



# CLES, a model framework for 21st century language certification?

Cédric Sarré

## ► To cite this version:

Cédric Sarré. CLES, a model framework for 21st century language certification?. Colpaert, J., Simons, M., Aerts, A. & Oberhofer, M. (eds) *Language Testing in Europe: Time for a New Framework?*, Antwerp: University of Antwerp., 2013, pp.194-200. hal-03924843

**HAL Id: hal-03924843**

**<https://hal.science/hal-03924843>**

Submitted on 5 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Cédric Sarré**

Université Paris-Sorbonne, Paris, France

cedric.sarre@paris-sorbonne.fr

## **CLES, a Model Framework for 21<sup>st</sup> Century Higher Education Language Certification?**

### **Bio data**

---

**Dr. Cédric Sarré** is a Senior Lecturer in English as a Foreign Language (EFL) and English Language Teaching (ELT) at Université Paris-Sorbonne, IUFM (School of Education), France. His research interests include ESP course development in online settings, the integration of technology – especially CMC – in language learning and teaching, language proficiency testing and teacher education. He has been a member of the CLES national board since 2005.

### **Abstract**

---

This paper aims at presenting the French Higher Education Language Certificate (CLES – Certificat de compétences en Langues de l'Enseignement Supérieur), a task-based language assessment created in 2000 by the Ministry for Higher Education and based on the CEFR. It is currently available in 10 different languages and at three different levels (CLES 1, CLES 2 and CLES 3) corresponding to levels B1, B2 and C1 of the CEFR. The format of the test is that of a scenario which puts test takers in a realistic situation with a specific mission to complete, the completion of the mission requiring students to read texts, listen to documents, write and speak/interact. In addition to the fact that CLES is available to university students for free, it also has a number of advantages over other better-known private language tests which will be presented.

After a brief presentation of the context in which CLES was developed, the theoretical framework, test specifications and national organisation of the certification will be examined. Then, we will discuss the validity, reliability and feasibility of the test, along with aspects that could potentially be improved, some of which as a direct consequence of the CEFR descriptors. Finally, the positive washback effect of the test on language teaching at university level – which far outweighs any of the issues raised – will be dealt with.

### **Short paper**

---

#### **Introduction**

Language teaching in French Higher Education for students specializing in subjects other than languages has always been very different from one university to the other - different contact times, different number of modules per degree, different ECTS credits attributed, etc. – mainly because the official recommendations from the Ministry for Higher Education have always been purposefully vague: the 2011 decree (which modified the 2002 original decree) setting out recommendations for Bachelor's degrees states that universities must give students the means to acquire language skills which will enable them to "read, write and speak in at least one foreign language" (Legifrance, 2011) and leaves it to the universities to organize the way they will provide training to reach these objectives as it doesn't give any detail. Interestingly, oral comprehension (listening) doesn't seem to be a skill worth developing in Higher Education as it isn't mentioned... As

for Master's degrees, they have to include training which will enable students to "validate their mastery of at least one foreign language" (Legifrance, 2002), but the official recommendations fail to mention how this validation should take place – with an exam, a language certification? – and at what level – as it isn't clear what "mastering" a language means... In spite of these very inexplicit recommendations which led to very different local organizations and, consequently, degrees, things have recently started to evolve towards common practices mainly thanks to the introduction of CLES (Certificat de compétences en Langues de l'Enseignement Supérieur), the French Higher Education Language Certificate.

CLES was created in 2000 as a result of the Ministry's proactive policy consisting in promoting language learning in Higher Education and was considered to be a possible answer to several problems:

1. The language tests available at the time<sup>1</sup> were not considered to be accurate enough indicators of a student's ability to communicate in real life situations in a foreign language, as most of them did not directly assess productive skills (writing, speaking, interacting) but offered extrapolated correlations about the test-takers' proficiency level in the non-tested productive skills from their level in the tested receptive skills. These extrapolations were not satisfactory for many prospective employers who ended up hiring graduates with high test scores who could not use the language in everyday work situations. This has recently been confirmed by Liao et al.'s 2010 study, commissioned by ETS, whose objective was to find correlations between the proficiency levels in the TOEIC listening and reading test and those in the TOEIC speaking and writing test. Indeed, the study concluded that "distinct aspects of language proficiency (...) cannot be adequately assessed by other tests" (p.11), in other words, it is not possible to accurately extrapolate language proficiency levels in specific skills (writing or speaking, for example) from the assessment of other skills (reading or writing), contrary to what was claimed for almost 30 years.
2. The language proficiency tests available were mostly provided by private companies. Resorting to commercial tests in Higher Education was – and still is – somewhat ethically problematic for those teaching in public/state universities, as (i) teachers then become instrumental to the commercial success of these companies through the use of public resources (teacher time and facilities for the administration of the test) and as (ii) subcontracting is a way of acknowledging that language teachers in this sector are not capable of testing their students' proficiency level, which couldn't be further from the truth.
3. None of the tests available at the time made it possible to assess students' language proficiency in academic contexts using the same format for several languages.

The CLES experiment was first launched in 2000: the first experimental phase took place between 2000 and 2002 and was followed by a second experimental phase from 2003 to 2005 which involved over 8,000 test-takers. Then, in 2007, CLES was officially recognized as a viable and operational language certification that could be implemented nationwide through a new ministerial decree. CLES is the result of the fruitful collaboration between language teachers and researchers who believe in the need for and viability of a non-commercial public language certification in French Higher Education.

### **CLES description**

#### **Theoretical framework**

As CLES was developed concomitantly with the publication of the Common European Framework of Reference for Languages (CEFR), it is grounded in the action-oriented

---

<sup>1</sup> The TOEIC, for example, rebranded "Listening and Reading", "Speaking and Writing" and "Four skills" in 2006, only existed in its basic version which did not assess productive skills.

approach: the basic principle of CLES papers is that they are presented in the form of a scenario, which means that all parts of the test are interlinked and that they all lead to the completion of a clearly stated mission within a specific context. This, of course, shows that CLES-takers are considered as "social agents, i.e. members of society who have tasks (not exclusively language-related) to accomplish in a given set of circumstances, in a specific environment and within a particular field of action" (CEFR, 2001, p.9). The scenario thus provides the "wider social context" (circumstances, environment and field of action) "which alone is able to give [tasks] their full meaning" (CEFR, 2001, p.9), as well as the realistic mission (macro-task) to accomplish. The action-oriented approach, when applied to assessment, implicitly refers to Task-Based Language Assessment (TBLA) which consists in "evaluating, in relation to a set of explicitly stated criteria, the quality of communicative performances elicited from learners as part of goal-directed, meaning-focused language use requiring the integration of skills and knowledge" (Brindley, 1994, p.74). Accordingly, assessment tasks are viewed as "devices for eliciting and evaluating communicative performances from learners in the context of language use that is meaning-focused and directed towards some specific goal" (Ellis, 2003, p.279). The basic assumption of TBLA is that linguistic competence (knowledge of vocabulary and grammar) is not enough to be able to achieve goals in real-life social situations as communicative competence also includes sociolinguistic, pragmatic and discursive skills (Mislevy et al. 2002). These are elements which are taken into account in the language performance assessment of CLES-takers.

### **Test specifications**

CLES scenarios start with the situation description, that is the realistic context that sets the stage for language use throughout the test and that explicitly mentions the role that test-takers have to take on, as well as the mission they have to complete. In order to complete their mission, which takes the form of written and oral language products, they are required to read and watch/listen to a set of authentic documents (text, video and/or audio files). The oral and written comprehension parts of the test enable test-takers to notice and extract from the documents essential content (ideas and lexis/structures) that will be useful for the completion of their mission: the comprehension documents thus serve as a source of input, just as the comprehension tasks serve the production tasks.

CLES allows students to certify their language proficiency at three different levels: CLES 1 (level B1), CLES 2 (level B2) and CLES 3 (level C1). CLES's relative youth, which is sometimes put forward as a drawback by its opponents, turns out to be a real upside as it stems from the CEFR and the Council of Europe's earlier publications: not only are the CLES proficiency levels based on the CEFR's common reference levels – unlike most language tests which had to subsequently try to find correlations between their own scoring systems and the CEFR's levels – but the marking schemes used for the production parts of the test are also based on the CEFR's language proficiency descriptors for each of the three levels. As for the contexts and themes chosen to develop CLES scenarios, they also have their source in the CEFR and the contexts of language use described in the global scale (CEFR, 2001, p.24): as level B1 is associated with "familiar matters regularly encountered in work, school, leisure" and "situations likely to arise whilst travelling in an area where the language is spoken", CLES 1 scenarios deal with situations connected to living and studying abroad; since level B2 mentions "complex texts on both concrete and abstract topics", "a wide range of subjects" and the ability to "explain a viewpoint on a topical issue giving the advantages and disadvantages of various options", CLES 2 scenarios deal with general topical issues, within the context of Higher Education, presented from different perspectives; finally, level C1 mentions "a wide range of demanding, longer texts" with "implicit meaning", "complex subjects" and the ability to use the language "for social, academic and professional purposes", which explains why CLES 3 scenarios deal with topics within the students' specialist field presented from an academic perspective.

In terms of skills tested, CLES 1 assesses language proficiency in reading, listening, writing and speaking, while CLES 2 assesses reading, listening, writing and interacting, and CLES 3 assesses all five skills. CLES is thus a complete language certification as it directly assesses all five skills. However, CLES is a new kind of language test as the assessment of receptive skills is item-based at levels B1 and B2 (not at level C1), whereas the assessment of productive skills is task-based (at all three levels). In addition, CLES is a multilingual certification: students' language proficiency can be assessed in 10 different languages (English, German, Spanish, Portuguese, Italian, Arab, Greek, Russian, Polish and Chinese) using the same paper format and topics regardless of the language. CLES is thus a tool that can play a part in the assessment of students' "plurilingual competence", as advocated by the authors of the CEFR (2001, p.133).

### **National organization**

CLES's non-commercial business model rests on a national network of over 50 accredited universities and Higher Education institutions, organized in nine regional groups: after their accreditation by the Ministry for Higher Education, universities receive extra funding for the implementation of CLES, as the objective is to make sure that CLES is offered to students free of charge.

The CLES network is coordinated by a national board comprising 12 members and is in charge of the entire certification process, from paper writing to paper rating, as the pooling of strengths and resources is at the heart of the process. As soon as a university becomes an accredited CLES centre, they enter the network and commit themselves to contributing to the running of the certification at the national level, starting with paper writing: test papers are (a) written by local teams following a strict test specification document (and after attending specific training), then (b) evaluated by a pair of experimented paper writers who make recommendations and ask for adjustments, and finally (c) validated by the national paper validation committee. Once validated, papers become part of the national bank and can be used in any CLES centre. CLES paper writers are paid by their own institution, the rule being that a CLES centre has to write one paper for every five certification sessions organized, which means that these costs can be integrated in local CLES budgets from the start.

Whenever a CLES centre wishes to organize a certification session, they have to "place an order" with the national coordinator who checks that they are accredited, that they have produced – or are in the process of producing – the correct number of papers, and then gives them access to a paper from the national bank (preferably not a paper written by the requesting centre). This just shows to what extent inter-institutional collaboration is at the basis of the CLES organization. Each member of the CLES network is bound to the others by a moral and professional commitment, not by commercial ties.

### **Discussion**

#### **Validity and reliability**

If CLES seems to possess the three primary qualities of a communicative language test as defined by Fulcher (2000) (as it involves performance, has an authentic communicative purpose and is scored on real-life outcomes), the issues that need discussing here are those traditionally associated to language assessment (CEFR, 2001, p.177), namely validity and reliability.

Validity first concerns test construct, i.e. the fact that "what is actually assessed (the construct) is what, in the context concerned, should be assessed" (CEFR, 2001, p.177). Given that CLES rests on the notion of communicative competence (see section II.1.), the four components of communicative competence are taken into consideration in the assessment, as illustrated in the CLES production assessment grids which clearly include linguistic, discursive, sociolinguistic and pragmatic descriptors. Besides, CLES aims at assessing language proficiency in academic contexts, which is exactly what it does given the scenario topics chosen and the true-to-life situations test-takers are put in (see

section II.2.). In this respect, CLES seems to be “a way of achieving a close correlation between the test performance, i.e. what the testee does during the test, and the criterion performance, i.e. what the testee has to do in the real world, and thus of ensuring the validity of the assessment” (Ellis, 2003, p.279). Another aspect of test validity is that of face validity, that is “the extent to which the test is perceived as acceptable by stakeholders, including testees” (Ellis, 2003, p.282): on this particular point, feedback on CLES couldn’t be better as both stakeholders and students recognize the relevance of the test’s format and content which are viewed as closer to real world situations than other tests, and, consequently, as a more accurate way of assessing language proficiency. However, CLES suffers from its youth and relative lack of recognition: its face validity isn’t as good among prospective employers and higher education institutions outside France, simply because it isn’t well-known.

When it comes to test reliability, CLES is nothing like psychometric tests made up of closed questions which prioritize reliability and objectivity by using statistical procedures extensively. Consequently, reliability is a key issue for CLES, as it is for any type of assessment, as it deals with “the extent to which a test measures a candidate’s proficiency in an error-free manner” (Ellis, 2003, p.310). In other words, what needs to be ensured is that a repeat test or a second rating would give the same result (the same measure of proficiency). As CLES, in the TBLA tradition, requires test-takers to produce language with a specific objective and in a particular social context and from which proficiency is measured, reliability mainly depends on “the accuracy of decisions made in relation to a standard” (CEFR, 2001, p.177) as human judgment is involved since raters have to make binary decisions (pass/fail) for each of the descriptors included on the evaluation grids. The problem here is clearly to limit rater subjectivity in order to approach objectivity, which can be done by providing “a rating scale, set of task requirements and marking criteria” (Milanovic, 2002, p.32), as well as “a brief description (...) of a typical performance” (ibid. p.33). Another idea is to “accompany descriptors of performance with actual examples of candidates’ work” (ibid.). Ellis goes further as he sets out a list of four possible solutions to enhance TBLA’s reliability (2003, p.311): making the test longer (to ensure provision of larger samples of language use), using two raters, training raters, and adjusting test scores thanks to statistical analysis. CLES has addressed the issue of reliability through several measures:

1. there are two versions of the evaluation grids – one simple one, and one which includes sample productions considered acceptable;
2. the answer key includes elements expected from a typical acceptable performance;
3. all raters have to follow specific training which aims at setting standards for the different task types (writing, speaking, interacting);
4. for the assessment of the oral interaction part of CLES 2, two raters are recommended.

### **Room for improvement**

Although the CEFR’s input has been invaluable in the development of CLES, its link to the CEFR is also a source for potential problems:

1. the CEFR descriptors sometimes lack elaboration, especially when it comes to linguistic competence associated to each level;
2. it seems difficult to establish clear cut-off points between the different levels when it comes to linguistic competence;
3. the multiplicity of scales for a given language activity in the CEFR also makes it difficult to get a global view of what a user/learner can do at a given level (in spite of the general scale).

This can be explained by the fact that the CEFR is “a point of reference, not a practical assessment tool” (CEFR, 2001, p.178), and it can also be considered as a work in



progress, as it states that there are "gaps in the descriptors provided", and that, for specific areas, "descriptors could presumably be written, but haven't been" (CEFR, 2001, p.37).

Another issue, although not directly related to the CEFR, is that of the conception of the test itself: CLES is a horizontal test, as it assesses language proficiency in the tested skills at a given level, unlike vertical tests which assess the test-takers' level on a continuum and grants them different proficiency levels for the different skills tested. The direct consequence is that test-takers need a pass for all the tested skills to get their certificate (a test-taker who has performed at level B2 in reading, listening and writing, but at B1 only in speaking won't be awarded the CLES 2 certificate). This obviously has a strong impact on CLES pass rates which are about 40% overall.

### **Conclusion**

With 400 CLES sessions organized in 50 CLES-accredited centres in 2011 and a total of 35,000 test-takers, the CLES dynamics seems to attract growing interest from both students and language professionals who, for the first time, feel that they are an important link in the certification chain and that they can make a difference. Indeed, as test development is an iterative process, CLES is constantly evolving as more data becomes available and as further adjustments thus appear necessary.

In spite of the issues raised (especially as regards test reliability and reliability of the CEFR descriptors), CLES's gain in validity and its tremendously positive washback effect on language teaching methods far outweigh its drawbacks, if we bear in mind the fact that examinations always have a direct or indirect effect on teaching methods and that this effect can be either positive or negative (Heaton, 1990). Indeed, in the past, language professionals in French Higher Education could be trapped in an exam preparation cycle which involved doing a lot of past papers for certain psychometric tests which were sometimes made a requirement for students to graduate. In this case, the testing procedure (psychometric test) had negative effects on classroom practice (test preparation through extensive use of past papers). CLES's washback effect, on the contrary, is very positive as (1) it has contributed to the introduction of the CEFR in Higher Education, (2) it has encouraged language professionals and institutions to offer language courses with clearer objectives expressed in terms of proficiency levels, (3) it has had an impact on classroom practice which now includes the implementation, more often than not, of the action-oriented approach and task-based language teaching (TBLT). For all these reasons, it is strongly believed that CLES could provide a model framework for 21<sup>st</sup> century Higher Education language certification development and increase its European dimension by developing partnerships with institutions outside France.

Still, more research is needed to seek satisfactory solutions to the problems associated with TBLA in general (Ellis, 2003, p.311) and CLES in particular. This is one of the missions of the CLES scientific committee whose task will be to initiate and support various research projects in years to come. In this respect, a new framework of reference for language testing could be very useful.

### **References**

---

Brindley, G. (1994). Task-centered assessment in language learning : The promise and the challenge. In N. Bird, P. Falvey, A. Tsui, D. Allsion & A. McNeill (eds), *Language and learning*, (pp.73-94). Hong Kong: Hong Kong institute of language in education, Hong Kong education department.

Council of Europe. (2001). *Common European Framework of Reference for Languages*.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

- Fulcher, G. (2000). The "communicative" legacy in language testing. *System*, 28, 483-497.
- Heaton, J.D. (1990). *Writing English language tests*. London: Longman.
- Legifrance. (2002). Arrêté du 25 avril 2002 relatif au diplôme national de master. Retrieved from <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT0000000771847&dateTexte=&categorieLien=id>
- Legifrance. (2011). Arrêté du 1er août 2011 relatif à la licence. Retrieved from <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000024457754&dateTexte=&categorieLien=id>
- Liao, C., Qu, Y. & Morgan, R. (2010). The relationships of test scores measured by the TOEIC listening and reading test and TOEIC speaking and writing tests. *TOEIC Compendium*. Retrieved from <http://www.ets.org/Media/Research/pdf/TC-10-13.pdf>
- Milanovic, M. (2002). *Language examining and test development*. Strasbourg: Language policy division, Council of Europe. Retrieved from [http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Publications/Language\\_examining\\_EN.pdf](http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Publications/Language_examining_EN.pdf)
- Mislevy, R., Steinberg, L. S. & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, 19 (4), pp.477-496.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.