

Europe and language learning: The challenges of comparable assessment Françoise Kusseling & Wilfried Decoo - Brigham Young University

Abstract

In 2010, the Council of Europe will conduct a European Survey on Language Competences (ESLC): high school students from all over Europe will be tested to measure how well they know other European languages. Proficiency levels of the *Common European Framework for Languages* (CEF, Council of Europe, 2001) will be used as the yardstick. The CEF is meant to help harmonize language learning programs across Europe in order to align assessment criteria.

This paper analyzes some of the challenges the ESLC faces, in particular those related to lexicon. Since the 1970s various endeavors, sustained by the Council of Europe, have tried to specify the functional vocabulary needed for certain levels, but without coming to an international consensus in the so-called Threshold publications. In the wake of the CEF, new efforts in various countries have led, and are still leading, to so-called Profiles or Referentials, which also define functional vocabulary for each CEF-level. However, a comparison of sources reveals major discrepancies between European languages. Sources give drastically different sizes for vocabulary for the same CEF proficiency level, from 400 to 3,300 for level A1; from 800 to 4,000 for B1; from 1,100 to 6,800 for B2; and from 3,300 to 30,000 for C2. Such differences destabilize the system of equivalences the Council of Europe seeks to establish with its new instruments. What the ESLC would evaluate would be less the language proficiency achieved by European pupils than the gap between the tests and divergent national criteria, or simply between the tests and divergent language learning materials.

The ESLC reflects Europe's desire to unify countries, while, at the same time, revealing the disparities we need to overcome.

1. Introduction

After the 2001 Common European Framework of Reference for Languages (CEF or, in some citations, CEFR or CEFR) publication followed a 2002 call to establish a "European language competence indicator" (ELCI) tied to the CEF (Commission 2005:7). It recommended the early learning of two foreign languages in schools and stated that "progress toward this objective can only be measured using reliable data on the outcomes of foreign language teaching and learning; this data needs to be based upon objective tests of language ability" (Commission of the EC 2005) in order to provide data "on which any necessary adjustments in their approach to foreign language teaching and learning can be based" (p. 5).

This initial project evolved into the *European Survey on Language Competences* (ESLC) under a consortium called SurveyLang responsible for specification of comparable tests in English, Spanish, French, German, and Italian, a tremendous undertaking let alone for one language and a clearly defined test group.

Since its development in the 1990s the CEF was intended to provide a "framework" to allow enhanced communication and better practices among professionals desiring to develop language programs best-suited to learners' needs and situations. It was not, however, intended, even though promoted afterwards as such, to become a standard for comparisons.

Fulcher (2008) sums up the issue: "The CEFR, despite the original intentions of the authors, is now being adopted as a tool in standards-based education in Europe. (...) It is therefore not surprising that there is pressure to adopt the CEFR, for in the wake of adoption comes the requirement to align curriculum and assessment to its scales, and be held accountable for outcomes" (p. 20). Thus the ESLC wants to assess the language proficiency of 15-year-olds and lead governments to align their foreign language programs. They intend to compare results across languages and countries using the CEF as a gauge.

Our purpose is to highlight a key component for the success of a comparable language learning assessment across Europe, that is: the lexicon. Defining the number of words known in a native language is a complex issue. The subjective estimation of candidates is a revealing exercise. One could ask a person the following questions:

- Could you quantify and write down the number of words you know in your first language?
- Would you be able to quantify and write down the words you know in your second language (L2)?
- What about the number of words you know in your second foreign language?
- Now would you be able to write down how many words students learn in a 101 or 102 language course?
- Finally, how many words would they own after three years of language study? Please keep these numbers in mind.

We expect answers to be wide-ranging depending on the definition of the concept word, depth or breadth of word knowledge, linguistic program goals, theme-based approaches, needs and skill levels of learners. Could we however come to some consensus that would allow comparisons?

We will, thus, first make some remarks on word monitoring, particularly in answer to the question: "How many words are monitored in L2 teaching?", then talk about the European survey on language competences (ESLC) scheduled for 2010 and the challenges this major evaluation effort faces, and finally propose some solutions.

2. How are words monitored in L2 teaching?

Whatever the importance of functional or communicative approaches, or whether one considers speech acts or collocations as the nucleus of language use or not, words as such are still the most tangible and quantifiable units to measure language levels attained. A number of publications discuss methods for vocabulary counting (e.g., Bauer & Nation 1993; Cowie 1992; Gardner 2007; Hazenberg & Hulstijn 1996; Meara 1996). The identification of dictionary lemmas gives a useful indication of word numbers, or word families (the base word and all its derived and inflected form, i.e., 800 word families equal 1,400 words) (Bauer & Nation 1993:253). Depending on the inflectional systems of various languages however, the ratio between word families and number of words can be quite different – e.g. French or German compared to English. Counting in lemmas facilitates interlanguage comparisons.

To present the topic of second language (L2) word monitoring we address the concept of “coverage” as a lexical criterion, and look at American and European proficiency guidelines in relation to number of L2 words known.

2.1. Coverage as lexical criterion

Coverage has been defined as the “percentage of words a learner knows when confronted with a text.” Research (Hu & Nation 2000) shows that 98% should be the desired coverage for reading comprehension, which would come to a maximum of one unknown word in 50. The next question is: How many words must be known by L2 learners for certain types of texts?, and, subsequently, To what extent will an L2 text used for didactic or evaluation purposes reflect individual lexical mastery?

2.2. Proficiency guidelines and number of words

The American Interagency Language Roundtable (ILR), the Council on the Teaching of Foreign Languages (ACTFL), and the Common European Framework of Reference for Languages (CEF) guidelines have defined levels of mastery as illustrated in Table 1 below.

Table 1: Learning progression of linguistic competences

ILR	ACTFL	CEF
5	Native	
4 / 4+	Distinguished	C2 Mastery
3 / 3+	Superior	C1 Eff. Operational Prof.
2+	Advanced - High	B2+ Vantage +
2	Advanced - Low & Mid	B2 Vantage
1+	Intermediate - High	B1+ Threshold +
1	Intermediate - Low & Mid	B1 Threshold
0+	Novice - High	A2+ Waystage +
0	Novice - Mid Novice - Low	A2 Waystage A1 Breakthrough

This table, however, begs the following questions: What words and what numbers of words might be attached to these levels? What do proficiency guidelines say about the number of words a student should master?

2.2 - Proficiency guidelines and number of words

Introductions to guideline-informed inventories usually circumvent criticism for their word choices with the caveat that their lists do not decide what a student has to learn at a certain level. Inventory users remain free to choose a collection of words based on needs. But conversely they do hope that developers, authors, and testers will use their inventories. The least one can say is that the number of words in each inventory is an indication of that targeted level.

The ACTFL guidelines, e.g., state for novice-level language learners: “Comprehend and produce vocabulary that is related to everyday objects and actions on a limited number of familiar topics and intermediate-level learners: Use vocabulary from a variety of thematic groups; Recognize and use vocabulary from a variety of topics including those related to other curricular areas.” (ACTFL 2001, p. 40) Similar descriptions, without precise figures, are used for other levels.

For our topic, the CEF guidelines are more important: "To what extent do they help us determine each lexicon related to a level?" The CEF guidelines suggest that a A1 “basic” user is equivalent to the ACTFL novice-low user: “Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.” (CEF, p. 112) And the novice-mid A2-level learner: “Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics.” (CEF, p. 112) The same imprecision continues up to the highest C2 level where one reads: "Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms." (CEF, p. 112) The principle is one of concentric development of lexical proficiency, but without further details.

In addition, the CEF states that "size, range and control of vocabulary are major parameters of language acquisition and hence for the assessment of a learner’s language proficiency and for the planning of language learning and teaching" (p. 150).

Under the header “specification”, users can determine "what size of vocabulary (i.e. the number of words and fixed expressions) the learner will need/be equipped/be required to control." Under "Lexical selection," the CEF clarifies that authors of materials have a number of options:

- [1]to select key words and phrases a) in thematic areas required for the achievement of communicative tasks relevant to learner needs, b) which embody cultural difference and/or significant values and beliefs shared by the social group(s) whose language is being learnt;
- [2]to follow lexico-statistical principles selecting the highest frequency words in large general word-counts or those undertaken for restricted thematic areas;
- [3]to select (authentic) spoken and written texts and learn/teach whatever words they contain;
- [4]not to pre-plan vocabulary development, but to allow it to develop organically in response to learner demand when engaged in communicative tasks. (pp. 150-151)

The CEF thus recognizes options which vary widely and run counter to a set core vocabulary foundation. However, considering the vocabulary selections in *Threshold* and related levels, it seems that option [1], with suboption a), has been the major procedure used: "select key words and phrases in thematic areas required for the achievement of communicative tasks relevant to learner needs." This approach leans on notions and functions. The exponents, as found in

Threshold, Waystage, and Vantage, provide "key words and phrases." In the CEF chapter on assessment, this preference is confirmed:

The set of content specifications at Threshold Level produced by the Council of Europe for over 20 European languages and at Waystage and Vantage Level for English, plus their equivalents when developed for other languages and levels, can be seen as ancillary to the main Framework document. They offer examples of a further layer of detail to inform test construction for Levels A1, A2, B1 and B2 (p. 179).

This paragraph seems to suggest an underlying aim of the CEF, namely that the "content specifications" in the "ancillary" publications are to be seen as compulsory, or at least strongly recommended and to be "examples" of what is to be included in European tests, rather than content that any educator or test designer can describe for a given situation.

Moreover, questions remain unanswered: "Are word choices for a given level to be made within a CEF-approved inventory?" If a selection covers less than the entire inventory, does it still match the level? Are users free to develop a completely separate inventory?

3. The European survey on language competences (2010)

With these levels and numbers in mind, we can better assess the Europe-wide testing project called the European survey on language competences (ESLC). Its purpose is to measure foreign language competence in each Member State to determine if 15-year old teenagers reach certain CEF levels. The tests will, first, assess reading, listening, and writing competences, in the five most commonly spoken European languages: English, French, German, Spanish, and Italian. They will take on various forms to allow for adaptive testing; testing instruments could have the form of computer-based tests, using open source software or paper and pencil tests.

Our research questions were: "What number of words should students know at each CEF-level to be able to pass these tests?", "What lexical coverage and number of words should students be tested on at each CEF level?" The CEF itself gives descriptions of behavior, not vocabulary, but refers to the lists of Threshold, Waystage, Vantage in English and its equivalents in the other European languages to be tested. CEF levels A1 through B2 are thus found in CEF descriptors, lists, profiles, and referentials provided by the Council of Europe and its authors.

Tables 2 through 6 show our research findings compiled through analysis of lists, profiles and referentials.

Table 2 - CEF A1 Vocabulary Range / Number of words

Description	Word number range	Difference
Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations	400 – 3,357	89%

When we align the description given for CEF level A1 and the word numbers we found through our research, we notice an 89% difference between the lowest and the highest word numbers quoted in the lists and referentials we examined.

Table 3 - CEF A2 Vocabulary Range / Number of words

Description	Word number range	Difference
Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations	400 – 3,357	72%

Table 3 reveals that a basic vocabulary repertoire can range from 400-3,357 words. We thus observe a difference of 72% between the lowest and highest quote.

Table 4 - CEF B1 Vocabulary Range / Number of words

Description	Word number range	Difference
Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations	400 – 3,357	89%

Here again an important difference of 89% which should characterize “sufficient” vocabulary allowing a learner to express himself with some circumlocutions on most topics.

Lastly, we see a steady trend of range difference at CEF level B2.

Table 5 - CEF B2 Vocabulary Range / Number of words

Description	Word number range	Difference
Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations	400 – 3,357	86%

The important range difference remains a constant across levels.

More answers have been suggested in various studies and sources, in particular the new Profiles and Referentials. Our Table 6 shows the existing wide word number range identified for CEF levels, mentioned in publications or counted in the lists. Counting by word families has been transposed into words, using the 1.7 ratio. Figures in parentheses indicate estimates based on proximate level figures.

Table 6 - Number of words identified for CEF levels

Authors	A1	A2	B1	B2	C1	C2
Van Ek & Alexander 1980		700	1,100-1,500			
Van Ek 1976			1,600			
Meara & Milton 2003	<1,500	1,500-2,500	2,750-3,250	3,250-3,750	3,750-4,500	4,500-5,000
Schmitt 2008, see also Nation 2006						15,000
Coste a.o. 1976			3,000			
Beacco a.o. 2004	1,000	1,700	(4,000)	6,800		
Rolland & Picoche 2008	3,357					
Milton 2006	(400)	800-1,000	800-1,000	2,000		3,300
Instituto Cervantes 2006	1,300	3,000	7,000	14,000	21,000	30,000
Bergan 2001		850	1,500	4,500		

The comparison exposes noteworthy differences. To explain these figures further:

Schmitt (2008), associating CEF descriptors and coverage criteria, assumes 8,000 to 9,000 word families are needed for level C2 (13,600 to 15,300 at the 1.7 ratio). He demonstrates about 15,000 words must be mastered to enable an individual to easily read news articles and average novels (see also Nation 2006; Adolphs & Schmitt 2004; Hazenberg and Hulstijn 1996). The C2-level for reading states: "I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialized articles and literary works" (CEF, p. 27). This 'Can Do' specification means a C2-level needs a minimum of 15,000 words.

Milton's (2006) figures come from results for the General Certificate of Secondary Education (GCSE) in Great Britain as it is correlated with CEF levels. They are lower than the others. Milton compares them with higher "EFL vocabulary sizes" by referring to EFL tests, as reported in Meara and Milton (2003).

The French B2 Referential (Beacco 2004) indexes 6,214 lexical items, not adding references to functions and notions. After sorting out homonymy and strong polysemy and adding grammar

items, the total reaches about 6,800 entries. This list consists of many low-frequency words whereas many high-frequency or basic words for daily communication are absent.

Rolland and Picoche (2008) worked on a proposal to systematically learn the French lexicon at level A1 and beyond. They kept the lexical items of "an absolute current usage or of first necessity" (p. 45, transl.) adding up to 3,357 word units. They state that such a corpus is needed for an A1-level – "first basic lexical content, which to us seems indispensable to be acquired by a learner aiming at reaching level A1" (p. 46, transl.). This limit needs to be disambiguated since it also describes levels beyond.

The Instituto Cervantes's *Plan Curricular* is a voluminous work. We received the following “per level” figures and the total from Dr. Montoussé Vega (personal mails to W. Decoo). We added the cumulative and the adapted columns. Since the Spanish listing counts single words, compound units (*lavaplatos, tocadiscos ...*), and structural combinations (*viajar en avión, viajar en barco, viajar en tren ...*) as separate entries, the figures are inflated. We applied a 20% reduction and rounded off the figures. The results in the comparative table 7 are therefore to be read as approximate figures. Even if a greater reduction were applied, the figures would still come out much higher than in other sources.

Table 7 - Words in *Plan Curricular*

	per level	cumulative	adapted
C2	9,919	36,680	30,000
C1	9,786	26,761	21,000
B2	8,271	16,975	14,000
B1	4,903	8,704	7,000
A2	2,192	3,801	3,000
A1	1,609		1,300
Total	36,680		

All these figures do not account for the number of instruction and training hours needed for the receptive or productive mastery of 1,000 or 2,000 words, the effect of the distance between an L1 and an L2 (or possibly an L3). Testing English proficiency with the same test among French- or among Spanish-speaking students will probably yield important differences that may not be the result of better or poorer attainment, all other factors being similar. Furthermore, what do these findings mean for test designers, publishers, and takers? We believe challenges facing European survey stakeholders can now be anticipated.

4. Challenges

The lexical predicament creates great challenges for the ESLC since it intends to assess reading and listening comprehension for CEF levels, across borders, in fairly short tests, detached from any given content. Students will have taken a few years of non-intensive language study, with

considerable lexical variability between them, even if theoretically students are at comparable CEF levels.

Fulcher (2004) explains that there is no obvious link between tests and the CEF since its scales are not founded on underlying theory and not attached to content specifications and, therefore, would be hard-pressed to “provide equivalence of meaning across tests” (p. 261; see also Weir 2005:298; Alderson 2007:661; Davidson & Fulcher 2007:233). For a CEF ‘Can Do’ statement such as “Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters to do with work and free time” (CEF, p. 74) the linguistic response can be wide-ranging.

Other challenges still face developers who want to compare students at a given level and to connect their test results to each different CEF level. Hulstijn (2007) argues it is impossible to define proficiency by CEF-level without matching and being able to discern between quantity (the number of proficiency factors a user masters) and quality (the degree of effectiveness, efficiency, precision, etc.). A learner may be on a B1 quantity level, but on a C1 or C2 level in terms of quality – or vice versa. Hulstijn closes by saying that “the task to develop and test a theory of language proficiency still remains on our agenda. Many language policymakers and educational professionals seem to ignore the need to conduct such research, pushing for further implementation of the CEFR and developing similar scales for other target groups” (p. 664; see also Davidson & Fulcher 2007 who voice similar concerns.)

Countries, and educators alike, need to agree on the number of words to be known per level. Once they have reached this agreement, they also need to decide which words should be taught at which level. Moreover, ESLC designers and planners should ask themselves to what extent the ESLC takes into account differences in school systems, methods, and objectives.

It is also important to note that CEF levels and their descriptors were initially meant for adult language learners. How does the vocabulary they need to learn differ from the one teenagers need to master? And, how applicable is this approach for 15-year-olds? For objective achievement testing, shouldn't the vocabulary be the one the learner has been taught?

CEF guidelines and testing experts recognize test criteria should be linked to learning materials and words that were actually taught. “Test rating criteria should be linked to the learner’s textbook” (Clifford 2006) and defined by criteria that are appropriate to the “requirements of the assessment task concerned” (CEF, p. 193)

However, because of what we observed on lexical coverage and guidelines above, the results of the ESLC will most likely identify

1. the gap between designed language tests and divergent national criteria (implicit expectations and decisions that have not been agreed upon across Europe) and
2. the gap between designed language tests and language learning materials (the varied didactic materials and methods used).

5. Proposals for solutions

The 2010 ESLC deadline is upon us and much still needs to happen to make this worthwhile evaluation effort reliable and valid. This is why we propose the following practical solutions:

1. the use of criterion-referenced tests on materials actually studied for valid and reliable language testing. Students will also be more confident and motivated if their achievement test targets what they actually *learned* over a certain period of time. The vocabulary of didactic materials presents much variability. They indicate that materials-independent tests will not assess language proficiency with any validity. We should remember that even one unknown word can render a test item unsolvable even when a student has all other skills necessary for the level being tested.
2. the development of study materials that follow a few quantifiable criteria, i.e., size and range of the lexicon for a specific CEF level to allow for more extensive comparisons. To this end, Figueras a.o. (2005) asks: "By which procedures can a person be assigned to one of the described levels on the basis of his or her test performance in such a way that the assigned level corresponds to the level as described in the CEFR? This is, in fact, the fundamental problem of quantitative measurement" (p. 271). The ability to quantify correctly is frustrated by the potential content variety within each descriptor of a certain level. Figueras a.o. concludes: "The effort necessary to scale the descriptors on a common quantitative scale has clear benefits, such as the development of a common vocabulary to compare levels across the cell boundaries in the qualitative grid" (p. 272). "A common vocabulary to compare levels" is possible for a specific learning need, a specific competence, a specific group, and a specific situation. Otherwise, the methodology would necessitate uniformity in content on a wide scale and would not be feasible given the basis of freedom and diversity which lies at the heart of European policy.
3. the availability of a Council of Europe guide on how to construct simple but equivalent test items, and finally
4. the involvement of commercial publishers and their authors in material-dependent test construction since most language materials are produced by commercial publishers. Stakeholders close to the schools may want to demonstrate the quality of their materials. An advantage of such local, material-dependent tests is that they would also conform to the actual progression of students. The carrying-out of practical solutions like these will take some concerted doing however.

6. Conclusion

The European foreign language movement struggles with a contradiction. On the one hand it does not want to impose a standardized approach to language learning, on the other hand it wants to harmonize and create a vast comparative assessment of learners. North (2007) claims that "there is no need for a conflict between the desire to have a central framework to provide

transparency and coherence and the need to have local strategies that provide learning goals specific to particular contexts." True, but then the implementation of a uniform European proficiency test for 15-year olds seems both contradictory and utopian.

We hope that our research has shown the need for more coordination and a more homogeneous determination of lexical sizes related to language levels. Such an approach could enhance language learning assessment efforts in Europe and elsewhere by focusing on the number and kind of words students should be expected to know at each competence level, and making sure students are being tested on the lexicon they have been taught.

References

- Adolphs, S., & Schmitt, N. (2004). Vocabulary coverage according to spoken discourse context. In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 39-49). Amsterdam: Benjamins.
- Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal*, 91 (4), 659-663.
- American Council for the Teaching of Foreign Languages. (2001) Program standards for the preparation of foreign language teachers. Hastings-on-Hudson, NY: ACTFL Materials Center
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6 (4), 253-279.
- Beacco, J.-C., Bouquet, S., & Porquier, R. (2004). *Niveau B2 pour le français. Un référentiel*. Paris: Didier.
- Bergan, S. (Ed.). (2001). *Language policies in higher education: A result of a round table debate of the Council of Europe's Higher Education and Research Committee*. Strasbourg: Council of Europe.
- Clifford, R. T. (2006). *Opening Remarks*. Bureau for International Language Co-ordination (BILC) Conference, Budapest, Hungary.
- Commission of the European Communities. (2005). *Communication from the Commission to the European Parliament and the Council: The European Indicator of Language Competence*. Brussels: Commission of the European Communities.
- Council of Europe. (2001). *Common European Frame of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Coste, D., Courtillon, J., Ferenczi, V., Martins-Baltar, M., & Papo, E. (1976). *Un niveau-seuil*. Paris: Hatier.
- Council of Europe. (2001). *Common European Frame of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. (2003). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching and assessment. Manual preliminary pilot version*. Strasbourg: Council of Europe, Language Policy Division.
- Council of Europe. (2005). *Reference supplement to the preliminary version of the manual for relating examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment*. Strasbourg: Council of Europe, Language Policy Division.
- Cowie, A. P. (1992). Multiword lexical units and communicative language teaching. In P. J.-L. Arnaud, & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 1-12). London: Macmillan Academic and Professional Ltd.
- Davidson, F., & Fulcher, G. (2007). The Common European Framework of Reference (CEFR) and the design of language tests: A matter of effect. *Language Teaching*, 40 (3), 231-241.
- Figueras, N., North, B., Takala, S., Verhelst, N., & Van Avermaet, P. (2005). Relating examinations to the Common European Framework: A manual. *Language Testing*, 22 (3), 261-279.
- Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly*, 1 (4), 253-266.
- Fulcher, G. (2008). Testing times ahead. *Liaison Magazine - Subject Centre for Languages, Linguistics and Area Studies*, 1, 20-24.

- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London: Routledge.
- Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28 (2), 241-265.
- Hazenberg, S., & Hulstijn, J. H. (1996). Defining a minimal receptive vocabulary for non-native university students: An empirical investigation. *Applied Linguistics*, 17, 145-163.
- Hu, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13 (1), 403-430.
- Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91 (4), 663-667.
- Instituto Cervantes. (2006). *Plan curricular del Instituto Cervantes. Niveles de referencia para el Español*. Madrid: Biblioteca Nueva.
- Language Policy Division. (2005). *Threshold level series: Reference level descriptions for national and regional languages*. Strasbourg: Council of Europe.
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35-53). Cambridge: Cambridge University Press.
- Meara, P., & Milton, J. (2003). *The Swansea levels test*. Newbury: Express.
- Milton, J. (2006). Language Lite? Learning French vocabulary in school. *Journal of French Language Studies*, 16, 187-205.
- Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *The Canadian Modern Language Review / La Revue canadienne des langues vivantes*, 63 (1), 127-147.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review / La revue canadienne des langues vivantes*, 63 (1), 59-81.
- North, B. (2007). The CEFR Common Reference Levels: Validated reference points and local strategies. Paper. *Policy Forum: The Common European Framework of Reference for Languages (CEFR) and the development of language policies: Challenges and responsibilities*. Strasbourg: Council of Europe Language Policy Division.
- Rolland, J.-C., & Picoche, J. (2008). Propositions pour un apprentissage systématique du lexique français au niveau A1 et au-delà. In J.-C. Beacco (Ed.), *Niveau A1 et niveau A2 pour le français. Textes et références* (pp. 43-277). Paris: Didier.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12 (3), 329-363.
- SurveyLang. (2008). *Inception report for the European Survey on Language Competences*. Retrieved November 12, 2008, from <http://www.ond.vlaanderen.be/obpwo/oproepen/eslc/>
- Van Ek, J. A. (1976). *The Threshold Level for modern language learning in schools*. Strasbourg: Council of Europe.
- Van Ek, J. A., & Alexander, L. G. (1980). *Waystage English: An intermediary objective below threshold level in a european unit/credit system of modern language learning by adults* (3rd ed.). Oxford: Pergamon Press Ltd.
- Weir, C. J. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing*, 22 (3), 281-300.