



Report

Sylvain Coulange*, Marie-Pierre Jouannaud, Cristiana Cervini and Monica Masperi

From placement to diagnostic testing: Improving feedback to learners and other stakeholders in SELF (Système d'Evaluation en Langues à visée Formative)

<https://doi.org/10.1515/cercles-2020-2015>

Abstract: Since 2012 an interdisciplinary and culturally heterogeneous team composed of more than 30 people has been engaged in the complex process of conceiving, designing and validating an online placement test with formative orientation called SELF (Système d'Evaluation en Langues à visée Formative), developed and already deployed in six different languages – Italian and English as pilots, followed by French, Mandarin, Japanese and Spanish. Its results are used to form groups and classes of similar ability, or to identify students' strengths and weaknesses in three macro skills (listening, reading, limited writing). In this report, we describe the steps the multilingual team is currently taking to transform SELF into a diagnostic test that will fulfill its original formative purpose and provide students and other stakeholders with more precise information about their performance. This can be done in two ways, by using the data automatically recorded by the online administration platform more thoroughly and by enriching user feedback with clear and informative graphics. This will enhance the validity of our test, and help close the gap between testing and learning.

Keywords: placement test, diagnostic test, data visualization, learner profiles, feedback

*Corresponding author: Sylvain Coulange, Université Grenoble Alpes, Grenoble, France, E-mail: sylvain.coulange@univ-grenoble-alpes.fr

Marie-Pierre Jouannaud and Monica Masperi: Université Grenoble Alpes, Grenoble, France, E-mail: marie-pierre.jouannaud@univ-grenoble-alpes.fr (M.P. Jouannaud), monica.masperi@univ-grenoble-alpes.fr (M. Masperi)

Cristiana Cervini: Università di Bologna, Bologna, Italy, E-mail: cristiana.cervini@unibo.it

1 Description of context

At Université Grenoble Alpes (France), courses in more than 20 foreign languages are on offer, in different modalities: face-to-face, blended or entirely through distance learning, and in semester-, year-long or intensive formats. In this context, and with thousands of students each year taking foreign languages as an elective, it is important to organize enrollments as efficiently as possible. Placement testing is an essential part of this process.

In 2012, a multilingual team started working on SELF, a placement test that would be available in several languages, and open to students from partner universities in France and abroad. The idea was to develop and document a methodology and tools that colleagues teaching other languages could later copy. The project was supported by a research grant¹ and is now operational in Italian, English, Japanese, Spanish, Mandarin, and French as a Foreign Language. SELF uses an item bank where each item is tagged by (among others) language, macro skill/language activity (listening, reading or writing), language focus (morpho-syntax, lexis, pragmatics etc.), discourse type, observed difficulty during pretesting, and CEFR level, determined by a standard setting procedure with a panel of experts. It is a semi-adaptive multi-stage test. The first stage (the initial testlet) is common to all test takers, but the items in the second stage depend on test takers' results in the first. Results in the second stage are used to refine the estimation of learners' level and arrive at placement results that are as reliable as possible.

SELF, however, is not just a placement tool. Its initials stand for “Système d’Evaluation en Langues à visée Formative”, i.e., a foreign language assessment system with formative orientation. Its goal is to help students realize where their strengths and weaknesses lie by giving them more precise feedback than just the group they should be placed in, in order to enable them to work, if desired, on their weak points and ultimately improve their foreign language skills. This feedback can be called diagnostic since “diagnostic tests seek to identify those areas in which a student needs further help. These tests can be fairly general, and show, for example, whether a student needs particular help with one of the four main language skills; or they can be more specific, seeking perhaps to identify weaknesses in a student’s use of grammar” (Alderson et al. 1995: 12). At present, the feedback is rather limited, providing information on students’ level in each skill targeted by the test (listening, reading, and basic writing skills), i.e., general diagnostic information in Alderson’s terms, but does not provide more specific feedback. Other stakeholders that could be given more information are institutions (language

1 Projet IDEFI Innovalangues (ANR-11-IDFI-0024, Masperi 2011).



Figure 1: First half of screenshot of the final user result page in SELF: Course placement information (proposed course level enrollment).



Figure 2: Second half of screenshot of the final user result page in SELF (level by subskill).

centers at partner universities), using SELF to place their students. The following will provide an account of the steps that are currently being taken to optimize user feedback in these areas.

2 Account of activity: Development of diagnostic feedback

Currently, SELF users receive placement information about the course level they should enroll in (see Figure 1), ranging from A1 to C1/C2. The system does not distinguish between C levels (expert users), and lower levels are divided into sublevels (B1.1 and B1.2 for B1, for example).

Users are also provided with information about their level in the three skills targeted by SELF (see Figure 2), and can thus see whether they have a balanced learner profile with similar levels in all three skills, or whether they need to focus their efforts more heavily on one skill, for example, depending on what their ultimate goal is. The CEFR explicitly mentions “uneven profiles, partial competencies”, and one of its achievements was to allow for this possibility and give instructors the tools to report it (Council of Europe 2001: 17).

This information can be downloaded and printed by students, but the interaction with the assessment system does not currently go any further. This is unfortunate, since there is potentially a lot more data stored by the system that could be used to provide more detailed information. The importance of detailed feedback has been emphasized by many researchers working on diagnosis: “for assessment

information to be used effectively it needs to be detailed, innovative, relevant and diagnostic, and to address a variety of dimensions rather than being collapsed into one general score” (Shohamy 1992: 515).

For this reason, we are currently developing user dashboards which will result in much more interactive feedback and provide a richer experience to users. A computerized system is ideal to provide this kind of experience, since data storage is automatic and storage capacity (for our purposes) is close to unlimited. What is needed is a way to convey the information to the user in a clear and useful way. We propose to do this by converting the data to visual information. Data visualization is an extremely powerful tool (Larson-Hall 2017; Tufte 2001), and it will allow us to provide learners with more precise information about their scores, give them access to the items they got right or wrong, show them how long they took to answer each question, and what item types they found easy or hard (see Figure 3). Since information about each test administration will be recorded, this will also help students track their progress over time and adjust their goals as a result. This might result in positive washback, if the provision of visual feedback motivates students to improve and see for themselves the upward trend in their results (as in online games where score boards fulfill the same purpose).

Let us now look more closely at Figure 3 (bottom section). As mentioned earlier, question items in SELF are identified not just by the skill they correspond to, but also by other characteristics such as their language focus and their discourse type. The language focus is the critical information that designers believe test takers need to understand in order to answer the question correctly. This might be an element of morpho-syntax (for example, identifying a past tense marker to understand that the time reference is past, when no other elements provide this information), or lexis (understanding a key term), or pragmatic intention (the illocutionary force of an utterance, for example, a refusal disguised as a question). The discourse type is the prevalent genre of the text the item bears on (narrative, informative, argumentative, etc.). Each item is linked to a text which test takers need to process to answer the question correctly. We hypothesize that familiarity with different discourse types is likely to affect success (Cervini and Jouannaud 2015). Since each test item is defined by these characteristics, we can calculate the percentage of items successfully attempted for each language focus and each discourse type and display this information with a spider chart representing the test taker’s strengths and weaknesses in each area. In our example (bottom of Figure 3), the learner is much better at working with informative discourse types than with narratives. A student majoring in science and destined to work with other genres may not feel this is a problem. However, given the centrality

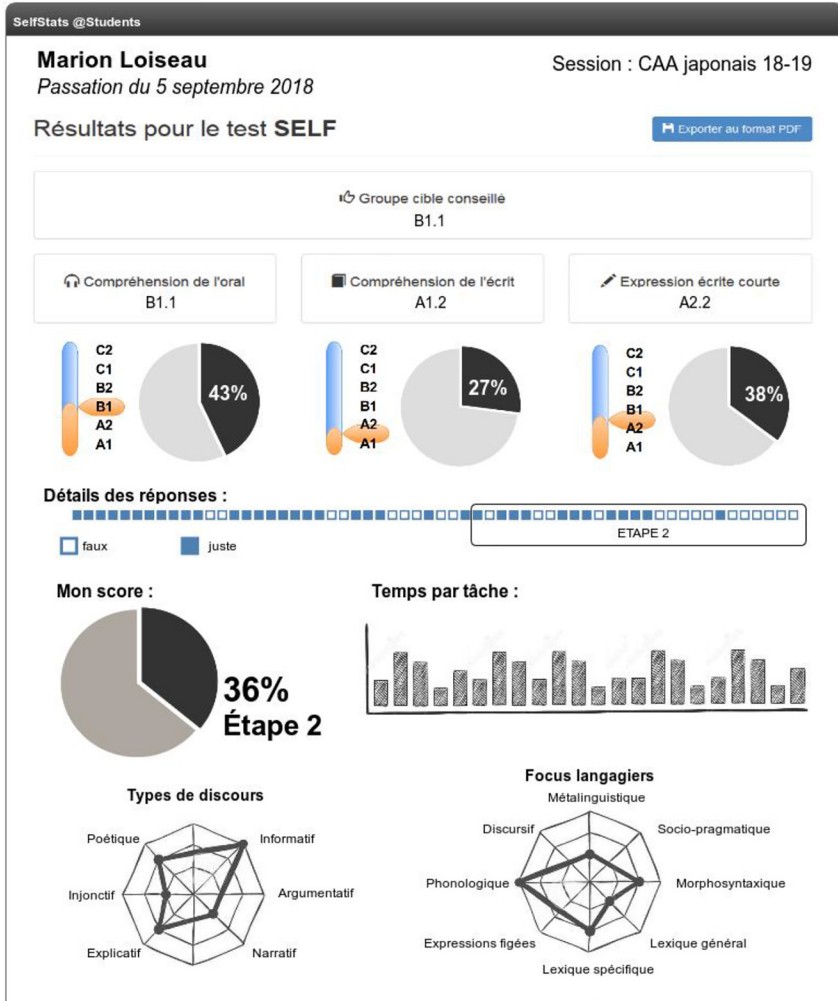


Figure 3: Prototypé of student dashboard.

of narratives to our human experience, they would probably be advised to allocate some of their language learning efforts to this area. The final decision, however, rests with the language learner, who is ultimately responsible for their own learning.

Dashboards will also improve practicality for other stakeholders, such as instructors or administrators. Practicality is one of the components of test usefulness according to Bachman and Palmer (1996). It is important for administrators in

particular to understand what the test is about or how it should be used, as they ensure that correct decisions are taken following test administration. Current test session feedback for groups (for example, for all science students wanting to take a Japanese course in the first semester) is provided in spreadsheet format (see Figure 4). It summarizes information from individual administrations: for each student, we get personal identification details entered when registering on the SELF assessment platform (first and last names, email address, degree prepared, major), as well as information about test administration (date, time taken, results in terms of both placement, and subskill level).

However, spreadsheets are not always easy to read and they do not summarize the information contained in them. Administrators (and teachers) looking at a spreadsheet cannot tell at a glance how well the group did. Additional manipulation is required to obtain information about the central tendency or spread for each of the variables contained in the spreadsheet. We intend to automatize the process and provide the results in visual format (see Figure 5). This is a prototype still under development, and it might need to be tweaked to make sure the interface is not too cluttered, as it has been shown that too much information is detrimental to the uptake of feedback (Goodman and Hambleton 2004). Once the prototype is developed, we will be able to pilot it with a sample of potential users. At present, we display for each group session the number of participants (top right), and, from top to bottom, a calendar showing the spread of administrations over time, leaf-plots (horizontal histograms) for group placement and skill results, and boxplots for time taken to complete each part of the test. Access to individual results would still be possible, and would be made interactive: clicking on a line representing one student's results (bottom of Figure 5) would send the administrator to this student's dashboard.

The dashboards presented above are the first step toward transforming SELF, currently mostly used for placement purposes, into a diagnostic test that will provide much more information for each test taker to act upon. This transformation is not new, and other researchers have tried to exploit the data collected during administration of large-scale proficiency tests, for example, to present diagnostic information to learners as well as general results (Buck and Tatsuoka 1998; Liu 2015). However, this procedure (known as cognitive diagnostic assessment or CDA) requires sophisticated statistical analyses, and its use has not been taken up on a wide scale. Our method is much simpler, and simply proposes a way to visualize the raw data recorded by the administration platform using informative graphics (Larson-Hall 2017). We use free and open-source software for this purpose: R for uploading, cleaning, selecting and analyzing the data, and JavaScript (with the D3js library) for the production of dynamic and interactive graphics.

SELF japonais v1 Nom d'utilisateur	LEA japonais 2017 Nom	Prénom	Email	Etablissement Filière	Spécialité	Année	Début	Durée totale	Score CE	Score CE	Score EEC	Test fini	Langue matre	Autre langue de référence
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	07/07/2017 11:32	5 En route vers A1.1 en cours	A1.2	A1.1	A1.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	07/07/2017 09:51	76 En route vers A1.1	A1.2	A1.1	A1.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	07/07/2017 15:58	172 En route vers B1.2	B1.2	A2.2	A2.2	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	07/07/2017 17:52	1 En route vers A1.2	A1.2	A1.1	A1.1	non	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	07/07/2017 15:44	1 En route vers A1.2	A1.2	A1.1	A1.1	non	français	français
				UNIVERSITE UPR Langues étrangères (LI débutant	Langues étrangères (LI débutant	LI débutant	08/07/2017 15:03	18 En route vers A2.1	A1.2	A1.2	A1.2	oui	français	anglais
				UNIVERSITE UPR Langues étrangères (LI débutant	Langues étrangères (LI débutant	LI débutant	08/07/2017 18:40	19 En route vers A1.2	A1.1	A1.1	A1.1 en cours	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	09/07/2017 11:14	63 En route vers A2.2	A2.2	A2.2	A2.2	oui	français	anglais
				UNIVERSITE Autre (non précisée)	Langues étrangères (non précisée)	1ère année	11/07/2017 15:39	12				non	français	français
				UNIVERSITE UPR Langues étrangères (LI débutant	Langues étrangères (LI débutant	LI débutant	13/07/2017 16:27	80 En route vers A2.1	A1.2	A1.2	A1.2	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	20/07/2017 15:27	44 En route vers B1.2	B1.2	A2.2	A2.2	oui	français	français
				UNIVERSITE UPR Langues Autre	Langues Autre	1ère année	20/07/2017 15:27	35 En route vers A1.2	A1.1	A1.1	A1.1	oui	français	créole
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	21/07/2017 16:31	18 En route vers A1.1	A1.1	A1.1	A1.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	28/07/2017 15:21	46 En route vers B1.1	A2.2	A2.1	A2.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	28/07/2017 15:23	56				non	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	27/07/2017 12:42	5 En route vers A1.1	A1.1	A1.1	A1.1	non	français	anglais
				UNIVERSITE UPR Langues étrangères (LI débutant	Langues étrangères (LI débutant	LI débutant	12/08/2017 16:04	65 En route vers B1.1	B1.1	A2.1	A2.1	oui	français	français
				UNIVERSITE Autre (non précisée)	Langues étrangères (LI débutant	LI débutant	13/08/2017 17:25	25 En route vers A2.1	A1.1	A1.1	A1.1	oui	français	français
				UNIVERSITE UPR Langage, lettres arts (1ère année	Langues, lettres arts (1ère année	1ère année	15/08/2017 16:39	15 En route vers A1.1 en cours	A1.1	A1.1	A1.1	oui	français	français
				UNIVERSITE UPR Langage, lettres arts (1ère année	Langues, lettres arts (1ère année	1ère année	18/08/2017 15:00	211 En route vers B1.1	B1.2	A2.2	A2.2	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	22/08/2017 11:15	23 En route vers A2.1	A1.2	A2.1	A2.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	23/08/2017 16:57	55 En route vers A1.1	B1.1	A1.2	A1.2	oui	français	français
				UNIVERSITE UPR Langues étrangères (LI débutant	Langues étrangères (LI débutant	LI débutant	24/08/2017 19:15	80 En route vers B1.1	B1.1	A2.2	A2.2	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	27/08/2017 09:13	79 En route vers B1.1	B1.1	A2.2	A2.2	oui	français	français
				UNIVERSITE UPR Langues étrangères (LI débutant	Langues étrangères (LI débutant	LI débutant	27/08/2017 15:23	108 En route vers A2.2	A2.2	A2.1	A2.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (LI débutant	Langues étrangères (LI débutant	LI débutant	28/08/2017 15:23	26 En route vers A2.2	A2.2	A2.1	A2.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	29/08/2017 16:13	122 En route vers A2.2	A2.2	A2.1	A2.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (LI débutant	Langues étrangères (LI débutant	LI débutant	29/08/2017 11:32	14 En route vers A1.1	A1.1	A1.1	A1.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	30/08/2017 11:51	121 En route vers A2.2	A2.2	A2.2	A2.2	oui	français	espagnol
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	30/08/2017 11:53	28 En route vers A1.2	A1.2	A1.1	A1.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	30/08/2017 14:33	35 En route vers A2.2	A2.1	A2.1	A2.1	oui	français	français
				UNIVERSITE UPR Langues Autre	Langues Autre	1ère année	30/08/2017 14:54	8 En route vers A1.2	A1.2	A1.1 en cours	A1.1	oui	français	français
				UNIVERSITE UPR Langues étrangères (1ère année	Langues étrangères (1ère année	1ère année	30/08/2017 18:23	24 En route vers A1.2	A1.1	A1.1	A1.1	oui	français	français

Figure 4: Current feedback to institutions.

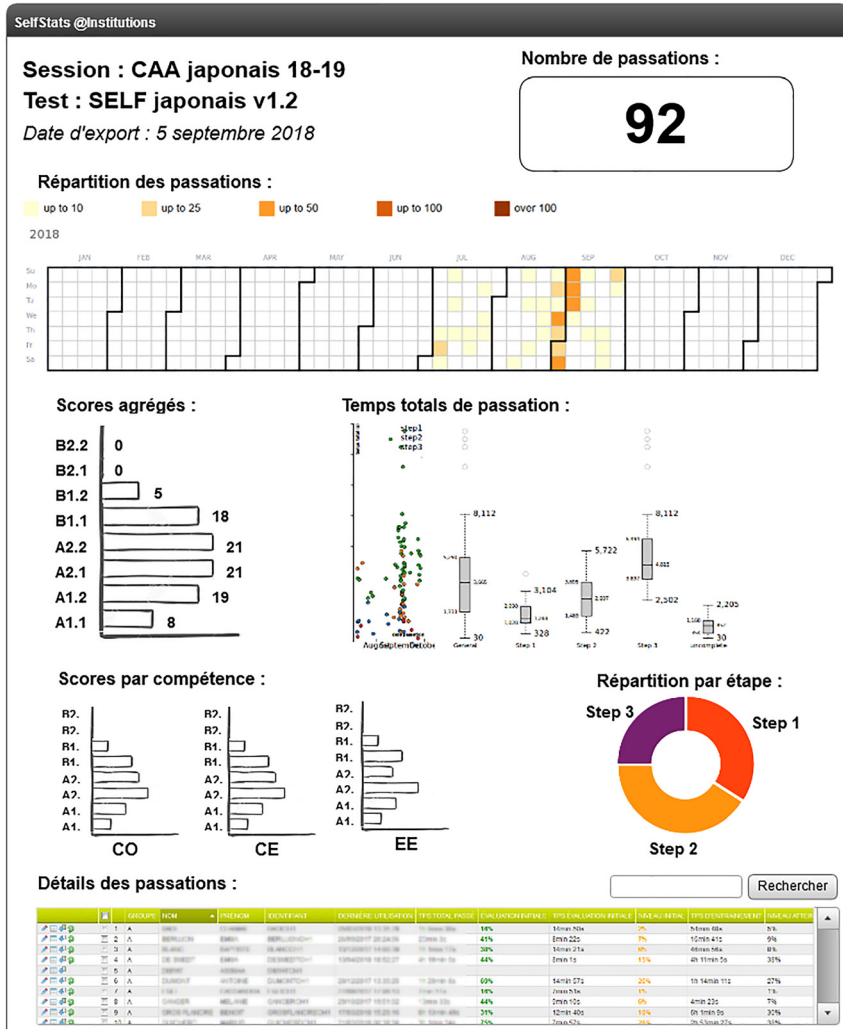


Figure 5: Prototype of dashboard for institutions.

3 Conclusion and future prospects

SELF is currently used by more than 25 French universities and language centers as a CEFR-based placement test. Since 2016, when it became fully operational, more than 90,000 students have taken the test in one (or more) of its six foreign languages. The results are used to form groups and classes of similar ability, or to

identify students' strengths and weaknesses in three macro skills. In this report, we have described the steps the multilingual team is currently taking to transform SELF into a more thorough diagnostic test that will fulfill its original formative purpose and provide students and other stakeholders with more precise information about their performance. This can be done in two ways: by using the data automatically recorded by the online administration platform more thoroughly and by enriching the feedback experience with it. Test administration data currently used for user feedback include total score and item characteristics in terms of level and targeted skill (so that a separate score is reported for each skill and scores are translated into corresponding CEFR levels). Additional data we intend to exploit include individual item scores, time spent on each item/ on the whole test, and more item characteristics such as language focus and discourse type. This additional information will be used to improve the feedback provided by the platform, in terms of quantity of information, usability, and comprehensibility. The use of clear and interactive graphics, as advocated by data analysts, will (we hope) help learners make sense of the results, motivate them to improve their skills, and lead them to make informed decisions based on the results displayed.

Once the prototypes have been developed and piloted with a sample of users, we will continue to explore other avenues for the provision of diagnostic feedback. One is the collection of more data through the development of more specific tests. Students with weak results in one area might want to take a further test targeting this area in more detail (for example, a phoneme discrimination test for students with weak results on items with a phonological language focus). The other avenue is to explore the use of more sophisticated data analysis techniques. One we are considering is the unsupervised clustering of students based on their answers using Latent Block Modeling (Brault and Mariadassou 2015). This will enable us to try to define learner profiles, and perhaps offer them feedback (semi)automatically. The final idea is to try to strengthen the link between assessment and learning by providing students access to online remedial modules based on their current learner profile (Alderson 2007; Masperi and Quintin 2014).

References

- Alderson, Charles, Caroline Clapham & Dianne Wall. 1995. *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Alderson, Charles. 2007. The challenge of (Diagnostic) testing: Do we know what we are measuring? In Janna Fox, Mari Wesche, Dreen Bayliss, (eds.), *Language testing reconsidered*, 21–40. Ottawa: University of Ottawa Press.
- Bachman, Lyle & Adrian Palmer. 1996. *Language testing in practice: designing and developing useful language tests*, vols. 1–1. Oxford: Oxford University Press.

- Brault, Vincent & Mahendra Mariadassou. 2015. Co-clustering through latent bloc model: A review. *Journal de La Société Française de Statistique* 156(3). 120–139.
- Buck, Gary & Kikumi Tatsuoka. 1998. Application of the rule-space procedure to language testing: examining attributes of a free response listening test. *Language Testing* 15(2). 119–157.
- Cervini, Cristiana & Marie-Pierre Jouannaud. 2015. Ouvertures et tensions liées à la conception d'un système d'évaluation numérique multilingue en ligne dans une perspective communicative et actionnelle. *ALSIC – Apprentissage Des Langues et Systèmes d'information et de Communication, Association pour le Développement de l'Apprentissage des Langues par les Systèmes d'Information et de Communication*. OpenEdition. <https://doi.org/10.4000/alsic.2821>.
- Council of Europe. 2001. *Common European framework of reference for languages*. Cambridge: Cambridge University Press.
- Goodman, Dean & Ronald Hambleton. 2004. Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education* 17(2). 145–220.
- Larson–Hall, Jennifer. 2017. Moving beyond the bar plot and the line graph to create informative and attractive graphics. *The Modern Language Journal* 101(1). 244–270.
- Liu, Heidi Han-Ting. 2015. The conceptualization and operationalization of diagnostic testing in second and foreign language assessment. *Working Papers in TESOL and Applied Linguistics* 14. 1–12.
- Masperi, Monica. 2011. *Innovalangues: Innovation et transformation des pratiques de l'enseignement-apprentissage des langues dans l'enseignement supérieur*. Rapport de recherche UGA, Université Grenoble Alpes. (hal-02000901).
- Masperi, Monica & Jean-Jacques Quintin. 2014. Enseigner à l'université en France, à l'ère du numérique: l'apport de dispositifs d'ingénierie innovants dans la formation en langues. In Cervini Cristiana & Valdivieso Anabel (eds.), *Dispositivi formativi e modalità ibride per l'apprendimento linguistico*, 61–80. Bologna: CLUEB.
- Shohamy, Elena. 1992. Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal* 76(4). 513–521.
- Tufte, Edward. 2001. *The visual display of quantitative information*, 2nd edn. Cheshire, Conn: Graphics Press USA.

Bionotes

Sylvain Coulange

Université Grenoble Alpes, Grenoble, France

Sylvain Coulange has been working as a pedagogical engineer for the Japanese team in the Innovalangues project (Université Grenoble Alpes) since 2015. He has taught French as a foreign language and is interested in the acquisition of second language phonology and speech processing.

Marie-Pierre Jouannaud

Université Grenoble Alpes, Grenoble, France

Marie-Pierre Jouannaud is a lecturer and teacher trainer at Université Grenoble Alpes. She was the English item writing coordinator for the SELF placement test developed by the Innovalangues project. She is interested in diagnostic testing and learner autonomy.

Cristiana Cervini

Università di Bologna, Bologna, Italy

Cristiana Cervini is an assistant professor in educational linguistics at the University of Bologna. She coordinated the development of the SELF placement test. Her present studies focus on assessment and evaluation and on CALL systems for hybrid and self-learning.

Monica Masperi

Université Grenoble Alpes, Grenoble, France

Monica Masperi is a senior lecturer in linguistics and didactics at Université Grenoble Alpes. She is the scientific director of the Innovalangues project, and her research focuses on Italian didactics, plurilingualism and the use of technology in language teaching and learning.