# CHANGING PERCEPTIONS IN LANGUAGE LEARNING AND TESTING

Julie Beverly, Eric Beverly, Mike Clarke and Sid W White

# Changing Perceptions in Language Learning and Testing

Julie Beverly, Eric Beverly Centre for Modern Languages Department of International Business

Mike Clarke and Sid W White Learning Support Advisors

> University of Plymouth Drake Circus Plymouth PL4 8AA

julie.beverly@pbs.plym.ac.uk

# Abstract

For substantial and sustained progress to be made in language learning there is a need for regular assessment and rapid feedback. This paper describes how a programme of computerised assessment, introduced in the first instance as a means of coping with what would otherwise have been an unmanageable marking load, has become as much a learning resource as an assessment tool.

It demonstrates how the choice of question types has evolved in response to students' performance, how the inclusion of graphics can facilitate learning and how the use of multimedia has enabled the assessment programme to be extended to include listening skills.

It reports on the feasibility of running sessions incorporating multimedia, on levels of student participation and student work rates and on student evaluations of the programme.

#### Why computer based assessment?

The discrete-point testing of lexis and linguistic structures has, for some time, been a part of the assessment schedules of modern language modules. In their survey of effective assessment practices in higher education, Dochartaigh and Schmidt found that grammar averages 10% of the assessment in first year language courses in French, German and Spanish and tends to be even more heavily weighted in *ab initio* courses.

This emphasis on the testing of vocabulary and grammar may appear somewhat reactionary. It is nonetheless true that, if we aim to achieve, in a relatively short time, a level of competence that will enable us to communicate effectively in another language, words have to be learned and an understanding of how the words work together has to be acquired. Without an adequate lexical base and without an understanding of how words can be adapted and strung together, attempts to communicate founder. However, neither lexical range nor grammatical competence can be acquired overnight and ways have to be found of encouraging students to work on these areas on a regular basis, hence the widespread use of regular progress checks or tests.

In the department of modern languages at Plymouth, weekly or fortnightly pen and paper progress checks were run to test students' knowledge of the lexical items and linguistic structures encountered in each unit of the language course. The progress checks were self-assessed: partly in order to minimise any sense of embarrassment or failure, and partly because lecturers did not have the time to take on an assessment load of this order. (One piece of marked work each week over 20 weeks for every language student, for example, amounts to 1000 pieces of work for every 50 students and this in addition to the standard assessments to measure competence in the four basic linguistic skills: of speaking and listening, reading and writing.) At the end of each semester summative tests were run made up of items taken from the progress checks.

The results were, as might be expected, mixed. Whilst the better students (the more diligent or better organised, as well as the more able) responded well to the regime, some students simply skipped the scheduled progress checks preferring to rely on the age old practice of learning or attempting to learn everything the night before the summative test.

The situation improved somewhat when these pen and paper progress checks were "computerised" using Question Mark for the Mac. The software allowed the conversion to be made with relatively few changes to the original format and, for the most part, the questions were of the Text Match or Fill in Blanks type as the intention was to test recall rather than recognition, active rather than passive knowledge. The feedback provided was, as in the keys provided for the progress checks, limited to the correct answer. The set-up was somewhat primitive: there was no access to a server and everything had to be done via diskettes but students now had an objective assessment of their performance and lecturers could monitor their progress more easily. Like the pen and paper progress checks this was, effectively, a voluntary activity. A resource which students could choose to make use of, or not, in preparation for the summative assessment. The novelty factors – doing things on computers, in their own time and as often as they wished - did lead to an improvement in the take-up rate but it was not markedly dissimilar to that for the pen and paper progress time.

Much more radical changes in the assessment programme have taken place since the university's acquisition of a site licence for Question Mark Perception.

#### From assessment tool to learning resource

The fact that the questions created in Question Mark for the Mac could not be converted for use in Question Mark Perception led to a complete review of the various question types and the creation of new questions and question databanks.

It had become clear, for example, that the Text Match question was not necessarily the best type for lexical items when students are required to supply the English equivalent for a word or phrase presented in the other language. The Italian expression *Ho mal di testa* (I have a head-ache), can, for example, be expressed in a variety of ways if one is prepared to allow variations such as: I have, I've, I've got, together with the range of spellings of head-ache. The permutations are not infinite but there are a large number and, although the Text Match question type allows an almost unlimited number of alternative outcomes, predicting them or editing the question to take account of the latest version is very time-consuming. To avoid penalising students for a less than perfect command of English and to minimise the need for question editing at a later stage, it was decided to make use of other question types such as Selection, Multiple Choice or Multiple Response and Hotspot whenever a response in English was required, unless the response or responses were entirely predictable.

The use of the Selection and Multiple Response question types has enabled the testing of multiple lexical items within a single question and more ground to be covered but with fewer actual questions. Thus a single question about clothing might include five or six different garments.

Multiple Choice and Multiple Response questions have proved effective in the testing of idiomatic expressions and Hotspot questions have contributed a "fun" dimension whilst at the same time enabling the precise targeting of lexical or linguistic elements.

Structured translations have been facilitated by the use of Selection and Multiple Choice questions and rational C-Tests have been constructed using the Fill in Blanks question type.

Essay questions have been inserted to enable students to comment on sessions.

The incorporation of graphics in all questions has permitted the contextualisation of both lexis and linguistic structures. Where, for example, the student is required to produce the target language equivalent of an English word or phrase the graphic or graphics is a concrete representation of the lexical item. However, in those cases where the student has to move from the target language to English, the graphic or graphics are simply illustrations and the relationship between the test item and the graphic is less explicit. Whenever possible the images selected are "authentic" in that they depict the country, the people or aspects of life in the country. The aim is to enhance the student's awareness or knowledge of the cultural context as well as facilitating the acquisition of vocabulary and the development of linguistic competence.

Instead of using formatting features such as bold, underlining or italics to highlight certain elements within a question, different colours are used according to the function of the text. Blue, for example, is used for instructions and contextual material and red for the targeted item. Green is used for the correct answer with key aspects highlighted in red. Purple is used whenever additional information is provided such as grammatical terms or key words. The feedback has been extended so that, as well as the correct answer, students are told where to look for more information and links are sometimes provided to other related websites.

When they were first introduced the computer based assessments were available as formative sessions to be used by students on a voluntary basis in preparation for the summative language tests at the end of the semester and take-up was patchy. Some students accessed them regularly and often, others made little or no use of them. To encourage take-up it was, therefore, decided to award some marks to the formative sessions and the average mark achieved on the formative sessions would contribute 5% of the module grade. To further encourage the use of the sessions as a learning resource they were made available in two forms: as issimo sessions and as corresponding *italiprova* ones. In the *issimo* sessions the questions were always presented in the same order and no time limit was set. Students were able to access these sessions as, when and where they wanted. In the *italiprova* session, on the other hand, the same questions appeared in random order and a time limit was imposed. The students were still able to access the *italiprova* sessions at a time and place of their choosing but these sessions could only be attempted once, within a limited period of time and it was the mark obtained on the *italiprova* sessions that contributed to the module grade. The access rate, as a result, rose considerably for all of the groups involved as indicated in the figure below:





number of sessions accessed in semester 2 in 2000 and 2001

The fifty-five students on the three modules for which there was a programme of computer based assessment accessed the formative *issimo* sessions in total 2521 times, an average of 45.84 per student. The sessions were accessed from different parts of the UK, France, Italy and Spain as well, of course, as the university campus.

Most students worked on the *issimo* sessions repeatedly until they achieved scores of 80% or more. They then attempted the *italiprova* session. Average percentage marks obtained by those who completed the *italiprova* sessions range from 69.63 to 98.50.

The summative language tests taken at the end of each semester were made up of items randomly selected from the *issimo* sessions and presented in random order. There are far too many variables for direct comparisons between the performance of this year's students on the summative language tests with those of previous years to be made. However, there are indications that there has been an overall improvement in students' performance. The average mark obtained this year by the group whose assessment programme most closely resembles the previous year's was, for example, higher by 15.79%:

minor 1	average mark for semester 2 language test
2000	63.46
2001	79.25

After each summative test students were given a questionnaire to complete. The responses for three of the questions are shown below:

How would you describe your use of the computerised issimo sessions	;?
---	----

	2000	2001
I made little or no use of the <i>issimo</i> sessions	18.18	10.9
I prefer to work with pen and paper and only used the <i>issimo</i> sessions to discover what would be in the language test	22.72	16.4
I used the <i>issimo</i> sessions regularly mainly as a way of practising for the language test	22.72	25.5
I used the <i>issimo</i> sessions regularly to help me learn the vocabulary and grammar	0	5.5
I used the <i>issimo</i> sessions to help me learn the vocabulary and grammar and as a specific means of practising for the language test	36.36	41.8

To what extent do you think your performance on the **Language Test** was affected by doing it on a computer?

	2000	2001
it was very much improved	0	14.6
it was improved to some extent	13.63	47.3
there was no noticeable difference	54.54	23.6
it was less good than usual	13.63	10.9
it was much worse than usual	9.09	1.8
other	10.08	1.8

Would you choose to take a test in this form again?

	2000	2001
yes	52.38	87.3
no	47.61	9.1
don't know		3.6



Which words would you use to describe the experience of doing the *issimo* sessions on computer? Tick as many as you wish.

	total		total
attractive	24	interesting	26
boring	3	irritating	4
clumsy	2	motivating	12
demoralising		pleasant	12
difficult	4	practical	37
easy	26	rewarding	10
enjoyable	15	stimulating	5
friendly	26	stressful	4
frightening	2	time-wasting	1
fun	9	unpleasant	
inflexible	9	unfriendly	

Other comments from students include the following:

less hassle than with paper, helps me using computers, gets me to learn the language, provides you with useful techniques for revising and it makes practice interesting, can work at your own speed, privacy, immediate feedback, checks that you are 100% accurate, the pictures help and it's less intimidating, accessible anywhere anytime, more environmentally friendly, pictures are stimulating and help to jog the memory, doesn't seem so much like an exam and stops you being so stressed out.

# Introducing audio

What follows is a description of the technical aspects of the use of sound files within an online computer assisted assessment.

The project under discussion required that sound clips be available to students of Italian as part of their on-line assessment. Perception software readily includes sound files of almost any description within a session. Insertion of sound files may be achieved through a graphics drive "wizard" or by inserting the relevant bit of HTML code directly into the script.

The sound files used were obtained from analogue tape originally recorded on a Panasonic (model) video camera using the camera's built in microphone. The record quality was low partly due to the quality of the microphone itself and also because of the speaker's movements which caused variations in the recording level. The recordings were then transferred to compact cassette and this further increased the background noise levels of the material.

Conversion to a digital format was achieved in two ways. Firstly, the material was captured directly from a Sony Walkman using a Creative Labs Sound Blaster Platinum Live 11 sound card. The recording was made using the sound card's dedicated Wave Table software. This method was problematical due to the low-level signal recorded onto the tape. Consequently, both the output level on the tape

player and the input level on the soundcard had to be at maximum which further degraded the sound quality of the material.

The second method for converting the analogue material to a digital format was to record the analogue material onto a CD using a Pioneer PDR 509 CD audio recorder. The CD was then captured to the sound card directly from the computer's in-built CD player. The output from the CD was noticeably higher quality than from the tape and therefore it was possible to obtain a cleaner recording from this source.

In both cases the material was captured at 44,100 HZ 16-bit stereo. Capturing in 16bit stereo is more memory hungry than in 8-bit but compression algorithms work more effectively to recreate the sound at a lower band width.

The recorded material used for the first batch of audio assessments was from an archive source and was not originally intended for conversion to a digital format. The process of obtaining recorded material could be improved upon by the use of the following:

dedicated studio or sound dampened venue for making the recording

lapel microphone so as to maintain a consistent distance between speaker and microphone

digital video camera or audio tape recorder that would take the material directly into digital format

# Editing the sound files

The raw material was imported into the Creative Wave table software as described above. From within the program the material was edited into suitable sections for publishing in the online sessions. The program is particularly suitable for the purpose as it allows easy duplication (cut and paste) of the material into new sound files. On completion the edited files were saved in a .wav format. This format, created by Microsoft, is a standard PC audio file format. It has the advantage that it is easily edited and in addition to the audio data it also contains information about the number of tracks, sample rate and bit depth.

The disadvantage of using .wav files in the context of online assessments is the large size of the file. The capture rate used to record the language material was the same as that used to record at standard CD quality. The analogue material was sampled at 44,100 samples per second. These samples are 2 bytes or 16 bits long and since the tape was in stereo, separate samples were taken for each channel. The number of bits that is required to sample each minute of sound is therefore:

44,100 samples/second x 16 bits/sample x 2 channels x 60 seconds/minute

- = 84,672,000 bits per minute
- = 10,500,000 bytes per minute
- = 10 mega bytes/minute

The large file size has two main implications. Firstly the audio files relating to the assessments will take up an unacceptably large amount of disc space on the server. Secondly, the files will take a long time to download to the students' machines and

this will be particularly noticeable if working from a site where band width is restricted.

It is, therefore, desirable to save the audio material in a format that uses a minimum of space. having been recorder end edited in .wav format the files were first converted from 16 bit stereo to 16 bit mono thereby reducing the file size by 25%. Following this the files were converted to the compressed MP3 format using Sound Forge 4.5 software from Sonic Foundry. MP3 encoders work by removing information that is not perceptible to the human ear. In doing this it is possible for the file size to be reduced by over a factor of 10.

#### The audio sessions

Three practice sessions were prepared to which the students had unlimited access. The first was similar in format to the kind of listening test often run in language laboratories. There was one sound clip, a recording of a short talk of approximately 4 minutes in length. The online session began with an explanation on screen of the procedure to be followed ie to listen to the recording and then to answer the questions that followed. To obviate the need for students to return to the first screen each time in order to listen to the recording, a copy of the sound file was included in each of the questions (8 in total). The operation of sound files on screen is very similar to the situation in a language laboratory. The tape can be played repeatedly and started and stopped at any point. A recap button as such does not exist but, as most students realised, the sound file on screen serves as a visual aid enabling them to locate the part of the recording they wish to listen to again with greater ease and speed than the fast forward/rewind and recap buttons of the language laboratory.

In the two subsequent practice sessions, the recordings were presented in their entirety only in the first explanation screen. Each recording was then broken down into three or four shorter clips for inclusion in the 8 to 10 questions. There were two reasons for this: firstly, to cut down on the amount of memory and, therefore, download time required and secondly to help the students to develop their aural skills by enabling them to focus on the relevant part of the interview or talk.

Graphics were included in all questions to supply a visual context for the subject matter and to "lighten" the screen. There were intended to facilitate understanding but none of the questions could be answered on the basis of the information contained within the graphics.

# The students' perspective

Some students encountered problems accessing the practice sessions. Only 4 of the 17 students in the group successfully accessed all three of them and 2 students did not attempt any of them. The average number of audio sessions accessed by these students was 5.27 whereas the average number of practice language sessions accessed by the same students was 65. The fact that no marks were attached to the practice audio sessions may have been something of a disincentive.

There were occasional problems with the network but generally speaking download times varied between a matter of seconds to one or two minutes. It is likely that the main factor was the relatively limited number of computers with a sound card available to them in open access areas. In the building in which they were most used to working, there were, in fact, none at all. Nonetheless, despite these difficulties, 10 of the 17 students performed better in their second listening comprehension test than they had done in the first one, which had been carried out in the standard way in a language laboratory, and the average mark rose from 55.16 to 68.77.

The summative audio assessment was conducted under supervised conditions and two members of learning and technical support staff were on hand to deal with any problems (of which there were none, in fact).

Reactions were, on the whole, favourable. The students liked it because they thought it was:

more practical the quality of sound was better easy to use easy to re-do questions clearer and guicker to use more personal and there was less interference from other students the speed could be controlled easily easier to see how far through the recording you are different you get your mark immediately the recording is broken down into shorter clips enabling them to concentrate on the relevant part easy to select the part you needed to listen to again the atmosphere seemed more relaxed it's very visual so you don't get bored and you can see where to go back or forward to on the bar

Given a choice between doing a listening test on computer or in a language laboratory 10 of the 17 would opt for a computer and 3 had no preference. Of the four who preferred the language laboratory, one is a self-confessed computerphobe, two are not keen on change and the remaining one had not realised that the sound file could be stopped and started at will.

# Accessing Sessions

At the centre of the Perception authoring software are the Question and Session databases. These databases enable the Perception author to store and organise assessment sessions. The choices made during the construction of questions and the introduction of questions into session databases are fundamental to the speedy delivery of student material. A session download time of between 2 and 5 seconds should be achievable provided that, during the session planning stage, consideration is given to the following:

#### **Question Blocks**

Ideally, sessions should be split into question blocks, each block having no more than the recommended maximum of fifty questions. The use of question blocks will enable quicker access to questions and initiate scoring early in the session. To enable a shorter download time and, therefore, faster access to the first of the questions, it is recommended that the first question block have the least number of questions (5 to 10).

#### **Selecting Questions**

Perception offers several options as regards the selection of questions. One allows the use of all the questions from a particular database, another allows the author to select each question and another allows questions to be 'randomly' selected. The size of the question database is a factor to be taken into consideration when planning a session using randomly selected questions. It should be appreciated that the searching of a large question bank can significantly slow the download process.

#### **Graphics and Multimedia**

Depending on the size of the file, the downloading of graphics, audio and video files can extend the download time significantly. It is therefore important that the size of graphics and multimedia files be kept to a minimum. Software such as Adobe Photoshop can reduce the file size of complex graphics and, consequently, the download time to students. Similarly, audio clips, which may take up large amounts of disk space and increase download time, can be reduced by compressing .wav files into mp3 files.

#### **Staggering the Start of Sessions**

Where large numbers of students are involved, consideration should be given to the number of students logging on to a session at any one time. Tests have shown that, in most cases, provided each group logged on at one-minute intervals, groups of twenty students were able to download a session within six seconds. Allowing more than twenty students access to the session at the same time increased the time for individual logon and, in some cases, caused the system to initiate warning messages asking students to resubmit a logon request.

# Conclusion

The evidence to date suggests that the experience of online computer based assessment in modern language modules is a largely positive one for students. The immediate availability of results and the accessibility of the sessions are clearly important factors but so too is the fact that they find this type of assessment not only less stressful than traditional exams but even quite fun.

For the lecturer there is, obviously, the advantage of a lighter marking load and, given that the Question Mark Perception software permits the generation of reports

for individual students, groups, single questions or sessions, there is also access to a range of information which permits a closer monitoring of students' progress. It is also possible to know much more about how, when and where students are working; what they are doing, what has been achieved and what needs further attention

# References

Dochartaigh, Pól Ó and Schmidt, Michael (1999) Effective Practices in Assessment in the Modern Languages