
13 I: Either one [C: Oh ok] which one would you like to speak about?

Kanae's limited L2 vocabulary knowledge, as suggested by her low score in Schmitt's Vocabulary Levels Test, is supported by her frequent use of Japanese in the interview to compensate for lexical deficiencies (approximately 19% of total words spoken). In contrast neither Takami nor the NS resorted to Japanese during their interviews, both apparently having sufficient resources to cope with the interactional demands placed upon them using exclusively English.

Self-correction refers to the participants' attempts to re-phrase something they have said mid-turn: a performance error indicating a lack of sufficient planning of the discourse. The NS self-corrects infrequently, and when he does it is never as a result of language difficulties but rather to re-organize (or *recast*, as Carter & McCarthy 2006: 173 term it) his discourse as in this example:

156 C: In Qatar however I er when I did my DELTA training [I: Mm] I met somebody who was working
157 in Qatar [I: Mm] and I think the British Council have closed down [I: Really?] their branch there
158 [I: Oh] I th I think yeah

This is a natural characteristic of spontaneous, unplanned conversation. Kanae, in contrast, self-corrects more often and when she does, the cause is usually localised problems with language *forms*, rather than issues with higher-level organisation of the

discourse. This suggests, as do the communicative competence tests, a lack of control of syntactical and lexical features in the L2:

127 *chigau jujitsu* true my mother [I: Mm hm] want to ah wanto me to go to Gaidai university [I: I
128 see] she ah her dream [I: Mm hm] isu attend attendance attendo airport airporto attendanto [I: Ah
129 ha] mm *mukashi kara no* her dream [I: Ah ha] *sore ga* me [I: Ah ha] buto "whoa" *nan te iu kana*

Figure 4.3 below shows a range of discourse features from the three conversations indicative of a high proficiency in English:

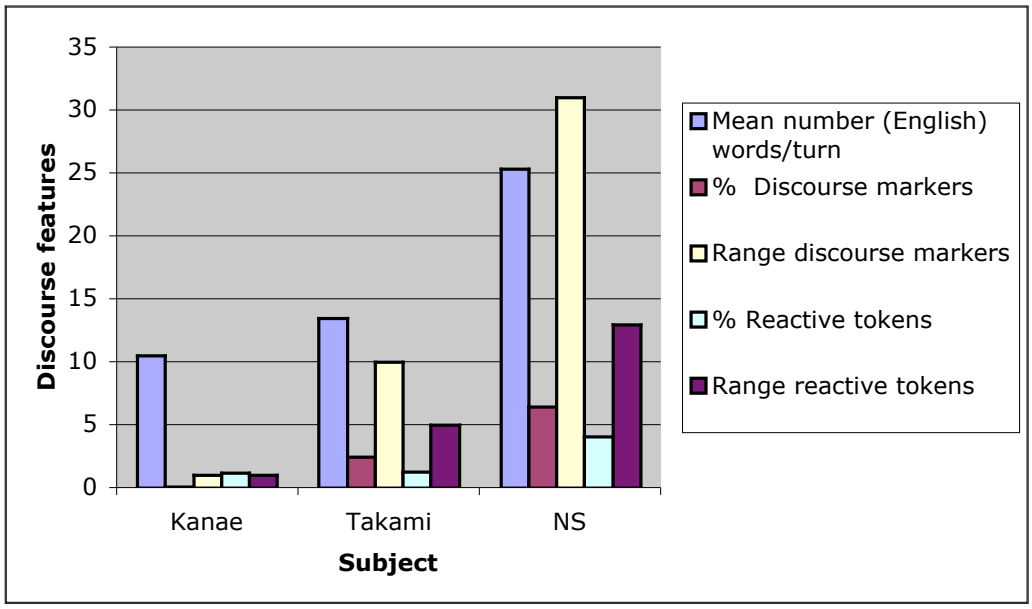


Figure 4.3 Discourse features illustrating high proficiency in the target language

As can be seen, the mean number of (English) words per turn increases with proficiency in a predictable fashion, and in line with estimates of communicative competence from the quantitative measures employed in this investigation. The same is also true of the number and range of discourse markers[10] or reactive tokens[11] used by the three subjects.

[10] *Discourse markers* are defined by Carter & McCarthy (2006: 208) as 'words and phrases which function to link segments of the discourse to one another in ways which reflect choices of monitoring, organization and management exercised by the speaker.'

These features all indicate an increasing command of English from left to right across the chart, with the ability to verbally express thoughts with precision and in greater depth, to effectively control lexical and syntactic structures, to plan and organize the discourse and to monitor and provide feedback on the interlocutor's turns.

The final chart in figure 4.4 below shows the mean speech rate of the three subjects in words per minute (wpm), estimated from the audio recordings of the IELTS interviews:

---

[11] *Reactive tokens* is the term used here to replace what are sometimes referred to as *back-channels* in the literature. The term 'back-channel' was first coined by Yngve in 1970 and refers "to the sounds (and gestures) made in conversation by the current non-speaker, which grease the wheels of conversation but constitute no claim to take over the turn." (Tottie 1991: 255). Back-channels fulfill several, simultaneous roles in discourse; they have a 'supportive function', signalling agreement and understanding and a 'regulative function', encouraging a speaker to continue with his or her turn. Over the years, the number of items covered by the term 'back-channel' seems to have expanded as researchers realised that lexical items and even short phrases could also be uttered by interlocutors in the course of a conversation without making a claim to the floor. Effectively, the edges between a back-channel and a turn have become blurred as Duncan & Niederehe (1974: 237) remark: "for some of the longer back-channels, particularly brief restatements, the boundary between back-channels and speaking turns became uncertain. On an intuitive basis, some of these longer back-channels appeared to take on the quality of a turn".

Clancy et al. (1996: 355) have relegated the word 'back-channel' to a sub-group and prefer instead to use the term 'reactive token' (RT) which they define as:

"A short utterance produced by an interlocutor who is playing a listener's role during the other interlocutor's speakership. That is Reactive Tokens will normally not disrupt the primary speaker's speakership, and do not themselves claim the floor".

They divide reactive tokens into five groups:
a) Back-channels ~ a non-lexical, vocalic form, serving as a continuer, display of interest or claim of understanding.
b) Reactive expressions ~ a short, non-floor taking lexical word or phrase produced by the non-primary speaker.
c) Collaborative finishes ~ a speaker's utterance is completed by a non-primary speaker.
d) Repetitions ~ a non-primary speaker repeats or 'echoes' a portion of a previous utterance.
e) Resumptive openers ~ 'back-channels' used by a non-primary speaker and immediately followed by a full turn from the same participant. Their function is to register the prior turn before taking the floor.

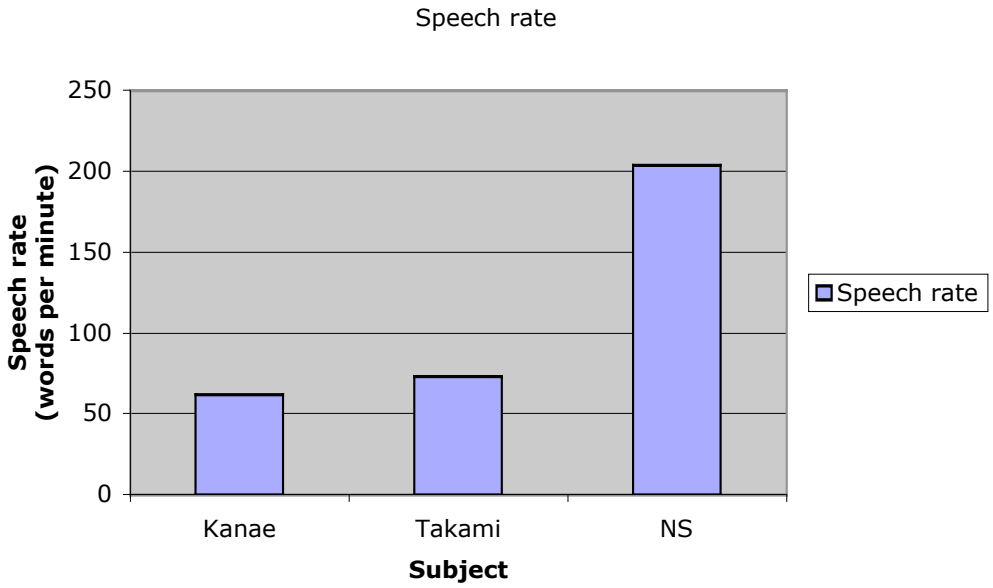


Figure 4.4 Speech rate of subjects in IELTS oral interview

As can be seen, the speech rate of the NS (204 wpm) was well above that of the two NNS students (62 wpm & 73 wpm respectively) so his superior results on the discourse features surveyed in figures 4.2 and 4.3 were achieved at the same time as producing much more language. Takami's greater proficiency in the L2 is only partially reflected in his over-all speech rate, which is only slightly higher than Kanae's. This is likely to be as a result of a greater emphasis on memorization and test-taking than productive skills and fluency within the Japanese education system. It highlights the importance of including a spoken component in the testing procedure in order to give a fair representation of students' over-all communicative competence.

The communicative competence tests (with the exception of the IELTS oral interview) were administered in one sitting for the pilot trialing and took approximately 3 hours for the volunteers to complete. Students were visibly exhausted afterwards and this suggested

that it would have been advisable to administer them over more than one day. For the main study it was therefore decided to divide the testing into two sittings, held during the first two periods of class time at the beginning and end of the year.

Summary of quantitative tests

The piloting of the quantitative tests suggested that they were sensitive to different levels of proficiency in the learners and were able to tap in to the different components of communicative competence. They also demonstrated that the time needed to implement the measures in the classroom was acceptable, given the practical limitations of the investigation. Table 4.5 below summarises how students' communicative competence or language skills were tested in the trial:

Competence / language skill	**Test**
A. Linguistic competence	• Pronunciation test • Grammar test • Vocabulary test • 'C'-test • Oral interview (phonology & vocabulary sections)
B. Strategic competence	• Oral interview (interactional competence section) • ss-ss role play (conversational management section)
C. Pragmatic competence (pragmalinguistic + sociopragmatic)	• DCT • Oral interview (body language, context appropriate vocabulary use sections). • ss-ss role play (conversational behaviour section)
D. Discourse competence	• Oral interview (interactional competence & phonology sections) • ss-ss role play (conversational management section)
E. Skills (listening, speaking, reading).	• Listening test • Oral interview • ss-ss role-play • 'C'-test

Table 4.5 Summary of measures to investigate learners' communicative competence & language skills

Table 4.6 summarises the key features of variables used in the quantitative tests and how these were operationalised:

Measure	Item type	Number of items	Measuring what?	Values
A. Listening Test	Picture selection, gap-fill, T/F & comprehension questions.	33	Listening comprehension of natural English in a variety of spoken genres.	Raw scores out of 33.
B. Pronunciation Test	Syllables, stress, phonemes, rhythm, intonation & sentence discrimination tasks.	60	Recognition of syllable stress, weak forms, individual phonemes, rhythm, catenation, sentence stress & intonation patterns.	Raw scores out of 60.
C. 'C'-Test	Gap-fill of texts with deletion of the 2nd half of every 2nd word, starting & finishing with intact sentences.	116	General L2 proficiency & reading comprehension of 4 paragraphs of increasing difficulty (Elementary – Upper Intermediate level).	Raw scores out of 116.
D. Grammar Test	Multiple-choice questions.	121	Comprehension of a wide range of grammatical structures.	Raw scores out of 121.
E. Vocabulary Test	Matching words to definitions.	120	Receptive knowledge of lexis from four different frequency ranges.	Raw scores out of 120.
F. DCT	Native speaker ratings of students' oral responses to scenarios.	16	Appropriateness of students' oral production of 8 request speech acts (on a scale of 1-5) to same or higher status interlocutors.	Mean scores for 5 native speakers (on a scale of 1-5) for 16 scenarios.
G. IELTS Oral Interview: 1. Phonology 2. Body language 3. Fluency 4. Vocabulary use 5. Interactional competence	Native speaker ratings of students' oral proficiency.	5	Oral proficiency of students (on a scale of 1-5) for 5 criteria: phonology, body language, fluency, context-appropriate vocabulary use & interactional competence.	Mean scores for 3 or 4 native speakers (on a scale of 1-5) for the 5 criteria.

H. ss-ss Role-play: 1. Conversational behaviour 2. Conversational management	Native speaker ratings of students' performances.	2	Approximation of students' role-play performances to native speaker norms on 2 criteria: 'Conversational Behaviour' (interpersonal distance, body orientation, touching, facial expressions, eye contact, gesturing & reactive tokens) and 'Conversational Management' (appropriate initiation & termination, turn-taking, conversational management, topic nomination & topical coherence, hesitation devices, discourse markers & conversational repair strategies).	Raw scores from 1-5 for the 2 criteria.

Table 4.6 Summary of key features of variables in the quantitative tests

4.3.2 Qualitative measures

The ethnographic measures employed in the trial were designed to complement the psychometric testing, by giving some insight into the learners' own perspectives on classroom events, that is, by providing an *emic* perspective. This was done in three different ways:

a) **Learner diaries**: Participants were asked to keep a learner diary (in their own time) for the duration of the trial, recording their impressions of classroom events; both what they believed had been studied during the lessons and their impressions of the input. Appendix IX shows the advice students were given for completing their diary entries. The learner diaries were collected in twice a term by the teacher and comments given, in order to encourage students to continue recording their thoughts throughout the period of investigation. As in the pilot study, significant

events from the diaries were recorded and recurring themes identified (see appendix X).

b) **Case studies**: At the beginning of the trial, 3 students from the control group and 3 from the experimental group (of low, average or high proficiency) were selected as case studies. They were interviewed 4 times during the course of the investigation, using their diaries as 'jumping off points' into a semi-structured discussion of classroom events. Interviews took place in either the classroom after lessons or, when time allowed, in the teacher's office, by appointment. Participants' comments were recorded and transcribed for analysis (see appendix XI). Unfortunately, 2 of the case-study students dropped out of the class at the end of the first term so complete data is only available for 4 learners.

c) **Recording of classroom interaction**: At various points during the investigation, learners were recorded interacting with each other during pair-work activities and samples of their discourse transcribed for closer analysis.

4.4 Testing procedures

All pre- and post-course quantitative testing (except for the student-student role-play) was carried out in the first and last 2 weeks of the trial. Students completed the multimedia DCT while the rest of the class was taking the grammar, vocabulary and 'C'-Test papers. Two computers and tape-recorders were set up, initially in the classroom, to streamline the process but learners were notably self-conscious 'talking to a computer' in

the presence of others. In addition, other students in the class were clearly listening to the DCT cues in an attempt to prepare their own responses, as these comments from the learners' diaries illustrate:

RM: I was very shy when I was taking speaking test today!! Because everybody was behind me and they could hear my speaking. I don't have self-confidence yet, so I was afraid what they were thinking. Almost all students in class 1 can speak so fluently. That's why I was nervous. I thought I could do better if I did it in another room, not in front of everyone. A girl in class 1, who doesn't take speaking test yet, said to me that she was able to listen the questions and she was thinking the answers a little, so when she take the test, she would prepare some answers. My turn was first, so I thought it's not fair!!

ES: [DCT] I didn't do the speaking test today, so I asked my friends who took that test. They said it wasn't so difficult, but they were nervous to take the test before classmates. I guess it would better to take an exam alone. We can display our real abilities because we can relax. Besides, we can hear questions and answers when someone takes the test before us. I think it's not fair.

In response to this, the computers were moved outside the classroom, into the adjacent corridor. This was by no means ideal either, since it was a cold location, and test-takers were disturbed by other students passing by (often noisily) in the corridors. Conditions were, however, identical for both control and experimental groups.

The IELTS oral interviews were conducted, by appointment, with one of the volunteer NS teachers in the university's staffroom. Again, this was not ideal since the room was often noisy, particularly during breaks, and the microphones on the video cameras picked up a lot of background noise, making the process of rating students more difficult. It also limited the privacy of the interviews in, what was already, a stressful situation. However, conditions were identical for both control and experimental groups.

The student-student role-plays were conducted at the beginning of the 2nd term (September 2004) and post-course (January 2005) in available classrooms within the university. Learners were paired off randomly and given only a few minutes preparation time, to avoid the inevitable attempt by students to plan their discourse together, and the role-plays video recorded for rating purposes.

The qualitative measures were conducted, as described in Section 4.3.1, throughout the course of the trial.

4.5 Data analysis

4.5.1 Quantitative measures

The statistical software package used for analysis of the pre- and post-course test results was SPSS for Mac OS X, version 11.0.4 (2005). Analysis of covariance (ANCOVA) on the individual dependent variables was conducted to investigate exactly what aspects of students' communicative competence showed differences, when comparing the control and experimental groups. Both general assumptions and those specific to ANCOVA, which underlie the parametric techniques used in this investigation, were first checked (see, for example, Pallant 2005).

4.5.2 Qualitative measures

The learner diaries were analysed using the 5-step procedure recommended by Bailey (1983; 1990) and Bailey & Ochsner (1983) (see chapter 3), which involved extracting and categorising 'significant events' from the diaries. For the case studies, the interviews with students were recorded and transcribed (see appendix XI), and relevant comments identified. Classroom interaction was also recorded and transcribed (see sections 6.1.3 & 7.2) but in this case, since the goal was a more careful analysis of the discourse, greater detail (hesitation devices, pauses, concurrent speech, etc.) was recorded in the transcriptions (for advice on transcription procedures, see Grundy 2000).

4.6 Training procedures

Training for both control and experimental groups focused principally on developing learners' listening and speaking skills since these were the areas of priority on the Communicative English Course participants were enrolled in. However, the *type of input* students were exposed to, the independent variable, differed significantly.

4.6.1 Control group

The control group worked methodically through the two selected textbooks, *Inside English* (Maggs, Kay, Jones & Kerr 2004) and *Face to Face* (Fuller & Fuller 1999), with some supplementation from other teaching resource books where it was felt necessary. These particular texts were chosen because they were judged to contain predominantly contrived texts, designed for pedagogic exploitation, which helped to create a sharp contrast between the input which the experimental and control groups received. Both books are organized topically, and claim to teach all four skills. Their content is summarized in table 4.7:

Textbook / Unit Theme	Content
Inside English	
Unit 1: Smile	Talking about happiness & stress, describing characters & faces, imperatives & phrasal verbs, /s/ & /z/ sounds
Unit 2: Celebrate	Talking about festivals & traditions, passives, stative/action verbs, expressions with *go*
Unit 3: Dance	Talking about going out, *for, since* & *been*, present perfect tense, informal language, weak forms
Unit 4: Call	Talking about telephones, offers & requests, polite questions, telephone language, register, intonation in requests
Unit 5: Review	Review of Units 1 – 4
Unit 6: Lifestyle	Talking about longevity, life predictions & food, future time clauses, *will* for prediction, food idioms, sounds & spelling
Unit 7: Animals	Talking about animals & their characteristics, relative clauses, conditionals, prepositions, homophones
Unit 8: Weird	Talking about coincidences or strange experiences, narrative tenses, *make/take* + noun structures, dipthong sounds
Unit 9: Wheels	Talking about transportation, *used to*, giving opinions, advice & suggestions, driving vocabulary, spoken forms of verbs
Unit 10: Review 2	Review of units 6 – 9
Face to Face	
Unit 1: All About the "Real" Me	Questionnaire about self, exchanging personal information, likes & dislikes
Unit 2: Friends Forever	Questionnaire about friends & friendship, talking about friends & personal experiences
Unit 3: Finding a "Special" Friend	Dating questionnaire, filling a dating profile form, discussing dating preferences
Unit 4: Shopping for Bargains	Shopping questionnaire, talking about shopping experiences, recognizing prices in English
Unit 5: Dreaming About Summer	Questionnaire about summer vacations, talking about travel
Unit 6: I'd Better Get a Job	Questionnaire about job preferences, answering job interview questions, completing job application forms, interview role-play

Unit 7: What Do You Think?	Asking & giving opinions about different topics, giving reasons for opinions
Unit 8: Rap, Rock, and Reggae	Questionnaire about music, expressing opinions about music & artists, understanding radio interviews
Unit 9: What a Character!	Questionnaire about character, talking about personality traits, identifying personal strengths & weaknesses
Unit 10: Money Matters	Identifying attitudes about money, talking about monthly budgets, numbers & prices
Unit 11: Situations in Life	Questionnaire about personal responses, talking about personal experiences, narratives & talking about past events
Unit 12: Facing the Future	Questionnaire about future desires, discussing plans for the future & the probability of future events

Table 4.7 Summary of syllabus for control group

4.6.2 Experimental group

The experimental group received predominantly (but not exclusively) authentic materials (as defined by Morrow 1977, in section 2.2) throughout the trial, taken from films, documentaries, 'reality shows', TV comedies, web-based sources, home-produced video of native speakers, songs, novels and newspaper articles. Materials were selected on the basis of their ability to highlight some aspect of communicative competence, along similar lines to those suggested by Celce-Murcia, Dörnyei & Thurrell (1997) in their 'principled communicative approach', and so, at times, textbook resources were utilized where it was considered expedient. The syllabus is summarized in table 4.8 and sample materials are given in appendix XII.

Theme	Content
1. Dictionary skills	Using a monolingual dictionary effectively
2. Listening to NSs of English	Stress-timing, linking, weak forms, accents from around the world (Ellis & Sinclair 1989: 56-57), 'Tom's Diner' (Suzanne Vega), 'This is the house that Jack built' (children's verse), phonemic charts
3. English pronunciation & intonation	Scenes from 'My Fair Lady' (George Cukor), 'as + adj. + as + noun' expressions, acting out scenes
4. Circumlocution strategies	What to do when you don't know a word (Ellis & Sinclair 1989: 39), describing unfamiliar objects, miming activities
5. Conversational repair strategies	Taking control of a conversation (Ellis & Sinclair 1989: 63), giving directions & practice using conversation strategies
6. Hesitation devices & British sociopragmatic conventions	Scenes from 'Big Brother' (UK reality show), practice using common hesitation devices, introductions in English, colloquial expressions, HW assignment - student Big Bother audition tapes, *in case* vs. *so that* (Naunton 1994: 32-33)
7. Starting conversations in English (sociopragmatic& pragmalinguistic conventions)	Extracts from 'Around the World in Eighty Days' (Jules Verne: 1873) & scenes from 'Around the World in 80 Days' (BBC TV series), strategies & expressions for opening up conversations with strangers, role-play activities, HW assignment – start a conversation with a stranger
8. Closing conversations in English (sociopragmatic & pragmalinguistic conventions)	Scenes from 'Annie Hall' (Woody Allen), scenes from 'Louis Theroux's Weird Weekends' (BBC TV series), practice closing conversations
9. Discourse intonation	Sentence stress, 'telling' & 'referring' (fall/fall-rise) tones, tone units, tonic stress (Bradford: 1988: 5-17)
10. Developing conversations in English	Interview with a musician (Falla 1994: 20-24), extract from 'Polite Fictions' (Sakamoto & Naotsuka 1982: 80-87), strategies for developing conversation in English, transition relevance places (TRPs), topic shift, paralinguistics of turn-taking, practice developing conversations
11. Listener responses (reactive tokens) & ellipsis in spoken English	Scenes from 'Secrets & Lies' (Mike Leigh), showing interest, surprise, understanding, agreement, etc. in English, ellipsis, colloquial language, practice using reactive tokens in conversation, discussing adoption, role-play - Hortense meeting her mother for the 1st time
12. Oral narratives & register in English	Scenes from 'Reservoir Dogs' (Quentin Tarantino), conversational story-telling skills (Jones 2001: 155-

	163), structure of oral narratives, strategies for making stories interesting (use of historic present, exaggeration, pitch range, body language), formal/informal register, colloquial language, role-play scenes, HW assignment – telling a personal story in an interesting way, taboo words
13. Formal & Informal registers in English (pragmalinguistic conventions)	Scenes from 'Fargo' (Joel & Ethan Coen), scenes from 'Fawlty Towers' (BBC TV series), article from *The Guardian* on The British class system, 'What Class are You?' quiz: http://www.pbs.org/peoplelikeus/games/index.html, features of formal & informal discourse in English, using intonation to show politeness, role-plays – checking in & making complaints in a hotel (politely or impolitely)
14. Listening to NSs of English II	Short documentary films from 'Video Nation' (BBC web-site): http://www.bbc.co.uk/videonation/ , transcribing NS English, Video Nation presentations
15. Body language	Extracts from 'How to Communicate Successfully' (Wright 1987: 35-42), extracts from 'Everybody's Guide to People watching' (Wolfgang 1995: 64-67), scenes from 'New Headway Video, Beginner' (Murphy 2002), facial expressions, eye contact, gestures, inter-personal space, touching, etc., HW assignment – interview a foreigner about common gestures in their country
16. Negotiating plans in English & common discourse markers	'Weekend Away' activity from 'Keep Talking' (Klippel 1984: 45-46), video of NSs planning a weekend away, natural ways to give opinions, agree & disagree in English, *could* & *would* modal auxilaries, *will* to confirm plans, present continuous to talk about fixed plans, discourse markers, role-play: planning a weekend away in Japan

Table 4.8 Summary of syllabus for experimental group

Most of the experimental materials were developed on a week-by-week basis by the author over the course of the 10-month investigation and a number of general guidelines were followed in this process:

Materials selection

In order to be considered for inclusion in the experimental syllabus, materials had to fulfill a number of criteria:

a) They had to include examples of specific target discourse features (both verbal and non-verbal) in order to raise learners' awareness of a wide variety of elements necessary for effective and appropriate communication in English and fulfill the course's aim of developing a broader range of communicative competencies.

b) They had to be judged intrinsically interesting for the Japanese participants in the trial, regardless of their language content. This was felt to be important to maximize learners' affective engagement with the texts, assuming that this would also encourage processing of the language at a deeper level.

c) They had to illustrate language variation in English and give learners exposure to different accents, genres, speech rates, registers, etc.

d) Audio-visual materials, taken from web-based sources or DVDs, were given priority over the more traditional, audio-only, materials typically used in language classrooms. This was because the visual element can provide learners with an enormous amount of additional pragmatic information: on the context in which the discourse is situated, the speakers' ages and social positions or the relationships between different participants (Brown & Yule 1983: 85). These contextual details dictate the kind of speech acts and NVC speakers employ and can be used to sensitise learners to language variation in English (pragmalinguistics) as well as the behavioural norms in the target speech community (sociopragmatics). Being able to watch the speakers as they talk can

also support learners' listening comprehension since the 'extra articulatory effort' (ibid: 86) on stressed syllables, reflected in mouth movements, facial expressions and gestures, provides valuable information on which words are 'content words' (those 'carrying' the key information).

e) The intensive language work with transcripts, planned for the post-listening stages of activities, meant that only relatively short scenes (up to approximately 10 minutes) were selected for inclusion. It was hoped that the material covered in class would motivate students to continue watching the DVDs in their own time in order to give them valuable extensive listening practice.

f) Heavily edited extracts, with frequent cuts, were avoided since the rapid changes in speaker and context that they produce was felt to increase the difficulty of comprehension.

g) Extracts from films, documentaries, etc., had to be able to 'stand alone' in the classroom. In other words, scenes had to make sense, when isolated from the rest of the work, without the need for excessive introduction or contextualization.

h) DVDs were given preference since they have several advantages over other types of audio-visual medium. Selected scenes can be quickly found during lessons since DVDs are organized into 'chapters' and fast-forward, rewind and search functions are superior to video. In addition, most DVDs also come with the option of English sub-titles, which can be used with particularly challenging material.

Finding authentic materials to fulfill all of these criteria can be a time-consuming process but once appropriate scenes have been identified and transcribed, they can be placed in a 'materials bank' and used over many years.

Transcription procedures

In order to facilitate the intensive language work suggested for the post-listening stages of lessons, and encourage learners to become 'mini conversational analysts' (as recommended in chapter 2), the spoken discourse has to be 'frozen' in transcripts. Decisions therefore have to be made in terms of how much detail to include in the written representation of speech. A 'thick description' of discourse (Geertz 1973) could include phonological, turn-taking, or NVC features, as well as the actual words spoken. However, too much information can be off-putting for learners so it was decided to keep the transcriptions simple on pedagogic grounds, unless there was a particular reason to include more detail.[12] Transcription lines were numbered for ease of reference during lessons and double-spaced to give students space for note taking on any discourse features of interest.

Task design

The materials were designed to include the pre-, while- and post-listening tasks typical in mainstream Communicative Language Teaching methodology.

Pre-listening tasks focused on raising learners' interest in the topics, introducing difficult vocabulary and providing any necessary cultural background knowledge. These kinds of activities help students develop schemas and scripts for the texts they are about to watch, which can support, or 'scaffold' (Bruner 1983), learning. In some materials, such as 'Freeze! Don't move!' (appendix XII), learners were first asked to write and role-play their own versions of scenarios occurring on the DVDs. This not only helped their comprehension in the listening tasks, but also encouraged them to 'notice the gap'

[12] There are some powerful arguments for *not* representing spoken discourse as written text, using sentences and traditional punctuation (see section 1.6.5), but these were put aside for the purposes of this investigation.

between their own interlanguage and the NS discourse, something which is believed to enhance the acquisition of intake (Schmidt & Frota 1986).

While-listening tasks normally focused on the meaning of the discourse first, before shifting to form, to avoid overloading learners' language processing systems. They also typically began with gist listening questions, followed by more detailed comprehension questions to encourage effective listening strategies.

Post-listening tasks were designed to 'revisit' the material in a new way. This involved a wide variety of task types, such as recycling vocabulary, focusing on target discourse features, or getting students to use target language in speaking activities.

The overall goal of the experimental materials was to provide learners with interesting, challenging and varied authentic texts as input, which would also raise their awareness of a range of discourse features important in the development of communicative competence. The syllabus attempted to maintain a balance between the different aspects of the communicative competence model (linguistic, pragmalinguistic, sociopragmatic, strategic and discourse competencies) although the specific needs of the participants meant that some areas were given more attention than others. Since language proficiency has both knowledge and skills elements associated with it, learners were also given numerous opportunities to listen to, or use, the target language in communicative tasks.

CHAPTER 5. RESULTS & DISCUSSION (PART I): QUANTITATIVE MEASURES

This chapter represents the first of three providing an overview of the results from the ten-month, classroom-based investigation into Japanese learners' development of communicative competence. Since a mixed methods approach was adopted, using both quantitative and qualitative techniques, the results have been organised along similar lines with this chapter discussing the quantitative dimensions and chapters 6 and 7 looking at the qualitative dimensions.

To begin with in this chapter, the results of the quantitative aspects of the investigation are summarised in section 5.1, before the individual measures are presented and discussed in detail in sections 5.2 - 5.9.

5.1 Univariate analysis of covariance (ANCOVA)

A one-way between-groups analysis of covariance was conducted to compare the effectiveness of the two different interventions designed to develop students' communicative competence. The independent variable was the type of intervention (textbook input or authentic input), and the dependent variables consisted of scores from the eight communicative competence measures after the intervention was completed. Participants' scores on the pre-intervention administration of the eight communicative competence measures were used as the covariates in this analysis. Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate.

After adjusting for pre-intervention scores, there were significant differences between the two intervention groups on post-intervention scores on: a) The listening component; b) The receptive pronunciation component; c) The receptive vocabulary component; d) The body language component of the IELTS oral interview; e) The oral fluency component of the IELTS oral interview; f) The interactional competence component of the IELTS oral interview; g) The conversational behaviour component of the student-student role-play; and, h) The conversational management component of the student-student role-play.

However, after adjusting for pre-intervention scores, there were no significant differences between the two intervention groups on post-intervention scores on: a) The 'C'-Test component; b) The grammar component; c) The DCT component; d) The pronunciation component of the IELTS oral interview; e) The vocabulary component of the IELTS oral interview.

Table 5.1 summarises pre- and post-course mean scores and standard deviations for measures of communicative competence as a function of input condition.

Source	Pre-course		Post-course	
	M	***SD***	***M***	***SD***
Listening				
Experimental group	16.71	4.66	21.95	4.83
Control group	15.36	3.92	19.23	3.81
Receptive pronunciation				
Experimental group	45.83	4.19	48.58	3.49
Control group	44.48	2.99	45.42	3.60
'C'-Test				
Experimental group	84.43	12.29	95.10	9.83
Control group	78.97	9.75	89.71	7.70
Grammar				
Experimental group	106.97	10.99	109.32	7.78
Control group	100.13	12.53	106.87	8.27
Receptive vocabulary				
Experimental group	73.42	11.90	85.19	9.59
Control group	75.0	8.32	78.13	8.24
DCT				
Experimental group	2.86	.40	3.05	.31
Control group	2.77	.30	2.87	.37
IELTS (pronunciation)				
Experimental group	3.82	.50	4.01	.53
Control group	3.51	.46	3.74	.43
IELTS (body language)				
Experimental group	3.80	.58	4.06	.57
Control group	3.39	.67	3.58	.53
IELTS (fluency)				
Experimental group	3.64	.61	3.84	.62
Control group	3.28	.57	3.39	.51
IELTS (vocabulary)				
Experimental group	3.70	.50	3.78	.56
Control group	3.39	.44	3.50	.37
IELTS (interaction)				
Experimental group	3.52	.71	3.75	.72
Control group	3.25	.78	3.24	.56
Role-play (conv. behaviour)				
Experimental group	4.17	.80	4.48	.54
Control group	3.26	.53	3.45	.55
Role-play (conv. management)				
Experimental group	4.06	.92	4.40	.62
Control group	3.26	.51	3.47	.55

Table 5.1 Pre- & post-course mean scores & standard deviations for communicative competence measures as a function of input condition

Table 5.2 summarises the analysis of covariance of post-course communicative competence measures as a function of input condition, with pre-course communicative competence scores as covariates.

Source	*df*	*SS*	*MS*	*F*	*Effect size*[1]
Listening	1	55.58	55.58	4.44*	.07
Receptive pronunciation	1	114.28	114.28	11.84**	.17
'C'-Test	1	116.37	116.37	2.69	.04
Grammar	1	1.12	1.12	.022	<.0005
Receptive vocabulary	1	922.82	922.82	14.81**	.20
DCT	1	.15	.15	1.73	.03
IELTS (total)	1	.72	.72	6.84*	.11
IELTS (pronunciation)	1	.16	.16	1.62	.03
IELTS (body language)	1	1.26	1.26	8.93**	.14
IELTS (fluency)	1	.86	.86	5.01*	.08
IELTS (vocabulary)	1	.27	.27	2.02	.03
IELTS (interaction)	1	2.29	2.29	10.25**	.15
Role-play (total)	1	3.10	3.10	17.58**	.25
Role-play (conv. behaviour)	1	3.44	3.44	17.74**	.25
Role-play (conv. management)	1	3.15	3.15	14.65**	.22

*$\underline{p} < .05$ ** $\underline{p} < .01$
[1]Eta squared

Table 5.2 Analysis of covariance of post-course communicative competence scores as a function of input condition (with pre-course communicative competence scores as covariates)

These results suggest strongly that, after statistically controlling for differences in proficiency levels between groups, learners receiving the experimental treatment (authentic materials) developed their communicative competence to a greater degree than those receiving the control treatment (textbook materials).

By performing separate statistical analyses on each of the various tests used in the trial, we are able to 'tease apart' the individual components of communicative competence and investigate exactly how the two groups differed from each other after the intervention. As we have seen, the results of the analyses show that the experimental group out-performed the control group on five of the eight measures: the listening test, the receptive pronunciation test, the vocabulary test, the IELTS oral interview and the student-student

role-play, while the 'C'-Test, grammar test and DCT showed no significant differences. The following discussion looks at each of these measures in turn.

5.2 Listening test

As the results in table 5.2 show, the difference between the experimental and control groups in terms of listening proficiency was significant after the intervention (p=.039), with a partial eta squared value of .07 indicating that 7% of the variance in post-course scores could be accounted for by the treatment (a moderate effect size). Figure 5.1 illustrates the changes in listening proficiency for the two treatment groups visually:

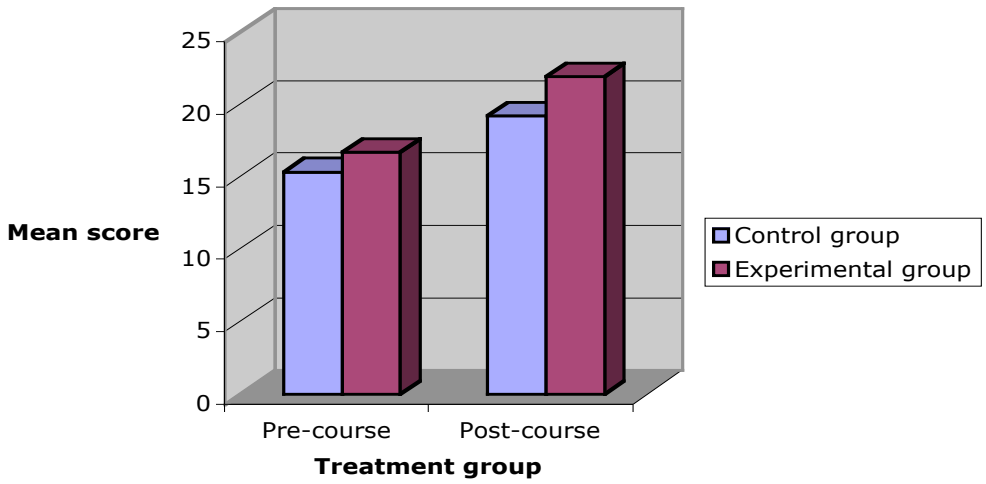


Figure 5.1 Mean scores on the pre- & post-course listening tests for both treatment groups

The quantity of listening input in both groups was similar over the 10-month study and, bearing in mind that a considerable proportion of the actual L2 listening practice students received in the classroom came from the same NS teacher, it is quite surprising to find any difference at all between the groups. There are two possible explanations for this.

The first is that the increased focus on phonological aspects of English (such as stress-timing, weak forms, linking and intonation units) in the experimental group succeeded in raising learners' awareness of these features of natural discourse and indirectly benefited their listening comprehension. Some students commented on this in their diaries:

TK: My listening skill has rather improved because I lerned the features of English. By focusing on the stressed word, the pitch, and something like that, I got to understand spoken English well.

The second possibility relates to the *quality* of the listening input learners were exposed to in the classroom. The textbook listening materials tended to have a slower speech rate and to display fewer features of natural, native-speaker discourse and therefore did not prepare the students for the listening test (which was natural-like) as well as the authentic materials. Again, some learners' diary entries suggest this as the source of improvement in their listening skills:

RM: In addition, when I watch movies and news in English and listen to them, I found my ears were getting used to the natural speed and I did understand.

The most likely explanation is that the difference in listening comprehension between the two groups was due to a combination of both these factors but this is speculation. Whatever the cause, the increased difficulty of the authentic listening materials used in the experimental group certainly did not appear to have a detrimental effect on learners' developing listening skills and this raises the question of how necessary it is to simplify listening texts for learners at intermediate to advanced levels of proficiency.

5.3 Receptive pronunciation test

The differences between the experimental and control groups in terms of their receptive comprehension of phonological features was highly significant after the intervention (p=.001), with a partial eta squared value of .17 indicating that 17% of the variance in post-course scores could be accounted for by the treatment (a large effect size). Figure 5.2 illustrates the changes in receptive pronunciation comprehension for the two treatment groups visually:

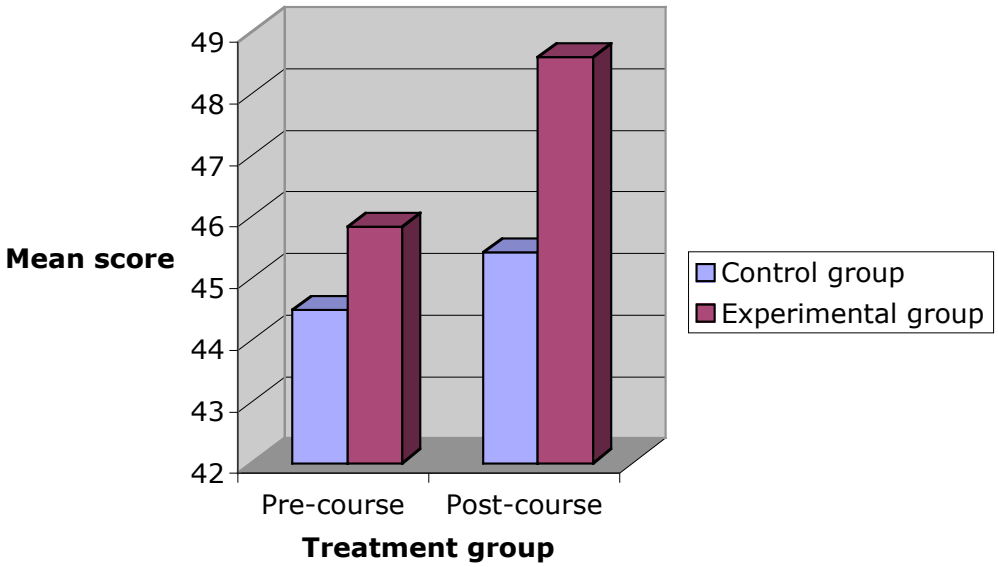


Figure 5.2 Mean scores on the pre- & post-course pronunciation tests for both treatment groups

Since more time was spent focusing on phonology issues in the experimental class, it is hardly surprising to see this difference and it supports the widely held belief that encouraging students to 'notice' features of the target language (whether grammatical,

lexical or phonological) can stimulate language acquisition (Schmidt 1990; Batstone 1996; Skehan 1998). In addition to the increased focus on phonology in the experimental group, the authentic input probably gave students more opportunities to see these features of natural language put into practice since, as mentioned in Section 5.1.1, the contrived textbook listening materials often presented phonologically distorted samples of the L2 in an effort to ease the process of comprehension.

5.4 'C'-Test

No statistically significant differences were found between the groups in terms of their performance on the 'C'-Test (p=.11). This is not particularly surprising since this test focuses primarily on students' reading skills whereas the intervention was designed predominantly to develop students' listening and speaking skills. Table 5.1 indicates that the mean scores on the 'C'-Test increased at a similar rate for both experimental and control groups (12.6% and 13.6% respectively) and this can largely be attributed to input the learners received from other classes over the period of the trial.

5.5 Grammar test

No significant differences were found between groups in terms of their performance on the grammar test (p=.88). Table 5.1 indicates that the mean scores on the grammar test increased only slightly for both experimental and control groups (2.2% and 6.7% respectively) and, again, this is not surprising since students' grammatical competence was already quite well developed, after around 8 years of English instruction largely focused on grammatical aspects of the language. The control group received more

grammar-focused input in their classes, since it was an integral part of the textbook syllabus (see table 4.5) but, even so, it did not lead to any significant increases in their grammatical competence and this is probably because the grammatical items covered were already familiar to the learners and therefore only served the function of reviewing old material. This highlights one of the problems for teachers using published course books in Japan, which is that learners tend to be quite advanced grammatically, but are not sufficiently prepared to cope with the listening or speaking materials associated with upper-intermediate or advanced textbooks produced for the international market. In other words, the Japanese education system, as it currently stands, does not produce learners who are balanced in terms of their communicative competence.

5.6 Receptive vocabulary test

The differences between the experimental and control groups in terms of their receptive comprehension of vocabulary was highly significant after the intervention ($p<.0005$), with a partial eta squared value of .20 indicating that 20% of the variance in post-course scores could be accounted for by the treatment (a large effect size). Figure 5.3 illustrates the changes in receptive vocabulary comprehension for the two treatment groups visually:

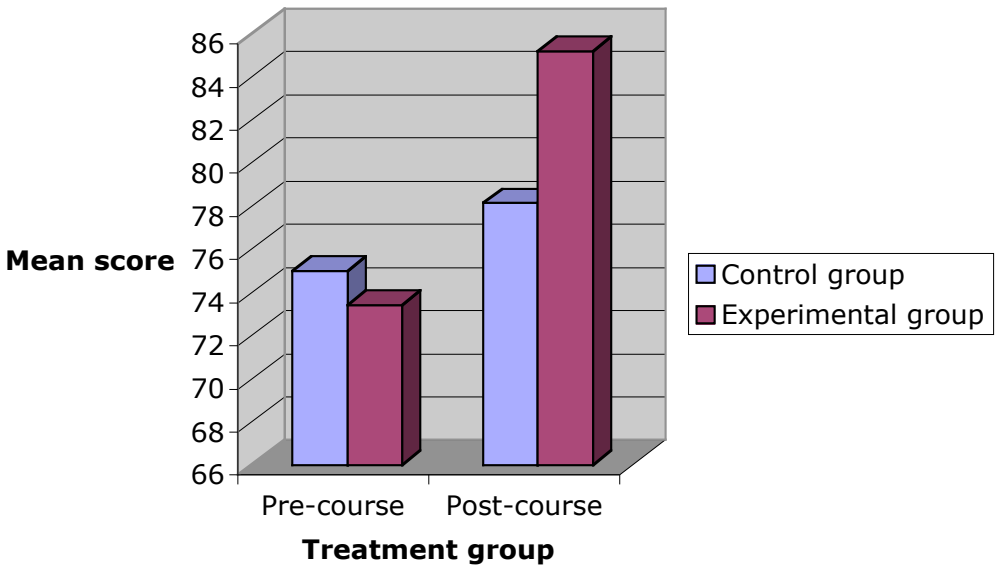


Figure 5.3 Mean scores on the pre- & post-course vocabulary tests for both treatment groups

The marked difference between the two groups post-course is quite surprising, bearing in mind that the treatment condition only accounted for around 33% of the formal English input students received over the 10-month investigation period (see Section 4.2.1), and suggests that the authentic materials were highly effective in developing learners' receptive comprehension of vocabulary. There are a number of possible reasons for this. Firstly, the authentic materials exposed learners to 'richer input' than the textbooks, with a greater number and wider range of vocabulary items, and therefore increased their chances of encountering and acquiring new words. Secondly, the predominant use of audio-visual materials in the experimental group (exploiting scenes from DVDs) meant

that new lexis was highly contextualized when it was presented to learners, and this is likely to have facilitated its acquisition. Thirdly, since it is hypothesized that the authentic materials provided more motivating input than the textbooks were able to do, it could be that learners' engaged more in the learning experience in the experimental group (see Peacock 1997 for support on this issue). Although students' levels of engagement with classroom materials were not measured directly in this investigation, the diary studies, reported on in Section X, do suggest that learners were highly motivated by the authentic materials. Subjectively speaking, from a teacher's perspective, I would also say that the general behaviour of students in the experimental group also supports the notion of higher levels of motivation. Experienced teachers are, I would argue, very attuned to their learners, and in-class observation of learner-behaviour by them can provide quite an accurate estimate of levels of engagement and motivation (this would be an interesting direction for future research on motivation to explore).

5.7 Discourse completion task (DCT)

No significant differences were found between groups in terms of their performance on the discourse completion task (p=.19). This was disappointing, and unexpected, since the experimental treatment included materials and tasks specifically designed to develop learners' pragmatic competence (see table 4.6 & appendix XI). There are at least three possible explanations for this lack of difference.

5.7.1 Hypothesis 1: A problem with the testing conditions?

Talking to a computer in the draughty corridors of a university, whilst being tape-recorded, doesn't exactly create the perfect conditions for the elicitation of speech acts and there is evidence from the learners' diaries that these testing conditions caused students a lot of stress (see also appendix IX):

RM: [DCT] I completely couldn't do my best in computer test because it was almost the last part of class and many people were in the corridor. I couldn't hear well and everything was disturbing me… unfortunately.

MT: [DCT] First I did a test with computer. I don't like this test, actually. I got upset and forgot every words for the moment. I didn't do good.

YK: Some people took speaking test [DCT]. It was quite hard too and I got so embarrassed when some students past by me. He was like "What's going on??"

SF: [DCT] I think speaking to your computer is shameful a bit.

It is quite likely then, that the testing conditions prevented learners performing to the best of their abilities. However, since both control and experimental groups were tested under identical conditions, we would still expect to see a difference if one really existed (as was found with the, equally stressful, student-student role-plays). This, therefore, is unlikely to be the reason for the lack of statistically significant differences on the DCT.

5.7.2 Hypothesis 2: A problem with the instruments used to measure pragmatic competence?

The second possibility is that the discourse completion task was simply not sensitive enough to detect changes in the learners' pragmatic competence. In an early attempt to design a reliable measure of pragmatic competence using a DCT, Cohen & Olshtain (1981: 130) concluded that their efforts had 'produced at best a crude measure of such

competence' and, although the DCT used in this study benefited from lessons learnt from trialing earlier models, perhaps it remains a blunt instrument? Every attempt was, however, taken to make the DCT a reliable measure of pragmatic competence:

a) A multi-media elicitation task was used so that students were able to see the context of the required speech act for each situation on the computer screen, at the same time as hearing the prompt. This was done to limit the extent to which respondents were required to fill in missing contextual details from their imaginations, which tends to distort the results.
b) Rather than asking students to select appropriate responses from multiple-choice answers, they were allowed to answer the prompts in their own words in an effort to produce a more sensitive measure.
c) Students were asked to produce oral, rather than written, responses to the situational prompts and these were recorded and transcribed for rating purposes. This process proved to be considerably more time-consuming, but was felt worthwhile because past research suggests that oral responses are a more realistic indicator of pragmatic competence than written responses, which tend to be shorter and less detailed.

The DCT did, however, have several weaknesses, which could have accounted for a failure to detect differences in learners' pragmatic competence. The native speaker raters were only given written transcriptions of learners' utterances, rather than the taped responses themselves, which meant that all the paralinguistic details, such as tone of voice, pitch changes or loudness (which are an important component of affective speech) were lost and learners were rated only on the actual words they used.

Secondly, the DCT only focused on the use of a single speech act (requests) with same-status or higher-status individuals, rather than a range of different speech acts. ILP studies to date have found no speech communities which lack speech acts for requesting, suggesting, inviting, refusing, apologising, complaining, complimenting and thanking (Kasper & Schmidt 1996), so perhaps these could act as a basis for DCT design.

Thirdly, although the NS raters received quite detailed instructions and guidance on the rating procedure (see appendix IV), their scores for the pragmatic appropriateness of students' responses differed markedly at times. An estimate of the DCT's reliability was obtained by comparing the author's post-course ratings with those of all other NS raters, using Pearson product-moment correlation coefficient. There was a positive correlation between the mean scores [r=.60, n=59, p<.0005] but a coefficient of determination (r^2) value of .36 means that the two variables shared only 36% of their variance, suggesting the DCT had low levels of reliability.

As a further estimate of the test's reliability, a native speaker's responses to the DCT were added in for the post-course rating as a control measure (the NS raters were not aware of this since all responses were rated blind). The raters *did* rate the NS's responses as the most appropriate, giving her a mean score of 4.35, well above the combined total mean of 3.1 for the students in the treatment groups. However, they sometimes varied dramatically in their ratings of the NS's responses: in some scenarios giving a score between 1 and 5 for the same utterance. Scenario 5 of the DCT displayed particularly wide variations so it might be worthwhile examining this in more detail to investigate the reasons for disparity in raters' scores. The DCT prompt for scenario 5 is shown below, along with the NS's response:

Scenario 5: You have to hand in an essay to the secretary. The secretary's office is closing soon and you are already running late. When you get to her office, two professors are standing in front of it. You ask them to let you through.

You say: *Excuse me could you please move aside so I could come in please?*

This response received the following rating scores from the 20 NS raters:

Rating score	Number of raters
5	3
4	4
3	8
2	5
1	0

Table 5.3: Rating scores for NS on scenario 5 of the DCT

It would seem that what is happening here is that the raters are compensating for the lack of paralinguistic information available in the transcriptions of the learners' utterances (in terms of exactly how they were delivered), by imposing their own interpretation of the original sentence stress and intonation as they read the responses. Exactly what they imagine the original utterance to have been like probably varies considerably with each individual rater, so that their final judgment of pragmatic appropriateness rests more on the *manner* they envisage the utterance being spoken in than *the words themselves*, thereby leading to the discrepancies seen.

Another illuminating example is the following post-course DCT response from NK, one of the highest proficiency students in the study, with a TOEFL score of 523. This student had lived in New Zealand for 3 years and would therefore be expected to have higher levels of pragmatic competence than the majority of his classmates, through 'the friction of (his) daily interactions: the pressure of not only making (himself) understood but also of establishing and maintaining smooth relationships with NSs in the host

environment' (Bardovi-Harlig & Dörnyei 1998: 253). Below, his response to Scenario 14 is shown:

Scenario 14: You have an appointment with a professor. When you arrive at her door, two of your friends are looking at her timetable and are blocking the door. You ask them to move aside.

You say: *Hey retard move your fucking arse*.

This response received the following rating scores from the 5 NS raters:

Rater	Rating
1	1
2	1 or 5
3	None (but added the comment: 'I think you made this up')
4	1
5	1

Table 5.4: Rating scores for NK on scenario 14 of the post-course DCT

Rater two's scoring here, of either 1 or 5, highlights another important issue for designers of pragmatic measures which is: precisely who is the group being addressed? Personally, I would have rated this response as a 5 because, for a young male university student talking to his peers, this phrase could be considered highly appropriate, with the use of insults and taboo language serving to reinforce their relationships. However, in this case, the raters (except for rater 2) seem to have imagined *themselves* as the recipients of this comment and therefore judged it as offensive. This, I believe, gets to the heart of the difficulty of measures of pragmatic competence: so much depends on the specific context of production, and on subjective notions of what constitutes an appropriate response, that raters are bound to disagree with each other. Perhaps multiple-choice responses, with

only 1 pragmatically appropriate answer and 4 NNS distracters, would result in a more reliable measure, as well as being much faster to implement?

In a final attempt to improve the reliability of the DCT, modified means were calculated for each respondent, by taking the 5 NS's ratings for pre- & post-course tests and removing the highest and lowest scores from each scenario. The results for the modified means were, however, identical to the original scores so it was concluded that the DCT had not detected any differences in the control & experimental groups.

5.7.3 Hypothesis 3: A problem with the pragmatics training?

A third possible explanation for the lack of significant differences between the control and experimental groups on the DCT is simply that the pragmatics training did not lead to the desired increase in students' pragmatic competence. Although the materials used with the experimental group did look at issues of register (see table 4.6), request speech acts were not dealt with specifically and it could be that development of pragmatic competence relies more on the memorization of specific fixed phrases for particular contexts than on real-time construction of utterances. If this were the case, more general advice in the classroom on producing pragmatically appropriate English might lead to higher levels of awareness of pragmatic issues, but it would not, in itself, necessarily translate into better performances on a DCT. There is some evidence from learners' diaries that this may have been the case:

MW: I also didn't do a good job in computer test [DCT]. Although I have learned how to speak Polite and Casual English, I could not use it properly.

SN: [DCT] The computer test I had today was terrible, I got disappointed at myself… I tried to show several skills I had studied through this course such as formal & informal way of speaking English, yet I could not. If I could take the test again, I'd love to!!

AO: I did the role playing and computer test. I wanted to use some expressions which I studied in your class, but when it came to say something, I could not do well. It was regrettable.

These comments do indicate that the pragmatics training resulted in, at least some of the learners in the experimental group, becoming more aware of register issues but perhaps more input, or more time, is needed before this awareness translates into improved performances on pragmatic measures.

Finally, it should be pointed out that, although the results from the DCT showed no statistically significant differences between the treatment groups, sub-scores on the IELTS oral interview and student-student role-play also measured pragmatic aspects and there is evidence from these measures that some aspects of the pragmatics training, such as opening and closing conversations, *were* effective.

5.8 IELTS oral interview

The differences between the experimental and control groups for the combined means of all 5 components of the IELTS oral interview were significant after the intervention (p=.011), with a partial eta squared value of .11 indicating that 11% of the variance in post-course scores could be accounted for by the treatment (a moderate effect size). However, the statistical analysis of the individual components of the IELTS test shown below, gives a clearer picture of changes in the learners' communicative competence over the 10-month trial.

5.8.1 The pronunciation component of the IELTS oral interview

No statistically significant differences were found between the groups in terms of their performance on the pronunciation component of the IELTS oral interview (p=.21). Table 5.1 shows that the pronunciation mean scores for both groups in the IELTS test did increase slightly over the trial period and this was to be expected since there was some pronunciation focus in both syllabi. The experimental group received more explicit training on phonological features of English, and this appears to have had a larger effect on their comprehension (with a highly significant difference seen in the receptive pronunciation test) than their production. Perhaps improvements in productive pronunciation require more intensive periods of training than the 9 hours or so learners in the experimental group received, or longer periods of consolidation than the 10-month duration of the trial.

5.8.2 The body language component of the IELTS oral interview

The difference between the experimental and control groups for the body language component of the IELTS oral interview was significant after the intervention (p=.004), with a partial eta squared value of .14 indicating that 14% of the variance in post-course scores could be accounted for by the treatment (a large effect size). Figure 5.4 illustrates the changes in appropriateness of students' body language for the two treatment groups visually:

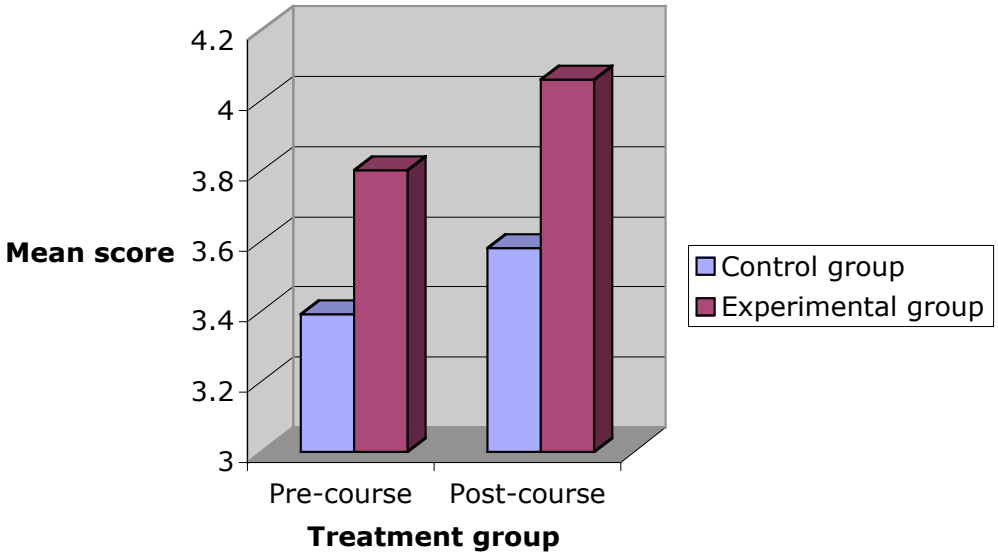


Figure 5.4 Mean scores on the pre- & post-course tests for the body language component of the IELTS oral interview for both treatment groups

This suggests that the sociopragmatic training (on facial expressions, gestures, eye contact & proxemics) implemented in the experimental group did successfully encourage learners to alter their behaviour towards the norms of the target speech community.

This is an important result because little empirical evidence exists (at least in the field of applied linguistics) for the success of this kind of training (although see Collett 1971 for a notable exception). NVC is widely recognised as being crucial for successful communication, and when people from very different cultures interact, sociopragmatic misunderstandings are frequent, and often serious, yet this area is rarely given any mention in current language teaching methodologies. If we want to take the notion of

communicative competence seriously in our profession, then it is crucial that we begin to broaden our training programs to include all of its different dimensions and not just those we are most familiar with. The results presented here suggest that NVC training in the classroom can improve learners' sociopragmatic competence and I would strongly argue for its inclusion when there are wide disparities between the students' culture and the target culture.

5.8.3 The fluency component of the IELTS oral interview

The difference between the experimental and control groups for the fluency component of the IELTS oral interview was significant after the intervention (p=.029), with a partial eta squared value of .08 indicating that 8% of the variance in post-course scores could be accounted for by the treatment (a moderate effect size). Figure 5.5 illustrates the changes in students' oral fluency for the two treatment groups visually:

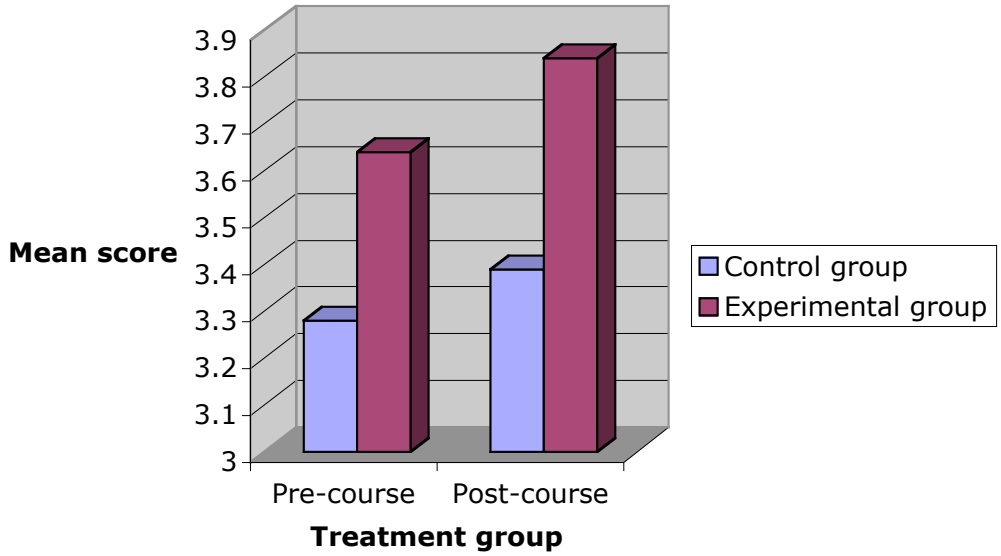


Figure 5.5 Mean scores on the pre- & post-course tests for the oral fluency component of the IELTS oral interview for both treatment groups

Table 5.1 shows that both the experimental and control groups improved in fluency over the 10-month investigation and this was anticipated since both groups were given numerous speaking opportunities in class: like any other skill, speaking improves with practice.

What is of particular interest here, though, is what led to the significant difference in fluency between the two groups. Since both treatments involved similar quantities of student talking time, the enhanced fluency in the experimental group is likely to have come from the explicit focus on conversational strategies (see table 4.6). Perhaps the learners were using hesitation devices more often or more effectively, thus appearing more fluent or perhaps the focus on discourse intonation and tone groups encouraged

learners to pause in more appropriate places in the discourse (for example, at TRPs). Without a quantitative analysis of the interview transcriptions, which is outside the scope of this investigation, it is impossible to come to any firm conclusions on this matter. What seems clear, however, is that some aspect of the explicit focus on conversational strategies had a beneficial effect on learners' fluency and this supports the use of awareness-raising, or 'noticing', strategies in the classroom (Schmidt 1990; Batstone 1996; Skehan 1998).

5.8.4 The appropriate vocabulary use component of the IELTS oral interview

No statistically significant differences were found between the groups in terms of their performance on the vocabulary component of the IELTS oral interview (p=.16). Examination of table 5.1 suggests that the mean scores for both groups changed very little over the period of the investigation on this criterion and there are two possible explanations for this. The first is that there *were* changes in the appropriateness of students' vocabulary use, but that these were difficult for raters to pick up while watching the video and, at the same time, grading all five sub-components of the IELTS exam. This is a distinct possibility since, as we saw in section 5.6, differences between the control and experimental groups on receptive vocabulary use were highly significant at the end of the trial. Furthermore, rating appropriate vocabulary use is extremely difficult in real time and, without the benefit of a written transcript, decisions on the appropriateness of each individual word have to be made instantaneously. I would argue that this imposes unrealistic demands on the raters. The second possible explanation is that the learners' productive use of new vocabulary lagged behind their receptive comprehension. This is a possibility since we might expect learners to require numerous encounters with new words before they feel sufficiently confident to begin using them in their own discourse. Perhaps the 10-month trial period was simply too short to detect these kinds of changes.

5.8.5 The interactional competence component of the IELTS oral interview

The difference between the experimental and control groups for the interactional competence component of the IELTS oral interview was highly significant after the intervention (p=.002), with a partial eta squared value of .15 indicating that 15% of the variance in post-course scores could be accounted for by the treatment (a large effect size). Figure 5.6 illustrates the changes in students' interactional competence for the two treatment groups visually:

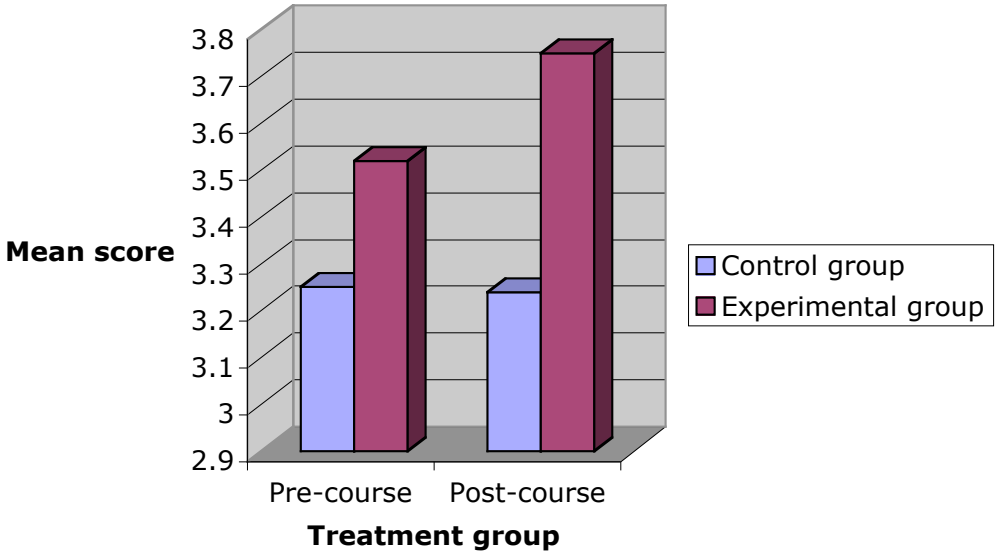


Figure 5.6 Mean scores on the pre- & post-course tests for the interaction component of the IELTS oral interview for both treatment groups

The mean scores in table 5.1 show that the control group did not change at all on this measure, while the experimental group improved 6.5%. Again, this suggests that the

explicit focus on conversational strategies (such as turn-taking, developing conversation, using reactive tokens, discourse markers or hesitation devices) in the experimental treatment was effective in bringing about changes in the learners' discourse competence.

This result also has potentially important implications for language teaching. Insights into the discourse strategies employed by NSs during conversation have only recently become available to teachers, thanks to the rapidly expanding fields of discourse and conversational analysis. Although some language textbooks and resource books have begun incorporating these insights into their designs (see, for example Nolasco & Arthur 1987; Dörnyei & Thurrell 1992), little empirical research currently exists to support the awareness-raising of discourse features in the classroom. Intuitively, it makes sense that they *would* benefit learners: the recognised value of 'noticing' on acquisition probably applies to all features of language, not just grammatical items, which are generally the focus of attention. The results presented here therefore support the incorporation of training in conversational strategies into the classroom.

5.9 Student-student role-plays

The differences between the experimental and control groups in both components of the student-student role-play were highly significant after the intervention for: a) the conversational behaviour scale ($p<.0005$), with a partial eta squared value of .25 indicating that 25% of the variance in post-course scores could be accounted for by the treatment (a large effect size) and; b) the conversational management scale ($p<.0005$), with a partial eta squared value of .22 indicating that 22% of the variance in post-course scores could be accounted for by the treatment (also a large effect size). This pronounced

difference between the two treatment groups was seen despite the fact that the period of investigation was half that of the other communicative competence measures (the first role-play was conducted in September, rather than April 2004).

5.9.1 The conversational behaviour component of the role-play

Figure 5.7 illustrates the changes in students' conversational behaviour on the ss-ss role-play for the two treatment groups visually:

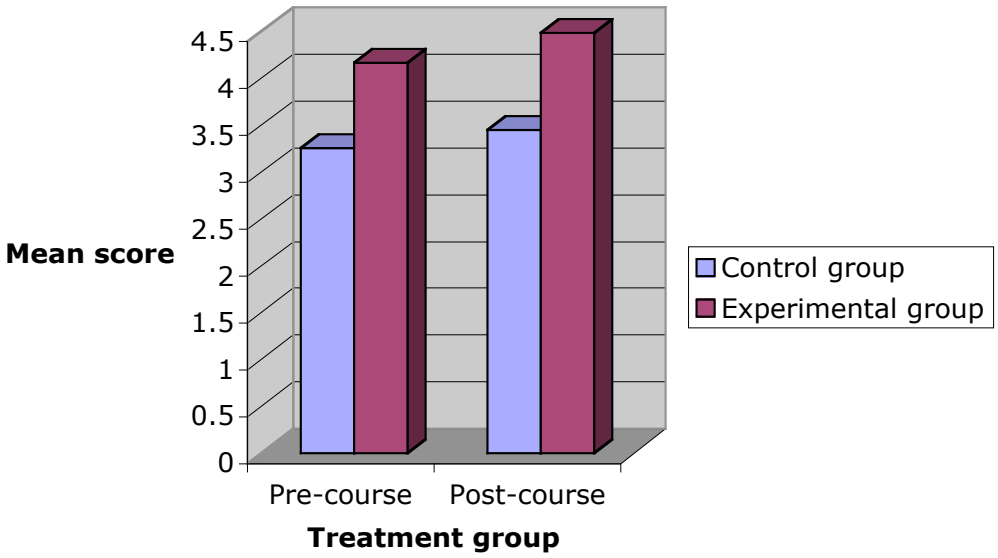


Figure 5.7 Mean scores on the pre- & post-course tests for the conversational behaviour component of the ss-ss role-play for both treatment groups

As appendix VII illustrates, the conversational behaviour measure focused on sociopragmatic aspects of the students' performance and was therefore similar to the

body language component of the IELTS oral interview, except that the interaction was NNS-NNS, rather than NS-NNS. The results provide further support that NVC training can produce a marked change in student behaviour, towards NS norms, in relatively short periods of time. The difference in performance between the two groups was largely anticipated since the experimental treatment involved specific advice on, and practice with, the features measured by this rating descriptor but it is encouraging to see that the learners were so readily able to incorporate these behaviours into their own productive repertoires. Because the role-play involved Japanese interacting with other Japanese, the changes in behaviour evidenced were obviously more for the benefit of the camera and the observing teacher than for each other. But this kind of 'play acting' is precisely what is required for learners to appear more competent, sociopragmatically, in a target culture. It is clear from the learners' diaries that some of them found it difficult to adapt their NVC towards Western norms:

SN: Body languages we employ are one of the clearest examples which show us how Japanese and British are different when we communicate. As I told you before, it is still a bit difficult for me to act like a foreigner. Japanese use by far fewer body languages than European or American, I think, of course, knowing how they behave when communicates is important in order to make a proper understanding what they really want to mean.

But, however strange the NVC patterns might feel to the students themselves, they are likely to appear entirely normal to the target population and, hopefully, will lead to a positive evaluation and greater levels of acceptance into the speech community.

5.9.2 The conversational management component of the role-play

Figure 5.8 illustrates the changes in students' conversational management on the ss-ss role-play for the two treatment groups visually:

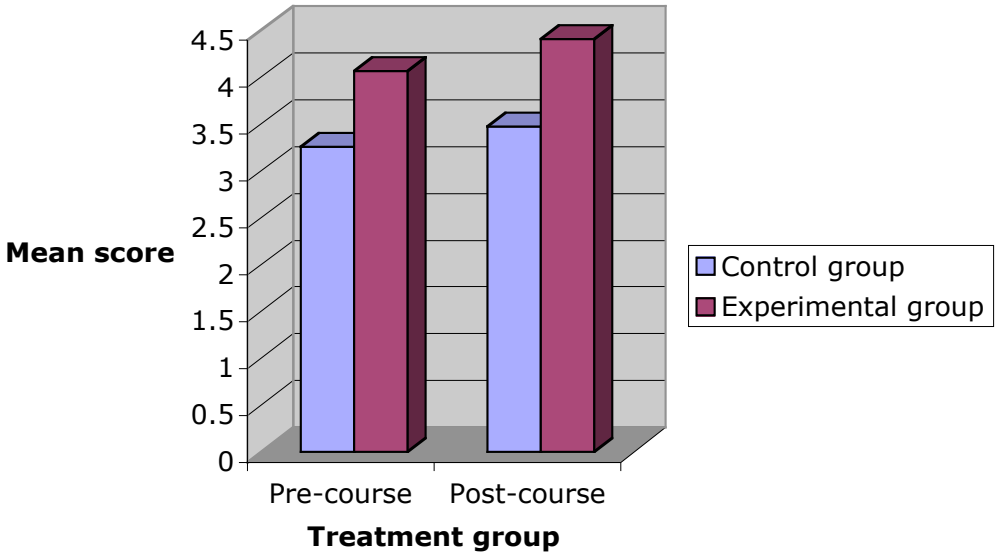


Figure 5.8 Mean scores on the pre- & post-course tests for the conversational management component of the ss-ss role-play for both treatment groups

As appendix VII illustrates, the conversational management component of the role-play focused principally on discoursal aspects of the students' performance, and was therefore similar to the interactional competence component of the IELTS oral interview. The role-play was, however, considered a more sensitive measure of students' discourse competence because learners had more opportunities to take responsibility for topical coherence and topic development in the absence of a NS interlocutor. The results for this measure strongly support the earlier conclusion that an explicit focus on conversational strategies can benefit learners' discourse competence. Whereas the control group were only given opportunities to speak together in pair or group-work activities, the

experimental group received specific advice on how conversation works in English, as well as practice incorporating these strategies into their own conversations.

The significant difference between the two groups in terms of their role-play performances therefore supports a language teaching methodology that aims at explicit awareness raising and practice of discourse features, rather than one that simply provides students with speaking opportunities, as traditional 'conversation classes' tend to.

5.10 Conclusion

The results in this chapter indicate that the experimental programme was significantly more effective in developing participants' overall communicative competence than the control treatment. This, therefore, validates the theoretical assumptions upon which the experimental syllabus was based and provides strong support for a change in the way that classroom language learning is conceptualised and implemented.

CHAPTER 6. RESULTS & DISCUSSION (PART II): QUALITATIVE MEASURES

The decision to include qualitative evidence (from learners' diaries and case studies) in the investigation was made in response to criticisms often directed at classroom-based research in the literature. These accuse many reports of being biased towards 'the visible' and blind to the social context of learning, leading to a distorted picture of classroom processes (Mehan 1979; Beretta 1986; van Lier 1988; Allwright & Bailey 1991; Nunan 1991; Alderson & Beretta 1992; Johnson 1995). My unique position as both researcher *and* teacher in this investigation meant that, not only was I not blind to the social context of learning, I was actually a *part of it*. This does, of course, bring its own complications. As I mentioned in Chapter 3, 'mixed methods' research requires the researcher to have essentially two heads, a 'psychometric head' and an 'ethnographic head'. The psychometric aspects of the investigation demand a researcher's more objective (etic) perspective on events, while the ethnographic aspects benefit from the teacher's more subjective (emic) perspective. In this Chapter then, as the use of the 1st person pronoun suggests, I relinquish all attempts at objectivity and describe events as my students and I experienced them, as a co-constructors of the social reality of the classroom.

6.1 Learners' diaries

6.1.1 Compliance rates

As Halbach (2000) reports, collecting data from learners is far from easy. He found that diaries were often of poor quality, overly brief and lacking in detail, and that compliance rates were less than 50% (see also Carroll 1994, for similar results). In this study,

however, the compliance rate was much higher (84.2%), although the diaries varied quite dramatically in quantity (from two to fifty pages) and quality (from extremely brief summaries, commenting on only the content of classes, to detailed and thoughtful notes on students' impressions of materials). The two diary extracts below (both from students in the experimental treatment group) illustrate differences in the quality and usefulness of responses received:

YS:
Nov. 30th (Mon) As usual. Everyone was talking excitedly.
Dec. 1st (Wed) I had a stomachache, so I didn't concentrate on the work.
Dec. 6th (Mon) Nothing especially. But I lost my paper. It's very kind of you to give us sub-paper.
Dec. 13th (Mon) I don't remember.

RM:
9.29.04 Today, we first checked our homework. According to the sheet, English conversation is a bit like playing tennis. It was interesting to know that I have to be competitive in English conversation. I agree that most Japanese people are likely to wait till a turn to talk comes to them. So I thought I have to try to be more active in the conversation when I speak English. After that, we learnt more about how to develop the conversation. I realized that native English speakers are really good at expanding the answers. As for me, I'm not good at it so I think I should try to find an interesting subject in the conversation and expand it. The last thing we did was to talk on a certain topic. Alex put us into a group. First I talked to Tomoe on holidays. I tried to expand the answer and asked as many questions as I could. Joe gave us a feedback on the conversation and he told me it was really natural so I was really happy.

This unfortunate reality imposes an unavoidable bias on the diary studies, in that the 'significant events' recorded in appendix IX tend to represent the views of the more motivated members of the class (or, at least, the more vocal and opinionated). YS, whose terse diary extracts are shown above, was an exceptionally introverted student who, despite my best efforts, did not respond at all well to any of the activities offered up in the Communicative English class. He gave the distinct impression that he would rather have been reading a good book than talking to other students, and was, in fact, frequently reprimanded for doing just that during lessons. My own view is that, because of his introverted personality, YS would be unlikely to have approved of any type of

communicative activity, whether a part of the experimental or control treatment. However, his voice is largely lost from this investigation because his rebellion against the classroom activities largely manifested itself in silence and withdrawal. Having said that, YS was by no means a typical member of the class and it is hoped that the students' comments discussed in this chapter represent a wide range of views from the classroom.

As Richards & Lockhart (1994: 18) remark, the discipline of diary keeping is 'a burden on the time and energies of participants' so methods need to be developed to encourage learners to respond in detail. Bailey (1990) and Porter et al. (1990) recommend leaving time at the end of class for learners to record their thoughts while events are still fresh in their minds and, although this is useful advice, it proved impractical in this study due to time contraints. Instead, I collected in and read the diaries twice a term, and responded to students' comments and questions with remarks in the margins, in order to encourage them to continue their diary keeping throughout the course.

In actual fact, learners did not need much persuasion to write diary entries, indeed many of them seemed to find relief in this outlet for their thoughts and feelings since the Japanese classroom imposes the restraint of 'social discretion' on its members. Lebra (1987: 347), in his important paper 'The cultural significance of silence in Japanese communication', describes this as the 'silence considered necessary or desirable in order to gain social acceptance or to avoid social penalty'. As a teacher in Japan, I am more than familiar with this form of Japanese reserve: it is the source of learners' reluctance to 'play by the rules' (or, at least *my* rules) in the classroom and answer my questions freely, causing lessons to grind to a halt, with an uncomfortable silence, as I wait for replies, and frequently leaving me with no option but to provide the answers myself. For this reason,

Japanese classrooms do not tend to follow the Initiation-Response-Feedback (IRF) model, reportedly typical of western classrooms (Sinclair & Coulthard 1992). The tension which this causes in Japanese classrooms, has also been noted by other foreign teachers:

> When I privately asked students whom I had come to know why they would "pretend not to know," why they would not answer in class, or would refuse to say anything, they usually said that they "were afraid of making mistakes," "were afraid of instructors," "thinking too hard," "I'm too nervous," "I feel tense." Others explained that being in the classroom is a "strained situation" or has a "strange atmosphere". Some students had a negative attitude toward those who answered in class: "a person who answers cannot be a nice person"; "such students are imprudent"; "students who answer are being bold." (McVeigh 2002: 99)

Lebra (ibid: 353) sees diary writing as a useful medium to release this socially imposed tension and his insights into the Japanese psyche help to explain the high compliance rates for diary keeping in this study:

> Since silence occurs in face-to-face interaction with another person(s), the frustrated silent 'speaker' may choose to break the silence through a sort of monologue which takes the form of writing. It is no coincidence that many Japanese keep or try to keep diaries – at least more than Americans to the best of my knowledge. In a type of psychotherapy called *Morita* therapy, which does not rely upon vocal communication between therapist and patient, the patient is required to keep a diary and the therapist writes his comment in the margin.

6.1.2 Collection procedure

Following Bailey (1983; 1990) and Bailey & Ochsner (1983), a five-step procedure for collection of diary data was used:

1. Diaries begin with a full account of the diarists' personal learning history (see Section 4.2.1).
2. Diarists are encouraged to be as candid as possible in their diary entries (see appendix VIII).
3. The initial database is revised for public consumption.
4. Patterns and significant events in the diary entries are identified (see appendix IX).
5. The data is interpreted and discussed.

As with any qualitative study, only small amounts of the total data could be included in the final account, making step 3 above particularly crucial since it determines what information from the diaries is chosen for public viewing. This is a subjective process, invisible in the final report and therefore open to researcher abuse or bias. However, since the diaries need to be 'distilled' into a digestible form for public consumption, it is a necessary step in the procedure.

The approach adopted in this study was to focus on recording comments from the diaries, which included learners' impressions, opinions and feelings on the classes, rather than just factual information. Appendix IX shows the complete record of significant events extracted from the learners' diaries and these have been organised into 'recurring issues' (Krishnan & Hwee Hoon 2002) identified in the diaries.

6.1.3 Recurring issues in learners' diaries

Learners' diary entries generally fell into one of four main categories, although the sorting process was sometimes complicated by the fact that some comments could simultaneously belong in more than one group. When this was the case, classification was determined according to what the writer seemed to be emphasising most. The four thematic areas of focus were: a) Comments relating to aspects of the input or activities learners 'noticed' in the classroom (components of the communicative competence model or language skills); b) Comments relating to the learning environment (the teacher, other learners, tension in the classroom, external concerns); c) Comments relating to aspects of the lessons learners found motivating or de-motivating; and d) Comments relating to the pre- and post-course tests or testing procedures.

Table 6.1 below shows the number of comments found for each of these recurring issues, for both experimental and control groups in the study.

Recurring issues	Experimental group (number of comments)	Control group (number of comments)
A. Noticing		
1. Linguistic competence	15	55
2. Pragmalinguistic competence	25	2
3. Sociopragmatic competence	60	7
4. Strategic competence	20	1
5. Discourse competence	34	0
6. Listening	25	16
7. Speaking	5	10
8. Reading	1	0
B. Learning environment		
1. The teacher	5	5
2. Other learners	10	8
3. Tension in the classroom	8	23
4. External concerns	4	13
C. Motivated/de-motivated by…?		
1. Motivating materials & activities	73	82
2. De-motivating materials & activities	15	26
D. Pre- & post-course tests	16	10
Total number of comments	316	258
Approximate number of words	15,880	8,670

Table 6.1 Summary of types of recurring issues found in learners' diaries

The patterns indicated in the recurring issues broadly follow those that would be expected from the different types of treatment the two groups received. The control group tended to notice features relating to linguistic competence (grammar, vocabulary or

pronunciation), reflecting the bias of the textbook materials towards this kind of information. The experimental group, on the other hand, noticed a much wider range of features, reflecting both the richer input that the authentic materials provided and the focus of the tasks designed to go with them. In terms of the 'four skills', the majority of the comments centred on listening and speaking skills development since these were the main focus in the Communicative English course provided to both groups.

A significant number of comments in students' diaries focused on the learning environment within the classroom and the relationships of learners with each other and the teacher. These highlight the critical role that the social context can have on learning, something which is particularly interesting since it is a moderating variable rarely mentioned in quantitative research reports.

Another common theme represented in the diaries relates to which materials and activities students either liked or didn't like. As Krishnan & Hwee Hoon (2002: 232) point out, these reflect a range of 'learning agendas' which can be usefully incorporated into curriculum planning to tailor courses more specifically to learners' needs. In this study, however, the treatment was pre-determined, so the comments were used solely to provide evidence of the success or failure of the two treatment conditions. The number of positive comments was similar for both groups but there were around twice as many negative comments (on the materials, activities or classroom environment) in the control group, perhaps indicating higher levels of overall dissatisfaction. This is supported by feedback from the only two participants in the trial who switched from the control group to the experimental group, KM and YN (they were therefore excluded from the quantitative analysis):

KM: Honestly I thought your class was boring in first semester because the class was mainly text activity. I thought "this is communicative English III so I don't want to study the same way as high school or communicative English I and II". I considered the class as easy class. Once I think so, it is difficult to keep high motivation to improve English skill in the class. I wanted more challenging and enjoyable class. In second semister, my mind was dramatically changed. Your class I were based on discussion and video activity without textbooks. This was really enjoyable and challenging.

YN: Class I [experimental group] and II [control group] are very different. Class I is more positive.

The final category in table 6.1 shows learners' comments on the pre- and post-course tests and these suggest that students were under considerable pressure whilst taking (at least some of) the tests. The different categories of recurring issues are investigated in more detail in the following discussion with illustrating example extracts. However, any readers wishing to reach their own conclusions can see the entire data set in appendix IX.

Noticing in the classroom: components of communicative competence

Noticing proceeded largely in line with expectations, reflecting the different kinds of input and activities exploited in the two treatment groups. The experimental group noticed, and commented on, a much wider range of features relating to different aspects of communicative competence since the authentic materials provided a richer source of input and the tasks also encouraged awareness-raising of them.

Fewer linguistic competence features were mentioned as there was less explicit focus on this area, but the most common ones referred to the use of informal language or to pronunciation issues:

EA: Also there were some slangs that I have never know like grand, love every minutes and so on. So this class was very useful for me to know and learn new sentences.

TY: I know very well that I always speak flatly when I speak English. I think that's why I don't distinguish the way I speak Japanese from the way I speak English. So when I speak English making the pitch range, I feel uncomfortable and strange anyway. I have to get rid of this bad habit and want to speak more fluently.

As in these two examples, the majority of students responded very positively to learning about colloquial language and English prosody: both areas largely ignored in the textbooks Japanese are likely to encounter during their studies at school or university. Formal, rather than informal, vocabulary tends to be presented to learners in their language models since it is widely regarded as having greater prestige, but this often results in students producing utterances, such as 'I have been associating with my girlfriend for a year', which are inappropriate for the contexts they most often use the L2 and can make them appear distant or aloof.

Pragmalinguistic features were often mentioned and these tended to be associated with issues of register or opening and closing down conversation:

SN: Did I tell you about my part-time job? I have worked at Ito-Yokado as a clerk of customer-service. Sometimes foreign people come to ask something to me. I can somehow catch what they say, and try to respond in a polite way, for of course they are customers. Then I freeze. How should I say? Is it ok and polite to say such a such? Of course in the end, I can manage it but some worries won't leave me. Now that I know some politer ways, they will be of some help.

TK: Today, we learnt how to close a conversation. For example, "Well… I've got to go", "OK then, thank you for your help!" and so on. Actually, I was surprised that there were so many ending clauses. Now I know them, so I want to use them when I talk with foreigners!

These examples are interesting because they illustrate a problem very specific to Japanese learners. As Loveday (1982) points out, Japanese tends to be quite codified and rigid, in terms of what language is appropriate for a given context. English, in comparison, tends to prefer more varied and individualised responses, illustrating a 'need for a more personalized reaction' from interlocutors (ibid: 7). This is likely to be the source of SN's

hesitancy in dealing with English speaking customers: he seems to be searching for that one, appropriate speech act and worries whether or not he has got it right, seemingly unaware of the fact that any of a range of responses would do just as well. It is also likely to be the source of TK's (and others') surprise at the range of possible closings in English conversation. The realization of this 'pragmalinguistic freedom' in the English language might very well relieve Japanese learners of some of the tension they feel in L2 communication, as well as leading to increased fluency as they self-monitor less. Furthermore, it could benefit learners' sociopragmatic competence by illustrating the value we put on individualism over collectivism and how this manifests itself in the language. HH's comments below suggest a growing awareness in the experimental group of language variation in English:

HH: I think my English skill has improved by studying in this class. We talked about many kinds of things, so I became able to speak correct English at each situations and make dialogues naturally. I also learned many ways to say, for example, Formal, Informal, Polite, Rude and Argue types. They are useful and very interesting for me.

Sociopragmatic features were mentioned frequently in the experimental group's feedback. Comments often related to aspects of non-verbal communication (NVC):

MiW: Today, we practiced how to introduce ourselves by using shake hands, hug, kiss one cheek. Japanese never do kiss in greeting. Therefore, I was ashamed to do so. However, I wanted to get used to the way of it!!

RI: I must not speak English as same as when I speak Japanese. I need a lot of gesture, eye contact, and a tone of voice… The way of speaking English is more exaggerated and overreacted!

ES: The role play was fun. I enjoyed playing with my partners. But I was still embarrassed to do eye contact. I'm not used to do that. I need to practice eye contact or facial gestures more.

TK: Body language is also the important part of conversation. Body language makes conversation more exciting. Shaking hands is the most common. When I met A's speaking partner [foreign students at the

university are often paired up with a Japanese student for language practice], I introduced myself to him and shook hands at the same time. I felt happy to use what I learned in the class immediately.

These comments suggest a desire on the part of students to adapt their NVC towards target culture norms, although they seem to find it difficult to do so. Reluctance to adapt behavioural patterns is often noted in the literature (Hinkel 1996; Siegal 1996) but the majority of students in this study seemed quite positive towards changing their NVC when communicating in English. Perhaps this is reflection of the novelty of the area for the learners, as well as the EFL, rather than ESL, environment for language learning in this study. Faced with different patterns of NVC every day whilst living abroad, may alter students' attitudes to these sociopragmatic features.

Learners also frequently commented on differences in styles of interaction between Japanese and westerners:

RM: [NSs planning a week-end away] Mark & Alison are very nice and funny husband and wife!! I found that their way to decide the plan was different from the way I did with my pertner. Both Mark and Alison gave their opinions clearly. In other words, they equally insist where they want to go and where they don't. Ours was like the one person mainly decide and the other just agrees
.

RM: In Japanese conversation, harmony is important and Japanese cultural aspects shows that sense of "pause" is also Japanese cultural beauty. I don't say which is better, English or Japanese. For me, both is interesting.

ES: [being assertive in conversation] Today's class was quite sporting. We were constantly talking during the class. I tried very hard to be dominant in my group, but I failed. My group's members were all talkative and Westernized. I couldn't be as bold as them. One of the reasons would be my background. My father is a traditional Kyoto person and my mother is an old-fashioned person grown in a countryside. I'm sure my family is completely typical Japanese family. If I talk as I did in today's class in my home, I might be chased out the home. My family won't allow me to talk like that. It is quite natural that I couldn't do in today's class because I've been grown in the traditional Japanese family for 20 years. Therefore, I need to change my character in family and classes.

TY: "Japanese conversation has been compared to ten-pin bowling". But I think there is a reason for that. Japanese tend to think "cutting in the conversation" can be selfish. I think that's why Japanese don't join the conversation very much. They care about how everyone look at themselves more than English speakers. So, today's class was really interesting because I could see the difference between Japanese and English speakers through the study of developing conversation.

RM: [starting conversations] It was obvious that when listener gives more information, the conversation is successful and I was surprised that Americans doesn't mind about speaking their personal affairs even about their divorcing. I think in Japan, we hide those personal aspects as much as we can. So I thought it was needed to make and decide the topics depend on the person we talk to.

These comments suggest an increasing intercultural awareness within the experimental group, which, it is hoped, would enable them to integrate into the target culture more easily. In other words, the authentic materials seem to be encouraging development of students' *Intercultural communicative competence* (Byram 1991: 19).

Issues associated with students' strategic competence were also common in the experimental group's diary entries. A number of the comments hint at the sense of shame Japanese seem to feel when there is a breakdown in communication:

MH: I'm really bad at communication strategies. When I was in U.S., I always had problems with it. I had to explain (describe) what it was like in the situation that I don't know the word. I tried hard, but took so long to explain or describe. So I was depended on my electronic dictionary. Today's lesson was useful for me. And it is true that we tend to run away from the problems like you said. I do too even I don't want to do that. I am learning many things in your class!!

TY: Today's class was really interesting for me. "A Tokyo Story" [teacher's personal story] was very fun. At the same time, I thought that there're many Japanese who do the same thing as the girl in the story [run away when there is a breakdown in communication]. I work at fast-food restaurant as a part-time job, I often see the people who in trouble like this. People who aren't good at English often tend to get panic easily.

TK: I have thought that asking again what you said is not good. I was afraid that if I asked repeatedly, partner might be angry or disappointed and think, "Do you really listen to me?" It was wrong idea. If I don't ask anything, the problems become deeper and deeper. Before that, I'm going to ask and confirm information. I think it makes the conversation more interesting.

MT: I also pretended to understand what my host family said when I was staying in America, and it was no good. And when I took an interview [IELTS oral interview], I did it once and I was confused afterward. So I stopped to ask what she mean. She never behave bad to me, and told me very kindly. At that time I realized that I shouldn't hesitate to humiliate myself. So today's class was very useful for me.

This sentiment is also reflected in the Japanese saying '*kiku wa ittoki no haji. Kikanu wa isshou no haji'*, which roughly translates as 'To ask once is shameful, but to *not* ask is

shameful for a lifetime' (meaning it is better to ask when you don't understand because the shame is less). This sense of 'humiliation', as MT puts it, tends to make Japanese extremely uncooperative participants in the process of negotiation of meaning when communication breakdown occurs. As a learner of Japanese, attempts to clarify misunderstandings or to seek 'scaffolding' assistance are regularly sabotaged by NS interlocutors who either refuse, or don't know how, to cooperate in the process of conversational repair. This considerably slows the foreign learners' acquisition of Japanese in everyday situations because the opportunities for learning, which instances of communication breakdown provide, are rarely taken advantage of.

The source of the Japanese speaker's sense of shame in these situations is something of a mystery but I would suggest three likely candidates. Firstly, the Japanese have a long history of cultural isolation, most clearly manifested in the *sakoku* ('closed country') policy from 1641 to 1853, when the death penalty was imposed on any foreigners or Japanese attempting to enter or leave the country. These isolationist tendencies have limited Japanese exposure to unfamiliar languages and cultures and have therefore also minimised the frequency of communication breakdown. Perhaps, then, part of the sense of shame comes from the rarity of communication breakdown in people's everyday conversation. Another possible reason stems from the Japanese desire to maintain harmony in conversation. Barnlund, in his well-known paper 'The public self and private self in Japan and the United States' (1975: 56), found that Japanese responses to the statement 'what I wish I were like in relationships' included: I try to be as polite as possible; I don't say all of what I think; I try to keep the conversation pleasant; I try to behave smoothly; I try not to disagree. This contrasts sharply with casual speech in

English where supporting responses tend to have the effect of closing down conversation, while confronting responses help to drive the discourse forward (Eggins & Slade 1997: 182). A final possible source of the difficulty Japanese have with conversation breakdown lies in the rigidity built into the language, referred to earlier in this section. With fewer linguistic choices available to express a given concept, perhaps some Japanese struggle to find ways to rephrase their ideas in any other way than one they have already used.

Another conversational strategy often commented on in the experimental group's diary extracts was the use of hesitation devices:

RI: When I used hesitation devices like "Erm", "let me think", "anyway" and so on, I seemed to be a person like an English native speaker. I would like to use these hesitations positively as I would speak English.

MiW: Today, we practiced how to use hesitation words. The English way of hesitate is different from Japanese. I need to be careful when I speak English because I usually say "E-to" or "Uh-n".

These indicate a positive attitude amongst students towards learning this kind of strategy and a desire on their part to speak 'naturally'. As MiW also illustrates above, many learners continue to use Japanese hesitation devices when using the L2, apparently unaware of how unnatural this makes their English sound. An explicit focus on this strategy in the classroom helps to counter this bad habit and because the range of hesitation devices is limited to a handful of utterances, students can be taught what they need to know very quickly.

Common issues in the experimental group relating to discourse competence included the use of 'listener responses' (back-channels), 'sign-posting language' (discourse markers) and topical development or cohesion:

YK: Today we studied various listener responses, and I thought that I was happy when listener widely responded to my conversation. I'm glad to know that there are a lot of listener responses.

YM: Today, we focused on 'sign-posting language' first. It was quite hard to classify into right categories. I think that is showing that I'm not familiar with them. There were some words that I didn't know such as 'If you ask me' and 'To put it bluntly', so it was a great opportunity for me to learn and practice using them!! I will surely try to use them when I speak English, which hopefully could make people impressed.

RI: I studied how I could keep conversation (going) naturally. It is effective for an interviewer to ask the opening question; how, what, where and when. Besides, a listener should expand his answers. If so, conversation will be kept (going) naturally. I could continue to talk with friends more naturally than I had thought. I think I want to try such conversation with any other people.

Again, the reaction of students to learning these discourse strategies was extremely positive, with most considering them 'interesting', 'more natural' or 'useful'. Japanese learners often experience problems developing conversation in English, as the extract from a university pair-work activity below illustrates. Here, <S1> and <S2> talk together about their hobbies:

1<**S2**> You you are good at pia playing piano?

2<**S1**> Yes (laughs) er can you cook?

3<**S2**> Er I'm good at peeling the apple skin

4<**S1**> Really? (laughs) Aah I can't do that

(Author's data 2002)

<S1>, rather than developing the topic of playing the piano put forward by <S2>, quickly abandons her attempt and switches to a completely new theme. This kind of discourse, displaying a lack of topical development/coherence, is a regular occurrence in student

classroom interaction and can have serious consequences in their relationships because it increases the interactional demands placed on the conversational partner. Native speakers in a target community are likely to quickly tire of this constant closing down of topics and look for new conversation partners, thus depriving students of opportunities to make friends.

The control group, receiving textbook input, not surprisingly commented on predominantly linguistic features in their diaries (pronunciation, grammar items and vocabulary):

YK: At last [finally] you made us pronounce correct way as usual. It's really helpful!!

MT: We learned the difference "Do you ever…?" and "Have you ever…?" I didn't know the difference so I was very happy to understand the difference.

SF: I learned the names of many parts of human's body. It increased my vocabulary a lot!!

HM: I learned some idioms. I like "my cup of tea". But, I prefer "my cup of hot chocolate". That's very pretty. I remembered the idiom now, so I want to use it when I find a guy who is my type.

The comments are usually positive and learners seem to have found much of the input valuable, but in terms of developing their overall communicative competence the materials were clearly limiting when compared with the experimental treatment.

The four skills

Most comments on the four skills in diary entries for the experimental group centred on listening activities. A number of students commented on comprehension difficulties with the authentic materials, more often due to the speech rate or unfamiliar accents than unfamiliar vocabulary:

MY: [My Fair Lady] I couldn't understand almost all of what character talked about. Their speed of talking is very fast for me. I barely recognize the contents of the movie by watching and guessing. I wanted English captions.

YK: Today, we listened to native English speaker's conversation. I could hardly understand what they said, so I was shocked a little. I haven't ever listened to the British's conversations, so I think I couldn't hear that. But I want to learn to listen to it, so I think this is a good opportunity.

ES: I didn't understand the conversation of Weekend Away between Mark and Alison well. I couldn't catch the phrases or vocabularies when I listened to the conversation. Therefore I was surprised to look at the tapescript that was composed of well-known vocabularies.

Some students believed that their listening skills had not improved because the activities had been above their level of competence:

ES: Although listening skill is still in low, I believe my speaking improved gradually. I don't think I gained something helpful for listening skill from the class… I couldn't catch up with the conversation in Big Brother or Video Nation. I hope I can find more proper materials to my level.

YS: I think my listening skill has not changed very much since before, there were many times I could not understand what people say in the films.

This view was not, however, supported by the results from the quantitative analysis, as we saw in Section 5.2. Others commented that, through exposure to authentic texts, they believed they were beginning to adjust to the natural speech rate of English:

RM: In addition, when I watch movies and news in English and listen to them, I found my ears were getting used to the natural speed and I did understand.

A number of students commented on the value of the 'listening strategies' training (listening for stressed words, weak forms, catenation, etc.) in helping them to cope with authentic listening materials:

SN: Today's class was really useful for me, for I'm not so good at pronunciation, and sometimes it's really hard for me to understand what a native speaker is talking because of their speaking speed or squashed sounds. I'm not going to say that I will be a parfect listener after today's class, but at least, it gave me a kind of tip about how I can infer the whole story from stressed words.

TK: My listening skill has rather improved because I lerned the features of English. By focusing on the stressed word, the pitch, and something like that, I got to understand spoken English well.

Comments from the control group's diaries focused mainly on listening and speaking skills. Some students commented on the difficulty of listening activities in the textbooks:

SF: I listened two women's speaking from a tape, and answered the questions. I could hardly understand them. I depressed a lot.

Other listening tasks were considered easy, but this seemed to boost students' confidence:

MN: Sometimes listening quiz was easy. But if all of the quizes were difficult, I think it demoralized us and we felt that we didn't want to do it anymore. So, it was good to challenge various types of listening!

There was unanimous support amongst the control group for the regular pair/group-work speaking activities, mirroring the preferences voiced in the Personal Learning Histories (Section 4.2.1):

SF: This course was effective to improve my listening and speaking in that it gave me chances to use what I had learned practically. Japanese students don't have much opportunity to listen and speak English. Hence, although they study English for a long time, they can't talk with foreigners, so I think it was lucky for me that I could attend your class this year.

The learning environment

Many comments, from the diaries of both experimental and control groups, related to the social dynamics at work within the classroom. These highlight the critical role that relationships (both teacher-student and student-student) have on learning and are all the more important because they are often ignored by researchers in reports of classroom research and also by teachers (see Dörnyei & Csizér 1998).

The teacher's personality and teaching style were clearly important considerations for some learners:

NN: I was a little nervous before you came to the class, but when I saw you, I felt relaxed because you were smiling. I enjoyed your first lesson and I like the peaceful atmosphere of your teaching, so I'd like you to keep it through this year.

RM: It's really nice to see you, Alex (Could I call you Alex??). Today was the first class of this course, so I've been wondering what the teacher is like. I was happy to know you are very friendly and smile all the time. That makes me enjoy, without feeling nervous.

SF: Hi Alex! Nice to meet you! Before class, I wondered if you were grim teacher. But you looked cheerful!! So I was relieved.

Other comments illustrate the powerful effects that learners' relationships with each other can have on learning in the classroom:

NK: Today we continued to talk about job. My partner was H. She is a new friend! It is nice to have a conversation with someone I've never talk in classes.

MK: Today, especially I could enjoy this role-play because my partners are very nice. When partners are bad this kind of activities become torture itself.

In this respect, the learners' goals in the classroom are quite distinct from the teacher's and tend to be more socially than linguistically oriented. This is also noticeable in

transcripts of classroom interaction such as the extracts shown below, which were recorded during the pilot study and are based on a series of lessons trialing the materials to develop learners' oral narrative skills (see section H, appendix XII). Here, two male students, S1 and S2, work together to construct a dialogue based on a factual account of a narrative event. The teacher's aim was to prepare learners for a listening task from Quentin Tarantino's film 'Reservoir Dogs', by presenting key vocabulary from the scene and also familiarizing them with the story outline. However, this goal is sabotaged to some degree by the students who, in their desire to socialize with each other, re-cast the task to suit their own needs:

S1, S2: Male students
US: Unidentified student in class
(…): Transcription remarks
[…]: Concurrent speech
italics: Japanese
bold: English translation

1 **S1**: *ha he ho ho nani ni shiyoka* A piece of paper
ha he ho ho what shall we do?
2 **S2**: A piece of paper
3 **S1**: *sou nanka omoshiroi hanashi wo tsukurashite morauyo*
Yes I'll make an interesting story
4 **S2**: *so omoshiroku shiyo omoshiroku naruka* (laughs) *konna shinkokuna kore ga omoshiroitte iunara*
Yes let's do that make it interesting? I don't understand how we can make
5 *ore wa rikai dekihin* we don't understand what why this is funny
this serious story interesting
6 **S1**: What do we now here it's my it's my it's my car (S2 laughs) I'm sorry and I want to ride ride pato
7 car once finish (pato car = patrol car)
8 **S2**: (laughs) Oh really oh that's terrible
9 **S1**: *Patokaate nihongo patroruka ka?*
Is 'pat car' Japanese? Is it 'patrol car' in English?
10 **S2**: *Ah pata* Patrol car [**S1**: Patrol car]
Oh that's patrol car
11 **S1**: (Checking dictionary) Be available patrol car *rashii chuuka*
patrol car-ish or
12 Police car *no hou ga iissu*
Police car would be better
13 **S2**: Mm police car sq squad car
14 **S1**: *nani sore* squad car
What's squad car?
15 **S2**: *iya soo ya issho ni kangaete*
Can't we think about this together?

16 **S1**: *Pair yaro*
We're working as a pair aren't we?
17 **S2**: Pair *detan yaro kangaetekure*
If you're my partner please think of something
18 **S1**: B work in pairs
19 **US**: *Omoshourokusuru*
Are you making it interesting?
20 **S2**: *sou sou sou*
Yeah

In lines 1 to 5, the students begin by discussing how they can turn the list of events into an interesting story[13]. S1 suggests an alternative sequence of events for the narrative in line 6, with the driver apologizing to the policeman for his bad behaviour and explaining that it was only so that he could get himself arrested and have a ride in a patrol car. This solution, as well as being witty, has the added benefit of avoiding the need to create a long, complicated dialogue in English (the significance of the word 'finish' in line 7) so S1 feels doubly pleased with himself and amuses S2 (line 8). In lines 9 to 14, the discussion turns to the word *pato car* (patrol car). The students are aware that often English loan words become distorted in Japanese, frequently being shortened so that, for example, *convenience store* becomes *conbini*. Consequently they guess (correctly) that *pato* is not a correct form in English. S1, after checking his dictionary, suggests *police car* as a more appropriate choice and S2 puts forward the alternative option of *squad car*. In line 15, S2 complains about S1's lack of cooperation in the pair work activity. The reason for this comment is difficult to interpret from the transcript alone but is in actual fact brought on by S1's continuing attention to his electronic dictionary, rather than S2. In lines 16 and 17, the pair continue to bicker over their contributions to the activity. In line 19, another student asks the pair if they have managed to make their story interesting

[13] The scenario is based on a policeman stopping a driver in the United States. He approaches the car with his gun drawn and nearly shoots the man in the car because he keeps reaching towards his glove box, rather than keeping his hands on the dashboard.

and S2 replies enthusiastically that they have. They continue, some 150 lines later, in a similar vein but now the story becomes twisted as it transpires that the driver has actually stolen the policeman's patrol car:

1 S2: And I approach and approached I approachedu it with a gun drawn *hajime nante kakarouka*
how shall we start?
2 **S1:** *koko wa nan* (incomprehensible) *futari de kaiwa hajimarun*
This part shall we have two people talking?
3 S2: *so so so* (laugh) *sakki no yatsuka* (laugh) *sakki no yatsuka aa yu chan yo ai ya 2 ban no yatsu*
That's right that's what we were talking about before the no. 2 guy
4 *zenzen haitte kitenai hitori de zura tte*
isn't talking at all is he?
5 S1: *kikijouzu yan* (S1 & S2 laugh)
He's a good listener
6 S2: aha uhu *mitai na*
Ah it looks like it doesn't it?
7 S1: (to the microphone) *kikijouzu* is means mm a person is good at listening to the story from someone
8 S2: (laughs) I walked around to the
9 S1: Ah no no no no good listener good listener (S2 laughs)
10 S2: Side and *zutto kore kaiteru dake yakedo iin*
I've just been writing is that ok?
11 S1: *iin janai no*
No problem
12 S2: Pointed my gun the at the driver *mazu* I say I say don't move *kono ato ni nan tte itte tara iin kana*
what should he say after that I wonder?
13 Why you stole why why *ka*
14 S1: Why
15 S2: (laughs) Why you stole why you stole *yana* my car patrol car? *eh? chigau ??*
eh? eh? is that wrong?
16 S1: Eh?
17 S2: nan yattakke *nande nande tottan yatta ore*
What was that? Why did he steal the car?
18 S1: *ha*
19 S2: *nande* patrol car patoca er *tottan yattakke*
Why did he steal the car?
20 S1: pat car *noritakatta kara*
Because he wanted a ride in a patrol car
21 S2: *sairen narashitakattan dayo mitai na soko noke soko noke oira ga toru mitai na*
He wanted to use the police siren to show how important he is "Get out of the way
22 *sonna ki na* Get out get out I'll I'm coming (S1 & S2 laugh)
everybody! I'm a VIP!"

This banter between S1 and S2 is more interactional than transactional in nature and helps them to strengthen their relationship. Re-casting the task in this way of course

means that the students are able to enjoy themselves a great deal, as is obvious from the transcript, but it also means that many of the teacher's original learning aims are not achieved. By the end of the activity, S1 and S2 had not clarified much of the target language, nor had they developed the necessary schemas and scripts to allow more top-down processing and ease comprehension difficulties when watching the Reservoir Dog's scene.

Japanese students tend to find it quite difficult to get to know each other (although they would like to), and generally avoid sitting down next to unfamiliar classmates unless directed to do so. This, unfortunately, puts more responsibility on the teacher for facilitating healthy relationships within the classroom. There are, however, differing attitudes to teacher intervention in this aspect of classroom dynamics. Some students enjoyed the opportunity to meet new people, while others resented it:

SN: Special secret advice I can give you is… to make boys & girls pairs when doing activities! Boys are always motivated by girls, girls, girls!

MI: I suggest you to change the students' seats every time, and have students have discussions with all the class members. Thanks to these, they could be partners with every student and know more each other.

MK: I like your class, because I can talk with my classmates and get to know them more and more. Many people are still shy, but I want to make as many friends as I can in this class!!

YS: We changed a person who sit together because some didn't want to sit their partner. So, we'll have to sit other person every class. But I don't want to do so. I like Y, and I want to sit with her. If we change our partner every class, I think that new problems will happen.

Many comments from both groups also suggest high levels of tension in the class, and the frequency of words such as 'tense', 'nervous', 'afraid' and 'embarrassing' is noticeable. This is something that I believe teachers in Japan tend to be largely unaware of, because these feelings are generally hidden from public scrutiny. The diaries seemed to provide

students with an outlet for their worries, which were most often associated with one of three general themes; ‘making mistakes’, ‘feeling inferior’ or ‘fitting in’:

MT: We see the scenario and checked the words. There was a lot of words that I’m not sure that the meaning is correct. It was embarrassing that I ask the meaning of a word that is easy and I should know, but it is not useful for me if I didn’t ask.

TM: And we talked about giving opinions. I think saying my opinion is important, but I was very shy. So, it’s difficult for me to say my opinions.

YS: Today is first lesson of this class. There are no person who I know. There are no person who know me. I’m very nervous.

KN: Today, I was very nervous and worried about my speaking skill. While I could speak when talking with my friends, I couldn’t in the class. A friend beside me spoke so fluently, so I was very depressed, envied him and accused myself of confidenceness [lack of self-confidence].

KK: I said “my favourite shopping place is DAIEI, supermarket” [a cheap supermarket in Japan] to my group mate. They laughed, but I was serious. I’ll say “my favourite is Costco” next time [a more socially acceptable answer].

To some extent, these factors are outside the control of the teacher and depend on the personalities that make-up any particular class. However, with a better idea of the social dynamics at work within the class (for example by close monitoring of learners’ interactions with each other or feedback through diaries) teachers can take measures to alleviate these tensions. It can be pointed out to students that mistakes are a natural part of the process of language learning and that, for teachers, they are useful because they help them to tailor the input to meet learners’ needs. The teacher’s style of error correction can also encourage a positive attitude in the class, by keeping feedback on mistakes light-hearted and non-judgmental. Employing a wide range of interaction patterns in the classroom with pair or group work and mingling activities can also encourage students to get to know everyone and improve relationships.

Finally, the influence of external concerns can be seen in some diary entries:

MK: I have lots of homework in other classes, so I usually go to bed around past 2:00. I'm tired.

NK: But to be honest, today I couldn't really focus on the class because the news of the final interview of studying abroad was noticed at p.m. 1:00. I was worrying it. But now I'm happy. I passed!!

MK: Today, we learned how to use "kiss" and "hug" and when we should use them. It was good class, but I couldn't concentrate. Because, I left my very important homework at home. During the class, my heart was full of this problem. I finished this homework by the all-night work, so I was regretting very much. I'm so, so, sorry for today's my behavior.

YSh: Today, it was typhoon, so some students were absent from school. I thought that I would be so, but I went to school. Before I arrived to school, I was tired, I was worried about the outside conditions during the class.

The social factors discussed above are powerful influences on the success or failure of any particular treatment condition, yet they are often both invisible and outside the control of the researcher. They are also likely to vary substantially from class to class and country to country: in Mexican classrooms, for example, the problem is more often getting students to *stop* talking to each other, while in Japan it is getting them to *start* talking. I therefore believe that it is important for classroom-based studies to be as explicit as possible about the social conditions prevailing during the research. This will allow the reader to judge for themselves the impact of these moderating variables on the results and also the relevance of the work to their own teaching context.

Motivation in the classroom

As table 6.1 shows, learners from the experimental and control groups seem to have been more often motivated than de-motivated by the activities and materials they were exposed to during the 10-month trial. Both types of treatment can therefore be regarded as being

successful in this regard, although not always for the same reasons. This illustrates how the actual materials used in the class make up only a part of the learners' overall evaluation. Other factors can be equally influential, such as how the teacher mediates between materials and students (designing tasks, omitting or supplementing to meet learners' needs), the amount of control the teacher exercises over classroom events, the atmosphere the teacher seeks to foster in lessons, the relationships between learners in the class, their attitudes to learning and so on.

The experimental group appeared highly motivated by the 'real world' aspect of authentic materials and how the content could be usefully applied in their own lives:

RI: All of the materials you gave me was practical. Sometimes I check them to remember how I should say something in formal situation, etc… My listening and speaking skills could be improved in your lesson. Before participating in your class, my English was terrible! It was very Japanese English and quite unnatural. However, I really think that now I can speak English more naturally than before thanks to your lessons. You taught me how I should speak English for native English. It was very hard but quite useful for me. I like your teaching! Your lessons are unique, interesting and active!

YM: I learned a lot of things that I haven't know before! Obviously, speaking has improved. I think that because I learned many techniques which make my English sound more natural. For example, it was great to learn 'listener responds' [reactive tokens], 'sign-posting languages' [discourse markers] and so on. Also, it was good to study polite English and formal English because I will definitely need them in the future.

Many comments display great enthusiasm for the use of authentic input for its own sake, which is hardly surprising given its potential to engage learners on a deeper level than anything most ELT materials writers are likely to produce in language textbooks, where their focus is often on using texts as vehicles to illustrate form, rather than to inform or entertain. Movies and songs were particularly appreciated and there are glimpses of their powerful emotional effects in the diaries:

RM: I definitely enjoyed watching videos. Every videos that you brought to our class was very interesting such as Big Brother and Babe. In these videos, there are full of things we can learn.

SN: Still on "Fawlty Towers": today we could watch the video with subtitles, so that I could catch every joke! Last class I caught most of them, but sometimes they, especially Basil, spoke too quick, for me to understand. However, every time I missed them, audiences in TV laughed! How frustrating! There must be something funny!

RI: I really felt that your class is very fun. The lesson of "Tom's Diner" was interesting. I could learn where I should put stresses or how to take rhythms while I enjoyed myself. I had hummed the song of Tom's Diner after finishing the class.

SS: [Annie Hall] I quite enjoyed watching the video, I liked Annie's character. 'Hi! Bye!' and 'la-di-da la-di-da…' are interesting. After this lesson, I kept saying those words to my friends (classmates) and they laughed at me.

Aspects of the lessons which seemed to de-motivate tended to be associated with difficult materials or activities which exposed learners to class scrutiny:

YM: I didn't like the class which focused on the parts that I'm not good at because they were hard… I understand I can't improve my English without practice but it was boring.

TT: Today, we watched the video [students' video auditions for Big Brother]. I was really embarrassed during I was saying something stupid on the screen. I was not sure, how many times I wanted to die. How many times I regretted that I didn't hire an assassine to kill the video. It was the longest three minutes in whole in my life. I think I aged some years.

In the control group, students were also motivated by 'useful' or 'interesting' input:

MN: We played customer and clerk [role-play]. It is very useful because we can use these conversations when we visit foreign countries.

KK: Today's class was usefull for our future. Good theme. I will learn how to write resume in case I apply job in other country.

TU: Today's theme was "work" which I really have interests… Since the topic was interesting, I had fun when we discussed it.

SF: We learned about iguanas a lot. I found out that they are vegetarians, triple in length in a year, and can live long relatively. I didn't like their ugly appearance, but I felt like having them as a pet after I heard they get on people's heads when they are frightened of something.

They were most commonly de-motivated by dull or unfamiliar topics and materials considered too easy:

YK: We talked about a party and dancing but actually I don't like dance or discos. So I was reluctant to talk about it and its topic was boring.

MK: Sometimes it is easy to do for university students. Get rid of too-easy materials.

YT: If the materials are more a little difficult, students will be motivated.

Comments such as these illustrate the difficulty of matching textbooks produced for an international market with the needs of a specific group of learners, both in terms of level and interests. Few university students in Japan go to nightclubs so, while this might be a stimulating topic to many young people, it was widely disliked here. Finding an appropriate textbook to match the group's level was made more difficult for two reasons. Firstly, poor streaming procedures, which rely too heavily on measuring grammatical aspects of English, mean that learners are often placed inappropriately and this can lead to dissatisfaction in the class:

TU: Actually, I didn't like the time to do groupe discussion, because some member of the class were too quiet to attend the discussion. When I asked them an opinion, they never returned or returned in Japanese. So, I strongly want university to devide students classes by more reliable examination (e.g. oral interview). I know it's tough, or maybe impossible, but I want it.

CM: Today, I was partnered with S. Her pronunciation is beautiful and she speaks English fluently, so, I was shy of speaking.

Secondly, because the Japanese school system teaches predominantly lexico-grammatical aspects of English, in order to prepare students for poorly designed university entrance exams (which focus on the same, traditional elements), university learners are typically

unbalanced in terms of their overall communicative competence. They often have extensive *knowledge* of English, and can perform very well on multiple-choice tests (as the results of the grammar test in table 5.1 illustrate), but poor *language skills* so that in face-to-face communication, they are wholly inadequate. This means that, as in this case, a textbook that meets learners' listening and speaking skills needs, will often not meet their lexical or grammatical needs.

Many of the comments above relating to motivational issues are perhaps unremarkable for experienced teachers, who through trial and error in the classroom have become sensitized to what motivates and de-motivates their students. They do, however, fit very well with stimulus appraisal models of language learning (Scherer 1984; Schumann 1997), which see input as being appraised by learners along five criteria:

i) Novelty: input has novel or unexpected patterns;

ii) Pleasantness: pleasant events encourage engagement & unpleasant events encourage avoidance;

iii) Goal/need significance: relevance of input to students' goals or needs;

iv) Coping mechanisms: ability of students to cope with learning events or to avoid and change outcomes;

v) Self or social image: compatibility of learning events with social or cultural norms.

Positive appraisals on these criteria are thought to encourage greater cognitive effort and engagement with input, leading to more learning, while negative appraisals result in avoidance and, therefore, less learning. While both treatment groups in the investigation evaluated their classes positively for the most part, I believe that the authentic materials

were better able to satisfy the appraisal criteria of novelty, pleasantness and goal/need significance and that this resulted in higher overall levels of satisfaction, increased engagement with the input and, consequently, more learning taking place within the experimental group. It might have been predicted that the last two criteria (coping mechanisms and self or social image) would be more positively evaluated by the control group since the textbook input they received was often less difficult and more in line with students' expectations. However, there is no real evidence of this in their feedback.

Many of the remarks made by learners in their diaries also support constructivist theories of language learning which see learners as actively selecting out information that has *personal meaning* from the varied input they are exposed to (Williams & Burden 1997: 21). What they identify as valuable is often information that they believe will help them reach their future goals (sounding 'natural', making foreign friends, being polite, and so on).

Communicative competence tests

Many of the worries, mentioned above, also surface in learners' responses to the pre- and post-course tests:

RM: I took an interview test [IELTS oral interview] by Mr Flynn last Fryday. I was a little bit nervous, so I'm afraid if I made mistake…

NM: [ss-ss roleplay] We had another work today. It was a conversation with a friend. Because you recorded our conversation on video, so I felt stressed. There were some silences in our conversation. It was difficult for me. I don't want to record. Do you have a plan like today's activity in this semester?

CM: The most stressful thing that we had today was roleplay [student-student role-play]. I couldn't concentrate on the other (grammar, word) tests since my mind was full of anxiety.

The desire to succeed led some students to attempt to cheat in order to improve their scores:

ES: Role-playing test was also terrible today. I and my partner's conversation was not active. Actually, my partner had asked me what topics should we talk about yesterday. However, our plan made no sense.

RM: A girl in class 1, who doesn't take speaking test yet, said to me that she was able to listen the questions and she was thinking the answers a little, so when she take the test, she would prepare some answers. My turn was first, so I thought it's not fair!!

This kind of behaviour is often seen in Japanese classes, but predominantly with speaking tests, where students seem to want to limit some of the unpredictability inherent in casual conversation by pre-planning or memorising. Perhaps this can also be explained by the more rigid and formulaic nature of the Japanese language discussed above. When learners speak English, they find themselves deprived of clear 'one context-one speech act' pragmalinguistic guidelines and this has a disquieting effect. Like a prisoner who has, for years, walked the same route around the exercise yard and is suddenly released might be unsure where to walk. This kind of control of peoples' behaviour, linguistic and non-linguistic, permeates all aspects of Japanese society. One of my favourite parks in Tokyo was a beautiful walled garden with a lake in the middle and carefully tended grass lawns and woods around it. However, the experience of visiting was always somewhat spoilt for me because, rather than being free to wander wherever the fancy took me, my route was imposed upon me by ropes, which restricted me to the gravelled paths, and signs, which dictated a clockwise circumlocution of the lake. Westerners, who tend to place a higher value on concepts such as 'liberty' or 'freedom of choice' than 'order' or

'social harmony', naturally rebel against this kind of imposition and I often chose to walk anti-clockwise around the lake, purely as an act of rebellion.

If these personal impressions, from an outsider looking in on another culture, are anywhere near the truth, there is a possibility that learners would benefit from an explicit focus on these deep-rooted sociopragmatic differences in the classroom. Perhaps in showing them that speaking English can be like wandering in a park, where they are free to sit on the grass, swim the lake or pick the daisies, we can free them from some of their anxieties. They will, of course, need new 'orienteering skills' to navigate this unfamiliar terrain, and this is why I believe training students to use communicative strategies is so important. These give learners the tools to cope with the less predictable nature of English discourse and the greater likelihood of communication breakdown.

6.1.4 Summary

The thoughts and feelings expressed by learners in their diaries are a fascinating insight into a very different perspective on classroom events from the one the researcher provides. They help to account for the results seen in the quantitative aspects of the study and ground them solidly in a research context, allowing others to better judge the relevance of the results to their own teaching contexts. The diary entries suggest the existence of a wide range of moderating variables, which are often invisible in classroom-based research, particularly when it only takes the researcher's etic perspective into account. Since these variables are largely outside the control of the researcher, and specific to the social context of the investigation, I would suggest that this kind of

qualitative evidence can play an invaluable role in determining the validity and reliability of quantitative results.

6.2 Case Studies

The case studies were designed to explore six learners' reactions to the experimental and control treatments in more detail, using their Personal Learning Histories or diaries as 'jumping off points' in face-to-face interviews. Three learners were selected from each group, representing high, medium or low proficiency students, based on their scores in the pre-course communicative competence tests, as shown in table 6.2[14].

Student	Treatment group	Proficiency level	Mean score on pre-course tests
MY	Experimental	High	82.1%
MK	Experimental	Medium	59.7%
YK	Experimental	Low	55.1%
MH	Control	High	75.3%
MU	Control	Medium	65.8%
HY	Control	Low	57.0%

Table 6.2 Summary of case-study students

Appendix X shows transcripts of the four sets of interviews conducted with the case study students in June, July and November 2004, and January 2005. A read-through of these indicates that, unsurprisingly, learners' comments here generally support the trends outlined in the diary reports (section 6.1). To avoid unnecessary repetition, only

[14] Two of the students (MK from the experimental group and MH from the control group) dropped out of the course at the end of the first term so data from the full 10-month investigation is only available for four of the participants.

comments that are seen as providing new insights into the learners or learning processes are discussed here.

6.2.1 Proficiency levels

The case study interviews give a sense of the wide range of proficiency levels within the control and experimental treatment groups, despite similar TOEFL scores (used as the basis for streaming students at Kansai Gaidai University). This suggests that the TOEFL test is providing a rather crude measure of learners' true communicative competence. This is probably because it places excessive emphasis on easily measured features of the language (such as lexico-grammatical knowledge) at the expense of less accessible features (such as listening and speaking skills), as one of the learners commented in her diary:

NN: I always think that the score of TOEFL isn't a good criterion of judging one's real English level, especially in speaking ability, so the test you gave… was more useful than TOEFL, I thought.

The difficulties of catering for different needs within the classroom are therefore compounded, unnecessarily, by ineffective streaming procedures.

Two of the case-study students, MY and YK from the experimental group, clearly illustrate the wide variations in proficiency (see appendix XI). MY is a 'returnee' who spent over five years living in Canada as a child and has native-like fluency and pronunciation, although, as she said, she lacked vocabulary:

24MY: And vocabulary, I don't think I have much vocabulary
25 because like in Canada I was very small so. When I take tests, you know in the
26 last parts there's like long stories and most of the hard vocabularies I can't really
27 understand so I think I have to work on that.

YK, on the other hand, is unable to produce any long turns in the interview at all, often speaking in a whisper, almost undetectable by the microphone and interspersed with long pauses, as if attempting to become invisible (a strategy often employed by less confident students in the classroom). By her own account, she struggled with lessons to the extent that she often could not even understand her teachers' instructions:

21 I: So what do you think of your classes this year?
22 YK: I don't enjoy them. I can't listen to the teacher so I don't know what to do.

Predictably, she had more difficulty coping with the authentic materials than MY and estimated that her initial comprehension of DVD scenes was around 20%, although she pointed out that the visual contextualisation of the films helped:

41 YK: Yeah I don't understand at all but the picture tell me.

YK's problems with the material seem to be more with the speech rates or accents since, when she read the transcripts, she realised that much of the vocabulary and grammar was already familiar to her:

44 I: So if you listen and read together is it easy to understand?
45 YK: Yes I can't listen to the native speaker. When I see the print I was surprised
46 because I didn't know they say.
47 I: Ah but when you read it you can understand?
48 YK: Yeah.

This suggests, again, that it is lack of exposure to natural English models that is the source of many learners' comprehension difficulties, rather than knowledge of the language system itself.

Because of YK's difficulties with the level of the class, we might have predicted that she would have expressed a desire for simpler, more controlled input but, surprisingly, she still stated a clear preference for authentic materials over textbook materials:

49 I: So this term we haven't used a course book. If you compare the things we have
50 studied this term with using a course book, which one do you think is better?
51 YK: I think text isn't needed because I want to improve my speaking and listening
52 skill. I think if I will use book it's hard to speak more smoothly with native speakers.

This stated preference for authenticity was common throughout learners' diaries and case study interviews. So what, exactly, was the appeal of authentic materials for learners?

6.2.2 The appeal of authentic materials

The attraction of authentic materials in the classroom seems to stem from two principle sources. Firstly, because they are seen as 'real', they are also considered more useful for learners' future communicative needs:

37MY: I think it's like more interesting and more new your stuff. I think
38 it's useful when you go abroad, learning slang so I thought it was good.

95 I: So it's quite difficult to understand [scenes from movies] so do you think we shouldn't use it?
96 MY: Ah but I prefer to watch because you are the teacher so usually you talk at plain
97 English or slowly but maybe you speak more faster when you talk to native
98 English speaker so we can know how the normal speed.

16 I: So you agree that the pattern of conversation in Japanese is different do you?
17 YK: Yeah.
18 I: So is that useful information?

19 YK: Yeah.
20 I: Why?
21 YK: To talk with English it is important.
22 I: So with listener responses you said 'I'm glad to know there are a lot of listener
23 responses' [response tokens]. Why did you write that?
24 YK: To use the listener response I can speak good like English.

These kinds of responses mirror those of Timmis (2002), who found, in his survey of learners' attitudes to 'native' vs. 'standard' English in 14 different countries, a continued desire amongst learners to conform to native-speaker norms.

A second reason for the preference of authentic materials is that the content itself is seen as more varied and interesting:

25 I: So if you compare what we've been doing with a normal course book, how do
26 they compare? How would you prefer to be studying English?
27 MY: I think it's more interesting like the things we're doing right now because we
28 don't just do listening or reading, it's not just one category of English, it's like all
29 kinds so I thought it was useful.

It would seem then, that although abandoning the 'discrete item' input of more traditional syllabuses has certain disadvantages (such as reducing the perceived order of, and control over, course content and complicating the process of evaluation), the use of authentic materials has many benefits for more advanced learners.

6.2.3 Challenge and support in the classroom

Mariani (1997) sees the level of challenge and support provided by materials and tasks in the classroom as critical elements for successful language acquisition. If materials are too easy, learners cannot develop their interlanguage effectively and become bored, while if they are too difficult or not adequately 'scaffolded', they are inaccessible and learners

become frustrated (see section 2.5.1). Many comments in the case study interviews suggest that learners were also very sensitive to this balance between challenge and support in the classroom:

138 I: Because as a teacher I've got 2 choices, either I could give you simpler listenings
139 like maybe from a book like this so you could understand everything or I could
140 give you listenings which aren't adapted for learners so maybe you can only
141 understand 60%. If you compare those 2 what would you prefer to get?
142MY: I'd prefer the 50% understanding one.
143 I: Why's that? Maybe you'd feel better if you understood everything.
144MY: Yeah but that means we already know all the stuff so there's not much to learn. If
145 you only know 50% and then we analyse and we understand the other 50% we
146 learn more.

50 MK: No but last year my class almost like that. The reading class is the teacher brings
51 the books and he translate the English to Japanese and we just wrote we just
52 memorised.
53 I: So you didn't like that?
54 MK: It's easy, I couldn't think my English improved by that reading class.

40 I: But for example in the class would you prefer easy listenings or listenings like
41 this if you could choose?
42 YK: This.
43 I: Really? Why is that?
44 YK: Because it isn't study if I can understand everything.
45 I: Oh so this is more of a challenge for you?
46 YK: Yes.

84 MU: Sometimes in the book [Inside English] the same thing is in the book what we
85 did last year so we repeating.
86 I: For example?
87 MU: Like present perfect.
88 I: So some of the grammar points?
89 MU: Yes yes.

There is, perhaps, nothing surprising in these comments, since we all expect a reasonable return on our investments of time, money and energy for the activities that fill our days. However, classroom activities are rarely assessed from this perspective, on their potential to challenge or support learners. When they are, we may reach the conclusion that

authentic materials, used intelligently with appropriate tasks, are often better able to meet learners' needs than internationally marketed textbooks.

6.2.4 Effects of learners on each other

Many of the comments from the qualitative data highlight how the attitudes to learning, which prevail amongst a particular group of learners, can have a powerful effect on the individuals within that group. One example of this, seen in the case-study interviews, relates to learners' willingness to use English in pair or group-work activities. With monolingual classes, such as those in this investigation, it is often difficult to persuade students to use the L2 with each other, and if the majority choose not to, the minority tend to acquiesce through a strong desire to fit in:

110 HY: Yes when I was 1st year student I was in level 6 class but many student speak
111 Japanese when native teacher told us to speak English. Now many student try to
112 speak English so this is good effect for me. So if I keep trying to speak English
113 my English will be better than now.

3 MU: Yeah my speaking has got better I believe because compared to class when I was
4 1st year people have a lot of motivation in this class so they spoke a lot of English
5 so I tried to speak.

55 MU: Some student try to speak Japanese when you don't watch them so you should
56 make rule like if you speak Japanese you pay.

Another example relates to learners' willingness to collaborate with each other in activities. ELT classes tend to favour pair and group-work, in an effort to promote negotiation of meaning and to develop listening and speaking skills, but learners are not always cooperative participants in this process:

45HY: some students make sentences by themselves but others try to
46 work together and most of partner didn't know do the work by myself or with
47 partner.

These influences on the classroom are often difficult for teachers to notice and are largely outside of their control. They help to explain how the same materials, taught by the same teacher can have such dramatically different results with different classes.

6.2.5 Lifestyle issues

Students often see higher education as a time to relax, make friends or earn money, after the pressures of high school and the 'examination hell' of taking university entrance exams. Since the risk of failing to graduate is quite low, many opt to devote more of their time to hobbies or part-time work than their studies:

1 I: So I just wanted to talk to you about things you put in your diary and about the
2 course and how you feel about it but I notice you've missed a lot of classes
3 recently. What's been happening?
4 MY: I've been working part-time job.
5 I: What a lot?
6 MY: I guess a lot it's kind of hard to keep up with the balance
7 I: What's your part-time job?
8 MY: I work at a karaoke bar.
9 I: So it's quite a lot of late nights is it?
10 MY: Mm.
11 I: So you have trouble getting up in the mornings?
12 MY: Yeah.
13 I: Are you missing a lot of other classes or just mine?
14 MY: I guess I'm missing some others.
15 I: How many hours a week are you working?
16 MY: About 30 or 40.
17 I: 30 or 40 hours a week? That's a lot.
18 MY: Maybe too much.

This is another issue largely outside the control of individual teachers, and the increasing difficulties universities are experiencing attracting students in Japan means they are reluctant to make changes which would decrease their popularity – as harsher examinations no doubt would. These factors, combined with the long commuting distances students often endure, mean that teachers are likely to continue to face the problems of tiredness and absenteeism in the future.

6.2.6 Summary

In general, the case studies raised similar issues to those identified in learners' diaries and supported the results reached in section 6.1. Although the interviews did, at times, provide deeper insights into learners' thoughts and feelings, they were less illuminating than had been hoped. This may have been due their rather hurried nature, squeezed into short lesson breaks or at the end of long days studying. It could also simply be the case that learners might have difficulty explaining (particularly in the L2) why they feel the way they do. Overall, participants seemed more comfortable, as well as more fluent, expressing themselves through the written medium.

CHAPTER 7 EVIDENCE OF INTERLANGUAGE DEVELOPMENT IN THE CLASSROOM: AN EXAMPLE

The results of the inferential statistical analysis carried out in chapter 5 indicated that the experimental group out-performed the control group on five out of the eight communicative competence measures: the listening test, the receptive pronunciation test, the vocabulary test, the IELTS oral interview and the student-student role-play. These gains are likely to reflect long-term, rather than short-term acquisition since, in most cases, the input stimulating their language development was received long before the final, post-course tests were carried out.

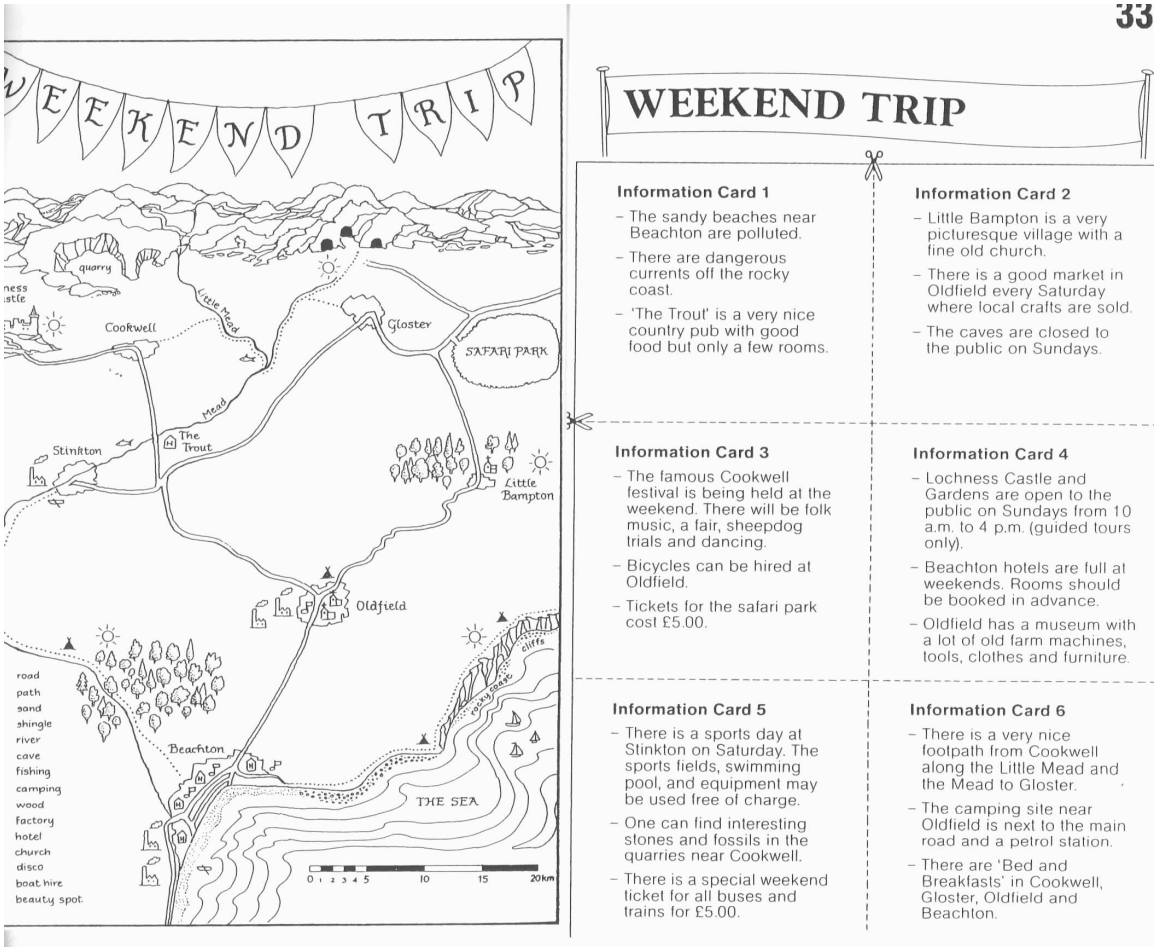
These quantitative tests are superior to qualitative investigations in that we can say, with reasonable certainty, that the differences between the two groups are real and can be generalised to the wider population (in this case, Japanese university students), which makes a strong case for the use of authentic materials in the classroom. Unfortunately, however, by condensing 10 months of classroom events down to a single set of statistical results, it is impossible to see *how* these changes in communicative competence occurred. This is where a more qualitative approach can help. Presumably, if we examine any discourse from the classroom in more detail, we will be able to identify examples of just how developments in learners' interlanguage came about. Of course, these examples will only reflect speech events for a few individuals in a few, isolated, instances so it is difficult to say with any certainty how representative they are of the whole group over the 10-month period of investigation. However, they might provide us with some insights, which we can tentatively apply more broadly.

7.1 The sample lessons

The sequence of lessons described in this chapter occurred towards the end of the 10-month investigation and focused on teaching learners appropriate ways to negotiate plans by giving opinions, agreeing and disagreeing, using the context of organising a weekend trip away by the sea.

7.1.1 The pre-listening stages

Students were given the following map of an imaginary coastal area, along with an information card containing further details, taken from the resource book 'Keep Talking' (Klippel 1984: 45-46):



After pre-teaching some of the lower frequency vocabulary (quarry, fossils, strong currents, etc.) students exchanged information from their cards in groups of six in order to familiarise themselves with the coastal area and the attractions available. They were then put into pairs and given the following instructions, before carrying out the role play:

"You have just arrived at the railway station in Beachton. It's 6pm. You're leaving Beachton again on Monday morning at 9am. Plan your weekend away together, deciding where you will go, where you will stay and what you will do. Write your schedule down".

At this stage, none of the target language for negotiating plans was presented since the aim was to see how well students could cope with the task at that point in time. One pair from the experimental group (Class 1), RI & MY, were tape-recorded during this first attempt at the role-play and their conversation transcribed.

7.1.2 The while-listening stages

In the next class, students had the opportunity to watch a video of a NS English couple, Mark and Alison, performing exactly the same role-play, organizing their week-end away (see appendix XII, section K). Because they were already very familiar with the context, the demands of the listening task were reduced. However, students still found it very difficult to complete the initial task, noting down where Mark and Alison decided to go or stay, as ES commented in her diary:

ES: I didn't understand the conversation of Weekend Away between Mark and Alison well. I couldn't catch the phrases or vocabularies when I listened to the conversation. Therefore I was surprised to look at the tapescript that was composed of well-known vocabularies.

There are a number of possible explanations for this difficulty. Firstly, unlike many conversations contrived for pedagogic purposes, the NS discourse is untidy: the two interlocutors make suggestions then drop them, make plans then change them, without, necessarily any chronological ordering. Next, the accents of the two speakers (British Midlands) were probably unfamiliar to most students and the speech rate was also quite rapid (approximately 137 wpm)[15], increasing the strain of comprehension. Lastly, some expressions used by the NSs, such as *that caught my eye* or *fairly grim* were probably unfamiliar to most learners. As discussed in chapter 2, these are all recognised challenges with authentic materials but, with the appropriate level of support, learners are able to cope with these difficulties.

7.1.3 The post-listening stages

The post-listening stages focused principally on a close analysis of the NS transcript and were designed to encourage students to 'notice the gap' between their own discourse and the NS's ~ something which is believed to encourage language acquisition (Schmidt & Frota 1986). In particular, the way in which the NS's language changed in the negotiation and confirmation stages of their planning was focused on (once the students had a thorough understanding of the text):

Negotiation stage	**Confirmation stage**
We could…	(So) we'll…
We'd have to…	
I would quite like to…	
It would be good to…	
What would you definitely want to do?	
Perhaps we should…	
Do you think we should…	
Shall we…?	

[15] This speech rate is only a rough guide since it does not take into account pauses in the interaction.

Learners were quickly able to identify the key structures from the NS transcript and also to recognise that their own ways of planning the same trip had been markedly different:

NM: we learned what kinds of verb we should use in planning a trip. When we did the activity, we didn't use "would", "could" and "should". We also didn't use the sentence "there is". This sentence is very common so I think it is very useful.

From careful monitoring during the role-play activity, it was clear that most students avoided using the model auxilaries (particularly *would*, *could* & *should*) whilst negotiating their trip. Instead, they tended to prefer to use present simple or *will* constructions, such as *We go to Cookwell* or *We will go to Cookwell*. Other common errors included the avoidance of *There is/are…* constructions (as NM, above, noted in her diary) and the over-use of the expression, *I want to…*

Of course, because the students were only communicating with other Japanese (and not NSs) and the transactional intent of their turns was clear, they were able to proceed with their planning smoothly, oblivious to the fact that, pragmatically, their language was often inappropriate. Using '*We go'* or '*We will go'* in the planning stages of a negotiation suggests an impertinent level of assuredness towards the way future events will unfold, which could well cause offence if the students were communicating with NSs (particularly people unfamiliar with other cultures), eliciting responses such as '*Oh we will, will we?*'. This is a typical characteristic of EFL (as opposed to ESL) learners, who do not have to cope with the 'friction of daily interactions' or the pressure of 'not only making themselves understood but also of establishing and maintaining smooth relationships with NSs' (Harlig & Dörnyei 1998: 253).

As well as a lack of pragmatic competence, students also often experienced difficulties at the discoursal level. Rather than jointly constructing their weekend trip, some students seemed to adopt a more Japanese style of interaction, mutually agreeing on a 'leader', who would then make all of the decisions and perform most of the topical actions while the other member of the pair would take on a more passive, supporting role. Again, this was something that the more aware students were able to identify for themselves, after seeing the NS model conversation:

RM: Mark & Alison are very nice and funny husband and wife!! I found that their way to decide the plan was different from the way I did with my pertner. Both Mark and Alison gave their opinions clearly. In other words, they eaqually insist where they want to go and where they don't. Ours was like the one person mainly decide and the other just agrees.

RI: [Alison & Mark's roleplay] I learned a lot of expressions and words I haven't used ever. For instance, I haven't said "definitely" though I know its word when I agree… I was also surprised that English people often say something together. In Japan, usually one doesn't say anything and listen while the other speaks. Therefore, in my case, it's very hard to interrupt other people.

This unequal sharing of the 'conversational leg work', as I have said, is quite a common characteristic of Japanese discourse (Sakamoto & Naotsuka 1982; Gilmore 1998) but is likely to be perceived negatively in an English speech community, with students being considered 'overly passive', rather than 'co-operative'. As I have argued throughout this work, the lack of pragmatic, strategic and discourse competence displayed by Japanese students can have serious effects on their ability to succeed in the L2 speech community and are frequently under-emphasised in language teaching.

The detailed analysis of the NS transcript allowed for valuable classroom discussion on a wide number of issues, including:

(i) The subtle differences between *can/could* and *will/would* lexical choices;

(ii) The reason '*I want to*…' might be considered impolite;

(iii) Identification of useful expressions, which could be 'poached' from the model dialogue (*There is/are…* ; *I fancy + verb –ing*; *I definitely want to…*; etc.);

(iv) The reason *will* (rather than the contracted form, *'ll*) can be impolite;

(v) The reason *will* is used in the confirmation, rather than negotiation stages;

(vi) The differences in the way Japanese and English conversation is constructed and the cultural expectations underpinning these differences.

The final stage in the sequence of lessons on negotiating plans was for students to perform the 'Weekend Away' role-play again with a different partner and try to improve the naturalness of their discussion, using what they had learnt from the NS model transcript. One of the students recorded in the 1st role-play attempt (RI) was again recorded with her new partner and this was also transcribed for closer analysis.

7.2 Evidence of interlanguage development

The aim of the sequence of activities described in section 7.1 was to raise students' awareness of target language items and L2 pragmatic/discoursal conventions and, therefore, stimulate noticing and language acquisition. The classroom events certainly encouraged the desired noticing, as extracts from the learner diaries show:

SN: When we made a plan, I faced a trouble. That is how to make proper use of "will", "be going to ~" and past tense such as "would" "could" and "should". I actually did not know the difference between them, that is, when to use which one. However somehow I finally could find out what "will" and "be going to ~" really means. I wonder why no Japanese teacher told us those important differences.

YM: Next, we looked at 'Weekend Away'. I was kind of surprised at various ways of making suggestions, giving + asking opinions and agreeing + disagreeing. It was great to learn the differences in meanings when we use 'would, should, could' and 'we'll, we're going to…' It was a bit difficult to understand them, but it made sense at the end, so now I can use them correctly. I think I will practice using them and expand the ways of talking in English.

The question then, was whether this noticing would result in any noticeable changes in the learners' L2 output. Often, when fluency (or 'freer practice') activities follow on immediately after presentation of new language, learners are unable to use the target forms and instead find other ways to complete the tasks (one of the common criticisms of the PPP approach). This suggests that repeated exposure, over a longer period of time, is necessary before new forms can be incorporated into learners' interlanguage systems and become available for productive use. However, in the sequence of lessons under investigation here, most of the grammatical patterns were already familiar to students. Rather than it being a *linguistic competence* issue, with the forms themselves causing them difficulties, it was a *pragmatic issue*, with problems matching the appropriate forms to contexts of use (the inevitable result of an over-reliance on memorisation of de-contextualised grammar forms in the Japanese education system). It might be possible in this case, therefore, to detect some immediate changes in the naturalness of students' output. In order to test this theory, both the pre- & post-treatment role-plays were transcribed for analysis and these are shown below:

'Weekend Away' Role-play Transcript (Pre-treatment)

(…): Transcription remarks
[…]: Concurrent speech
…: Pause (unspecified length)
italics: Japanese

RI & MY

1 RI:	When we we arrive at 6pm Beachton and maybe we we want to have
2	dinner
3 MY:	Mm yeah so
4 RI:	So mm
5 MY:	Ah so we might go to [**RI**: Mm] either a restaurant
6 RI:	Restaurant ah this is nice pub good food they they serve good food but it's
7	far from
8 MY:	It's it's I think it's really far [**RI**: Oh yeah] because this is 20 km (looking

at the scale on the map) which means it's like very very far (laughs)
RI: So really maybe around this area Beachton I have dinner we we have dinner
MY: At a restaurant near here and after we we might go to the disco
RI: Disco (both laugh) really?
MY: 'Cos there's lots of disco (both laugh)
RI: You you want go to disco? [**MY**: Mm] yes let's let's go let's go after dinner
MY: Mm
RI: (Writing down) Ah Friday night… pm dinner in Beachton and after… until er all night?
MY: No now
RI: Disco in Beachton and where where should I we [**MY**: Stay] stay?
MY: Er (incomp.) this place is full of people so we have to
RI: Oh so so you have booked booked?
MY: Yeah we've booked already
RI: Thank you
MY: So we stay at this Beachton hotel
RI: And so Saturday Sunday morning where
MY: Saturday morning
RI: Mm?
MY: Saturday morning
RI: Yes
MY: We
RI: I want mm I want to go to Cra Cro
MY: Cookwell
RI: Ah I want to… er where which places do you want?
MY: Mm maybe we might go… Do you have a car?
RI: No ah er there are the special special weekend ticket for all buses and trains for five five dollars [**MY**: Five dollars five] pounds? Maybe very cheap so we can use
MY: The bus?
RI: Bus or train
MY: And go to… [**RI**: To] Oldfield and we OK we use the bus and go to Oldfield O-Oldfield… and then we borrow a bicycle [**RI**: Oh] (both laugh)
RI: OK so go to Oldfield… to train? By train?
MY: By bus
RI: Bus bus and hire
MY: Hire bicycle
RI: Bicycle and by bicycle where?
MY: We go to Little Bampton
RI: Ah yes I want to go… very nice scenery there (writing down) go to Little Bampton by bi by bicycle
MY: And see see the village and church
RI: Yeah
MY: And then we we go on a bus and go to Cookwell ah *touii kana* (both laugh)
RI: Cookwell? Ah
MY: Maybe we go to Cookwell on Sunday
RI: Yes
MY: And
RI: But I want to also on Sunday I I w I want to visit cave
MY: Cave?
RI: Cave
MY: But it's closed on Sunday
RI: Oh closed?

65 MY: So maybe we'll go to the cave today on Saturday
66 RI: But there are no hotels around there
67 MY: But there's B&B in Gloster [**RI**: Oh so you] so maybe maybe we use a bus
68 to the… so do you want to go to the safari park?
69 RI: Mm not so much
70 MY: OK then we will use a bus and go to the cave [**RI**: Cave] use the bus
71 RI: Go to the cave
72 MY: Go to the cave
73 RI: By bus and after
74 MY: After that
75 RI: Visiting the er we go to Gloster?
76 MY: Gloster and stay at B&B… and…
77 RI: Sunday? Er
78 MY: Sunday we… maybe we can… maybe we can walk to Cockwell Cookwell
79 [**RI**: Oh walk?] because there's a path
80 RI: Oh yes (both laugh) it's a little far but
81 MY: A a little
82 RI: It's OK (**MY** laughs) it's very healthy [**MY**: Yes] (**RI** laughs) yes got go **83** go to
84 MY: Go to Cookwell and [**RI**: Cookwell] and the festival
85 RI: By walk
86 MY: Cookwell festival
87 RI: Ah yes it may be fantastic
88 MY: And…
89 RI: (writing down) Festival in Cookwell
90 MY: And then er
91 RI: Maybe
92 MY: Go to castle? [**RI**: Mm] And see the… go to the beauty spot in the castle
93 RI: Ah but to 40 4 pm
94 MY: Eh he until before [**RI**: Ah] 4pm
95 RI: And there is there time to visit?
96 MY: So
97 RI: Maybe OK maybe we have time… castle
98 MY: Mm and then… and stay at the Trout
99 RI: Ah OK Trout
100 MY: Then yeah
101 RI: The Trout and we will have nice dinner [**MY**: Yeah] and fun
102 MY: Yeah and then the next morning we go back to Beachton end in Beachton
103 RI: Oh yes by bus? Train
104 MY: Maybe go by [**RI**: Bus] bus and then [**RI**: Leave?] leave Beachton
105 RI: OK… write down? (both laugh) Almost all of places we will visit
106 MY: Mm?
107 RI: We will visit almost all place
108 MY: Yes
109 Teacher: Busy weekend?
110 RI: Yes
111 MY: Yes very busy

'Weekend Away' Role-play Transcript (Post-treatment)

RI & MW

1 RI: So
2 MW: So on Friday 6pm Beachton?
3 RI: Mm this
4 MW: There?
5 RI: Yeah
6 MW: And Monday 9 9am Beachton?
7 RI: Where are we going to stay?
8 MW: Ah on
9 RI: Friday night?
10 MW: On Friday night… so Friday night
11 RI: Maybe in Beachton [**MW**: Yes] there there are a lot of B&B [**MW**: Ah]
12 there but many people stay there
13 MW: So we should stay Beachton on Friday night
14 RI: Mm if we stay Beachton we we should book in advance book Beachton…
15 we so we stay in Beachton on Friday night
16 MW: Ah what… what could we do [**RI**: Mm] on Beachton? Only staying or?
17 RI: Oh ah there's there's a disco [**MW**: Ah disco] disco many disco so we we
18 enjoy disco
19 MW: Ah yes… fine it's fine so [**RI**: Yes] on Friday night let's stay at Beachton
20 and Friday night how about going disco?
21 RI: Yes that's good
22 MW: Yes you are very good at dancing so [**RI**: (laughs) No] it's nice
23 RI: So so on Saturday morning… Saturday morning how about Saturday
24 morning what what are you going to? [**MW**: Ah] What do you wanna go?
25 MW: Ah this is a beauty spot this this
26 RI: Ah not so much
27 MW: Not so much?
28 RI: Better than
29 MW: Better than Little Bampton
30 RI: Little Little Bampton very picture picture village picture I can't explain
31 but
32 MW: Ah has a church?
33 RI: Yeah beautiful church and nature beautiful nature
34 MW: But actually I'm not interested a church (RI laughs) interested in church
35 RI: Yes how about safari park?
36 MW: Safari park ah safari park is there a… safari park?
37 RI: Just only 5 pounds [**MW**: 5 pounds?] 5 dollars?
38 MW: Ah it's very cheap
39 RI: Cheap cheap yeah we could go there
40 MW: Ah do you want to go safari park?
41 RI: Actually not so not (both laugh) because a little cheap so maybe I think
42 it's not so interesting
43 MW: How how about visit Loch Ness Castle?
44 RI: Yes yes Sunday but open [**MW**: Ah but] Sunday just Sunday [**MW**: Just
45 Sunday?] yes so Sunday [**MW**: Sunday] on Sunday we could go visit
46 [**MW**: Loch Ness Castle] Loch Castle so ah
47 MW: So… so
48 RI: And Cook in Cookwell there's a festival
49 MW: Festival really?
50 RI: Shopping we can we can go shopping interesting maybe
51 MW: Hopefully I like to go shopping

52 RI: Mm yes maybe you you like shopping
53 MW: Yes so I suppose you wanna go Cookwell?
54 RI: Yes so on Saturday [**MW**: Mm] we we could we'll go there [**MW**: Yeah]
55 Cookwell and enjoy festival
56 MW: Oh yes enjoy festival Saturday go to Cookwell
57 RI: And on Saturday night where we where are we going to stay?
58 MW: Stay?
59 RI: Maybe near
60 MW: Cookwell is Trout
61 RI: Trout or yeah Trout the Trout is very nice food er the Trout has very good
62 nice food
63 MW: OK I see so Trout on Saturday night at the [**RI**: Trout] yes let's stay Trout
64 RI: Yes but only a few rooms so
65 MW: Oh yeah I see you mean [**RI**: if I yes] we should book early?
66 RI: Yes yes before before staying
67 MW: Mm OK good idea
68 RI: And Sunday
69 MW: Sunday morning let's go to Loch Ness Castle
70 RI: Yes
71 MW: Good idea
72 RI: And anything else? How how about erm quarry
73 MW: Quarry?
74 RI: Quarry quarry there is stones and fossil
75 MW: Ah OK fossil? Ah
76 RI: Or erm also we could visit the cave
77 MW: Cave? Ah
78 RI: Ah but Sunday will be closed
79 MW: Ah closed ah sadly we couldn't visit on Sunday
80 RI: Ah… ah in Oldfield there's there's a lot of old farm
81 MW: Old farm? Ah
82 RI: Clothes and furniture [**MW**: Ah] we could buy
83 MW: Yes sounds nice
84 RI: Mm but but this place also a petrol
85 MW: Mm? Petrol station
86 RI: Station?
87 MW: Camping… ah could could we camping on Oldfield?
88 RI: Mm camping?
89 MW: Camping let's
90 RI: Ah yes
91 MW: Er do you like camping?
92 RI: Yes yes
93 MW: Ah how about on
94 RI: Sunday night?
95 MW: Yes camping?
96 RI: Camping (laughs)
97 MW: On Sunday night
98 RI: OK but but we have we have to leave [**MW**: Yes] on Monday
99 MW: Night er Monday
100 RI: Early morning
101 MW: Early morning ah
102 RI: So ah
103 MW: Ah no no doubt I maybe I couldn't get up early (both laugh)
104 RI: Yes so [**MW**: so] we we had better stay Beachton
105 MW: Again
106 RI: Yeah on Saturday night
107 MW: On Saturday yes

108 RI: If you you wanna go camping [**MW**: Camping?] ah how about Friday
109 night or
110 MW: Ah
111 RI: Ah no no no Saturday night?
112 MW: Saturday night?
113 RI: Saturday night
114 MW: But Saturday night is Trout
115 RI: Cookwell?
116 MW: Cookwell
117 RI: So Cookwell and on Saturday night go back over here er very com
118 complex (laughs)
119 MW: I'd rather stay the Trout [**RI**: Mm] on Saturday night
120 RI: We we couldn't
121 MW: Mm *ee yo*
122 RI: On on Sunday how about anything else?
123 MW: Ah on Sunday
124 RI: What do you wanna?
125 MW: Stinkton is only
126 RI: Mm
127 MW: Just castle
128 RI: Oh
129 MW: Ah footbath
130 RI: Footbath what what is footbath?
131 MW: Yeah
132 RI: Footbath
133 MW: Name of
134 RI: Like hot spring no? [**MW**: Ah] I guess (asking teacher) What what is a
135 footbath?
136 MW: Footbath
137 Teacher: Footpath? [**MW**: footpath] For walking [**RI**: Ah walking]
138 MW: Walking
139 Teacher: Yeah not for not for cars
140 MW: Ah yes
141 RI: So path walking do you like walking?
142 MW: Ah actually I don't like walking but recently I I seldom playing sports or
143 doing doing… walking
144 RI: So so er er so you think you you have to [**MW**: Yes] do sport?
145 MW: Ah yes… mm
146 RI: So mm after visiting the castle we can we could walking [**MW**: Walking]
147 we could go we could walk ah but very far
148 MW: Ah mm… mm ah ah how after visiting castle how about how about
149 walking to Beachton?
150 RI: (laughs) Ah very far
151 MW: Is it not?
152 RI: Not so possible [**MW**: Yeah] ah (laughs) Do you wanna try?
153 MW: Mm no it's too
154 RI: Maybe we mm we can go
155 MW: Ah I wanna do boat… boat hire?
156 RI: Ah boat hire ah and you [**MW**: Ah] use maybe we can use this boat [**MW**:
157 Ah] so so after visiting castle
158 MW: After visiting the castle
159 RI: By foot?
160 MW: Ah yes let's go to Stinkton and
161 RI: Stinkton and we hire we could hire the boat
162 MW: Boat and back to Beachton
163 RI: Beachton

164 MW: Ah yes that's nice
165 RI: Yes
166 MW: So
167 RI: Yeah great
168 MW: So remember on on Friday night we we will stay Beachton and going
169 disco
170 RI: And yes disco enjoy disco and Saturday morning we we'll
171 MW: We'll go
172 RI: Ah Cookwell
173 MW: And we will enjoy
174 RI: Festival festival
175 MW: Festival and shopping
176 RI: And on Saturday night [**MW**: Saturday night] we could we will [**MW**:
177 Stay Trout] stay Trout and enjoy dinner
178 MW: Yeah and so Saturday morning we'll visit Loch Ness Castle and after
179 visiting
180 RI: We we'll walk [**MW**: Stinkton] Stinkton and hire [**MW**: Hire the boat] the
181 boat to Beachton
182 MW: Mm to Beachton
183 RI: And stay
184 MW: And stay Beachton [**RI**: On Sunday] to sleep (both laugh)
185 RI: OK
186 MW: Mm yes it's nice

An initial read-through of the two ss-ss role-play transcripts does not reveal any particularly remarkable differences between them: both seem to achieve the task quite successfully, in transactional *and* interactional terms, without resorting to L1 use to any degree. This was largely anticipated since the pair recorded on the 1st attempt (RI &MY) were both two of the most proficient English speakers in Class 1 (the top level in the university). MY, one of the case-study students, had also spent 5 years living in Canada so was distinctly superior to most other students (see appendix XI). It is worth bearing in mind, then, that this example represents one of the better 1st attempts in the class.

A more detailed analysis of the transcripts, however, begins to reveal some significant differences between the pre- and post-treatment attempts. Firstly, the 2nd attempt is almost twice as long (1024 words vs. 665 words), indicating a more elaborated negotiation, which closely approximates the length of the NS discussion (979 words). Secondly, the speech rate increases from approximately 59 wpm in the 1st try, to 69 wpm in the 2nd try,

suggesting increasing fluency in performing the task. Again, this was anticipated since task performance will naturally tend to improve with repetition (Skehan 1998). To what extent the observed improvement in these particular discourse features was a result of simply repeating the task, as opposed to the awareness-raising activities carried out in class, is impossible to say because there was no control group.

In order to further investigate differences between the two role-play attempts, key target language for negotiating plans was identified and extracted from the transcripts and compared with the language used by the NSs in their discussion. A total of 93 words/phrases were recovered from the three transcripts and the number of occurrences of these in each role-play attempt is shown below in Table 7.1[16]:

Target language (negotiating plans)	**Occurrences NS Version**	**Occurrences NNS Version (1st)**	**Occurrences NNS Version (2nd)**
1. Could (all occurrences)	12	0	16
2. That's…	11	0	2
3. There's (lots of) + noun	10	3 (4.5)	5
4. Yeah	10	7 (10.5)	12
5. We/you could (n't)	9	0	13
6. OK	8	7 (10.5)	5
7. Would/('d) (all occurrences)	7	0	1
8. Can (all occurrences)	6	3 (4.5)	6
9. We'll + infinitive (decision stage)	6	0	4
10. Should (all occurrences)	5	1 (1.5)	3
11. We should + infinitive	5	0	3
12. Mm	4	12 (18)	20
13. Mm hm	4	0	0
14. They have (n't)…	3	0	0
15. You can/can't + infinitive	3	0	0
16. Do you think we should + infinitive?	2	0	0
17. I fancy + verb-ing	2	0	0
18. Shall we + infinitive?	2	0	0
19. That would be + adj.	2	0	0
20. We need to + infinitive	2	0	0

[16] i) Adjusted figures for the 1st NNS role-play, multiplied by 1.5 to take into account its shorter length, are shown in parentheses.
ii) Asterisk '*' before phrases indicates dispreferred choices for either grammatical or pragmatic reasons.

21. We/you can either… or…	2	0	0
22. We're + verb-ing	2	0	0
23. We're here/there	2	0	0
24. What do we know about + noun?	2	0	0
25. All we want to be doing is + verb-ing	1	0	0
26. Are we going to + infinitive	1	0	0
27. Could (n't) we + infinitive?	1	0	2
28. Could do.	1	0	0
29. I can live without that	1	0	0
30. I do want to + infinitive	1	0	0
31. I think we'll probably be + adj.	1	0	0
32. I would quite like to + infinitive	1	0	0
33. If it's going to be + adj., then…	1	0	0
34. It'll be + adj.	1	0	0
35. It's got + noun	1	0	0
36. It's nice	1	0	2
37. Let's + infinitive	1	2 (3)	5
38. Perhaps go to + noun	1	0	0
39. So the attractions are…	1	0	0
40. So you definitely don't want to + infin.	1	0	0
41. That's OK	1	0	0
42. That's sorted then	1	0	0
43. The + noun + is unlikely to be + adj.	1	0	0
44. There are + noun	1	2 (3)	1
45. We may as well + infinitive	1	0	0
46. We/you have to + infinitive	1	1 (1.5)	2
47. We'd have to + infinitive	1	0	0
48. Well I definitely want to + infinitive	1	0	0
49. We've got to + infinitive	1	0	0
50. We've only got the option of +verb-ing	1	0	0
51. What are the things that we really want to do?	1	0	0
52. What do you + state verb	1	0	2
53. What would you definitely want to + infinitive	1	0	0
54. You decide	1	0	0
55. You said you definitely wanted to + infinitive	1	0	0
56. *(Maybe) I/we + infinitive	0	0	3
57. *(Maybe) we can + infinitive	0	3 (4.5)	5
58. *…do you want?	0	1 (1.5)	0
59. *Go to + noun	0	2 (3)	1
60. *Hopefully I like + verb	0	0	1
61. *I want to/wanna + infinitive	0	6 (9)	1
62. *Maybe you like + noun?	0	0	1
63. *We might go to + noun	0	3 (4.5)	0
64. *We will + infinitive	0	4 (6)	3
65. *We'll + infinitive (negotiating stage)	0	1 (1.5)	1
66. *Which places do you want?	0	1 (1.5)	0
67. *You want + infinitive?	0	1 (1.5)	0
68. Actually + -ve expression	0	0	3
69. Anything else?	0	0	2
70. Do you like + noun?	0	0	2
71. Do you want to + infinitive?	0	1 (1.5)	4
72. Good idea	0	0	2
73. Great	0	0	1
74. How about + verb-ing?	0	0	10
75. I think it's + adj.	0	1 (1.5)	1
76. I/you/we have to + infinitive	0	1 (1.5)	2

77. I'd rather + infinitive	0	0	1
78. Is there + noun?	0	1 (1.5)	1
79. It may be + adj.	0	1 (1.5)	0
80. It's OK	0	1 (1.5)	0
81. Not so + adj./adverb	0	1 (1.5)	5
82. Not so much	0	1 (1.5)	2
83. Noun + has + adj. + noun	0	0	2
84. Shall (all occurrences)	2	0	0
85. So I suppose you want to + infinitive?	0	0	1
86. So really maybe…	0	1 (1.5)	0
87. Sounds nice	0	0	1
88. We had better + infinitive	0	0	1
89. What are you going to + infinitive	0	0	1
90. What could we + infinitive?	0	0	1
91. Where should I/we + infinitive?	0	1 (1.5)	0
92. Where/what are we going to + infinitive?	0	0	3
93. Yes	0	12 (18)	35

Table 7.1 Number of occurrences of target language for negotiating plans in NS & NNS role-plays

Of course, placing too much emphasis on the significance of these figures would be dangerous since, in all three role-plays, talk was being created contingently, turn by turn, by the interlocutors. Each of the three speech events was unique to the particular context and the particular participants involved and it is impossible to say with any surety how typical these features would be in a larger population. For example, the NS interaction could have been affected by any number of factors, including:

a) The husband-wife relationship of the NSs;

b) Their nationality or regional location;

c) Gender differences;

d) The 'make-believe' aspects of the role-play;

e) Performing for the benefit of a video camera; and so on,

As Schegloff (1993b: 117) makes clear, the jury is still out with respect to the relevance of quantitative approaches to conversational analysis:

It *may* turn out that there is nothing both distinctive and defensible to be gained from quantitative studies of talk-in-interaction. Should that happen, we *could* face a curious irony [….] In contrast to much of the subject matter of the social sciences – which has been taken to be fundamentally *in*determinate at the level of individual occurrences and orderly only at the aggregate statistical level – conduct in talk-in-interaction *could* then appear to be demonstrably orderly *at the level of the singular occurrence only* and, in effect, *not* orderly in any distinctive, relevant, or precisely determinable way in the aggregate. As of now, I do not believe this conclusion is warranted; the proper grounding and payoffs of quantification have not yet been thoroughly explored.

The following observations are therefore made in the light of the knowledge that they may be illustrating idiosyncratic differences that cannot be generalized to a wider population. My own gut instinct however, is that this is not the case but readers will have to come to their own conclusions.

7.2.1 Comments on the transcripts

The first characteristic noticeable from table 7.1 is the large variety of target language expressions used by the NSs to negotiate their weekend away. However, many of these only occur once or twice in the transcript and are therefore unlikely to be noticed by students, unless made salient through orphographic means (**bolding**, *italics*, CAPITALISATION, etc.) or explicit focus in the classroom. As discussed in chapter 2, this is typical of authentic language, which has not been contrived specifically to present target forms, and can be viewed as either a strength or a weakness, depending on your point of view. Fewer forms, repeated more in a text might increase the chances of noticing and uptake by learners. On the other hand, a wider variety of phrases might increase the chances of students discovering new forms, which appeal on a personal

level, and stimulate acquisition (for example RI, in her diary, comments on the expression *definitely*). In this case, despite the wide variety of phrases employed, it is the collocations that are likely to be new for students, rather than the lexis itself so the text is unlikely to cause excessive difficulties.

Next, focusing on dispreferred target language expressions produced in the NNS transcripts (indicated by *), the number of occurrences drops from around 33 in the 1st role-play to 15 in the 2nd attempt. In particular, instances of '*I want to + infinitive*', or '*we will ('ll) + infinitive*' are reduced and this suggests that the noticing work done in the classroom on the negative pragmatic consequences of using these forms has had an effect.

Use of the modal auxiliaries *could*, *would* and *should* also increases markedly from the 1st role-play attempt to the 2nd (from 1 to 20), with *could* in particular being used much more post-treatment. This also provides strong support for the notion that the work done in the classroom on these forms has had a direct and beneficial effect on learners' output.

The phrase '*How about + verb-ing*', to make suggestions, is not used at all in the 1st role-play but has 10 occurrences in the 2nd attempt. Interestingly, this language did not show up in NS transcript either, but instead was elicited onto the whiteboard in the class as an appropriate form to use when planning. This illustrates how input for acquisition can originate not only from materials brought into the classroom but also the work that goes on around them, with students or the teacher offering up other alternatives to those seen in the texts.

The back-channels *Mm* and *Mm hm* are also interesting because whereas the NSs use both forms sparingly and in equal proportions, the NNSs in both the 1st and 2nd role-plays

use only *Mm*, but use it a lot with 18 and 20 occurrences respectively. A similar pattern is seen with *yeah* and *yes* where the NS transcript shows only 10 examples of the first form and none of the second, while the NNS transcripts show a preference for *yes* and have occurrences of both forms totaling 28 and 47 respectively. These differences seem to reflect the 'highly affect-laden interactional style' (Clancy et al. 1996: 381) typical of Japanese discourse and are in line with the higher levels of back-channelling reported by many researchers (Clancy 1982; LoCastro 1987; Yamada 1997; Gilmore 1998). If this does reflect transfer of L1 interactional patterns onto the L2, it is not something that necessarily needs changing since it would be unlikely to have a negative impact on students' ability to communicate effectively with English NSs.

7.3 Conclusion

The closer analysis of classroom events in this chapter provides some useful insights into the processes at work in this investigation and helps to explain the statistically significant differences seen between the experimental and control groups in the analysis of covariance. The authentic audio-visual materials exploited in this class can be seen to be a rich source of input for learners, providing them with information on lexico-grammatical, pragmatic and discoursal features of natural conversation rarely highlighted in language textbooks. Students were able to identify differences between their own output and the native speaker model and use this information to make beneficial changes in their 2nd role-play attempts, which demonstrate increasing communicative competence. Through many small steps forward, such as those illustrated here, the learners in the experimental group were able to develop a broader range of communicative

competencies than their peers in the control group, where the textbook input available for acquisition was predominantly lexico-grammatical.

CHAPTER 8. CONCLUSIONS, STRENGTHS, LIMITATIONS & IMPLICATIONS OF THE STUDY

8.1 Conclusions

This longitudinal, classroom-based study investigated the effects of authentic input vs. textbook input on Japanese university students' developing communicative competence over a period of 10 months at Kansai Gaidai University, Japan. By taking a pragmatic approach and incorporating both quantitative and qualitative techniques into the investigation (often referred to as *mixed methods* or *multi-strategy research*), it was hoped that the study would more effectively illustrate the complex and multi-faceted nature of language learning in the classroom. This has, I believe, been the case with insights emerging which would have remained hidden, had either a purely quantitative or qualitative approach been taken.

8.1.1 Quantitative measures

At the end of the treatment period, one-way between-groups analysis of covariance (ANCOVA), which statistically controls for differences in proficiency levels between participants, indicated that the experimental group had out-performed the control group on five out of eight measures of communicative competence, with statistically significant differences on the listening test, the receptive pronunciation test, the receptive vocabulary test, the oral interview and the student-student role-play. No statistically significant differences were seen between the two groups on the 'C'-Test, the grammar test or the discourse completion task. For the 'C'-Test, this was not surprising since it is primarily a measure of students' reading skills: an area that was not focused on in either group to any

extent. For the grammar test, the control group were expected to improve more than the experimental group because they received greater quantities of grammar input in the textbooks used. The fact that there were no significant differences between the two groups can probably be accounted for by the fact that their grammatical competence was already well developed at the start of the investigation (with mean scores of 83.4% and 89.1% respectively on the grammar test), which meant there was less room for improvement in this area. In addition, some students in the control group commented in their diaries that much of the grammar input in the textbooks was already familiar and perhaps, therefore, did not provide enough challenge. This highlights one problem with internationally marketed language course books, which is that they do not always meet the needs of a specific group of learners. In this case, although the speaking and listening tasks provided an appropriate level of challenge, the grammar tasks often did not.

The lack of any statistically significant difference on the discourse completion task was surprising since the experimental group received considerably more pragmatic input over the course of the investigation and were therefore expected to develop more pragmatic competence. Although they *did* demonstrate an increasing awareness of pragmatic issues in their diary entries, this did not seem to translate into more appropriate request speech acts on the DCT. Three possible reasons were suggested for this:

i) A problem with the testing conditions.

ii) A lack of reliability of the testing instrument (the DCT).

iii) Ineffective pragmatics training.

The most likely causes were a combination of ii) and iii) and this has implications for both research and language teaching (see section 8.3 below).

It should be pointed out that although the DCT results showed no difference in pragmatic competence between the two groups, the conversational management component of the student-student role-play did indicate that the experimental group were better able to initiate and terminate their conversations appropriately; an area focused on explicitly during the trial. This suggests that some aspects of the pragmatics training were successful.

Modifying variables were carefully controlled so that the two groups were, firstly, as similar as possible (nationality, age, gender & English proficiency) and, secondly, received identical treatments (learning context, teacher, lesson length) *except for the type of input* (the independent variable). This means that we can be reasonably confident in attributing the superior development of communicative competence in the experimental group to the authentic input and its accompanying tasks. These results were anticipated for two reasons:

i) The 'richer' authentic input was better able to illustrate different aspects of communicative competence, which could then become the focus of tasks.

ii) Tasks were designed to highlight a wide variety of discourse features, not normally focused on in language textbooks, and then to give students practice in using them. By encouraging learners to, first, notice new features of the L2 and, then, to use them in their output, two important conditions for language acquisition were satisfied.

I believe these findings have important and wide-ranging implications for language teaching (see section 8.3.2).

8.1.2 Qualitative measures

The qualitative data for the investigation came from three principal sources: a) learner diaries, documenting what participants believed had happened in classes and their reactions to the materials or activities; b) case studies with six students of high, medium, or low proficiency from the experimental and control groups, using diary comments as 'jumping off' points into more detailed, semi-structured interviews; and c) a closer description of a series of lessons focusing on developing learners' negotiation skills.

The learner diaries were extremely effective in eliciting students' thoughts and feelings about the classes and the compliance rates of 84.2% were much higher than those often reported in the literature. This was probably due to the unique classroom environment which tends to prevail in Japan, where social constraints impose silence on students. The diaries allowed learners to engage in a dialogue with the teacher that would have been impossible otherwise, and many of them seem to have appreciated the chance to voice their own thoughts and feelings through this medium.

Significant events were extracted from the learners' diaries and organised into four categories of 'recurring issues':

i) Comments relating to aspects of the input or activities learners 'noticed' in the classroom (components of the communicative competence model or language skills);

ii) Comments relating to the learning environment (the teacher, other learners, tension in the classroom, external concerns);

iii) Comments relating to aspects of the lessons learners found motivating or de-motivating;

iv) Comments relating to the pre- and post-course tests or testing procedures.

Including the students' own emic perspectives in this way provided insights that would otherwise have been unavailable to the researcher, and often complemented the quantitative results. The diary extracts relating to i) above suggested that the experimental treatment had been successful in raising students' awareness of the different components of the communicative competence model: linguistic competence, pragmalinguistic competence, sociopragmatic competence, strategic competence and discourse competence. This 'noticing' appears to have facilitated the acquisition of both linguistic and paralinguistic features often inaccessible to learners through traditional language textbooks and helps to account for the statistically significant differences between the experimental and control groups in the post-course communicative competence tests. Comments relating to ii) illustrated how social goals in the classroom can often override pedagogic goals and represent an important modifying variable in classroom-based research. Since the social context is not something that can be controlled by the researcher, I believe it is essential to be as explicit as possible about the learning environment within which quantitative studies are grounded. Comments relating to iii) provided support for stimulus appraisal models of language learning (Scherer 1984; Schumann 1997), which see input as being appraised by learners along five criteria: novelty, pleasantness, goal/need significance, coping mechanisms and self or social image. Positive appraisals of input are believed to encourage greater cognitive effort and greater engagement, leading to more learning, while negative appraisals result in avoidance. Although both treatment groups appeared to be quite motivated by the classes they received during the trial, evidence from the learners' diaries and case-study

interviews suggests a strong preference for authentic materials in the experimental group. It was postulated that the reason for this was that they were better able to satisfy the appraisal criteria of novelty, pleasantness and goal/need significance (Schumann 1997) and that this resulted in higher overall levels of satisfaction, increased engagement with the input and, consequently, more learning taking place. Finally, comments relating to iv) suggested that learners found the speaking tests (either the oral interview, DCT or student-student role-play) particularly stressful and sometimes employed strategies for reducing the inherent unpredictability of these tasks, such as trying to hear DCT prompts before taking the test, or pre-preparing conversation topics for the role-play.

The results of the case studies, in contrast to the diary reports, were rather disappointing since they provided few additional insights into the learners' thoughts and feelings on classroom events. This was explained in terms of Japanese students' preference for written, as opposed to spoken, modes of communication and the rather rushed nature of the case study interviews, which were often 'squeezed in' at the end of classes. Having said this, one important point to emerge from the one-to-one interviews with case study students was their marked preference for authentic materials over textbook materials, despite the challenges this posed. Even YK, the least proficient learner in the experimental group, was quite insistent on this issue. It would seem then, that despite theoretical objections to the use of authentic materials in the classroom, particularly by Widdowson (1994; 2003; etc), learners themselves are clear about the advantages they provide. Perhaps it could be argued that learners sometimes don't know what's good for them, but I would see this as academic arrogance: learners are better placed than anyone to say what input best meets their specific interlanguage needs and language goals.

Finally, chapter 7 provided a closer description of a series of lessons aimed at developing students' ability to negotiate plans appropriately in English. This illustrated how using authentic materials with the experimental group allowed learners to 'notice the gap' (Schmidt & Frota 1986) between their own discourse and that in the native speaker model. The sample role-play transcripts, recorded at the beginning and end of the sequence of lessons, suggested that the awareness raising activities had been successful in developing learners' communicative competence in a number of ways:

i) Linguistically, by expanding their vocabulary repertoire;

ii) Pragmatically, by encouraging more appropriate choices of modal auxiliaries at the different stages of the negotiation;

iii) Discoursally, by encouraging learners to jointly negotiate their plans and share the 'conversational leg work'.

By focusing in on some sample lessons in this way, the emphasis shifted more towards the processes involved in language acquisition, in a way that complements the 'product bias' of the quantitative analysis, and allowed us to see *how* the statistically significant differences seen between the experimental and control groups might have come about.

8.1.3 Reconciling quantitative & qualitative measures

The use of both quantitative and qualitative approaches in classroom-based research inevitably reveals more about language learning than either approach could do singularly. However, they present the researcher with two distinct views of what took place, which do not necessarily sit comfortably side by side. In this sense, they demand some form of reconciliation at the end of the study. In chapter 3, I quoted Bryman (2006), who gives a

comprehensive list of possible reasons for using mixed methods and these provide a useful framework for illustrating how the two approaches complemented each other in this trial:

a) *Triangulation*: Combining quantitative and qualitative approaches in this investigation certainly enhanced the validity of the study. The diary comments and case-study interviews from the experimental group suggested that students were noticing a wide variety of features from the authentic input and this increased awareness helped to explain the statistically significant differences between the two groups at the end of the investigation. Furthermore, the closer description of a sequence of lessons in chapter 7 allowed us to focus in on the processes of language acquisition at work during the trial. This showed how highlighting the differences between the learner and NS discourse resulted in more pragmatically and discoursally appropriate language being produced. If such small steps forward were occurring throughout the trial, it is easy to see how the experimental treatment might have led to superior results in the post-course communicative competence measures. Similarly, the use of quantitative approaches enhanced the validity of the qualitative results, which, although providing interesting insights into individual participants, cannot themselves be generalised to the wider population. In this way, the two approaches are similar to different lenses on a microscope: qualitative methods are like high-powered lenses, focusing in on events and allowing us to see the details, while quantitative methods, on the other hand, are more like low-powered lenses which show us the whole picture (see Onwuegbuzie & Leech 2005: 383).

b) *Offset*: This reason is closely tied up with the rationale for triangulation discussed above since both approaches have strengths and weaknesses, which can be offset by combining them together. In this case, the descriptive power of the qualitative methods complemented the generalising power of the quantitative methods.

c) *Completeness*: Combining quantitative and qualitative research without doubt provided a more comprehensive account of events in the classroom. The diary reports, case study interviews and transcripts of classroom interaction allowed us to see things from the learners' (emic) perspective and showed how, often, social goals took precedence over pedagogic goals for them. By providing a clearer picture of the social context in which the investigation took place, other researchers are better placed to judge the relevance of the results to their own teaching environments.

d) *Different research questions*: The quantitative methodology was primarily concerned with discovering real differences between the two treatment groups which could be generalized to the wider population and sought to answer the question: Is the 'richer' input provided by authentic materials, combined with appropriate awareness-raising activities, better able to develop a range of communicative competencies in learners than textbook input? The qualitative research, on the other hand, was primarily concerned with the context of learning and how the learners themselves viewed events, in order to answer the question: What caused the observable differences in communicative competence between the two treatment groups?

e) *Unexpected results*: The diary comments were extremely useful in helping to explain the lack of any statistically significant differences between the two treatment groups on the DCT, suggesting that the experimental group had greater pragmatic awareness by the end of the course but were unable to apply this in real time to produce more appropriate requests.

f) *Sampling*: Results from the pre-course communicative competence tests were used to select case study students of high, medium and low proficiency so that learners' views were more fairly represented.

g) *Utility*: By providing a more comprehensive picture of what took place in the classroom, practitioners have a clearer idea of how the results relate to their own learning contexts.

h) *Diversity of views*: By combining methods, a wide variety of different participants' views could be incorporated into the investigation: the researcher's, the teacher's, the learners' and the NS raters'.

8.2 Strengths & limitations of the study

8.2.1 Strengths of the study

By considering many of the criticisms commonly levelled at classroom-oriented research at the experimental design stage (see chapter 3), I believe that this study was able to avoid many of the potential pitfalls seen in earlier work. Firstly, the investigation was carried out entirely in genuine classroom contexts (Nunan 1991 estimated that this was the case in only 30% of studies he analysed), which means that the results have higher

external validity and other readers are more easily able to generalise the findings to their own pedagogical contexts.

Secondly, the longitudinal study was continued for a comparatively long time period (10 months) with a reasonably large sample size (62 students), increasing the likelihood of detecting real changes in learners' communicative competence. Alderson & Beretta 1992 and Ritchie & Bhatia 1996 both criticise many longitudinal studies for being too short, something which is understandable given the financial or time constraints often associated with research projects, but which threatens the validity of results.

Next, the use of both quantitative and qualitative (mixed methods) approaches in the investigation, although extremely time-consuming to implement, provided a more comprehensive picture of what happened in the investigation and enhanced the findings in the ways outlined in section 8.1.3.

The investigation also benefited from the fact that relatively few constraints were imposed on the researcher/teacher by the educational institutions involved in the project. With ethical considerations coming increasingly to the fore in many countries, it is often impossible for teachers to use their own students for research purposes. This, in my opinion, is not only excessive, it also increases the difficulty of finding out what is really happening in the classroom. For example, in the longitudinal investigation reported on here, it is highly unlikely that an outside researcher with limited access to the university, teacher and learners involved would have been able to provide as detailed an account as that given here. Research is necessarily intrusive because that is what is required if we are to get anywhere near the truth: a photojournalist operating in a war zone intrudes into other peoples' misery when he or she takes pictures of a family grieving their dead.

Should they too be asked to limit their activities and respect the rights of their subjects? If they did, our understanding of the realities of war would be severely curtailed. Although, in the context of classroom research, I do, of course, support participants' rights to anonymity or to opt out of the investigation altogether, I feel that restrictions being imposed in some parts of the world now (particularly the United States) go too far.

The classrooms used in this particular investigation had the advantage of being well equipped, with video/DVD players and projectors installed in all rooms. This was particularly important for the experimental group, who regularly used audio-visual materials to focus on sociopragmatic features of the input, such as NSs' non-verbal communication. Obviously, in many language learning contexts these kinds of facilities are not available and this potentially limits the relevance of this work for many teachers. Since the exploitation of authentic materials in the classroom has largely become possible because of technological advances such as DVD and the internet, institutions with access to only chalkboard and textbook resources would find it very difficult to implement the kind of syllabus recommended here.

8.2.2 Limitations of the study

As discussed above, the classroom-based research carried out in this study aimed to provide as comprehensive a picture as possible of what happened during the trial. Although I believe this enhanced the findings considerably, it also brought with it a number of problems. Firstly, the time-consuming nature of collecting so much data means that it is harder for others to replicate: few researchers are likely to have the unrestricted access to learners and resources that I was fortunate enough to be given.

Secondly, the sheer quantity of data generated by examining so many different aspects of the learning context creates difficulties in the writing-up stages. There is a certain amount of tension between providing a comprehensive account of events while still remaining within the accepted limits for a thesis of around 80,000 words.

The study's central hypothesis, relating to students' development of communicative competence, also brought with it similar difficulties, created by the demand for comprehensiveness. Since the communicative competence model is made up of five distinct components, it required a large number of different tests in order to be able to assess it fairly (eight in total, or thirteen including the various sub-components). Again, this makes the study more difficult to replicate since the pre- and post-course tests were very time-consuming to implement and noticeably strained the good will of the participants.

Since it appears from the literature available on communicative competence that the various components can be developed separately from each other to some degree (see chapter 1), it is obviously important to try to measure all of them in order to gain an accurate picture of each student's true proficiency level. This, however, leads to a further complication with the statistical analysis: that of the controversial subject of the Bonferroni adjustment. This states that when an experimental design looks at the effect of an independent variable on a number of dependent variables and multiple comparisons are made between 2 groups, there is an increased risk of finding significant results purely by chance (i.e. of making a Type 1 error). By analogy, the more rolls of the dice we have, the more likelihood there is of getting a six. To avoid this, some statisticians recommend applying a Bonferroni adjustment, which sets the alpha value at a more stringent level,

calculated by dividing the selected alpha level by the number of comparisons made. So, the more comparisons we make in an investigation, the less likely we are to reach statistical significance because the selected alpha level becomes more and more difficult to reach. For example, in this study the alpha level was set at .05 (a standard figure for the social sciences) and this led to the conclusion that 5 of the 8 measures of communicative competence showed statistically significant differences. However, if a Bonferroni adjustment were made, the alpha level needed to reach statistical significance would become .00625 (.05 divided by 8), meaning that only three of the eight measures would be considered significant (the receptive pronunciation test, the receptive vocabulary test and the student role-play). Now, if each of the sub-components of the IELTS oral interview and student-student role-play were *also* regarded as separate comparisons, the total number of dependent variables would rise from 8 to 13, and the alpha level would then need to be set at .0039 (.05 divided by 13). In this scenario, only five of the thirteen measures would be considered significant (the receptive pronunciation test, the receptive vocabulary test, the conversational behaviour & conversational management components of the student role-play, and the interaction component of the oral interview).

It can be seen then that exactly which results are considered significant in a study is open to different interpretations, depending on whether or not a Bonferroni adjustment is applied. This creates something of a paradox in social science research because the more comprehensive we try to be (making multiple comparisons between groups), the less likely we are to reach statistical significance (if a Bonferroni adjustment is applied). Thus

it would seem that while mixed methods approaches encourage more comprehensive testing, statistical procedures, incorporating Bonferroni adjustments, encourage less.

Having said this, there is still a great deal of controversy surrounding the issue of Bonferroni adjustments, and many researchers believe that it should not be applied when assessing evidence about specific hypotheses. Perneger (1998: 1236), writing in the British Medical Journal, claims that the view widely held by epidemiologists is that 'Bonferroni adjustments are, at best, unnecessary and, at worst, deleterious to sound statistical inference'. He justifies this position on a number of counts:

i) The study-wide error rate applies only to the hypothesis that the two groups are identical on all the variables (the universal null hypothesis). If one or more of the *p* values is less than the set alpha level (which is true in this study), the universal null hypothesis is rejected. In this case, we can say that the two groups are not equal for all the variables, but we are unable to say which, or even how many, variables differ. This information is of little interest to researchers who want to assess each variable in its own right.

ii) Bonferroni adjustments imply that comparisons between groups need to be interpreted differently, depending on how many tests are carried out. If results are, cynically, 'sliced up like salami' (ibid), with one *p* value published at a time in journals, or if less 'helpful' results are jettisoned at the writing-up stage, significance can, theoretically, be massaged in the direction the researcher desires. This defies common sense: evidence in

data should be interpreted based on what it shows, not on how many other tests are performed.

iii) Using Bonferroni adjustments increases the likelihood of making a type II error (stating that there are no significant differences between groups when, in fact, there are); they do not, therefore, guarantee a more prudent interpretation of results.

Adjustments for multiple tests were, in actual fact, originally intended to aid decision-making in repetitive situations, such as identifying defective products in factories, where all the items being compared are identical. It was not intended to be used for assessing evidence about specific hypotheses and for the above reasons, I have chosen to ignore it in this study.

Since the chosen research design imposed considerable demands on the author – designing materials for the experimental syllabus, teaching both treatment groups for the duration of the investigation, implementing the pre- and post-course tests and conducting the qualitative studies – some things which would have benefited the study were left undone. For example, had the teacher documented his own personal impressions of events in a diary throughout the trial, it would have provided a useful comparison to those of the learners. In addition, more recording and transcribing of student-student interaction during the trial may also have told us more about the processes at work in the classroom which resulted in the statistically significant differences observed between the two groups. Regular recording in the classroom would have had the added benefit of familiarising students with the equipment more, possibly leading to less self-monitoring in their speech. As it was, the learners were tape recorded so rarely that they remained

very conscious of the microphones and often provided a running commentary on their discussion for the benefit of the tape, as these two extracts of classroom interaction from the pilot study indicate:

S1: *Ariehan* (speaking into the microphone) *ariehan* means impossible (S2 laughs)
Impossible **impossible**

S1: Moment sorry I'm sorry [S2: Don't worry] (speaking into microphone) I'm sorry Hige sorry. Koizumi sorry Harakatashi Harakatashi is government person and it's a it's the time is taishou.
S2: (laughs) *zatsugaku ya zatsugaku ippai haitteru*
This tape is full of interesting things

Next, the case-study interviews might have produced better results had they been carried out in a more relaxed environment; as it was, they were often squeezed into the breaks between lessons and were therefore rather rushed. Finally, had more time been available, it would have been preferable to have designed a DCT specifically for this investigation, testing a wider range of speech acts such as those for requesting, suggesting, inviting, refusing, apologising, complaining, complimenting and thanking which have been found to occur in all speech communities explored to date (Kasper & Schmidt 1996).

Another weakness of the investigation relates to difficulties experienced in training NS volunteers for the interviewing and rating tasks. Work practices in Japan mean that teachers very often rush from one university to another during the day and have very little time available for training sessions. This meant that, at times, their understanding of what was required of them was not as thorough as I would have liked, threatening the reliability of the testing instruments. For example, in the IELTS oral interviews, some interviewers adopted a fairly confrontational style while others were more relaxed and friendly. This undoubtedly affected the learners' responses and, quite possibly, their

overall rating on this measure. Had more time been available for training, the NS volunteers could have watched some model interviews and rehearsed their interviewing technique to ensure consistency across the board.

A high mortality rate (of almost 33%) and student absenteeism were other difficulties that had to be contended with during the trial. Fortunately however, the number of students in each group did not drop so low as to invalidate the inferential statistical analysis. This is always a risk in the classroom-based research and is largely beyond the control of the researcher. The only way to safeguard against this problem is to include as many participants in the trial as possible although, again, this is something we often have no control over.

It is important to remember that the results reported in this investigation represent a comparison between authentic materials and language textbooks selected by the author. The textbooks used with the control group were chosen for their relatively contrived, more traditional content, in order to create a sharp contrast in the input received by the two treatment groups. Of course, had more progressive textbooks been used, such as Natural English (Gairns & Redman 2002), quite different results might have been seen. Similarly, the quality of the materials designed by the author for the experimental group is another important consideration. When generalising out to a wider population, therefore, readers must, to some extent, judge for themselves the importance of the findings and their relevance to other teaching contexts.

8.3 Implications

8.3.1 Implications for future research

In an editorial in the first volume of the Journal of Mixed Methods Research, January 2007, Tashakkori & Creswell discuss the importance of establishing the value of mixed methods research, and ways in which it can provide greater insights than quantitative or qualitative approaches alone. I believe that this work illustrates very clearly the benefits of a multi-strategy approach to classroom-based research and contradicts the views of purists such as Smith (1983) and Smith & Heshusius (1986), who contend that quantitative and qualitative methods cannot be combined on the basis that they stem from different ontologic and epistemologic assumptions about the nature of research (Bryman 1984; Onwuegbuzie & Leech 2005; Dörnyei 2007).

In this study, a more pragmatic approach was adopted and methods were selected on the basis of their potential to throw light onto different aspects of language learning, rather than their theoretical affiliations. As discussed in section 8.1.3, combining qualitative and quantitative methods within the same study allows us to zoom in and zoom out on particular areas of interest, and the results from one approach are often able to help explain those from the other. I believe, therefore, that future classroom-based research would benefit more from combining quantitative and qualitative methods in order to better explain the complex and multi-faceted nature of language learning:

> The social phenomena that we study 'on the ground' in the real world are unarguably complex, dynamic, and contextually diverse. The need to understand these phenomena, to make sense of contemporary social problems and to find promising solutions to them remains pressing, if not urgent. We therefore need to use *all* of our methodological expertise and skills in this endeavour for contemporary understanding of social

issues. We need to marshal *all of* our multiple ways of knowing and their associated multiple ways of valuing, in the service of credible and useful understanding. (Greene, Benjamin & Goodyear 2001)

Although the results of this investigation suggest that the authentic materials and their associated tasks were more effective than the language textbooks in developing learners' communicative competence, more empirical longitudinal studies in other contexts are needed before we can make any strong claims. Even if the differences found in this study are confirmed elsewhere, it is highly likely that results would be influenced by the nationality or proficiency level of the participants, the kinds of materials or tasks included in the comparison, the attitudes and abilities of the teachers who deliver the courses, and other modifying variables. Only through repeat trials, where these variables are very carefully described, will the picture become clearer.

Earlier studies have shown that the various components of the communicative competence model *are* amenable to training (see chapter 1). However, as far as I am aware, this is the first investigation to have attempted to describe the effects of input on all five components simultaneously with the same group of learners. For practical reasons, this is important because it is a more faithful representation of the realities of language learning in the classroom. In addition, as yet little is known about how the different communicative competence components interact with, and influence each other. Nor do we know how much emphasis to put on each area of communicative competence at different stages of a learner's language development. I have argued, for example, that Japanese students would benefit from a focus on strategic competence at a much earlier stage in their learning, since the number of communication strategies needed is relatively

limited and increased strategic competence would benefit their confidence and maximise their exposure to comprehensible input.

Finally, further work is needed to develop reliable and practical tests to assess communicative competence in a comprehensive manner. Although measures of linguistic competence (grammar, pronunciation, vocabulary, spelling, etc.) are well established, this is not the case for the other components of the model. For example, the DCT used in this study was extremely time-consuming to implement and proved to be a rather unreliable measure of learners' pragmatic competence.

8.3.2 Implications for language pedagogy

The results of this study suggest that not only were the authentic materials used in the investigation better able to develop a broader range of communicative competences in learners, they were also strongly preferred by the participants in the experimental group. It has also been suggested that language textbooks, published for the international market, are less effective for a number of reasons:

i) Although their precise syllabuses with clear aims give the perception of comprehensiveness, they often only develop a very narrow range of communicative competences since lexicogrammatical features tend to dominate the content.

ii) Their graded syllabuses, moving from what is considered 'easy' to 'difficult' content, give the impression that we know more about second language acquisition than is actually the case. Items which are easy to analyse, such as 3rd person 's', can be hard for students to learn and others which are more

difficult, such as the Spanish subjunctive in *Que te vaya bien*, can be readily acquired as fixed expressions. In addition, since learners' interlanguage systems develop idiosyncratically and difficulty is largely affected by the extent of our background knowledge of a topic, grading materials to suit a particular group of learners is extremely hit or miss.

iii) The 'lock-step' approach to language learning, where items are presented and 'learned' thoroughly before moving on, is inconsistent with SLA theory and tends to lead to dull activities or materials.

iv) They are often less interesting for learners since the topics are not selected to meet their specific needs or interests (as was seen with the dance theme in the control group). According to Schumann (1997), three important components of our 'stimulus appraisal system' (which determines to what extent we engage with input) are novelty, pleasantness and goal/need significance. Authentic materials chosen by teachers and learners themselves are more likely to be positively evaluated and lead to sustained deep learning (see chapter 6).

v) Controversial (but often more stimulating) issues are avoided since they risk causing offence, and therefore reducing sales, in some cultures.

vi) Input in course books is often contrived by material writers with language aims taking precedence, which can result in dull texts, lacking in naturalness.

vii) Both teachers and learners are less committed to materials forced onto them than those they are able to select themselves.

How should we respond to this in the classroom? I would argue for the need for a paradigm shift, as suggested by Woodward (1996), but the critical question is what exactly we shift *to*? One possible solution is to abandon attempts to organise content around a structural syllabus (particularly at higher proficiency levels) and to instead provide learners with 'rich' samples of authentic input, with the potential to develop a wider range of communicative competencies and better meet their specific needs and interests. The communicative competence model could be used to inform the syllabus, ensuring that learners' linguistic, strategic, pragmatic and discourse competencies were all developed appropriately. This is precisely what the study reported on here attempted to do and, as we have seen, the results appear to be very promising.

However, a number of difficulties stand in the way of implementing this kind of syllabus. Firstly, it assumes that teachers and learners have access to a range of authentic materials in both spoken and written modes: something that is rapidly becoming more feasible with the widespread availability of DVDs and the internet, but is still not possible in many parts of the world.

Next, it assumes that language teachers have a deep understanding of the factors affecting learners' communicative competence and the necessary expertise to identify and exploit authentic materials to effectively meet their needs. Sadly, this is often not the case, particularly in Japan where native speakers of English are often automatically considered language experts and, within the university education system, any Master's degree is generally considered adequate. NNS teachers may also be disadvantaged since it is naturally more difficult for them to judge pragmalinguistic or sociopragmatic appropriateness. Current pre-service teacher training programmes for ESOL, such as the

ubiquitous CELTA course, still tend to emphasise lexicogrammatical features above all others, perpetuating the status quo.

Furthermore, a communicative competence-centred approach using authentic materials can be extremely time-consuming to implement. Initially, it requires some kind of needs analysis to determine how best to develop learners' overall competence and decisions need to be made in terms of how much emphasis to place on the different components of the model. As discussed in chapter 1, students' needs are likely to vary considerably depending on factors such as their proficiency level, their previous learning experiences, differences between their own culture and the target culture, and so on. For example, Japanese students, unused to 'negotiating meaning', are more likely to need a focus on communication strategies than European students. After this, appropriate materials, which can illustrate the identified target features, need to be collected (by teachers or the learners themselves). As I pointed out in section 4.6.2, criteria for selection of authentic materials to use in the classroom can also be very demanding. Once materials have been collected, transcripts of audiovisual material need to be prepared and tasks designed to highlight or practise the discourse features of particular interest. These can, of course, be shared amongst teachers in a particular institution to reduce the workload. In fact, it is advisable to construct a 'materials bank', indexed to identify topics or target features covered and skills practised. The texts collected in this way can be used to build up a corpus, using software such as Wordsmith Tools (Oxford University Press). This has several benefits for the users:

i) It allows word frequencies to be calculated to help ensure tests are a fair reflection of material covered in a course (Willis 2003 has some interesting ideas on this kind of 'pedagogic corpus');

ii) Keywords can be used to compare the corpus with a reference corpus, such as the BNC, in order to establish how representative the language illustrated in it is of a particular target speech community;

iii) The texts can be tagged to allow rapid identification of materials which contain target vocabulary, grammar items, speech acts or other discourse features;

iv) Concordance lines can be generated from the corpus to provide further illustrations of particular language points or discourse features (hesitation devices, discourse markers, etc.) in the classroom.

The final difficulty associated with this kind of approach is the design of tests, which are both a fair reflection of the course content and practical to implement. Since the syllabus is not pre-conceived, but is rather co-constructed by participants during the course, what takes place in the classroom cannot be predicted beforehand and must be examined retrospectively. Fair assessment can therefore only occur if tests are designed specifically for each course. In addition, any syllabus that aims to develop a broad range of communicative competences in learners must also endeavour to test them too. If we continue to assess only those features of the language that are easily measured (often lexicogrammatical items), the 'backwash effect' will ensure that students and teachers remain firmly focused on them at the expense of other areas. This remains problematic since reliable measures of strategic, pragmalinguistic, sociopragmatic and discourse

competencies have not been established yet. Strategic competence is difficult to assess because instances of communication breakdown are usually infrequent and unpredictable in conversation so tests would probably need to involve some kind of elicitation. Pragmalinguistic competence is also extremely difficult to assess, as we have seen in this investigation. DCTs are very time-consuming to implement and native speakers often disagree on the level of appropriateness of students' responses, when judging pragmatic rather than grammatical features. Perhaps some form of multiple-choice test would be practical though: for example, students could see various scenarios played out on video and then select the best utterance to complete the scene from a range of choices varying in pragmatic appropriateness. This would mean that they could all be tested simultaneously and the scores quickly calculated. Measuring sociopragmatic competence normally involves analysing students' behaviour in a particular context. Role-play scenarios are probably the best method for assessing this area, with either the teacher rating pairs as they perform or video recording to rate at a later stage. I normally record the role-plays so that students can watch them and rate each other's performances in class, an activity which they often find enjoyable and which encourages them to focus on this area of communicative competence. Discourse competence is difficult to assess because it requires analysis of longer stretches of spoken or written discourse. With writing samples, this means focusing on the overall structure of a piece of work and assessing how cohesive/coherent it is or how well it approximates the generic model. With spoken samples, it means analysing conversational turns for cohesion or coherence or identifying whether longer turns (for example in oral narratives) include all the obligatory parts. This is very difficult to do accurately without transcripts of the

conversation (something which is, of course, impractical in the classroom) and also requires a high degree of language awareness from the assessor. Despite the difficulties in assessing learners' overall communicative competence outlined here, this is an issue which needs to be engaged with and discussed so that practical solutions can be found.

A final point emerging from the qualitative aspects of this investigation was that social dynamics within the classroom play a critical role in language learning. If this is the case, teachers obviously need to make a concerted effort to understand the social forces at work in their classes, and to manage them in ways which maximise learning. This is difficult to do in any learning environment but is particularly problematic in Japan since:

a. Students tend to be shy and often avoid interaction with other members of the class (other than their best friends) or the teacher if they are able to;
b. They are unused to taking responsibility for their own learning and tend to rely on the teacher to dictate what to do, how to do it and who to do it with;
c. Social pressures in Japanese classrooms discourage students from expressing their opinions or feelings openly: often in this study I was surprised by diary entries which expressed great enthusiasm for lessons, despite the fact that students had remained pan-faced and apparently unimpressed throughout the class itself.

I believe that, in reality, teachers really have very little idea about what is actually going on in their classrooms. So many of the states and processes which affect language acquisition are hidden from view: students' needs and feelings; their learning goals; their interlanguage systems; the relationships between different members of the group; their

preferred learning styles and their overall communicative competence. In addition, teachers are normally so preoccupied with the logistics of running a lesson, that they have little time to focus on the social dynamics of the class. Monitoring student interaction during pair or group-work activities is a common technique employed to assess what is happening but this only provides 'snatches' of discourse which actually tell us very little. Recording and transcribing classroom interaction, as we saw in chapters 6 and 7, can be extremely illuminating but it is unrealistic to expect busy teachers to do this on a regular basis. If we want to better understand the processes at work in the classroom, I believe the best option available to us is the increased use of learner diaries. As we have seen, they are practical to implement and often provide insights which, if acted upon by the teacher, can greatly enhance language acquisition. They also show students that we are interested in their opinions and are willing to change classes to meet their needs, both of which are likely to increase motivation.

8.4 Summary

This study was inspired by the belief that language learning has the potential to achieve much more if it can only break out of the mould in which it has been set for so long. Insights emerging from a wide variety of fields, such as discourse or conversational analysis, pragmatics and ethnology, have deepened our understanding of what language is and what it means to be communicatively competent and, as McCarthy & Carter (1994) say, the landscape has indeed changed forever. Unfortunately, the practical implications of all of this are yet to make any real impact in ESOL classrooms. For my own part, I am quite convinced about the need for a paradigm shift in language learning

but the question is to what extent learners, teachers, institutions and publishers resist or facilitate change: there are great pressures at work which function to maintain the status quo. I hope that this investigation will encourage teachers to explore new ways of doing things and put the fizz back in to their classrooms.