What do tests of listening comprehension test? – A retrospection study of EFL test-takers performing a multiple-choice task

Wu Yi'an *Beijing Foreign Studies University*

A test-taker's performance in a listening comprehension test is seen to be a function of at least two variables: his or her listening comprehension ability and the test method. To know how this trait-method unit works, it is necessary to look into the test-taking processes of test-takers. Accordingly, employing an immediate retrospective verbal report procedure, a study was conducted among Chinese EFL test-takers. The test format investigated was multiple-choice. The results identified the subjects' listening processes leading to comprehension and comprehension breakdowns and captured the effect of the multiple-choice format on the subjects' test performance. The immediate retrospection research methodology as applied to the test of listening comprehension was explored.

I Introduction

A language test-taker's performance is a function of two constant variables: his or her language ability and the test method.[1] A test is thus referred to as a 'trait-method unit' (Cambell and Fiske, 1958: 81), a union of the particular trait to be measured with a particular measurement procedure (Messick, 1988; cf. Dunkel *et al.*, 1993). To explore how this trait-method unit works, it is necessary to examine the test-taking processes of testees. But methodological difficulties with examining the thought processes of subjects taking a test impose such constraints that we are still far from knowing the answers. Applying an immediate retrospection verbal report procedure, the present study was designed for two broad purposes: to investigate the test-taking processes of EFL testees in a listening comprehension test using multiple-choice (MC) format; and to evaluate

---

Address for correspondence: Wu Yi'an, Department of English, Beijing Foreign Studies University, Beijing 100081, China; e-mail: yianwubw@public.bta.net.cn
[1]Variables that can potentially affect test performance also include such personal attributes as the testee's cognitive style and random factors such as the testee's emotional state at the time of taking a test. For an overview of these factors and how they differ from the method variable, see Bachman (1990: 163ff.).

immediate retrospection as a research method for accessing such processes. Specific research questions will be formulated after a brief consideration of the nature of listening comprehension, the introspection methodology and relevant research findings.

II Theoretical assumptions

1 An information-processing perspective on listening comprehension

Leading researchers in the psychology of language and in artificial intelligence (e.g., Forster, 1979; Marslen-Wilson and Tyler, 1980; Anderson, 1983; 1990; Garnham, 1985; Aitchison, 1989) have generally agreed upon the postulation of a language understanding system (LUS), composed of several language processors and a general problem-solver (GPS).[2] The language processors have access to the lexicon and, on receiving linguistic stimuli, conduct lexical, syntactic and semantic processing respectively. The GPS has access to one's general knowledge and is responsible for decision-making in comprehension by collecting and analysing information from the subsystems.

This proposed comprehension system is useful in that it provides a plausible framework for us (1) to conceive of language processing as systematic information processing, and (2) to explore relations between different levels of processing.

Broadly, what the framework implies is that the function of the LUS involves the activation of two types of knowledge: linguistic and non-linguistic. For listening comprehension, linguistic knowledge consists of phonological, lexical, syntactic and semantic knowledge, which serves as 'abstract cues or instructions' for making sense of the input (Bransford and McCarrell, 1977: 389). Non-linguistic knowledge refers to general knowledge and beliefs about the world. It is generally postulated that, while linguistic knowledge is activated in the bottom–up manner of processing, constraining the accessing of general knowledge, general knowledge is used in a top–down manner, facilitating the interpretation of the incoming input (e.g., Clark and Clark, 1977; Garrod, 1986; Brown, 1990). Exploration into how the two types of knowledge are employed in second language listening comprehension has led to a body of empirical research, whose findings are, however, controversial (Rubin, 1994: 211). Among the few

---

[2]Differences, however, are found in what the proposed models posit as the processing mode of the subsystems. For an account of the serial model and the interactive model, see Forster (1979) and Marslen-Wilson and Tyler (1980) respectively. For parallel distributed processing, see McClelland *et al.* (1986) and Qinlan (1991). As will be seen, these differences hardly concern the present study.

theoretical postulations broadly addressing the question, the more plausible seem to be Stanovich's 'interactive-compensatory model' (1980) which claims that deficiency at any level of processing can be compensated for by any other level of processing (in Brindley and Nunan, 1992), and Brown and Yule's principles of 'local interpretation' and 'analogy' (1983: 60ff.; cf. 234ff.).

From an information-processing point of view, comprehension is subject to limitations of human memory capacity. In language comprehension, human working memory performs two functions: storage of information for later retrieval, and processing (Just and Carpenter, 1987; 1992). When the task demands are high, as in a test of listening comprehension, often because of both storage and processing needs, the computation will slow down and some partial results from working memory processing may be forgotten (1992: 122–24). This can perhaps account for the fact that EFL listeners often seem to be able to hear everything but either forget what they hear easily or cannot process what they hear into meaning relationships.

In terms of processing, a distinction is drawn between automatic and controlled processes (Anderson, 1983; Bialystok, 1990). While the former occur rapidly and with little effort, the activation of the latter requires the attentional control of the listener. Insufficient automaticity in task performance invariably makes the subject's attention the controlling factor (McLaughlin *et al.*, 1983; Nagle and Sanders, 1986). In an EFL test-taking situation, a large proportion of the available resources in working memory is taken up when the testees struggle with language difficulties. The result is slowing down the processing, which causes forgetting some partial processing results, or simply leaves no time for processing meaning.

It is assumed that listening and reading comprehension share much the same processes and that L2 listening comprehension does not differ in kind from L1 listening comprehension (cf. de Jong, 1984; de Jong and Glas, 1987). Comprehension breakdowns can occur due to lack of any of the knowledge sources, and/or failure in any of the levels of processing, and/or lack of adequate automaticity in processing.

2 Immediate retrospection

Recent developments in cognitive psychology see a revival of the introspection method, particularly in the form of concurrent and immediate retrospective verbal reports, as a valid means to obtain data on individuals' thought processes in task performance (Faerch and Kasper, 1987; Ericsson and Simon, 1993).

The basic theoretical assumptions underlying the processing model

for verbal reports are that 'human cognition is information pro-cessing' (Ericsson and Simon, 1993: 11) and that a cognitive process can be seen as a sequence of states of heeded information or thoughts (Ericsson and Simon, 1987). The model assumes an information-pro-cessing system of three types of memories: several sensory stores, a short-term memory (STM) and a long-term memory (LTM), each with its own capacities and accessing features. Processing occurs when information is accessed through a co-ordination of these mem-ories at work. The sensory organs receive information, portions of which are recognized as it resides very briefly in sensory stores. The recognition process associates sensory stimuli with existing infor-mation in LTM,[3] which results in encoding. The encoded information can reach STM via two routes: through the recognition process or from LTM by the association process, the latter of which is more time-consuming than the former. What information finds its way to STM is controlled by the central processor (CP) and this is the infor-mation heeded, information available for verbal reporting (Ericsson and Simon, 1993).

Immediate retrospection can be expected to provide evidence about thought processes, since a subset of the sequence of thoughts occur-ring during performance of a task is stored in LTM in the form of a retrievable trace of connected episodic memories. After the com-pletion of the task, there remain in STM retrieval cues that allow effective retrieval of that subset of the sequence of thoughts. Retro-spective reporting involves retrieval of these episodic memories and verbalization of their content. Given a task duration between two and 10 seconds, little difference is predicted between concurrently and retrospectively reported information concerning the same cognitive process (1993: xvi, 149).

In response to challenges to the introspection methodology,[4] exten-sive studies have been conducted to investigate the effect of verbal reporting on one's thought processes and completeness of such reports. A general lack of effect on the subjects' thought processes is reported from studies comparing the performance of verbal report subjects to that of silent controls (e.g., Ericsson and Crutcher, 1991). Studies of effects from retrospective verbalization have yielded con-sistent results, though it is found that the type of instructions given for such reports is crucial for guaranteeing that no alteration of the

---

[3]There is reason to believe that the long-term memory is active and developing, and therefore constructive. See Bartlett (1932) for the original idea, Brown and Yule (1983) for a discussion of the idea in relation to natural discourse comprehension, and Dechert (1987) for empirical support of the position.

[4]A detailed review of major challenges and responses is documented in Chapters 2 and 3 of Ericsson and Simon (1993).

course and structure of the cognitive processes occur. Incompleteness of verbal reports is expected on two accounts. First, the automatic, intermediate stages of acts of recognition are not available in STM. What can be reported is the result of the recognition process. Second, owing to the limited capacity of STM, its transient contents get lost quickly, so that their retrieval is impossible. In interpreting verbal reports, inferences are typically made from what the individual reports as the input or/and output of his or her thought processes.

In light of the information-processing perspective of listening comprehension discussed above, it seems that much of the linguistic processing, being largely automatic, is not available to retrospection. What can be captured are the bit of linguistic input and the knowledge that it triggers and is retrieved from the LTM and how these work together to produce a response to a comprehension question. As such, verbal protocols are believed to be particularly illuminating concerning how the two broad types of knowledge work to produce comprehension and possible causes of comprehension breakdowns, linguistic as well as otherwise.

A further problem with retrospection concerns the temporal separation between processing and reporting, which can result in the subjects' tendency to elaborate and even include rationalizations and/or justifications generated after the actual thought processes have completed. Accordingly, a distinction is drawn by Ericsson and Simon between Level 1 or 2 verbalization[5] on the one hand and Level 3 verbalization on the other. The former consists of responses by the subject to the instruction to verbalize *thoughts*; the latter to instructions to verbalize *specific information*, for example, reasons and explanations. Level 3 verbalization is believed to induce changes to one's cognitive processes. To secure Level 1 or 2 verbalization, improved instructional procedures are called for, which help subjects focus on the task and thoughts directly leading to their task performance.

In conclusion, immediate retrospection can be expected to yield valid data of the thought processes of subjects in task performance, given that appropriate probing procedures are observed. Interpretation of such data, however, should take into account the levels of verbalization elicited and the inherent inadequacies of the data produced using retrospection.

---

[5]Level 2 verbalization is different from Level 1 verbalization only in that the former involves recoding of thought content. It follows that in language performance the distinction is often blurred.

III Relevant research findings and research questions

Research into listening comprehension is limited in scope, owing to its inaccessible nature. Existing research findings, though informative, cannot all claim to be valid and generalizable, often due to problems in design. Given limited space, what is reviewed below is only the research in comprehension that is most relevant to the present study.

1 Research on method effects

The effect of test method on testees' performance has been in the main investigated using the psychometric approach. In this approach, manipulations in testing methods are made and indication of the differences which such manipulations can make is observed in the testees' proficiency scores. Klein-Braley (1983), for example, found that different methods of deletion used in cloze tests of reading caused a difference in the test scores of advanced EFL students. Alderson (1983) and Bachman (1983), reported in Shohamy (1984) further explored the causal factors for the differences in the testees' scores and claimed that the manipulation of different methods in cloze testing resulted in a difference in the abilities to be measured, and hence the difficulty level and factor structures of the tests concerned. Shohamy (1983; 1984) expanded the scope of investigation to cover speaking through interviewing and reporting, and reading comprehension using the MC, open-ended questions, and summary testing formats. Her findings further confirmed earlier claims quoted above. While this research testifies to the effect of test method on test performance in general, thus contributing to our understanding of test construct validity, the product-oriented research methodology falls short of bringing to light *how* the methods employed work to produce an effect on test performance and therefore have little to offer to test design and the interpretation of test scores.

The mentalistic approach to test method has been little explored, especially with listening comprehension, for obvious reasons. Buck (1990; 1991), in his extensive research on testing listening comprehension, however, used the introspection method with six Japanese EFL testees, investigating, among other things, the effect of open-ended short-answer comprehension questions on the measurement of listening comprehension. The subjects' verbal protocols revealed the complex ways the test format influenced the measurement (1991: 85), which could not have been uncovered by the psychometric approach. His major findings include test unreliability caused by a shortage of time, response evaluation and implementation problems. The problem with this part of Buck's research seems to be the lack of a plausible

way to sift the data out of the long and verbose protocols from the point of view of his research purpose, so that not much is provided regarding how the test method itself actually influenced the subjects' processing. It appears that no research findings are yet available concerning the effect of the MC format on listening comprehension following the mentalist approach.

2 Research on the trait

Research into listening comprehension abilities has generally focused on (1) comprehension breakdowns due to deficiency in linguistic abilities, and (2) the interaction between linguistic and non-linguistic knowledge. Kelly (1991) had advanced French speakers of English transcribe and/or translate BBC radio news recordings and other listening comprehension texts and counted the number of errors made on perceptual, lexical and syntactical accounts respectively. By allowing the subjects to listen at their own pace and to relisten when necessary, Kelly was able to control the general knowledge variable and successfully isolate lexical ability as the major factor affecting the subjects' listening comprehension. This strength of the study, however, could readily be considered a weakness, because the experimental procedure rendered the task far removed from real-life listening wherein general knowledge is necessarily involved.

More research has been directed at how linguistic and non-linguistic knowledge works in second language listening (or reading), with inconsistent findings. Wolff (1987), O'Malley *et al.* (1989) and Lund (1991), applying the procedure of recall protocols, found that their subjects activated their general knowledge to rescue comprehension breakdowns due to difficulties in linguistic processing. On the other hand, Conrad (1985), using a cloze test, claimed that poorer listeners relied more on syntax than on contextual semantic cues. Along the same line, Bacon (1992, in Rubin, 1994) found that listeners resorted to bottom–up strategies on more difficult input. As Rubin observes, these studies indicate the complexities of listening processing. Moreover, they reveal the helplessness of indirect measurement for tapping such complex processing, and point to the fact that in the area of listening comprehension we are still rather ignorant.

3 The research questions

Given the theoretical positions taken for the study and the status of the field as briefly reviewed above, the study aimed to provide an answer to the following questions:

1) How do the subjects employ both linguistic and non-linguistic knowledge in performing on MC listening comprehension tasks?
2) What effect does the MC format have on the subjects' performance?

In addition, the research procedure will be reported systematically with a view to exploring and evaluating the immediate retrospection methodology.

IV Method

a Subjects: The subjects were 10 Chinese volunteers, selected from 18 recruited in Cambridge University according to the following three criteria:

1) They should have acquired intermediate-level proficiency or above in listening.
2) A range of levels was sought, for comparing the listening processes of subjects of different levels in proficiency.
3) The subjects should be motivated and able to talk in the way required in the verbal reporting procedure.

A pre-test was administered and informal conversations followed each of the pre-test sessions. Ten subjects were then selected. Three native English speakers, all Cambridge University students, were used for checking the suitability of the test for the study.

b Materials: The test consisted of a 3.5-minute recorded radio interview (see Appendix) and six MC questions, chosen for two reasons. First, MC has been the most widely used test format in EFL testing. Second, with five out of the six items asking 'why' questions, all levels of processing would be required, which was conducive to observing how processing linguistic and non-linguistic knowledge works interactively to produce comprehension.

c Procedure: The written responses of 74 Singaporean Chinese to the test were examined, to facilitate predicting the behaviour of the experimental subjects and the design of the interview procedure for the study.

Two pilot interviews were conducted, using a free interview format, to explore the methodology and gain experience. It was found that creation of a non-threatening atmosphere was essential in the experiment proper.

A pilot study was then conducted among four of the Chinese subjects. Based on this experience, the following procedures were

observed in the main investigation which involved the other four subjects:

1) Probing, which consisted mostly of 'what', instead of 'why', questions, started immediately after the subjects' initial retrospection was completed for each task.
2) As in a real test, the subjects listened to the input twice: first to the whole interview while completing the MC tasks; then to the interview section by section, which was immediately followed by retrospection.

It was decided that the results and discussion of the study would be based on the protocols of the four subjects in the main investigation.

V Results

An examination of the protocols showed that two types of verbalization could be identified. The first type (I) consisted of Level 1 and Level 2 verbalizations in Ericsson and Simon's (1993) terms, which allowed straightforward inferences of the subjects' thought processes. This type of verbalization was typically found at the beginning of each section of the subjects' reports, and sometimes following the researcher's probing, which triggered cues that enabled the subjects to have access to contents of their LTM. Verbalizations of the latter kind were often preceded by such linguistic markers as '. . . first I wanted . . .', '. . . actually I firstly', '. . . in the first time, I looked [at] . . .' The second type (II) consisted of Level 3 verbalizations, which were thoughts generated or elaborated and which often resulted from instructions for the subjects to explain their thoughts. As will be seen, Level 3 verbalizations were especially informative about possible causes of the subjects' comprehension breakdowns. Each excerpt from the protocols below is labelled I or II to indicate the type of verbalization involved.

In the following presentation of the results, the question and the MC options for each item are first given.[6] A table reporting the subjects' answers follows, wherein their changed answers are in parentheses. Following each table are the selected, most relevant parts of the subjects' retrospection protocols. Observations are provided to supplement the selected data as an initial discussion of the results on an item basis. Excerpts from the protocols are in italics. Strongly stressed words are underlined. Words taken from the input or the questions are in quotation marks.

---

[6]For ease of reading and interpretation, the selected excerpts have been slightly edited, merely leaving out redundancies, without altering the essential meaning.

Item 1: Andy believed that people liked their music because
    A. it made them feel relaxed.
  *B. it was what they wanted at the time.
    C. it was in a traditional style.
    D. it reflected their unhappiness.

Table 1 Subjects' responses for Item 1

Subject	AP	BL	CQ	DG
Response	B	B	D	D(B)

Data for the correct answer:[7]
  (1) AP: *they said something like that (B) and – because what they sang is what the people like . . . it was what they wanted at that time . . . actually – firstly – when I – look through these questions I thought – might be B (I) but I have to get some evidence from their talk to say OK it must not be – A, C, and D – so it must be B* (II)
  Observation: AP verbalized the message (output) he got. Previewing the question and the options helped in the sense that it guided him regarding what to listen to and what to expect. His knowledge about pop music helped him predict the answer.
  (2) BL: *they describe – a lot of theory about music – like exciting – so that is they want to get* (I)
  (following probing) *because they said – the music exciting so – they do not feel relax I don't think it's relax – exciting not means relax – and + traditional music is – not exciting so I think it's not this one (C) + + I can't explain this one (D)* (II)
  Observation: BL chose the right answer for the wrong reason. It seemed that she heard the prosodically salient 'exciting' and merely guessed B, using her knowledge about pop music. She also used her faulty word knowledge, that 'exciting' couldn't mean 'relax', and her belief about music, that 'traditional music' was not 'exciting', which led her to reject A and C.

Data for the incorrect answer:
  (3) CQ: *when people say – the music is – exciting so I think it maybe – reflect their – unhappiness – people's unhappiness* (I)
  (following probing) *I'm not very listen clearly – what they say – because I do not get used to this style – before I practised TOEFL –*

---

[7]Unlike English, Chinese does not have a tense system, and acquiring the tense system poses a great problem for Chinese EFL learners. This is reflected in the protocols wherein many present tense verb forms are used when the past tense forms would be appropriate.

only one man and one woman – this time – all are men so that –
sometimes I cannot – identify – who is Andy and Sam who is the
reporter (II)
Observation: CQ didn't get the message from the input. He picked
up 'exciting' and merely guessed.
(4) DG: *D – he + he didn't say anything about that – but + +*
something very nearly the D – and first I want to choose that (I)
I can tell you it's not make people feel relaxing (II)
oh yes it's the right time + that one (pointing to B) + *because the*
music become the right time at the right moment (II)
Observation: DG was not sure what to choose while listening even
though she ticked D. During the retrospection she was still trying
to figure out the right answer. Given more time now, B triggered
her memory of a crucial utterance from the input and it was then
that she decided on B.

Item 2: According to Sam, they became successful because
    A. they went to another country.
    B. they changed their style of clothes.
    C. they had an honest attitude.
  *D. they did what their manager told them to do.

Table 2 Subjects' responses for Item 2

Subject	AP	BL	CQ	DG
Response	D	(D)B	D	(B)

Data for the correct answer:
(5) AP: *this definitely D because he said they – just follow the*
manager's idea (I)
(following probing) *the first time I – looked – the four answers*
and I tried to heard the country – or because they – wear some
special kind of clothes but they didn't change their clothes quite
often – that means B is not the answer – and A is not the answer
and 'they had an honest attitude' is not relevant – because – it's
nothing to do with attitude or something . . . so that (D) would be
definite (I)
(6) CQ: *they say something about what they do and then they say*
this is what manager ask me to do (I)
(following probing) *the next choice maybe C because – they say –*
the audience – appreciate – their – attitude . . . but I'm not very
sure (II)
Observation: It seems that had CQ understood 'rebellious' in 'we

projected a sort of "rebellious" attitude', he would not have considered C.

Data for the incorrect answers:

(7) BL: (chose D first) *I just listen manager's idea I think oh they followed manager's idea . . . this time* (second listening) *I hear without manager idea* (I)

I think they describe different people have different style different attitude . . . so from this I think they always change style of clothes (I)

Observation: BL must have taken 'was our manager's idea' in the text for 'without manager's idea', which suggested problems in segmenting words from sound streams. Her decision on B was likely the result of guessing, based on the salient words she picked up, such as 'different' and 'style', with which she matched B. Her belief about pop music is believed to have a role in her decision-making.

(8) DG: (didn't tick any while listening) *it's difficult to say + + I didn't heard – very clear about that they give something + + which young people like – like the clothes and things + so maybe I'll get B in the examination* (II)

(following probing) *they say something like the manager told them to do . . . I'm not sure* (II)

Observation: It appeared that what DG picked up was not adequate for her to process into a message related to the question and she probably tried to guess the answer. She reported hearing B, but it appeared that B was more congruent with her belief about pop music.

Item 3: Why does Andy think they have been successful for so long?

   *A. Because they haven't changed their style.
   B. Because they use a lot of new technology.
   C. Because fashions in music have not changed.
   D. Because their friends and families have supported them.

Table 3   Subjects' responses for Item 3

Subject	AP	BL	CQ	DG
Response	A(D)	C	(C)B	C(B)

Data for the correct answer:

(9) AP: *he said – some people change their style – whenever there are some new fashion come in – so they just keep changing but – they* (Andy and Sam) *didn't change anything and they just – keep*

their style – but he also said he got some friend and family support but that (A) might be – the main reason (I) . . . *but for the pop music – singer I think they should change their style quite often because . . . otherwise you're – pass some kind of period* (II)
(following probing) *but I have to listen to that* (base my answer on the input – the researcher's interpretation) (II)
Observation: When his comprehension based on what he heard from the input was incompatible with his belief about pop music, AP wouldn't allow the latter to dominate his decision-making. He thought D was possible, probably because he picked up the more salient 'family' and 'friends' which activated a familiar script, but failed to process the utterance 'We're like family friends to them now'.

Data for the incorrect answers:
(10) BL: *I'm not sure I think maybe C – because they said – other group – fashions – always change – maybe they do not change fashion – they always give one – style always* (I)
(following probing) *actually I think C – is same as A – I think styles with fashions the same meaning* (II)
(11) CQ: *the first time I choose between A and C – because I heard something about – the style – but later I choose C . . . because – I think – they change the fashion and the second time I heard – change fashion is the style of other band – instead of theirs – they attribute their success – to they use – a lot of technology – so I choose B* (I)
(12) DG: *first I choose this one (C) because I heard – Andy said they're not change a lot . . . he said something about another group – they changed fashion . . . then he also said they used a lot of new – technology* (I)
(following probing) *I'll change* (my answer to B) *because – he didn't say – they hadn't change – it's not + it looks like – quite a 100% change or not change . . . it's not like this one (C) – haven't changed – I don't think that's true* (II)
Observation: The three subjects all heard that the band didn't change their style, but none of them even considered A, probably because A was incompatible with their knowledge about pop music. They appealed to C, partly because they mistook it to mean that the band's fashion has not changed, which seemed again counterintuitive. The protocols showed that when the subjects' linguistic knowledge was inadequate, they'd rather rely on their non-linguistic knowledge. In addition, they all had problems distinguishing between A and C.

Item 4: Why did Andy and Sam become friends?
    A. Because they both wrote music.
    B. Because they were neighbours.
   *C. Because they both played guitars.
    D. Because they had the same sense of humour.

Table 4   Subjects' responses for Item 4

Subject	AP	BL	CQ	DG
Response	C(B)	(B,C)	(B)A	B(A,C)

Data for the correct answer:
   (13) AP: (chose C while listening) *I'm not quite sure if it's B or C – they talked these two things – because they were neighbours – they knew each other – in the middle school or something . . . but they also talked about they got the guitar in the same year and played – just together* (I)
(following probing) *I think it might be C because . . . they first mention this (C)* (II)
   Observation: AP must have mistaken 'we met the others in the neighbourhood' for 'they were neighbours', which led him to suspect B was the answer. There was obviously an element of guessing in AP's choice of C.

Data for the incorrect answers:
   (14) BL: (couldn't make a decision while listening) *they said they met – in the school – they play guitar to – gather . . . they are very serious – not humour . . . Andy wrote tune but Sam wrote – I didn't know* ('lyrics') (I)
I don't think they are neighbour – they became neighbour later – because they met in the school – that means – they're not neighbour (II)
   (15) DG: (circled B while listening, still trying to figure out the answer while retrospecting) *he said Andy write the tunes and I + write* ('lyrics') *+ oh – and then they put together* (Not knowing 'lyrics' made her choose A) (I)
it didn't say anything about they're – neighbour – they say they play with the neighbourhood . . . I got it (B) wrong (II)
they both got the guitar at the birthday – then – it is the reason they become friends . . . I'll choose that (C) (II)
   Observation: DG and BL more or less heard quite a few discourse fragments from the input, but seemed undecided about which option to choose. This could be because, due to lack of automaticity in processing, the capacity of their STM was overloaded so that

processing the literal meaning into relationships was not possible. This perhaps explains why they found several options plausible. When more time was available during the retrospection, DG retrieved what she heard while viewing the options again and was able to get the correct answer.

(16) CQ: *first time I choose B – and later I think – they did not say they are neighbour they only say – they know each other when they are in the school – it doesn't mean they are neighbour* (I)
I think it should be – A because they both wrote music (I)
(when asked if he heard that) *because I cannot find answer to other three* (II)
(when asked if they both played guitars) *I did not heard – such thing* (II)
Observation: Not knowing 'lyric' probably led CQ to consider A.

Item 5: What procedure do they follow for writing songs?
   *A. There is no fixed order.
   B. Sam writes the music.
   C. They both write words and music.
   D. Andy always has the first idea.

Table 5 Subjects' responses for Item 5

Subject	AP	BL	CQ	DG
Response	A	A	A	A

Data for the correct answer:
   (17) AP: *'it depends' on something – that means they don't have any – rules* (I)
   (18) BL: *he use a lot of 'sometimes' – so no fixed order I think* (I)
   (19) CQ: *the first sentence he say is 'it depends' – that means no fixed order – it varies from time to time – and – later they say sometimes this way and sometimes that way that means there's no fixed order* (I)
   (20) DG: *it just say 'sometimes' – normally – Andy write – the music and Sam write the word – but sometimes it's not – sometimes they – do another way* (I)
   Observation: They all picked up relevant, salient linguistic cues and inferred the correct answer. However, although DG chose the right answer, her understanding of 'fixed order' was incorrect.

Item 6: They have stayed together because
   *A. they tolerate each other.
   B. they never argue.
   C. they both like touring.

D.they owe each other money.

Table 6 Subjects' responses for Item 6

Subject	AP	BL	CQ	DG
Response	A	D	A	D

Data for the correct answer:
  (21) AP: *merely they said they can solve the problem together* (I)
  (22) CQ: *if – there's something wrong then they found some way to solve it – that means they can – tolerate each other* (I)

Data for the incorrect answer:
  (23) BL: *B is not right because they have a big row sometimes + they said they need a lot of money* (I)
  (when asked if they owe each other money) *no – I think there's a sentence – I'm not sure it's 'owe' – it's some word like this – this sentence is in er each other's pockets* (I)
  24 DG: *they spent a lot of time with each other's pocket – like – they owe each other's money* (I)
  Observation: It seemed that neither BL and DG knew 'tolerate' nor 'put up with', which probably led them to abandon A. Having obviously failed to process the last utterance of the input, they merely guessed their answer from the discourse fragment 'spend so much time in each other's pockets'.

VI Discussion

1 Research question 1: How did the subjects employ both linguistic and non-linguistic knowledge in performing an MC listening comprehension task?

The protocols clearly indicate that listening comprehension is essentially a process of making sense of the linguistic input in the light of relevant non-linguistic knowledge and the purpose of listening (cf. Brown and Yule, 1983; Brown, 1990). In an MC listening comprehension test, the questions and optional answers serve as the imposed and therefore shared purposes for listening and exert a constraining impact on the listeners' listening processes.

In their attempt to make sense of the input, the subjects invariably activated their non-linguistic knowledge through the linguistic cues abstracted from the input including the options. Excerpts (1)–(3), (7)–(12), (23) and (24) involve obvious cases of such activation. Two tendencies presented themeselves regarding how the two types of knowledge worked in an MC testing situation.

First, partial success in linguistic processing often forced the subjects to activate relevant general knowledge in the light of the questions and options concerned in order to make sense of what was heard, often the prosodically salient words (e.g., excerpts (2), (3) and (6)–(12)).[8] In such cases, non-linguistic knowledge was accessed, as it were, to 'compensate' for the subjects' linguistic deficiency in order to cope with the task (cf. Stanovich, 1980). It should be pointed out, however, that such compensation did not necessarily guarantee arriving at the correct answers.

Second, partial success in linguistic processing could also lead the listeners to allow their belief to override what was correctly abstracted through linguistic processing ((10) and Observation following (12)). On the other hand, competent linguistic processing made AP confidently reject the lure of his belief about pop music in his decision-making (9). Here again, access to general knowledge provided no guarantee for reaching the correct answer.

For EFL listeners, therefore, linguistic processing is basic in the sense that failure or partial success in it may result in their allowing the activated non-linguistic knowledge to improperly dominate their decision-making. Further, it may induce groundless guessing ((23) and (24)). Linguistic processing is basic also in the sense that competence in it constrains but does not preclude non-linguistic knowledge activation ((1) and (9)). A distinction can in fact be made between the compensatory and facilitating functions of non-linguistic processing. The present study shows that the former function is more associated with less successful linguistic processing and/or less able listeners, whereas the latter goes with more advanced listeners and/or more competent linguistic processing.

In the light of the above findings, earlier efforts to relate linguistic and non-linguistic processing to difficulty level of input (Wolff, 1987; Bacon, 1992) or to the proficiency levels of listeners (Conrad, 1985), or to link the amount of interaction between the two types of processing with communication breakdowns (O'Malley *et al.*, 1989), seem to be misdirected. It appears more plausible to conceive the listening comprehension process as an activation of linguistic and non-linguistic processes in a parallel manner for all learners with input of all difficulty levels (cf. Jensen and Hansen, 1995: 103). Differences, however, are found in processing modes or strategies. It is

---

[8]Partial success in linguistic processing refers to (a) hearing and successfully processing only salient words in a stream of discourse fragment (e.g., (2) and (3)), and (b) hearing many discourse fragments without successfully processing them into meaningful relationships (e.g., (14) and (15)), most probably owing to limited capacity of short-term memory and lack of automaticity in processing.

suspected that product-oriented research methodology, recall proto-
cols in particular, is responsible for the findings incompatible with
those of the present study.

The protocols also revealed failure in processing the linguistic
input, i.e., in segmented sounds into words ((7) and Observation) and
lexical ignorance/deficiency ((2), (3), (6), (14)–(16) and (24) and
Observation), as direct causes of comprehension breakdowns.

2 Research question 2: Did the MC format have an effect on the subjects' performance and, if so, what was the effect?

First, viewing the questions and options seemed to facilitate pro-
cessing for the more advanced listener AP in that it helped the subject
form anticipations of the incoming input and provide foci for listening
((1) and (5)), but not for the less able subjects.

Second, misinterpretation of the options seemed to have partially
caused the subjects to select incorrect answers. As is shown in (10)–
(12), the subjects obviously didn't know the difference in meaning
between *A and C of Item 3, which was partly accountable for their
selection of C. Further, unknown words in the options, 'tolerate' in
Item 6 for example, partially turned the part concerned into a test of
vocabulary or reading and therefore threatens the construct validity
of the test.

Third, the format allowed much uninformed guessing. Uninformed,
for the guessing was often a result of too much dependence on non-
linguistic knowledge activated through only partial linguistic pro-
cessing and through the lure of the options. Examples are found in
(2)–(4), (7)–(12), (15), (16), (23) and (24).

Fourth, uninformed guessing sometimes led to the subjects' selec-
tion of the right answer for the wrong reason, as is shown in (2)
and (20).

We may perhaps conclude that, while the MC format favours the
advanced listener, it adds difficulty for the less able listener, and that,
owing to its allowance for much uninformed guessing, the construct
validity of the test is left open to question.

3 Immediate retrospection

Immediate retrospection yielded verbal reports of the input and output
of the subjects' thought processes, from which valuable information
about how the subjects processed the input was obtained. However,
the quality of the verbal reports seemed to depend on the probing
procedures employed. The free interview format could easily produce
Type II verbalization data. Carefully planned probing procedures and

sensible interpretation of the data obtained would presuppose task analysis based on a sound model of listening comprehension and experience in probing. Timing in task design is also crucial for obtaining valid data. Selection of subjects for retrospection was important. Given similar educational background and shared first language, the subjects should be motivated to talk as required, and have little difficulty in expressing themselves in the language chosen for reporting. To get insight into how the linguistic and non-linguistic knowledge is at work in listening comprehension, it was found necessary to have both advanced and intermediate level listeners as subjects.

The research method seems to lend itself to the investigation of second language listening processing. This is because the listening task demands the listener to pay such great attention to the input that much of what one hears still remains in his or her memory during the retrospection. The retrospection task, moreover, forces the listener to monitor what is going on in his or her mind while listening. Owing to the demanding nature of the tasks and the limitation of the processing capacity of short-term memory, presumably much of the processing cannot be automatic and is therefore readily available for reporting. In addition, reporting what is going on necessarily includes information of what goes wrong. In this respect, immediate retrospection also yields data illuminating possible causes of comprehension breakdowns. As Brown (1995: 42) cogently remarks,

> We can learn rather little about the processes of comprehension when they flow comfortably . . . We have an opportunity of learning rather more where understanding is difficult to come by, where interpretation is only partially achieved, or where an attempt to communicate results in misunderstanding.

The fact that immediate retrospection worked in this study is not to be taken to suggest that the method is without its limitations; rather, it shows that informed employment of the method can help us access a certain part of listening processing more fruitfully than is possible with product-oriented methods. Indeed, the success or failure of a research method, and a research paradigm as well, depends much on the matching or mismatching between the research purpose and the attributes of the particular method/paradigm chosen (Larsen-Freeman and Long, 1991: 14).

VII Conclusion

Linguistic and non-linguistic knowledge was activated in a parallel manner for all the subjects with input of various levels of difficulty. Processing of linguistic knowledge was shown to be more basic, constraining the activation of non-linguistic knowledge. Non-linguistic

knowledge was activated to 'compensate' for linguistic deficiency or to facilitate linguistic processing, with the 'compensation' mode more associated with partial success in linguistic processing and the less able listeners, and the facilitating mode more with competent linguistic processing and more advanced listeners. 'Compensation', however, more often than not provided no guarantee for arriving at the correct answers.

The MC method posed threats to the construct validity of the test in two ways: it favoured the more advanced listener, but put the less able at a disadvantage; and it allowed much uninformed guessing and resulted in the subjects' giving the correct answers for the wrong reasons.

Immediate retrospection, employed properly, was shown to be a promising research methodology for accessing some of the essential EFL listening comprehension test-taking processes.

Acknowledgement

VIII References

Aitchison, J. 1989: *The articulate mammal: an introduction to psycholinguistics*, third edition. London: Unwin Hyman, 201–40.

Alderson, J.C. 1983: The cloze procedure and proficiency in English as a foreign language. In Oller, J.W., editor, *Issues in language testing research*, Rowley, MA: Newbury House.

Anderson, J.R. 1983: *The architecture of cognition*. Cambridge, MA: Harvard University Press, 261–305.

—— 1990: *Cognitive psychology and its implications*, third edition. New York: Freeman, 358–98.

Bachman, L. 1983: Performance on different item types in cloze passages with fixed ratio and random deletions. Paper read at the 18th TESOL Annual Convention, Toronto, Canada.

—— 1990: *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bacon, S.M. 1992: Authentic listening in Spanish: how learners adjust their strategies to the different input. *Hispania* 75, 29–43.

Bartlett, F.C. 1932: *Remembering: a study in experimental and social psychology*. Cambridge: Cambridge University Press.

Bialystok, E. 1990: *Communication strategies: A psychological analysis of second language use*. Oxford: Basil Blackwell, 116–38.

Bransford, J.D. and **McCarrell, N.S.** 1977: A sketch of a cognitive approach to comprehension: some thoughts about understanding what it means to comprehend. In Johnson-Laird, P.N. and Wason, P.C., editors, *Thinking: readings in cognitive science*, Cambridge: Cambridge University Press, 377–99.

Brindley, G. and **Nunan, D.** 1992: *Draft bandscales for listening-IELTS projects 1992: Project 1*. National Centre for English Language Teaching and Research.

Brown, G. 1990: *Listening to spoken English*, second edition. London: Longman.

—— 1995: *Speakers, listeners, and communication*. Cambridge: Cambridge University Press.

Brown, G. and **Yule, G.** 1983: *Discourse analysis*. Cambridge: Cambridge University Press.

Buck, G. 1990: *Testing second language listening comprehension*. Unpublished PhD dissertation, University of Lancaster, England.

—— 1991: The test of listening comprehension: an introspective study. *Language Testing* 8 (1), 67–91.

Cambell, D.T. and **Fiske, D.W.** 1959: Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56 (2), 81–105.

Clark, H.H. and **Clark, E.V.** 1977: *Psychology and language: an introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.

Conrad, L. 1985: Semantic versus syntactic cues in listening comprehension. *Studies in Second Language Acquisition* 7, 59–72.

de Jong, J.H.A.L. 1984: Listening: a single trait in first and second language learning. MS version.

de Jong, J.H.A.L. and **Glas, C.A.W.** 1987: Validation of listening comprehension tests. *Language Testing* 4 (2), 170–94.

Dechert, H.W. 1987: Analysing language processing through verbal protocols. In Faerch, C. and Kasper, G., editors, *Introspection in second language research*, Clevedon: Multilingual Matters Ltd, 96–112.

Dunkel, P., Henning, G. and **Chaudron, C.** 1993: The assessment of an L2 listening comprehension construct: a tentative model for test specification and development. *Modern Language Journal* 77, 180–91.

Ericsson, K.A. and **Crutcher, R.J.** 1991: Introspection and verbal reports on cognitive processes – two approaches to the study of thought processes: a response to Howe. *New Ideas in Psychology* 9, 57–71.

Ericsson, K.A. and **Simon, H.A.** 1987: Verbal reports on thinking. In Faerch, C. and Kasper, G., editors, *Introspection in second language research*. Clevedon: Multilingual Matters Ltd, 24–53.

—— 1993: *Protocol analysis*, second edition. Cambridge, MA: MIT Press.

Faerch, C. and **Kasper, G.** 1987: From product to process – introspective methods in second language research. In Faerch, C. and Kasper, G., editors, *Introspection in second language research*. Clevedon: Multilingual Matters Ltd, 5–23.

Forster, K.I. 1979: Levels of processing and the structure of the language processor. In Cooper, W.E. and Walker, E.C.T., editors, *Sentence processing: Psycholinguistic studies presented to Merrill Garrett.* Hillsdale, NJ: Lawrence Erlbaum, 27–85.

Garnham, A. 1985: *Psycholinguistics: central topics.* London: Routledge.

Garrod, S. 1986: Language comprehension in context: a psychological perspective. *Applied Linguistics* 7 (3), 226–38.

Jensen, C. and **Hansen, C.** 1995: The effect of prior knowledge on EAP listening-test performance. *Language Testing* 12 (1), 99–119.

Just, M.A. and **Carpenter, P.A.** 1987: *The psychology of reading and language comprehension.* London: Allyn and Bacon, Inc, 453–82.

—— 1992: A capacity theory of comprehension: individual differences in working memory. *Psychological Review* 99 (1), 122–49.

Kelly, P. 1991: Lexical ignorance: the main obstacle to listening comprehension with advanced foreign language learners. *International Review of Applied Linguistics* 29 (2), 135–49.

Klein-Braley, C. 1983: A cloze is a cloze is a question. In Oller, J.W., editor, *Issues in language testing research.* Rowley, MA: Newbury House.

Larsen-Freeman, D. and **Long, M.H.** 1991: *An introduction to second language acquisition research.* London: Longman.

Lund, R.J. 1991: A comparison of second language listening and reading comprehension. *Modern Language Journal* 75, 196–204.

Marslen-Wilson, W. and **Tyler, L.** 1980: The temporal structure of spoken language comprehension. *Cognition* 8, 1–72.

McClelland, J.L., Rumelhart, D.E. and **the PDP Research Group** 1986: *Parallel distributed processing: exploration in the microstructure of cognition*, Volume 2, *Psychological and biological models.* Cambridge, MA: MIT Press.

McLaughlin, B., Rossman, T. and **McLeod, B.** 1983: Second language learning: an information-processing perspective. *Language Learning* 33 (2), 135–58.

Messick, S.A. 1988: Validity. In Linn, R.L., editor, *Educational measurement*, third edition. New York: MacMillan, 3–8, 77–96.

Nagle, S.J. and **Sanders, S.L.** 1986: Comprehension theory and second language pedagogy. *TESOL Quarterly* 20 (1), 9–26.

O'Malley, J.M., Chamot, A. and **Kupper, L.** 1989: Listening comprehension strategies in second language acquisition. *Applied Linguistics* 10 (4), 418–37.

Qinlan, P. 1991: *Connectionism and psychology.* Hemel Hempstead, Hertfordshire: Harvester Wheatsheaf.

Rubin, J. 1994: A review of second language listening comprehension research. *Modern Language Journal* 78, 199–221.

Shohamy, E. 1983: The stability of oral proficiency assessment on the oral interview testing procedure. *Language Learning* 33, 527–40.

—— 1984: Does the testing method make a difference? The case of reading comprehension. *Language Testing* 1 (2), 147–70.

Stanovich, K. 1980: Towards an interactive compensatory model of individual differences in the development of reading fluency. *The Reading Teacher* 33.

Wolff, D. 1987: Some assumptions about second language text comprehension. *Studies in Second Language Acquisition* 9, 307–26.

Appendix: Tape-script

Interviewer:	My guests today are Andy and Sam, who as everybody knows are members of 'The Heaters', a band that has been famous throughout the world for over twenty years. I'd like to start by asking you both what you attribute your initial success to.
Andy:	Well, it's . . . er . . . difficult to tell. I suppose luck played a large part. You could say, I guess, that we were in the right place at the right time. When we started, the public were, well . . . sort of . . . fed up with the sort of music that was going on. It was all a bit dull, . . . rather bland. What they wanted was something more exciting, something that could get them going and, well . . . we provided that.
Sam:	I think another thing is that we had a very clear image that people could easily get hold of. I mean, we had our own style of clothes and we projected a sort of . . . mmm . . . 'rebellious' attitude that young people could really identify with. To be honest that wasn't really us, it was our manager's idea, but we went along with it and, well . . . it really worked. It made us new and different and I guess that's why we took off.
I:	So why do you think this success has lasted so long? I mean, we all know that groups come and go, they're here today and gone tomorrow, so why do you think you've outlasted the others?
A:	That's a difficult one. I suppose you could say that our audience has grown with us. We haven't really changed, we've just grown older and so have they. I mean, other groups try to follow fashion, to copy anything that comes in, but when that fashion goes out, they go out with it. We've stayed faithful to our roots and although we obviously use the new technology available in music, we've basically stuck to our original style. I think the fans appreciate that. We're like family friends to them now, you could say (*he laughs*).
I:	Another thing I'd like to ask you is, how did you meet and how did the band get started?
S:	Well, Andy and I met when we were at school together. We both got our first guitars for our 14th birthdays and then we started hanging around together. To start with we were just messing around but then we started to take it more seriously. Andy was always writing tunes and I was writing lyrics, so we just sort of . . . put them together. We met the others in the neighbourhood and asked them to join us, and since they weren't

I: doing anything else, they did. In the early days we rehearsed in my parents' garage, and that's how it all started.

I: How do you write your songs?

A: Well, it depends, but certainly I always write the music and Sam always writes the words. But, the . . . sometimes I have an idea for a tune first and Sam puts lyrics to them and sometimes it's the other way round. And then sometimes the two things just sort of happen together. I guess you could call it, . . . er, inspirations (*he laughs*).

I: Has it been difficult for you to stay together? I mean, groups are always splitting up. Have you ever come near to that or have you just always got on well?

S::   Well, . . . um . . . obviously we've had our rows, sometimes big ones. It's inevitable when you spend so much time in each other's pockets, especially when you're on tour, which is . . . you know . . . tough. But, . . . we put up with each other most of the time. And if there is a big problem we can usually find a way of sorting it out.

A: (*laughing*) Yeah, we have to. We need the money.