TS/05/04

Choose a passage which has not previously been used in a reading comprehension test. Design a range of questions to test comprehension of it by a group of learners of your choice. Justify your choice of passage and the questions you have set in both theoretical and practical terms.

1.   INTRODUCTION

Teacher designed tests are a common feature of most EFL classrooms, and are one way of assessing learners' language abilities. Designing a test for a specific group of learners may be 'a matter of problem solving, with every teaching situation setting a different testing problem (Hughes 1989: ix)', but nevertheless it is important to be clear about the purpose for which a specific test is to be used and to make sure the test designed is appropriate for that purpose. A test of reading comprehension can be seen as testing a particular skill (reading, as opposed to writing, listening or speaking), but such a test could also be designed as a collection of tasks reflecting activities normally performed outside the testing situation, thus in theory enabling the tester to 'demonstrate how test performance corresponds to non-test language use (Bachman and Palmer 1996: 78).' From this perspective, consideration of the authenticity of text and task become crucial.

This essay explores the development of a test of reading comprehension for use with pre-intermediate learners. The first section looks at general factors to be taken into account when designing language tests. The next section describes the stages I went through in designing a reading comprehension test for use with a particular group of learners. The essay then describes the process of question writing and discusses modifications made as a result of piloting the test, and the last section evaluates how effective the final version of the test was seen to be. Although the final questions were in Finnish, for reasons of clarity all test questions referred to in the context of this essay are in English.

2.   DESIGNING TESTS

This section will briefly consider some of the factors which a test designer should take into account when designing a language test. These considerations apply to classroom tests as well as tests to be taken by large numbers of students. Bachman and Palmer (1996: 164) point out that 'design, operationalization and administration need to be carried out for every test…develop[ed]' but the difference is in the amount of detail and resources involved.

2.1. Testing for a Purpose

'[T]he primary purpose of tests is to measure (Bachman and Palmer 1996:19)', but it is the purpose of the measurement that should determine the focus of any specific test. When using tests to obtain information about learners, four main purposes can be identified (Hughes 1989: 9-14, Bachman and Palmer 1996: 97-98), as shown in table 1. It is important to ensure that the test being designed is suited to its purpose, as inferences about language ability, and possibly far-reaching decisions about a candidate's future may be based on the results.

Type of Test	Purpose
1. Proficiency Test	To assess general ability in a second language.
2. Achievement Test	To evaluate how much a learner knows from a defined amount of course or class work
3. Diagnostic Test	To identify a student's strengths or weaknesses in specific areas of language.
4. Placement Tests	To determine which would be the most appropriate class, stream or level in which to place a student so that subsequent language teaching is appropriate to their needs.

Table 1: Types of Language Test and their Purpose

2.2 Specifications

In order to assess if a test measures what it intends to measure (that is, it has construct validity) a set of specifications should be written as part of the overall test design. These include information about content, format, timing, criterial levels of performance (the required level of performance for success) and scoring procedures (Hughes 1989: 49-51). Lynch and Davidson (1994: 732) in their 'criterion-referenced test development' approach emphasize the importance of the specification format as a 'flexible tool that test developers can reshape to respond to specific testing requirements.' They highlight the importance of what they call the mandate (the reason for the test) for writing test specifications and assert that the test specifications

together with the actual task-writing is an effective way of 'clarifying the criterion being tested (ibid: 730)'. Brown (1994: 387) observes that for classroom tests a specification can be 'a simple and practical **outline** of [the] test' (original emphasis), derived from the test objectives.

2.3   Authenticity

The term authenticity, as used in the context of testing, can be understood to mean the degree to which a given task and set of materials corresponds to 'real life' tasks and interactions. It is one of the test qualities used by Bachman and Palmer (1996: 23-5) in their model of usefulness, and they suggest (ibid: 19) that test tasks which 'provide higher degrees of authenticity…' may be of particular interest to teachers. Authenticity of text and task at the very least can be seen to add face validity to a test. Spence-Brown (2001: 465), among others (for example Bachman and Palmer 1996, Kirschner et al 1996), sees authenticity in testing as having a much wider application, and includes within the paradigm of authenticity the interaction of the test takers with the task and also assessment criteria and procedures. Hughes (1989: 15) adds a reminder that however authentic a given test is designed to be, first and foremost for the candidate it will be a test. Nevertheless authenticity remains an important parameter in test design.

2.4   Pre-testing

Pre-testing is a recognised part of test development and facilitates the gathering of information for any necessary revisions, both to the test and its administration. Even with the rigorous development procedures for a global test such as the TOEFL (Test of English as a Foreign Language) reading test described by Peirce (1992: 669-673), pre-testing is included as an integral part of the process (ibid: 677-680). Brown (1994: 389) observes that in the classroom situation trialling a test is seldom a realistic possibility, but recommends a careful final edit as a substitute procedure.

Having briefly examined some of the parameters affecting language test design in general, the next section looks at the process of developing a test of reading comprehension for use in the classroom.

3. DESIGNING A READING COMPREHENSION TEST

The reading test to be discussed was developed as part of the assessment of a course module concerned with the topic of travelling. The test can be considered an achievement test in so far as it is an evaluation for a specific module, but it could also be considered a proficiency test, as it measures a learner's ability to gain information from an authentic text (that is, a text originally written for a purpose other than language teaching) through reading.

3.1 The Context: the Learners

The learners that I work with attend a special educational institute in Finland which provides vocational upper secondary and adult education. Most of the students are in the age range 16-22, and have special needs, for example, difficulties with reading and writing. As a consequence many students seem to be outwardly unmotivated for further English studies, having already had six years of compulsory English in primary and secondary school. Nevertheless, they are required to take 3 credit units of English during their three year vocational course. The classes are small and heterogeneous, but a sizable majority of students could be considered as falling within the pre-intermediate range.

3.2 Considerations in Text Selection

The first thing to do when developing a test of reading comprehension is to decide what the test is going to measure. Spolsky (1985: 181) observes that 'how we go about measuring something is dependent on what we think we are measuring.' Hughes (1989: 116-117) discusses both macro-skills (for example scanning a text to locate specific information) and micro-skills (for example using context to guess the meaning of unfamiliar words) as having relevance for the assessment of reading ability. He concludes that ability to demonstrate mastery of the macro-skills also implies mastery of the micro-skills. I was interested in developing a test to measure how well my students would be able to access information from an authentic text using the scaffolding of the test questions, and therefore I felt the integrative approach to testing reading comprehension used in this assessment was appropriate.

The text chosen for the test was from the tourist information magazine *Time Out 2005/6: London for Visitors*. The parts of the magazine used were three pages from

the *essential information* section from the end of the magazine: *Getting Around, Resources* and *Emergencies* (the full text can be seen in Appendix 1). The text and this particular part of it were chosen as I thought it representative of the type of information my students may come across when on holiday outside Finland, and containing the type of information they might need. The main purpose of *Time Out* is to give readers information, and it is written so that information on a particular topic is located under headings and subheadings. For example, under the heading *public transport* information can be found about the following: travelcards, the London underground, buses, rail services and water transport, all of which are clearly indicated through the use of a bold type face.

Hughes' advice on text selection (1989: 119) includes using passages which contain 'plenty of discreet information' if scanning is to be tested. He also suggests considering giving candidates 'a good number of fresh starts' by using a number of passages. I considered the passages selected from *Time Out* to give plenty of scope for scanning questions, detailed reading and fresh starts, while maintaining the central theme of travelling. This kind of text also seems to give the opportunity for testing reading by what Jafapur (1985: 197) writes about as 'short context technique', which he claims measures reading skill rather than anything else, and taps 'relevant real-world reading behaviours (ibid: 205).' The latter claim seems to equate with task authenticity, which was one of the reasons for choosing this text.

3.3   Aspects of Text and Task Authenticity

The degree of resemblance between a passage used in a test and the original text from which it was taken has implications for claims of text authenticity within the context of a test. Hill and Parry (1994: 257) highlight the fact that many authentic texts used in reading comprehension tests are not used in facsimile form, denying readers cues such as type face and format. Although the text used in the reading comprehension test in its final form amounted to three pages, I felt this was justified as the form of the text is as the original, and this is what the students would have to tackle in the real world. To make the reading more manageable the test was divided into sections, so the students only had one page at a time from which to find the answers, and as discussed more fully in section four, the questions provided the direction from which

to scan for the information required. The students therefore only needed to read a small amount of text to find the answers.

Task authenticity in a reading comprehension test also relates to the language that the test questions are posed in. Hughes (1989: 129) makes the point that the questions should always be less demanding than the text itself. Kirschner et al (1996: 88) and Hill and Parry (1994: 253) advocate the questions being written in the test takers' own language where possible when this mirrors what they would be doing in real life. Initially I had decided to set the questions to the reading comprehension test in English because this was more practical for me, but during the development process I realised Finnish was more appropriate as this would mean the test would then only be assessing students' understanding of the text and it would also reflect how the students might approach such a task in 'the real world' if required to do so. An additional consideration was the practical one of students accessing and staying with the test and not giving up.

3.4   Specification

Writing a specification for a reading comprehension test can focus the task of choosing an appropriate text or texts, and according to Lynch and Davidson (1994: 732) is critical in task development; in this case writing the questions in order to test reading comprehension. The specification for the reading comprehension test under discussion was used for developing the test questions and the final version can be seen in table 2. As mentioned by Lynch and Davidson (1994: 730) writing the questions also '[feeds] back to the elaboration of the test specification', and they suggest that the test specification 'also provides a detailed record of evidence for judging how well the test items…match what the test claims to be measuring.'

Specification for a reading comprehension test

- The purpose of this test is to assess the ability of a pre-intermediate learner to obtain accurate information from an authentic written text. It can be important for students to be able to access relevant information from an informational text written for native speakers, using the techniques of skimming and scanning.

- The text and the questions will relate in some way to travelling.

- Both the text and the tasks will be authentic, that is, replicating as closely as possible what a learner may be expected to do in the 'real world'. In Bachman and Palmer's terms (1996: 18) 'target language use' (TLU) domain or tasks.

- There will be instructions at the beginning of the test in Finnish to explain the purpose of the test, and how the student should go about doing the test.

- The reading comprehension questions will be in Finnish to reflect TLU.

- The questions will either be multiple-choice or require a word or words for the answer which can be found within the text. The learner is required to select the correct response or provide the appropriate word or words from the text as an answer.

- The test will last no longer than 45 minutes and allow time for slower candidates to complete within this time.

- The scoring will be one or two points for each correct answer, depending on the amount of information required, and the question format.

Table 2: Final Specification for a Test of Reading Comprehension

4.   THE PROCESS OF QUESTION WRITING

Once the text was selected, the next stage was to write questions which reflected the test design. As mentioned above, it became an iterative process whereby the test specification was also modified as a result of piloting the draft items. This section begins by presenting the original version of the test questions, and then looks at how the questions and the format were modified in response to observations and comments received when piloting the test.

4.1   The First Draft

Initially the whole test comprised four sections (one to a page) with the first section conceived as an orienting, but also authentic task. The contents page of the *essential*

information of an earlier issue of *Time Out* was chosen (the contents page in the 2005/6 issue was significantly smaller and more general) and the instructions directed the test takers to choose the heading they would look under to find certain information, with the first item presented as an example. The other three sections in the original draft related to three separate pages in *Time Out 2005/6*. The contents page can be seen in appendix 2 and the original questions from all four sections in table 3 below.

CONTENTS

Which section would you look under if you wanted to find:

What?	Where?
Example: a map of the underground	Books and maps
1. somewhere to stay	
2. an internet café	
3. the Finnish embassy	
4. a dentist	
5. a place to leave your luggage while you spend the day in London	
6. the place to contact to try and find the parcel you left on a train in London	
7. a church to go to on Sunday	
8. a holiday job in London	

GETTING AROUND

1. Which is the nearest airport to central London: Gatwick or Heathrow

2. Where can you find information on public transport in London?

3. Where can you buy a travel card in London?

4. If you were going to London for the day with an adult friend and two children under the age of 10, which travel card would you buy?

5. Why?

6. Which section would you look under to find out how much it costs to travel on the London Underground?

7. How much would it cost you to hire a bike for a day?

8. If you hired a bike you would need to pay £100 deposit. Explain what you think **a deposit** is.

RESOURCES

1. Where could you go to send an e-mail home if you were visiting London for the day?

2. What are the normal opening times for post offices in Britain?

3. Where would you be expected to tip in Britain?

4. How much should you tip?

EMERGENCIES

1. If you call 999 (or 112 from a mobile phone) in an emergency, what will the operator ask you?

2. Where could you go to get emergency dental treatment?

3. When is the emergency dental care service open?

4. What time should you arrive to make sure you get treatment the same day?

5. If you needed dental treatment and were going by public transport, where should you go to?

Table 3: Original Questions for the Reading Comprehension Test

When designing questions for the students to answer I focused on the text pages and tried to consider what kind of information they would realistically need to look up if visiting London, and therefore what sort of questions they would be asking themselves. The first question was deliberately written as an 'easy question', which I expected almost all the test takers to get right, so that they would have the confidence to realise they could extract information even from a text which at first sight could appear rather daunting. To be able to answer this question test takers would have to find a particular heading and sub-heading, and compare two short sections to extract the appropriate information.

Which is the nearest airport to central London: Gatwick or Heathrow?

Most of the other questions in the first draft were short answer questions, but a choice of possible answers was not provided. The only guidance the test taker had was in the question itself. The questions were almost all asking the test takers to find specific information, and there were between one and four questions per item of information. This was felt to be realistic, in that often one particular item of information is required, such as a phone number, but on other occasions several pieces of linked information could be needed, for example a place together with opening times and information as to how to get there.

Where could you go to get emergency dental treatment?

When is the emergency dental care service open?

What time should you arrive to make sure you get treatment the same day?

If you needed dental treatment and were going by public transport, where should you go to?

4.2   Pre-testing

The first draft went through two stages of revision, based on pre-testing. The initial version as described above, with questions in English, was given to three family members and a colleague, all of whom are fluent in English. From the first pre-test several things emerged that needed changing. These are outlined below:

- The length of the test: in its original version there were too many questions. My colleague reported difficulty keeping concentration,

and this would probably be even more of a problem for students, whatever their motivation.

- <u>Unbalanced sections:</u> the sections contained different numbers of questions; it would be a more balanced test if there were the same number of questions for each section.
- <u>Unexpected answers:</u> some of the questions were interpreted differently than was expected by the test writer as shown by the answers given, and this indicated revision might be in order.
- <u>The scoring key:</u> this needed revising because of the variation in the answers.

As a result of the above a thorough look at the questions was undertaken, and all the questions in which the wording was ambiguous or less than clear were changed or withdrawn. Examples of the modifications and changes that were made are presented in the next sub-section.

A second version of the test, still comprising four sections was then trialled with three students, all at roughly pre-intermediate level, male and aged 17. In addition to the changes made in the questions themselves, the questions were now given in both English and Finnish. From this pre-test further changes seemed to be indicated in the following areas:

- <u>Language of the questions:</u> using both languages was too cumbersome, and the students found it irritating to search through the 'amount of question material' to find the question in the language they felt comfortable with.
- <u>The contents section:</u> this appeared to cause confusion as the students found it difficult and did not really understand the point of it. Apparent problems with this section seemed to affect their attitude to the rest of the test.

As a result of the second pre-test, the contents section was dropped completely, as it was apparent that it didn't perform the orienting function anticipated, and possibly reduced performance on the rest of the test, as well as adding to the length of the test.

The language for the instructions and questions was also changed to Finnish for reasons of practicality and authenticity. Some of the other modifications made are discussed in the next sub-section.

4.3   Developing the Questions

Kirschner et al (1996: 89) express clearly the obligations a test writer has when developing a test. 'It is the test writer's task to define, identify and subsequently remove any potential difficulties inherent in the test questions.' This implies the importance of pre-testing in the process of test development. It may also mean checking the test specifications and considering how well the tasks (questions) are reflective of the specifications. The specifications themselves may also need to be reconsidered.

As pointed out above it became apparent the test was too long, and consequently the number of questions needed to be reduced. As the minimum number of questions for a section in the original test was four, I decided to reduce the number of questions to four for all three sections. This meant in practical terms that each test taker would have three pages of authentic text clipped together, and three pages with questions separately clipped together, individual pages of text and questions corresponding. I thought this would give the students enough opportunity to show their ability in scanning and reading for detail, and also give, in Hughes' terminology, several 'fresh starts (1989: 119)'. In the final version (see table 6) the first section, *Getting Around,* gave three fresh starts, the second section, *Resources,* gave three fresh starts, and the final section, *Emergencies,* gave two fresh starts.

It also seemed that the format of some of the questions was posing difficulties for the test takers. Kirschner et al (1996: 89) express it this way; 'test questions constitute a communicative interchange between the test writer and the test taker.' In this case miscommunication was occurring, which indicated a change in the style or format of the questions was necessary. Looking at the questions and the answers from the pre-test more closely the question form did not seem direct enough for the test takers to find the information the test writer was seeking. This was also reflected in the way the questions were presented, and in addition implied changes in the marking may be necessary, as a simple and clear marking scheme is easier to operate for a busy

classroom teacher. A comparison of the way the questions were changed for the first section can be seen in table 4.

VERSION 1	FINAL VERSION
GETTING AROUND	GETTING AROUND
1.Which is the nearest airport to central London: Gatwick or Heathrow	1.Which airport is nearer to central London?
2.Where can you find information on public transport in London?	Gatwick □ Heathrow □
3.Where can you buy a travel card in London?	2.Where can you buy a travel card in London?
4.If you were going to London for the day with an adult friend and two children under the age of 10, which travel card would you buy?	3.If you were going to London for the day with an adult friend and two children under the age of 10, which travel card would you buy?
5.Why?	Day Travelcard □ One-day Family Travelcard □
6.Which section would you look under to find out how much it costs to travel on the London Underground?	Three-day Travelcard □ Oystercard □
7. How much would it cost you to hire a bike for a day?	4. Which section would you look under to find out how much it costs to travel on the London Underground?
8. If you hired a bike you would need to pay £100 deposit. Explain what you think **a deposit** is.	Using the system □ Underground timetable □ Fares □

Table 4: Comparison of Questions from the Section 'Getting Around', in the First and Final Version of the Reading Comprehension Test.

Questions 2, 5, 7 and 8 were omitted in the second version as questions requiring similar information were already in the test and the scope for answers was too diverse. In the case of question 8, the item in question on reflection was thought to be too difficult. The format of questions 4 and 6 in the original was changed to multiple choice. In this way a focus for looking for the answers was provided, but evidence of reading with understanding would still be necessary in order to arrive at the correct answer. Question 3 in the original was still thought to be valid, as to get the correct

answer the student would simply have to write (copy) one or more options from the text, once the correct part of the text was identified.

One question in the *Emergencies* section was completely changed as I was unable to formulate it with clarity in relation to the answer I was seeking. I also realised the information in the text was not easy to locate on the page. The change can be seen in table five.

FIRST VERSION	FINAL VERSION
If you call 999 (or 112 from a mobile phone) in an emergency, what will the operator ask you?	According to the text, if you lose your credit card, which of the following should you do? Report the loss to the police □ Phone the 24 hour services □ Inform your bank □ All of these □

Table 5: Example of a Question Substitution

From a TLU stance it is probably more likely a tourist would want help with a lost credit card than make an emergency telephone call. The test taker in the final version is directed to the options, but has to make a decision between them. This would seem to be realistic. The final version of the entire test can be seen in table 6 below, and the marking key in appendix 3.

NAME_____   DATE_____

Reading Comprehension Test

This test is assessing your ability to find information from '*Time Out*' 2005/6. *Time Out* is a magazine written for visitors to London.

There are three sections in the test. The questions are in Finnish.

Look for the answers to the questions from the page which has the *same heading* as the *questions heading*.

GETTING AROUND

1. Which airport is nearer to central London?

   Gatwick   □
   Heathrow   □

2. Where can you buy a travel card in London?

3. If you were going to London for the day with an adult friend and two children under the age of 10, which travel card would you buy?

   Day Travelcard   □
   One-day Family Travelcard   □
   Three-day Travelcard   □
   Oystercard   □

4. Which section would you look under to find out how much it costs to travel on the London Underground?

   Using the system   □
   Underground timetable □
   Fares   □

RESOURCES

5. Where could you go to send an e-mail home if you were visiting London for the day?

6. Do Post Offices in Britain open on Saturdays?

    Yes   □
    No   □

7. Where would you be expected to tip in Britain?

8. How much is a normal tip in Britain?

    10 per cent   □
    15 per cent   □
    20 per cent   □

EMERGENCIES

9. According to the text, if you lose your credit card, which of the following should you do?

    Report the loss to the police   □
    Phone the 24 hour services   □
    Inform your bank   □
    All of these   □

10. Where can you go to get *emergency dental treatment* in London?

    Charing Cross Hospital □
    Guy's Hospital   □
    Royal London Hospital □
    St Thomas' Hospital   □

11. What time should you arrive to make sure you get seen the same day?

    Before 11 am   □
    At 11 am   □
    After 11am   □

12. If you need emergency dental treatment and you have to go by public transport, at which station should you get off ?

    Guy's Hospital □
    London Bridge □

Table 6: Final Version of the Reading Comprehension Test

Having looked at the question writing and re-writing in some detail, the last section approaches the questions of reliability and validity within the context of this classroom test, before looking again at authenticity in the light of piloting and using the test.

5.   DISCUSSION

The test of reading comprehension was finally used under test conditions with nine students from three different vocational classes. This section discusses the results of these tests and considers the implications for the future development of reading comprehension tests for use in the classroom with similar students.

5.1   Reliability and Validity

Reliability and validity may not be at the top of a classroom teacher's agenda when planning a test, but it is still important to take them into consideration. Reliability is concerned with consistency of measurement and is 'an essential quality of test scores (Bachman and Palmer 1996: 20).'   While no test can be considered completely reliable it may be possible to '…minimize the effects of those potential sources of inconsistency that are under our control through test design (ibid).' Hughes (1989: 38-41) offers some practical guidelines for increasing reliability which include: making sure there are no ambiguous items, providing clear and explicit instructions and writing a detailed scoring key. Even though my test had been through two pre-test versions, while observing the students taking the test and when marking the tests I was still concerned that some of the test items showed evidence of lack of clarity, if not ambiguity. For example, in the *Emergencies* section, question 10 asks:

Where can you go to get *emergency dental treatment* in London?

In the English version, emergency dental treatment is written in italics, thus highlighting which heading the test taker is looking for. The words in the question correspond exactly to the subtitle in the text under which the answer can be found. In translation the test taker doesn't get the benefit of these exact words, and consequently some of the test takers became frustrated and confused as to whether the

relevant information was there at all. This could be a case where 'authenticity' (in this case giving the questions in the test takers' language) mitigates against test performance. The test writer would have to consider if there is a right solution in this case.

The revised versions of the test had the effect of simplifying the marking key (see appendix 3). Most items are unequivocal and score 1 point. For questions to which several items were possible, the test taker is credited with one point if only one item is given, but could get two points if two or more items were offered. All acceptable answers are detailed in the key. Some questions are designed to make test takers look at quite small differences, and not necessarily go for the obvious answer. For example, in the *resources* section, question eight asks:

>*How much is a normal tip in Britain?*

In this case, although 15 per cent is mentioned in the text, and is perhaps more easily noticed as it appears as a figure, ten per cent is the required answer. This is specified in the key and the test taker would have to show evidence of accurate reading to get the correct answer.

If a test has construct validity it should measure the ability it is said to measure. Hughes (1989: 26) states that construct validity is generally unproblematical in a direct test of reading ability. This may well be a reasonable assumption for a teacher-written test for assessing reading comprehension in the classroom. I am however left with one concern in this regard which relates to the particular students I work with. This is the tension between using an authentic text in facsimile form and the fact some of my students have reading difficulties relating to the physical parameters of reading in any language. This may mean the format of the text itself could be responsible for apparent reading comprehension problems in English as measured by this test, which may not exist if the text size and density were different.

5.2   Authenticity Revisited

One of the main cornerstones of this test from my point of view as a test writer/teacher was that of the authenticity of the material and of the tasks developed from it. However, mention has to be made of how the test situation itself affects the concept of authenticity from the perception of the test takers. Spence-Brown (2001:

475) notes that '…it has often been observed that in a test the implicit rules of the testing game will over-ride those of the explicit task in determining behaviour and evaluation.' I noticed that when the students were taking the test, first and foremost it was perceived as a test. The very word test on the first page served to orientate them. They were aware that test behaviour means, for example, quiet individual work and no conferring. The same material used as a classroom activity may be perceived as more authentic because of the absence of pressure that inevitably surrounds a test, and a student could have the choice of whether to work alone or with someone else. Peirce(1992: 682) makes the point that '[the] meaning [of the text] derives from the interaction between the text, the test taker, and the testing situation in which the text is read.' She argues that a test is of itself an authentic social situation which is recognised as such by test takers (ibid: 685). Authenticity cannot be considered as an absolute term but I am still persuaded that in a test of reading comprehension an authentic/unmodified text is justified. It is also worth taking the time to make the questions as relevant to the text and the test-takers in terms of authenticity as possible, even if the test takers themselves do not give it as much weight as they might in a situation in which assessment was not involved.

5.3   Test Results

Of the nine students who took the test only one was unable to do anything with it. With help, in the form of a classroom activity this would have undoubtedly been possible for her, but not in the form of a test. All other students were able to work with the test as given, and it was seen as being within their capabilities. The instructions seemed to relate reasonably well to the format, and despite there being seven pages in two 'booklets' there wasn't any major confusion and eight students completed the test satisfactorily. I felt that this kind of result confirms the validity of the test for use in my classroom, and I would use it or a similar test to assess students' reading comprehension in the future.

6.   CONCLUSION

I found the task of developing a test of reading comprehension for use in the classroom forced me to differentiate between a test and an activity. In the classroom both could be seen as fairly interchangeable. '…I think any test could just be an activity and any activity could be made into a test…you just evaluate one in a certain

way (the test) and the other is used for the learning process (personal communication from a colleague 2006).' The process of developing this test made me think carefully about the different stages involved and realise the crucial importance of planning to try and 'assure that the test will be useful for its intended purpose (Bachman and Palmer 1996: 86).' The model of mandate leading to the interchange between writing the specification and the task or item provided a useful framework on a practical level, but also increased my understanding of the process of test development as a whole. I now have a clearer understanding of how the process of test writing for the classroom compares with developing tests to assess large numbers of students world-wide, being essentially the same, but on a smaller scale, and using the resources available to suit the particular circumstances.

REFERENCES

Bachman, L.F. and Palmer, A. S. (1996) *Language Testing in Practice.* Oxford: Oxford University Press.

Brown, H.D. (1994) *Teaching by Principles.* New Jersey: Prentice-Hall.

Hill, C. and Parry, K. (1994) 'Assessing English language and literacy around the world', in Hill, C. and Parry, K. (eds) *From Testing to Assessment.* London: Longman.

Hughes, A. (1989) *Testing for Language Teachers.* Cambridge: Cambridge University Press.

Jafapur, A. (1987) 'The short-context technique: an alternative for testing reading comprehension'. *Language Testing* 4: 195-220.

Kirschner, M. Spector-Cohen, E. and Wexler, C. (1996) 'A teacher Education Workshop on the Construction of EFL Tests and Materials'. *TESOL Quarterly* 30: 85-107.

Lynch, B.K. and Davidson, F.(1994) 'Criterion-Referenced Language Test Development: Linking curricula, Teachers and Tests'. *TESOL Quarterly* 28: 727-743.

Peirce, B. N. (1992) 'Demystifying the TOEFL Reading Test'. *TESOL Quarterly* 26: 665-689.

Spence-Brown, R. (2001) 'The Eye of the Beholder: authenticity in an embedded assessment task' *Language Testing* 18: 463-81.

Spolsky, B. (1985) 'What does it mean to know how to use a language? An essay on the theoretical basis of language testing'. *Language Testing* 2: 180-91

The Time Out Guide: London for Visitors (1991/2) 111. Produced in co-operation with the London Tourist Board and Convention Bureau.

Time Out: London for Visitors (2005/6) 118-124. London: Time Out Guides Limited.

Appendix 1

Text for the Reading Comprehension Test: taken from *Time Out 2005/6*

Appendix 2

Contents Page from *Survivial, Time Out 1991/2*

This was not used in facsimile form in the earlier versions of the test as it comprised only a small part of the page from which it was taken.

CONTENTS

Appendix 3   Marking Key (final version)

	MARKING KEY	
<u>Getting Around</u>**:**	1.Heathrow Airport	1 point
	2. Tube and rail stations, London Travel Information Centres, Shops that display the sign	1 point for any or all
	3.One-day Family Travelcard	1 point
	4. Fares	1 point **4 points maximum**
<u>Resources</u>	1. A cybercafe, big stores, public library, Cybergate, easyInternet café.	1 point for any of these. 2 points for more than one.
	2. Yes	1 point.
	3. Taxis, minicabs, restaurants, Hotels, hairdressers, some bars.	2 points for the whole list, 1 point for incomplete list.
	4. 10 per cent	1 point **6 points maximum**
<u>Emergencies:</u>	1. All of these	1 point
	2. Guy's Hospital	1 point
	3. Before 11 am	1 point
	4. London Bridge	1 point **4 points maximum**