Piccardo, E. & North, B. (2017). Developing phonology descriptors for the Common European Framework of Reference (CEFR). In M. O'Brien & J. Levis (Eds). *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Calgary, AB, August 2016 (pp. 97-109). Ames, IA: Iowa State University.

DEVELOPING PHONOLOGY DESCRIPTORS FOR THE COMMON EUROPEAN FRAMEWORK OF REFERENCE (CEFR)

Enrica Piccardo, OISE-University of Toronto, Université Grenoble-Alpes – France
Brian North, Eurocentres, Eaquals

While pronunciation is the object of an increasing number of studies, the translation of this growing body of research into teacher-oriented resources is still at its initial stages (Derwing & Munro, 2015). Often still playing *de facto* a gatekeeping role in society, the importance of phonological competence is being more and more recognized, with explicit instruction increasingly appearing in the L2 classroom, and a focus on the definition of valid assessment criteria for pronunciation proficiency. However, there is still a lack of validated, concrete descriptors to support practitioners and learners alike. Although the Common European Framework of Reference for Languages (CEFR, 2001) presents a solid conceptualization of phonological competence, this did not aptly translate into the existing scale for *Phonological control*. In the context of a Council of Europe's project to update the CEFR descriptors, a new analytic scale was developed to replace the existing holistic one. This article reports on the methodology used for developing and validating the descriptors for *Overall phonological control*, *Sound articulation*, and *Prosodic features*. The rationale behind the scale is discussed together with the complex relationship between language proficiency and phonological competence, in the light of the notion of intelligibility.

CONTEXT

The *Common European Framework of Reference for Languages: Learning, teaching, assessment* (CEFR) (Council of Europe, 2001) is an influential language policy document in Europe and beyond, and its descriptor scales, empirically developed in a Swiss National Research Project (North & Schneider 1998; North 2000; Schneider & North 2000), are used as a source for the development of language standards. Among the wealth of scales offered by the CEFR, only one scale deals with phonological competence, the phonological control scale (Council of Europe, 2001, p. 117). Since the publication of the CEFR in 2001, unlike the other scales included in the document, the scale for phonological control has attracted a great deal the criticism, particularly from researchers interested in both phonology and the CEFR (see in particular Horner, 2010, 2013, 2014, Frost & O'Donnell, in press, Harding, 2016). This does not come as a real surprise, though, as problems had been encountered with the scaling of descriptors for phonology in the Swiss research project previously mentioned, partly because it "can involve an implicit negative concept: that of accent. Less accent is good, more accent is bad" (North 2000, p. 238). Furthermore, in the Rasch analysis to calibrate the descriptors, phonology did not 'fit' well in the construct of the rest of the descriptors for spoken English, and then 'misfitted' completely when other languages (French and German) were added.

The need for the creation of a new illustrative descriptor scale for phonological control to be integrated in the CEFR appeared both timely and critical. The opportunity to address

this issue was offered by a 2013–2016 Council of Europe project aimed at extending the CEFR illustrative descriptors. The focus of the overall project was to provide scales for some areas that had not been covered in the 2001 publication, particularly mediation. The conceptualization of mediation addressed aspects of language teaching pedagogy that had been present but underdeveloped in the CEFR (Piccardo, 2012) and took into consideration new, related perspectives that have been emerging in language education.

One underdeveloped area was precisely phonology. Although the CEFR presents a solid conceptualization of phonological competence, albeit necessarily adapted to an audience of language educators and not of phoneticians, this conceptualization was not adequately operationalized in the existing scale. Needless to say the existing scale did not incorporate the new vision of phonology that has developed since the mid-1990s when the CEFR research was undertaken, as will be discussed below. The project reported on aimed to address some of these issues and to provide realistic, concrete descriptor scales to support practitioners and learners alike. Work was undertaken in four phases: analysis of existing instruments, literature review, development and initial feedback, qualitative and quantitative validation.

ANALYSIS OF EXISTING INSTRUMENTS

Naturally the starting point was an analysis of the treatment of phonology in the CEFR. Phonological competence has an important place in the descriptive scheme of the CEFR. It is mentioned in the overview of **communicative competences** (Section 2.1.2, p. 22) and in the description of the **communicative language processes** (Section 4.5, pp. 90–91), when it comes to execution (Section 4.5.2) and especially Production (4.5.2.1):

> "The production process involves two components: The *formulation* component takes the output from the planning component and assembles it into linguistic form. This involves lexical, grammatical, phonological (and in the case of writing, orthographic) processes which are distinguishable and appear (e.g. in cases of dysphasia) to have some degree of independence but whose exact interrelation is not fully understood. The *articulation* component organises the motor innervation of the vocal apparatus to convert the output of the phonological processes into coordinated movements of the speech organs to produce a train of speech waves constituting the spoken utterance, or alternatively the motor innervation of the musculature of the hand to produce hand- written or typewritten text."
>
> (Council of Europe 2001, p. 91)

The main coverage then comes in the description of **communicative language competences** (Section 5.2), where it is detailed as follows:

> "5.2.1.4 Phonological competence involves a knowledge of, and skill in the perception and production of:

- the sound-units (*phonemes*) of the language and their realisation in particular contexts (*allophones*);
- the phonetic features which distinguish phonemes (*distinctive feature*s, e.g. voicing, rounding, nasality, plosion);
- the phonetic composition of words (*syllable structure*, the sequence of phonemes, word stress, word tones);
- sentence phonetics (*prosody*):
    - sentence stress and rhythm;
    - intonation;
- phonetic reduction:
    - vowel reduction;
    - strong and weak forms;
    - assimilation;
    - elision. (Council of Europe 2001, p. 116-117)

In addition, phonological control is listed among the 12 qualitative categories relevant to oral assessment (Section 9.4, p. 193). Finally, there is the descriptor scale already mentioned, which provides one descriptor per level from A1 to C1 (p.117). The strengths and weaknesses of the treatment of phonology in the CEFR can thus be summarized as follows:

Strengths

The description of phonology in the CEFR is thorough and sufficiently broad to allow a revision and extension of the descriptors in order to capture recent reflections and developments in second/foreign language education. Learnability (and consequently 'teachability') of phonology is envisaged. The integration of phonology among relevant categories of assessment is clearly mentioned, especially in the appendices.

Weaknesses

The existing scale *Phonological Control* does not capture this conceptual apparatus, and appears unrealistic with regard to accent and progression (particularly in an apparent leap between B1 *Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur* (actually calibrated to B1+: North 2000) and B2 *Has a clear, natural, pronunciation and intonation.* (actually calibrated to B2+: North 2000). This latter descriptor also suggests a native speaker standard as the criterion for 'natural,' the only CEFR descriptor to compare to the competence of a 'native speaker.' Finally, the scale mixes stress/intonation, pronunciation, accent and intelligibility without providing clear indication of progression in any of these factors specifically.

A number of studies have been published on the CEFR phonology scale (Cauvin, 2012; Galaczi, Post, Li, & Graham, 2011; Horner, 2010, 2013, 2014; Isaacs & Trofimovich, 2012; Harding, 2013). Horner examines which pronunciation features have an impact on communicative efficiency, questioning the possibility of establishing hierarchies and analyzing the aspects of pronunciation that are "irritating" (and for whom), difficult to

acquire (and in what order), and that have an impact on intelligibility. With regard to intelligibility, he identifies as core features, factors such as word stress and accurate reproduction of phonemes and sentence stress, while he considers intonation, rhythm, and phonetic reduction as more peripheral. Although he refers to English language phonology in his articles, these considerations can apply across languages, albeit with individual variations. Another important study is Trofimovich and Isaacs, in which the authors underline how the CEFR combination of "descriptions of easily understandable speech and a noticeable foreign accent in the same band descriptor" (2012, p. 914) can be problematic and stress the need to "disentangle accent from different aspects of communicative effectiveness, including comprehensibility."(2012, p. 914). Finally, Harding (2016) criticises the usability of the current scale. He used mixed-methods research to investigate the construct underlying the scale, particularly its orientation towards a nativeness principle or a comprehensibility principle respectively.

Other language proficiency scales, however, are not noticeably better than the CEFR scale in their treatment of phonology. In assessment, phonology is rarely a feature to be separately assessed, being generally subsumed under "speaking" or "fluency". In the TOEFL iBT Test, for instance, references to pronunciation are to be found under the category "Delivery", which includes in the same descriptor four different, albeit overlapping, features (fluency, pronunciation, intonation, and intelligibility). In the Hong Kong Diploma of Secondary Education (HKDE), specific descriptors are provided organized around 5 levels of "pronunciation and delivery." Only the IELTS test defines 9 bands of "pronunciation." Then there is the issue of the way that progression is achieved in scales for phonology through the substitution of qualifiers like "some", "a few", "many", "most", etc. The inadequacy of such an approach has long been acknowledged (Alderson 1991; Champney 1941,). Consistent interpretation becomes impossible unless raters are trained to interpret the descriptors in the same way, which they will then tend to do without referring to the actually wording. This issue is particularly acute in the IELTS scale (e.g., "full range of pronunciation features", "wide range", "range", "limited range", "some features"). Unfortunately, the scales produced by researchers trying to improve on the CEFR phonology scale (e.g. Frost & O'Donnell, in press; Horner 2013) tend to continue in the same vein.

LITERATURE REVIEW

In addition to studying papers concerning the CEFR phonology scale, a broader literature review was conducted (using www.teacherreference.com, www.mla.org; ERIC, and Google scholar). The aim was to explore the way phonology is dealt with, from a pedagogical point of view in second/foreign language education; and to identify resources and concepts to inform the development of descriptors.

Strikingly few of the articles found focused specifically on the teaching, learning, and/or assessment of phonology. The research tradition in phonology has often been only peripherally relevant to applied linguistics (Derwing & Munro, 2015). Indeed, as Munro & Derwing (2011) point out in their research timeline on accent and intelligibility, there is a fundamental incongruence between research interests and pedagogical ones. The focus in research on accent and accuracy instead of on intelligibility has been detrimental to the development of pedagogically-oriented work on pronunciation. "Identifying native-

like production as the central goal in pronunciation teaching inevitably leads to the conclusion that pronunciation is probably not worth teaching because of the limited likelihood of achieving that end." (Munro & Derwing, 2011. p. 317). Even though intelligibility has been seen as a priority in language instruction for a long time (the authors mention Sweet's handbook dating back to 1900!), conceptualization of the notion is a relatively recent process, even though "it is intelligibility – rather than native-like pronunciation – that is most critical for successful communication in an L2" (Munro & Derwing, 2011, p. 316–317).

Although there has been a growing body of such research in the last few years (e.g. Saito, 2012; Lee, Jang & Plonsky, 2015; Thomson & Derwing, 2014), research and reflection on the role of phonology in second/foreign language education is still limited, and so are resources to support teachers. In a seminal article, Derwing and Munro described the situation as follows: "The study of pronunciation has been marginalized within the field of applied linguistics [and] as a result, teachers are often left to rely on their own intuitions with little direction" (2005: 379). Derwing and Munro (2015) report, on the basis of a wide range of studies, that "teachers are hesitant about systematically teaching pronunciation" (p. 78), that they feel a "need for access to more professional development" (p. 80) and that "[t]he curricula in the various programs in which the teachers worked did not focus on pronunciation" (ibid.). Teachers do not teach pronunciation because they "lack confidence, skills and knowledge" (MacDonald, 2002: 3). They have received little or no specific training (Breitkreutz, Derwing & Rossiter, 2002; Burgess & Spencer, 2000; Derwing & Munro, 2005; Baker, 2011).

Derwing & Munro list a series of pronunciation-related phenomena that have been identified by various researchers as crucial to the success of any communicative activity. These include "listener factors, such as familiarity with foreign-accented speech, willingness to communicate, and attitudes toward L2 speakers" (Derwing & Munro, 2005, p. 392). However, "little research has explored which linguistic features of speech are most crucial for intelligibility and which, while noticeable or irritating, merely contribute to the perception of an accent" (Trofimovich & Isaacs, 2012, p. 906). This is unfortunate "given the well-established fact that accent is partially independent of comprehensibility and intelligibility and that the latter two are more important to successful communication" (Derwing & Munro 2015, p. 168). Some studies are starting to focus on assessment, in order to address the fact that proficient speakers can be unfairly penalised due to the mismatch between their level of functional proficiency and their level of phonological competence. Isaacs (2008) and Trofimovich & Isaacs (2012) highlight the centrality of intelligibility and attempt to provide rating scale specifications that move beyond the native speaker standard and bring light to what are often just intuitive impressions (Isaacs & Thomson, 2013).

DEVELOPMENT AND INITIAL FEEDBACK

The literature review helped to identify categories that could inform scales based on developments in pronunciation research and teaching. It was also insightful in thinking through the internal progression of phonological competence as well as its relationship to progression in language proficiency. In drafting descriptors, the aims were (a) to revise the general scale for phonological control to provide a snapshot of phonological

competence; and (b) to supplement this with sub-scales to help teachers and learners to identify areas for improvement. The following core areas to inform descriptor production were identified:

- **Articulation** (including pronunciation of sounds/phonemes);
- **Prosody** (including intonation, rhythm and stress –word stress/sentence stress – and speech rate/chunking);
- **Accentedness** (accent and deviation from a 'norm');
- **Intelligibility** (i.e., actual understanding of an utterance by a listener) and **comprehensibility** (i.e., listener's perceived difficulty in understanding an utterance).

It proved difficult to establish a clear distinction between these aspects, since they overlap and influence judgment in an inextricable way. It was therefore decided to operationalize them in only three scales: *Overall phonological control*; *Sound recognition and articulation*; *Prosody (intonation, stress and rhythm)*. Intelligibility was the key factor for discriminating between levels and explicit mention of accentedness was made at all levels in the *Overall* scale in order to raise awareness of the non-causal relationship between accent and phonological competence.

First drafts were piloted in a one-hour workshop with eleven English and French language experts from different European countries, in order to check the clarity and appropriateness of the descriptors, the coherence of the specific aspects they intended to assess and the coverage of relevant issues. The participants were asked in pairs to work on two worksheets. One contained mixed descriptors from both the prosody/intonation scale and the sound articulation scale, presented in random order; the other contained the descriptors for overall phonological competence, again in random order. Participants were asked to identify the descriptors belonging to each of the scales, assign the descriptors of each identified scale to the six CEFR levels, and provide qualitative feedback on the scales and/or the individual descriptors. The results were encouraging: only one or two sound articulation / prosody descriptors were misplaced per worksheet, which appeared to confirm the clarity and consistency of the scales. The overall descriptors were correctly matched to the intended CEFR levels with no disagreement. The final task, ranking the sub-scale descriptors by level, generated more discussion, but produced almost 70% of correct ranking. After completion of the tasks, an extensive discussion followed in which participants provided qualitative feedback on both the descriptors and the scales. These insights were very helpful in revising them.

The revised scales were then shared with five other experts working on phonology in relation to the English and French languages. Reactions were very positive, some consultants giving very detailed feedback. The specific feedback on each descriptor that was received was collated and compared in order to inform revision. On the basis of this feedback, and through wider consultation within the authoring group of the main descriptor extension project, the descriptors were again revised, translated into French, and prepared for the validation process.

QUALITATIVE AND QUANTITATIVE VALIDATION

The validation took place in two phases between January 2, 2016 and February 7, 2016. In a first—qualitative—phase, the descriptors were rated for quality before being matched the intended categories and levels. In the second—quantitative—phase, data from judgements of the CEFR level of the descriptors, plus from the use of the descriptors as a checklist to assess performance on videos, was used to calibrate the descriptors to the CEFR levels. Some 250 respondents completed the survey in Phase 1. They were asked to: (a) identify the category of the descriptors; (b) rate the descriptors for clarity and for pedagogical usefulness, and (c) assign the descriptors to CEFR levels. The first task had a tricky element, in that descriptors were included (e.g. on accent) that did not fit under either of the two categories given: 'Sound articulation' and 'Prosodic features.' They were intended to land in a "Can't decide" column. Only one descriptor was not assigned to the correct category, though another three were not as clearly identified with their category as was desirable. Only two descriptors were rated at less than 80% for the two quality criteria: clarity and usefulness. In terms of assignment to level, only three descriptors were assigned to a wide range of level. As a result of the exercise, five descriptors were dropped and several others reformulated in preparation for Phase 2.

In the quantitative phase, 272 respondents took part in the survey. The first task was to assign each of 34 descriptors to a CEFR level, by answering the following question: *At what CEFR level do you think a person can do what is defined in the descriptor?* The second task consisted of assessing video performances with a rating scale applied to each descriptor.

Assigning Levels

The data was subjected to two analyses: 1) a collation showing, for each descriptor, the percentage that rated the intended level, and 2) a Rasch analysis (Linacre, 2014) to link the ratings to the mathematical (logit) scale underlying the CEFR levels (North 2000). In the simple collation, 28 descriptors were assigned to the intended level by 40% or more of the respondents. Five descriptors seemed to be relatively evenly divided between two levels, e.g. *Can articulate the majority of the sounds of the target language reasonably clearly in extended speech* straddling B1 and B2, thus coming out as B1+. Three descriptors, all concerned with sound recognition, were spread across a range of levels. In the Rasch analysis, the scale produced was very long, from 4.20 logits to -4.60 logits. Because the scale length and slope was very similar to the scale produced in the Swiss research project (North, 2000), straightforward anchoring to the cut-offs between CEFR levels set in that project seemed appropriate. However, only one CEFR item was available as an anchor. Therefore, three more descriptors adapted from Cambridge English Language Assessment criteria for A2, B1 and B2 were also used as anchor items, each anchored to the value of the midpoint of the band of proficiency for the criterion level concerned.

Assessing Performance

The second task was to assess three video performances with 27 descriptors (i.e. excluding the seven descriptors for sound recognition). The videos used were those developed by the Centre international d'études pédagogiques (CIEP) based in France (http://www.ciep.fr/en) in 2008. Respondents chose whether to rate performances in English, French, German or Spanish. There were 872 responses from the 272 participants, but some 200 were excluded for misfit, leaving 667 responses in the data. Two analyses were carried out: one free (unanchored) analysis and one anchored to the North (2000) scale with the same anchor items as in the first task. Again, the scales produced from both analyses was long: from 4.81 logits to -4.96 logits for the unanchored analysis and from 4.96 (high C2) to -4.52 (just below the A1 cut-off) for the anchored one. However, in line with the recommendation of the *Manual for relating assessments to the CEFR levels* (Council of Europe, 2009), a second standard-setting method was also used. This was a simplified form of the Bookmark Method (Council of Europe, 2009, p. 77–81). Each of the five team members independently selected the point for the cut-off between levels on the unanchored scale. The cut-points selected for A2/B1, B1/B2 and B2/C1 by the different team members were identical. Four chose the same cut-point for A1/A2, leaving the C1/C2 cut-point as the only one requiring discussion.

Treatment of Accent

The main issue that emerged in the analysis concerned treatment of accent. The implicit native speaker model of the existing B2 descriptor *Has acquired a clear, natural, pronunciation and intonation* appeared to have instilled the unrealistic expectation that user/learners at the C levels would not have any accent, This B2 CEFR descriptor had actually been calibrated at 2.53, a very high B2+, within the margin of error to the C1 cut-off (North 2000). Yet research has demonstrated that that accent remains a feature of the speech of many people with even a very high level of language proficiency. It is not the "naturalness" of native speech that is essential: it is intelligibility, which is usually not the same thing. Not surprisingly, the statements about accent, intended to be the second part of descriptors at the B and C levels on the *Overall* scale, tended to come out lower than the level intended. One C1 descriptor, however, had been retained intact as a "double barrelled" descriptor with its proviso about accent still attached: *Can articulate virtually all the sounds of the target language; some features of accent retained from other language(s) may be noticeable, but they do not affect intelligibility at all*. This descriptor was calibrated as intended, demonstrating that when statements about accent were added as detail, the effect of rating at a lower level was reduced. Nevertheless, as a result of the responses, the C2 statement about accent was completely rewritten as follows: *Intelligibility and effective conveyance of and enhancement of meaning are not affected in any way by features of accent that may be retained from other language(s).*

Final Decisions

Some descriptors were interpreted in different ways in the two tasks. Deciding which level to finally assign to descriptors was therefore, as always, a question of judgement, based on all the evidence available. There were four pieces of information for most descriptors to guide the decision (two from each task), although the seven descriptors for

sound recognition were not used in the assessment task, and therefore had only two information points. In 15 cases, all information points agreed with the intended level and in another five cases, three of the four agreed with the intended level, giving 20 items for which the calibration could be confirmed with confidence. With a further seven items, the information was mixed and for five descriptors it was definitely different to what was intended. One, for *Sound recognition*, was dropped without discussion since it was distributed across all levels. Of the seven others with mixed information, three were accepted as calibrated at a "plus" level, between the criterion levels—but were later dropped from the final scale. Two more were amended slightly in order to make clearer the intended level, and the final two, dealing with accent, were retained where originally intended. The five remaining descriptors, calibrated clearly from what was originally intended, were completely rewritten once the rest of the descriptors on the scales had been completed, in order to provide a coherent scale describing all categories at all levels. The survey had worked well, producing a good, long scale. After some discussion and consultation, however, it was decided to drop the descriptors on sound recognition. The resulting scale is available in the extended version of the CEFR illustrative descriptors (available on the Council of Europe's website). It is presented as an analytic grid in three columns, overall phonological control on the left and the two sub-scales on the right.

CONCLUSION

The need for new descriptors of phonological control aligned with current research in phonological competence, and developed following a methodologically sound process including qualitative and quantitative research, informed the project reported in this article. The aim was to reflect the solid, comprehensive conceptualization of the tenets of phonological competence in the CEFR text. The development, which replaced the original CEFR scale for *Phonological control* and supplemented it with two more scales for *Sound articulation* and *Prosody (intonation, stress and rhythm)* respectively, offered the opportunity for ground-breaking work which hopefully will provide the basis for teachers to include appropriate objectives for phonology in their planning and to develop assessment criteria appropriate to the level(s) concerned.

ABOUT THE AUTHORS

Enrica Piccardo is Associate Professor at OISE – University of Toronto and at the Université Grenoble-Alpes France. A CEFR specialist, she has collaborated with European Institutions on international projects. Her monograph *From Communicative to Action-oriented: a Research Pathways*. (2014) available online is being used in teacher education worldwide. She is the principal investigator of two SSHRC funded research projects, QualiCEFR and LINguistic and Cultural Diversity REinvented (LINCDIRE) and of a Council of Europe funded project QualiMatrix. Her research focuses on assessment, plurilingualism, and creativity in language education. Recent publications include a *TESOL Quarterly* article (2013), a co-edited *The Canadian Modern Language Review* issue (2015), a book chapter on plurilingualism and heritage languages (2014) CUP and one on creativity (2016) De Boek.

enrica.piccardo@utoronto.ca

Brian North is a researcher in language education. He has extensively collaborated with international organizations, particularly the Council of Europe, as a project coordinator and consultant. He spent most of his career at Eurocentres, the Swiss-based foundation that has been an official NGO to the Council of Europe since 1968. In his PhD thesis he developed the levels and descriptors for the CEFR. He then co-authored the CEFR itself, the prototype European Language Portfolio, the Manual for relating examinations to the CEFR, Eaquals' Core Inventories for English and French, and the European Profiling Grid for language teacher competences (EPG). He was Chair of Eaquals from 2005 to 2010. Currently he is coordinating the project to extend the CEFR illustrative descriptors.

bjnorth@eurocentres.com

REFERENCES

Alderson, J. C. A. (1991). Bands and scores. In J. C. A. Alderson and B. North (Eds.) *Language testing in the 1990s*. (pp. 71-86). London: Macmillan.

Baker, A. A. (2011). ESL teachers and pronunciation pedagogy: Exploring the development of teachers' cognitions and classroom practices. In. J. Levis & K. LeVelle (Eds.). Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference, Sept. 2010. (pp. 82-94), Ames, IA: Iowa State University.

Breitkreutz, J., Derwing, T. M., & Rossiter, M. J. (2002). Pronunciation teaching practices in Canada. *TESL Canada Journal, 19,* 51–61.

Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System, 28,* 191–215.

Cauvin, E. (2012). Le CECRL et l'évaluation de la compétence phonologique: point de vue à partir d'un corpus sur l'anglais : Evaluer avec le CECRL. *Les Langues modernes 106*(1), 33-40.

Champney, H. (1941). The measurement of parent behavior. *Child Development 12*(2), 131–66.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment (CEFR)*, Strasbourg: Council of Europe available online at: http://www.coe.int/t/dg4/linguistic/cadre1_en.asp

Council of Europe (2009) *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*, Strasbourg: Council of Europe http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379-398.

Derwing & Munro (2015). *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research*. Amsterdam/Philadelphia: John Benjamins.

Frost, D. & O' Donnell, J. (in press). Evaluating the essentials, the place of prosody in oral production. In J. Volin (ed). *Pronunciation of English as a Foreign or Second Language*

Galaczi, E., Post, B., Li, A. and Graham, C. (2011) Measuring L2 English phonological proficiency: Implications for language assessment. In J. Angouri, M. Daller and J. Treffers-Daller (eds) *The Impact of Applied Linguistics: Proceedings of the 44th Annual Meeting* (pp. 67–72). Downloaded 24.04.17 from: http://www.baal.org.uk/proceedings_11.pdf

Harding, L. (2013). Pronunciation assessment. In C.A. Chapelle (ed.) *The Encyclopedia of Applied Linguistics.* Oxford: Blackwell. . doi:10.1002/ 9781405198431.wbeal0966

Harding, L. (2016). What do raters need in a pronunciation scale? The users' view. In T. Isaacs & P. Trofimovich (Eds.), *Interfaces in second language pronunciation assessment: Interdisciplinary perspectives* (pp12–36). Bristol, UK: Multilingual Matters.

Horner, D. (2010). *Le CECRL et l'évaluation de l'oral: Anglais*. Paris: Belin.

Horner, D. (2013) Towards a new phonological control grid. In E. D. Galaczi & C. J. Weir (Eds.), *Studies in Language Testing: Vol. 36 Exploring Language Frameworks. Proceedings of the ALTE Kraków Conference* (pp. 187-204). Cambridge: UCLES/Cambridge University Press.

Horner. D. (2014) And What about testing pronunciation? A critical look at the CEFR pronunciation grid and a proposal for improvement. in L.M Rupp & R. van den Doel (eds.), *Pronunciation Matters* (pp 109–124). University of Utrecht.

Isaacs, T. (2008). Towards Defining a Valid Assessment Criterion of Pronunciation Proficiency in Non-Native English-Speaking Graduate Students. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes, 64*(4), 555–580.

Isaacs, T & Thomson, R. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10, 135–159.

Isaacs, T. and Trofimovich, P. (2012) Deconstructing comprehensibility: identifying the linguistic influences on listener ' L2 comprehensibility ratings. *Studies in Second Language Acquisition 34* (3), 475-505.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics36*, 345-366.

Linacre, John M. (2014) *Winsteps: Rasch-model Computer Program*, Chicago: MESA Press.

MacDonald, S. (2002). Pronunciation—views and practices of reluctant teachers. *Prospect, 17(3),* 3–18.

Munro, M. J., Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching 44(3)*, 316–327

North, B. (2000). *The Development of a Common Framework Scale of Language Proficiency*, New York: Peter Lang.

North, B and Schneider, G (1998) Scaling descriptors for language proficiency scales. *Language Testing 15*(2), 217–262.

Piccardo, E. (2012). Médiation et apprentissage des langues : Pourquoi est-il temps de réflechir à cette notion ? *ELA : Études de Linguistique Appliquée 167*, 285–97.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46*(4), 842-854.

Schneider, G. & North, B. (2000). *Fremdsprachen konnen: was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschatzung der fremdsprachlichen Kommunikationsfahigkeit*, Nationales Forschungsprogramm 33, Chur/Zurich: Wirksamkeit unserer Bildungssysteme, Verlag Ruegger.

Thomson, R. I. & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: a narrative review. *Applied Linguistics 36,* 326–344

Trofimovich, P & Isaacs, T. (2012). Disentangling accent from comprehensibility, *Bilingualism: Language and Cognition 15* (4), 905–916.

EXAM-RELATED RESOURCES

Hong Kong Examinations and Assessment Authority (HKEAA). Hong Kong Diploma. of Secondary Education Examination. Available online: http://www.hkeaa.edu.hk/en/hkdse/

HKEAA. Hong Kong Diploma of Secondary Education Examination. English Language Level Descriptors. Speaking Descriptors. Available online: http://www.hkeaa.edu.hk/DocLibrary/HKDSE/Subject_Information/eng_lang/LevelDescriptors-ENG-Speaking.pdf

IELTS SPEAKING: Band Descriptors (public version). Available online: https://www.ielts.org/pdf/SpeakingBanddescriptors.pdf

Educational Testing Service (2014). Independent speaking rubrics/Integrated speaking rubrics. (TOEFL iBT Test). Available online: http://www.ets.org/s/toefl/pdf/toefl_speaking_rubrics.pdf, and scoring guides-Speaking. Available online: http://www.ets.org/s/toefl/pdf/toefl_student_test_prep_planner.pdf

Trinity College London (2015) Integrated Skills in English (ISE) Specifications – Speaking & Listening ISE Foundation to ISE III. Available online: www.trinitycollege.com/resource/?id=6298

Centre international d'études pédagogiques (CIEP) (2008). DVD "Productions orales illustrant les 6 niveaux du Cadre européen commun de référence pour les langues". Available online: http://www.ciep.fr/ressources/ouvrages-cederoms-consacres-a-levaluation-certifications/dvd-productions-orales-illustrant-les-6-niveaux-cecrl