



City Research Online

City, University of London Institutional Repository

Citation: Marcoul, I., Campbell, A. and Erich, S. (2012). Constructing Assessment: An Investigation into the Effectiveness of Online Diagnostic Tests to Assess Linguistic Competence. *Learning at City Journal*, 2(2),

This is the unspecified version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <http://openaccess.city.ac.uk/1639/>

Link to published version:

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

Constructing Assessment: An Investigation into the Effectiveness of Online Diagnostic Tests to Assess Linguistic Competence.
Isabelle Marcoul, Anna Campbell and Svenja Erich, School of Arts.

Abstract

This research analyses the creation and use of online diagnostic tests to assess the linguistic competence of students at City University London. It examines the choices and approaches to designing this type of test and the effectiveness of the whole assessment process. Moreover, it highlights the importance of creating assessments testing the way learners mainly construct their knowledge regardless of the discipline or subject assessed. Finally the study recommends for the design of online assessments not to be led by the technology but to test the approaches learners use to map their own knowledge.

Key words online diagnostic test, language learning and assessment, mapping and constructing knowledge

Context and outline of the study

Before they can join a language class the students at City University need to be assessed and assigned to the language courses appropriate for their level of linguistic competence which ranges from beginner to advanced levels (Appendix.1). Traditionally the tests were handed out in a printed format and marked by language lecturers. Administratively this has always been a complex and lengthy process as only a few days are available to test, enrol and assign the students to a language course¹. In 2011, to speed up the whole assessment practice, these language diagnostic tests were set up and corrected online with little need for human interaction.

This research looks into the impact and effectiveness of this type of online assessment and design. It discusses the need to create online examinations reflecting the way learners map and mainly construct knowledge as a process of learning. Taking a constructivist perspective for its theoretical framework, the study examines the results of the online tests to see whether their design has correctly assessed the learners' competence. For this the research compares the tests' results with the courses and levels which the students finally attended. It assesses whether there were discrepancies between the outcomes of their assessments and their actual levels. It also uses the results of the students' self- assessment done online to see whether this influenced their performance in the tests. Both the analytical framework and the data are then used to further discuss assessment online and draw conclusions for this study.

¹ In September 2011 the number of students taking this online test exceeded 1000 and allowed students to enrol to language modules either as part of their electives or as free standing modules

Discussing Second Language Acquisition

The chosen approach to the design of these tests was based on our understanding of teaching and learning a second/foreign language. We recognise that the context in which a foreign language is taught and tested is linked and situated in the real world. Learning a second or foreign language is not some kind of knowledge randomly picked up or some kind of abstract knowledge secluded in the realm of an abstract world. The learning process involves the development of discourse processing, conceptualisation, mastering interactions with other people, sign systems, and the production of variable contents all performed in real life contexts. While codes, contexts, and interactions must be distinguished in theory, in practice they interact holistically and therefore our diagnostic tests needed to reflect all these parts together. Overall the tests were designed to examine the production of a language combined with the capacity to communicate and intellectually map and construct meaning. From this perspective assessments needed to test students' cognitive and intellectual skills which form linguistic competencies generally expected in academic disciplines.

Constructing knowledge

The method we use to teach a second language is called communicative approach and asks the students to practise the language while learning about its rules, structures and mechanisms. For this the learners need to build their own understanding of the language and develop their own mapping process to build their own knowledge. This approach is by nature constructivist and involves students' participation in class and the use of a variety of cognitive skills including automatic processes, memorising as well as logic, reasoning and reflection to effectively and academically use newly integrated knowledge. This complex combination of skills and production of the language ultimately leads to the development of the learners' self-expression in the target language.

From the constructivist perspective learning becomes an active process of constructing and mapping rather than just acquiring knowledge. The teaching function becomes a supporting process where students and teachers construct and make sense of what is learnt, rather than a communication of knowledge transmitted down from teacher to student. As quoted in Duffy & Cunningham (2003) and stated by Von Glaserfeld (1989,p.134): "instead of presupposing, knowledge is a representation of what exists, knowledge is a mapping, in the light of human experience, of what is feasible".

Constructivism has a variety of terminology and concepts, often perceived as not very helpful or practical and too general as a philosophy unable to provide enough precision for instructional decisions in relation to assessments. (Rust, et al. 2005). Nevertheless this does not mean that we should disregard this approach. A constructivist stance is useful in helping to examine the learners' capacity to construct new knowledge instead of focusing on the reproduction of old knowledge. It best reflects the process of building knowledge so essential in the "mastering" of a second language. We assess the students' capacity to map and use their understanding and intellectual constructions, to ultimately produce their own version of the foreign language which is the focus of our assessment. In practice, the

learner's engagement in his or her learning is essential. This does not mean that testing the body or content of knowledge specific to each academic discipline is not important but the choice of a constructivist approach for assessment emphasises the need to test the way students build and process knowledge and consolidate factual knowledge with intellectual use. This approach to online assessment does not need to be confined to the testing of language competence, it can be used in a wider context than discussed here e.g. for cognitive assessment, performance assessment and portfolio assessment (Reeves, 2000). Ultimately it avoids the focus of testing language competence based on simply memorizing sentences or texts in the target language without understanding how the language itself is used and meanings are produced.

Format and content of online diagnostic test reflecting the construction of knowledge

In practice, to test the learners' capacity to construct their own version and production of the language the diagnostic test focuses on reading, writing, speaking and listening comprehension with task-based assessments. The main target is to diagnose strengths and weaknesses, to measure progress and to inform teachers of what needs to be covered again in class (Ashton and Wood, 2006). As discussed earlier it was important to avoid testing only the memorisation of facts or the production of a piece of the language learnt by heart. Therefore the online assessments needed to include complex simulations reflecting how students learn and where the learners can change input variables or make changes to output variables (Thomas et al. 2004).

Designing online language tests with time and technological constraints

From the beginning, the lecturers in charge of designing the content of the tests, and the Education Support Team, took into consideration the time and technological restrictions of this project and found that assessing writing, listening and speaking skills online remained impossible without the help of a lecturer to mark the tests. As it was essential to get the results of these assessments within a few hours after completion we could only design an online test requiring no human interaction in the marking process. Likewise, audio or video files could not be used as a large number of headsets would have been needed in the computer laboratories for the students to take the tests. So, overall for practical reasons the tests had to be designed in a way that they could make the entire process of assessment fast, and easy to organise.

However, this type of restriction did not stop those involved in the design from linking test content to the learners' process of constructing languages.

The teams involved researched the type of questions or design to create such online diagnostic tests (See Appendix 2) and after consideration of the various formats the following sections were selected for all the diagnostic tests:

- A *generic text* with a wide range of linguistic difficulties was chosen to assess the students' comprehension and capacity to decode its content with multiple choice questions.
- Another section called *Use of Language* required the students to select the

most appropriate answer to a series of questions. For this the students had to understand the meanings but also the concepts behind the questions.

A few examples to illustrate this choice of questions:

What would you like as a main course?

A sorbet with strawberry

Six oysters

Steak and kidney pie with chips

In this type of exercise the learners need to understand the vocabulary requiring short term and long term memory but also the cultural context and meaning in which the words are used.

Pick the odd word out:

Hat

Cap

Cat

In this example, the learners need to make a choice based on the concepts shared between some of the words and not on similarity of sound or pronunciation. This type of exercise requires analytical skills and processing capacity on the learners' part. It goes beyond the memorisation of some vocabulary and highlights the capacity to conceptualise.

As mentioned earlier we did not want to assess the production of writing requiring human correction so for this specific skill, two types of exercises, one with multiple choice questions and one with fill the gap sentences, were adopted. In the first one the students were asked to select sentences with *grammatical and spelling errors* at various levels of difficulties reflecting the different levels of competence tested. In the second exercise the students had to fill the gap with a word or expression in sentences which assessed their understanding of either a concept or a structure.

For example:

This dangerous breed of dogs can't by law go out without a

Following the completion of these exercises the students were then asked to assess the register of the language used in parts of a text. For this they had to judge whether the content had been written in a colloquial, formal, academic or spoken language style. The analysis of these styles show the understanding of language registers found either in oral or written expression:

- Colloquial use of the language generally implies the use of slang
- Formal styles are mainly found in writing or used in a formal capacity, For example, how do you do? Dear Sir, Sincerely yours etc. In some language the use of You highlights the formal You (How do you Do?), the familiar You (How are you doing John?) and the You referring to a group (How are you – you mean how I am? – No You and your friend?)
- Academic style focuses on analytical presentation and use of the language
- Spoken style generally covers grammatical differences between written and spoken language. An example is “I cannot” in its written format becomes “I can't” in its oral version.

The design assumption was that the learners with higher levels of competence would be those who could identify the variations in the text registers more successfully.

At the end of the test the students were also invited to assess their own levels of competence so we could see whether there were discrepancies between the actual levels and the students' perceptions of their own linguistic capacities.

Decisions for technological and administrative processes

Given the needs and constraints of this assessment process it was agreed to use the following online tools to assess and collect information.

Virtual learning environment

The university has a virtual learning environment called Moodle. This is a website where each module of a course has its own webpage. After discussing the issue of diagnostic testing it was decided that this would be done through Moodle as all students have access to it.

Googleforms

In previous years the students filled out their personal details by hand on their test and this had to be manually inputted into a spreadsheet by an administrator. As over one thousand students took the test in 2011/12 this has amounted to many hours of administration work so it was decided to look for an online solution to this need to collect data. An online service called GoogleDocs was chosen. This is a free and easy to use service that allows the user to create a form to collect data. This form is completed online (in this case it was added as a weblink in Moodle) and the data is collected in a spreadsheet which can be downloaded in excel format. A form for each language was created to collect data including name, student registration number, course etc.

Moodle quiz

As discussed above, an online solution to diagnostic testing was needed. Moodle quiz was chosen as it is a flexible tool allowing for assessment that is marked automatically. This was the favoured option in order to provide the students with immediate information about the level they had attained and the class they were to attend. Each of the five languages had a Moodle quiz of 100 questions. The staff involved in creating the quiz brought their quiz questions to one of two training sessions where they were taught how to use the quiz tool. They received technical support to add the questions to the Moodle quiz. All the quizzes were tested to check for errors before being made available to students.

Moodle module

The quizzes were added to a Moodle module called Language courses diagnostic tests. This was a specific module set up to house the tests. The enrolment for this module needed to be handled differently as most enrolment into Moodle modules at City University London is done through SITS (the student record system at City University). However, this was not appropriate in this case as language modules can be taken by all students and staff so an enrolment key was set up. This was a password that students could use to enrol on and access the module, and was

provided to students at the language fair and on the guidance note about how to access the tests. It was also sent out to those that took the test at a later date.

Data and Analysis

In this particular project we aimed to find out whether our choice of design and content efficiently tested the students' academic performance and language proficiency. For this we compared the data of the test outcomes and the actual classes with their respective level in which the students stayed to study the language. To finely grain the analysis of the data we focused on any discrepancies and results in the different sections of the diagnostic tests to ascertain whether some parts were easier or harder to complete and therefore influenced the final results of these tests. It was also essential to find out whether some sections of the tests could increase the number of right answers and distort the levels of competence by inflating the test results.

Effectiveness of the test

In the German test 86% of students were placed in the correct course, followed by French with 76% and Spanish with 72% (Table 1). Overall, the discrepancy between the test results and the students' actual level was minor.

Language	French	German	Spanish
Number of student sample	64	29	43
Correctly diagnosed level	76%	86%	72%
Incorrectly diagnosed	24% of which 18% were borderline discrepancy and 6% wrong	13% of which 10% were borderline discrepancy and 3% failed	28% of which 16% were borderline discrepancy and 12% were wrong

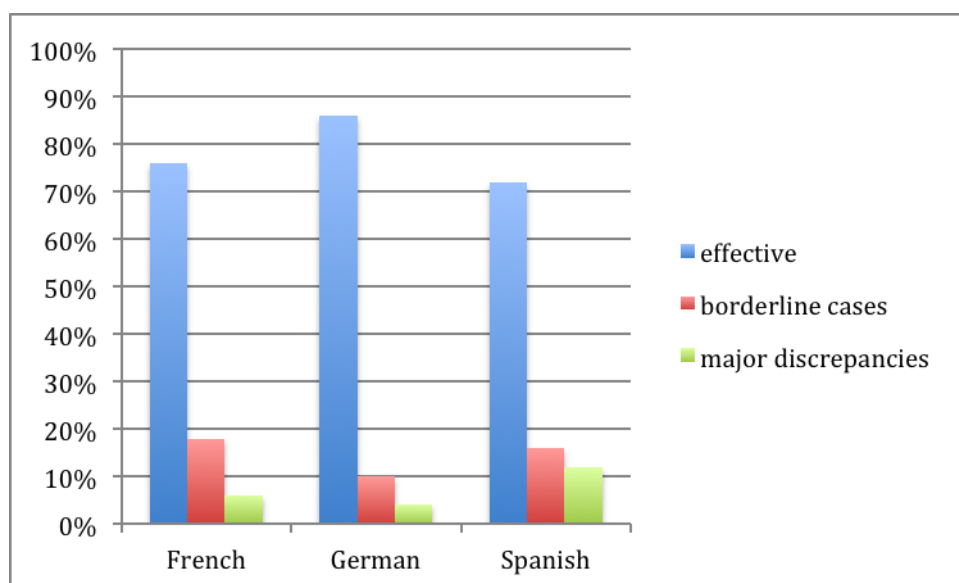
Table 1 Percentages for the test results

In the first instance, the German tests appear to be most effective. However, the data available from these tests came from a smaller number of students compared with the French and Spanish tests. The samples for the number of students for each language were based on those who took the diagnostic tests and attended the first term of the language courses. The attendance data enabled a view of whether or not the students changed courses and/or levels. In this instance with less data for German we have assumed that the results for French and Spanish were more representative of the reality of online assessment.

The French test accurately assessed the lower levels as very few students found it necessary to go from lower intermediate to upper intermediate or vice versa (see Appendix 1 for levels). This lack of swapping classes between lower or upper level courses indicates that the students remain in their assigned courses as they were right for them. The Spanish test was less successful than the French test as more students in this target language went to a level other than their test result suggested.

Although anecdotal, the migration of these students can be explained by the timetable clashes they have encountered throughout the year. The lower level courses do not take place on the same day as the more advanced ones, so when a student enrolled in a lower intermediate level course could not attend the classes due to a clash in his or her timetable, we have noticed that students moved to a course at another level taught at another time or day that suited the students better. Therefore we have concluded that this migration phenomenon is a confounding variable.

Beside minor borderline discrepancies, the tests have been effective in accurately assessing the students (see Bar chart).



Bar chart: Effectiveness of the language diagnostic tests for French, German and Spanish students in 2011

Another explanation for this migration between classes beside timetable issues is the choice of a level of competence and its course as an attempt to join a less demanding course. The most affected courses were those which are not part of the students' degrees but offered as free standing modules. These are chosen by students who want to keep learning the language for their own pleasure or to preserve their existing level of proficiency. So instead of progressing to a more difficult level of competence they chose to stay in a less demanding course. This has been observed amongst those who had borderline results between lower and upper intermediate levels.

Expectations and results – sections of the test

The students' performance in the individual sections of the test showed that in most of the exercises their results consistently reflected the expected results for each level of competence. It is however important to note that there were two exceptions:

- In the gap filling exercise, where students had to call on their active knowledge, all students clearly underperformed (by 20-40%) compared to their average result. This was largely expected, as the capacity to conceptualise is one of the most challenging tasks in the tests as well as in the learning process of mapping knowledge.
- In the exercise, where students had to assess the register of a written part of a text, we had only expected students with high levels of proficiency to do well. This turned out to be an incorrect assumption. The majority of the students at lower levels of competence did very well in this exercise to the point that this distorted some of the test results. We concluded that this might account for some of borderline test results making it difficult to separate lower from upper intermediate levels of competence.

Self-Assessment

In addition to the language test, the students were required to complete a short online form where they had to assess their own level of competence. The description for each level is based on the Common European Foreign Language (CEFL) framework defining the various levels of linguistic competence in a very practical manner. (See Appendix 1) The purpose of collecting this type of data was to compare the students' feedback with their attainment in the diagnostic test. In this part of the diagnostic tests the students were simply asked to indicate the level they believed they had in the language they chose to study by selecting the description that reflected their perceived levels of competence.

More than half of the students (52% for French, 53% for Spanish) and close to half, 48% for the German courses accurately assessed their own level (table 2). Those who wrongly assessed their level of competence thought that their linguistic proficiency was below their actual level and only a small minority of students (on average 5% in each language) overestimated their competence. When we compare the data there is evidence that students with borderline test results were more prone to accurately assess their level and joined a course at a level in which they felt comfortable.

Self Assessment	French	German	Spanish
Number of student sample	64	29	43
Correctly self assessed	52%	48%	53%
Incorrectly self assessed	48%	52%	47%

Table 2 Percentages for the self assessment

Although the Self-Assessment questionnaire did not overall provide finely grained information to interpret borderline test results it gave an insight into what may influence students' choice of courses and how they define their own language proficiency by constructing it.

Conclusion

Setting up the boundaries of what needs to be assessed requires a robust understanding and knowledge of the learning processes involved in the discipline taught. Linking constructions and perceptions of learning to the assessment, whether influenced by our social practices or linked to learning in the taught discipline with its culture, remains vital to adequately assess our students. This inclusive approach which links learning context with intellectual and cognitive skills reflects the nature of learning and therefore its assessment should also mirror this learning processes.

However different assumptions are made about the acquisition and the content of knowledge, often indicating different epistemological presuppositions about the nature of academic knowledge and learning. With the use of technology to test the learners' complex academic performances the temptation is to restrict examinations or tests to superficial solutions more influenced by the technology rather than the pedagogy.

By constructing a genuine common understanding of what is assessed and, defining the learning process demanded by the discipline are far more positive steps than having the technology dictate the format and content of the tests themselves.

Technology offers access to a world where time and space are no longer a real presence or at least a constraining one. In this case the use of online tests made the whole process of assessing our students' proficiency a much simpler procedure with rapid results facilitating the enrolment of a large number of students into our courses. The efficiency of the diagnostic tests was made possible by combining exercises focused on assessing the quality of the intellectual mapping produced by the students with the technology.

References:

Ashton, H. & Wood, C (2006). Use of online assessment to enhance teaching and learning: the PASS-IT project. *European Educational Research Journal*. 5(2), pp.122-130

Bennett, R. E. (1998) *Reinventing Assessment: Speculations on the Future of Large Scale Educational Testing*, Educational Testing Service.

Available:<http://www.ets.org/research/researcher/PIC-REINVENT.html> (13 Jul 2011)

CAA Centre (1999) Designing effective objective test questions: an introductory workshop. Available: <http://caacentre.lboro.ac.uk/d1docs/otghdout.pdf> (12.June 2012)

Commun European Framework for Languages (CEFL) Available:
http://www.coe.int/t/dg4/portfolio/?m=/main_pages/levels.html (12 June 2012)

Duffy, T. & Cunningham, D. (1996). Constructivism: Implications for the design and delivery of instruction. In *Handbook of Research on Educational Communications and Technology*. (ed D. Jonassen), Macmillan, New York.

Maclellen, E. (2001) Assessment for learning: the different perceptions of tutors and students. *Assessment and Evaluation in Higher Education*, Vol 26 No.4, pp307–318.

Appendix 1

Beginner: Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she know and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help

Lower Intermediate: Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.

Upper Intermediate: Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

Advanced: Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field or specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

Advanced Plus: Can understand a wide range of demanding, longer texts and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.

Appendix 2

Question type:

These are a few examples of what is overall used and was useful for the diagnostic tests we were looking to create:

Question types

True/false

While mostly used to assess knowledge, true/false questions can in fact be used to assess knowledge, comprehension and application levels

Matching

Mostly used to assess basic knowledge but can also be used to assess comprehension

Multiple Choice Questions (MCQ)

These are flexible and can be used to assess all levels. They are particularly suitable for knowledge, comprehension, application and analysis

Multiple Response Question

Can assess the same range of levels as MCQs but have the potential to create more difficult questions within each category

Ranking questions

These are well suited to assessing application and analysis

Assertion

These are generally suitable for knowledge, comprehension and analysis

Assertion reasoning questions can be marked by a computer and test very complex thought. They combine elements of multiple choice and true/false question types to allow testing of more complicated issues. (CAA, 1999).