KARIN VOGT, Heidelberg

Assessment: Washback of the CEFR and PISA

1. The Common European Framework of Reference as a Language Policy Document

Although the Common European Framework of Reference (CEFR) was only published in 2001, it has a history that goes back to the attempts to formulate a European credit scheme for language learning in the 1970s in the form of the well-known Threshold Level (van Ek 1975; Trim 2007). The document, published during the European Year of Languages 2001 consists of nine chapters dealing with aspects as diverse as the approach adopted for the framework, the common reference levels themselves, language use and the language user, language teaching and learning, the role of tasks, curriculum issues and assessment. The scales, i.e. the arrangement of descriptors in ascending order, and descriptors, i.e. descriptions of language proficiency, with which language proficiency is to be described across languages, however, are at the heart of the document. It is this part which is both most widely acknowledged and discussed.

In fact, having been developed as a common basis for language syllabi, curriculum guidelines, examinations or textbooks across Europe, the CEFR has become the standard reference document for learning, teaching and testing languages in Europe. It can be considered as a milestone in European language policy, with the political impetus of encouraging mobility and promoting understanding and cooperation. More practical rationales include the introduction of a European Language Portfolio (ELP) that would help record progress in acquiring foreign languages, including informal learning environments and partial competences. Another important rationale of the document is the harmonization of educational achievement in general and the European-wide recognition of foreign language tests and certificates in particular. The effect of the CEFR on European and regional educational language policy has been extraordinary, impacting on foreign language education, syllabi and curricula, not only throughout Europe but well beyond it, e.g. in Taiwan (Wu and Wu 2010).

It is interesting to note, however, that there seems to be a discrepancy between the reception and discussion by policy makers and scholars and the attention it receives on the part of frontline teachers. A study in Germany (Vogt 2011) and one international study (Valax 2011) have shown that teachers tend not to engage with the document itself and have no precise knowledge of its contents. This lack of familiarity paved the way for uncritically taking over notions from the CEFR. Leupold and Porsch (2011), in their study of teachers of French in the German context, have shown that teachers are not familiar with CEFR-related language policy documents such as the National Educational Standards (Bildungsstandards).

The lack of familiarity with the CEFR on the part of practitioners might explain the tendency to interpret the basically descriptive document in a prescriptive way, which happens with certain stakeholders. Many authors have pointed out the flaws of
the document in the scholarly literature (Bausch et al. 2003; Fulcher 2004; Alderson et al. 2006; Alderson 2007; Harsch 2007; Vogt 2011) so that this discussion will not represent the focus of this article. Rather, the impact of the CEFR in terms of wash-back, in other words, the effect the document itself and subsequent related measures have had on the teaching of English as a foreign language, will be analysed and discussed with regard to different areas of testing and assessment, with a focus on the German educational context.

2. Impact of the CEFR on a European and International Level

This part of the article can only discuss selected aspects and is far from being comprehensive. Nevertheless it tries to summarize the different ways in which the CEFR has had an effect on educational policy and educational reality.

The Council of Europe advocates an action-oriented approach to language learning and teaching that views the individual as a social agent who uses language to perform actions. The action-oriented approach with functional-notional elements has been embraced throughout in foreign language policy, but not by foreign language teachers on the grassroots level, and this is true in several European countries. The CEFR levels themselves are to be found in many language policy documents, e.g. curricula, school leaving certificates or syllabi. Many member states have incorporated CEFR levels as learning objectives for the foreign language classroom but have apparently not fully implemented the entire document. Scholars and policy makers work on the basis of the CEFR levels, so in this way one aim of the Council of Europe, the enhancement of European co-operation and the establishment of a common yardstick for the description of language proficiency levels, has been achieved.

The impact of the CEFR on international language testing has been considerable, particularly the pressure to align existing (sometimes highly prestigious) standardized exams to the CEFR and its proficiency levels. Work of a testing experts group to produce a manual of how to relate exams to the CEFR has taken several years and came to its conclusion in 2009 (Council of Europe 2009; case studies in Martyniuk 2010). The linking procedure recommended in this document is predominantly based on social moderation, with standard setting used as a method to claim a link between tests and the CEFR (Tannenbaum and Wiley 2004). Summarizing the impact of the CEFR, Little (2006) observes that the focus of the application with regard to the CEFR has been on language testing and much less on other domains such as teacher education, curriculum and course design etc. where projects have been carried out on a local and limited scale.

At least on a language policy level, the impact of a CEFR spin-off, the European Language Portfolio (ELP), has been considerable throughout Europe. Adapting a tripartite structure and pursuing a pedagogic and reporting function at the same time, the ELP is very obviously linked to the CEFR. “I can do” descriptors in the language biography section are derived from the CEFR. To date (November 2011), there are 119 accredited ELPS for target language learner groups in primary, secondary, vocational, tertiary and adult education, designed for various contexts, e.g. very young learners aged 3, primary school, lower and upper secondary and adult education with different target groups such as refugees in Ireland (Little and Lazenby Simpson 2004).
Self-assessment in the ELP is used to record language learning experiences as well as cultural competence that may have been acquired inside and outside of educational institutions. This reporting function is supplemented by a pedagogical function for which self-assessment is seen as one component of learner autonomy and thus represents a key to lifelong learning (Little 2006). Just like the CEFR, the ELP has been fully embraced by policy makers but the degree of implementation at the chalk face is considerably lower, as recent research has shown (Tsagari and Vogt 2011).

3. Impact of the CEFR on a National and Regional Level

PISA

The first of the series of PISA ("Programme for International Student Assessment") studies was undertaken in 2000, focusing on reading literacy and with mathematical and scientific literacy as further domains. Commissioned by the OECD, PISA is the most comprehensive survey of learning outcomes at the end of compulsory school education (Baumert et al. 2001). Among more general objectives like providing reliable indicators of schooling success, one of the declared aims of the OECD and the UNESCO is to proceed to an outcome-oriented approach in education (OECD 1999). This outcome-oriented approach to language learning and assessment has played a major role in the post-PISA development.

Two results sparked a public debate in Germany that was not experienced on this scale in other participating countries. First and foremost, the overall performance of 15-year-olds in Germany was below the OECD average, something which a nation proud of its educational system found hard to digest. Second, the range of the individual results was bigger than in any other country, meaning that while 9% of the tested population performed on the highest competency level, 10% of the pupils did not reach even the lowest level. Altogether, 23% of the representative sample could only read on an elementary level, compared to an OECD average of 18% (Baumert et al. 2001).

National Educational Standards

Following the mediocre PISA results, policy makers quickly took action. Among other major general educational reforms that were almost frantically brought under way, the formulation (and eventual assessment of its outcomes) of unified National Educational Standards (NES, German Bildungsstandards) for core subjects in Germany was agreed on in 2002 by the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder (KMK). The NES, implemented in the curricula of all Länder, are part of the overall educational monitoring strategy of the Standing Conference, which links international studies like PISA with centralized tests of the achievement of national standards as well as comparative tests (Vergleichsarbeiten) which evaluate the Länder-based performance of individual schools (Sekretariat 2006). For English as a first foreign language, NES were published for the different school-leaving certificates issued at the end of grades 9 and 10 (Sekretariat 2003; 2004). They are geared to CEFR levels, although their underlying competence model is slightly different (Rupp et al. 2008). Although formulated in terms of desired outcomes in three domains (intercultural competence,
communicative competence and methodological competence), the formulation of the linguistic competences is at the heart of the document also because the CEFR scales and descriptors are largely used (albeit in an adapted way) to describe learners' average learning outcomes in terms of language proficiency after nine or ten years of schooling.

NES and the attached illustrative test tasks have been the subject of much scholarly criticism (cf. Bausch et al. 2005). The following remarks will only focus on the role of the CEFR in the document, though (for a more detailed critique cf. Quetz and Vogt 2009). Again, a descriptive frame of reference of the original document was taken over for a normative document, to support an outcome-oriented approach to foreign language teaching and learning. The weaknesses of the scales and descriptors such as inconsistencies, vagueness, and lack of an underlying learning theory were perpetuated in the process of mapping proficiency levels and descriptors to the CEFR descriptors. In the NES, descriptors (or levels) are only available for linguistic competences, with other competences like intercultural competence or strategic competence de facto being neglected.

In sum, an outcome-oriented approach to language learning initiated by the large-scale assessments (PISA, TIMMS) was taken further by the NES. In the aftermath of the NES, new forms of establishing whether learners had achieved the stipulated proficiency level were devised. In line with the national educational monitoring strategy, different measures were taken, among them the monitoring of outcomes via standardized tests designed by the Institut für Qualitätsentwicklung im Bildungswesen (IQB), founded in 2004 as an institution financed by all 16 Länder, and a tendency towards central final school-leaving examinations in the different Länder.

International large-scale tests like PISA or later DESI for English would evaluate the success of the educational system in its entirety. Results of standardized tests for VERA, comparative tests, would give feedback on the school level, also feedback for (further) improvement of teaching processes. The role of the IQB herein is to provide suitable test items that would comply with quality criteria of standardized testing. Individual feedback, e.g. from VERA tests, would improve the learning processes on a learner level. In all cases, the test results would ideally provide hints on how to improve both learning and teaching the foreign language.

**VERA and Other Comparative Tests**

Comparative tests (VERgleichsArbeiten) were developed in order to check whether learners would be able to attain the competence levels stipulated by the NES. They are undertaken one or two years before the actual end of schooling in order to give learners and teachers the possibility of altering their learning and teaching processes. Test items are devised by the IQB while the different Länder are responsible for the organization, data analysis and reporting procedures. The Länder also decide on the length and determine the test items to be used in the tests, which are also referred to by different names in the particular states.

The tests items for English are professionally designed by test experts, trained item writers and teaching experts and they are piloted; in other words, they comply with academic standards for standardized tests and provide data on the language proficiency of pupils (for detailed results cf. Köller et al. 2010). They try to take into
account educational realities by including more open test formats, not exclusively
closed formats like multiple choice or true/false items. There are a few disadvantages,
however. In general, test tasks in standardized tests do not always reflect communica-
tive purposes. Besides, VERA in English only covered receptive skills (reading com-
prehension and listening comprehension) in 2009, probably for pragmatic reasons.
Also, these tests can only cover linguistic aspects of foreign language proficiency.
Other important aspects such as intercultural competence or strategic aspects of for-

teign language learning are not considered in the tests although they are part of the

competence model of the NES. In case of a washback, this would be a fatal develop-
ment because teachers would teach the “relevant” aspects only in order to best prepare
their learners for the test, and a teaching to the test would take place, with the main
goal of a communicative classroom being severely neglected. Weskamp (2009, 269)
maintains that effective tests that would evaluate standards-based learning outcomes
in a German tradition have to take into account the criteria of authenticity and interac-
tivity. They are to consider the needs and interests of the learners as well.

To avoid an unwanted washback effect, to better take into account the realities of
the communicative foreign language classroom and in order to bring assessment and
learning closer together, tests ought to be embedded in more complex learning tasks.
Tasks would contextualize tests better and make them more transparent for learners,
thus increasing their face value. Besides, tasks linked with appropriate tests would
ensure that the necessary competences are acquired little by little. The IQB has com-
missioned an expert group to devise, pilot and evaluate appropriate tasks that would
gradually build up the desired competences. Work in this expert group has revealed
the importance of formative assessment procedures (i.e. assessment employed during
the learning process) to monitor competence development. And indeed, a strong focus
on summative assessment (i.e. assessment that summarizes the development of learn-
ers) by way of large-scale testing can be demotivating for both teachers and learners,
as experience from the UK has shown (Gardner 2010). This is particularly true when
alleged standardized tests, e.g. designed as centralized school-leaving exams or as
comparative tests administered by the Länder, do not comply with quality criteria of
tests, as has happened in different Länder in Germany (e.g. Quetz and Vogt 2009).

Even when standardized tests are good tests, they need to be taken seriously by
teachers and learners. At the moment, this is not always the case when learners realize
these tests “do not count” for their grade and teachers feel hassled with the organiza-
tional burden without seeing the benefits. In this context, Weskamp (2009, 269) main-
tains that standards-driven educational reforms will only succeed if they include the
teachers’ positions and their subjective theories. So far, this has not quite happened,
which jeopardizes the success of the whole undertaking because a vital group of
stakeholders has been largely ignored (Vollmer 2006). This view is corroborated by
the findings of Leupold and Porsch (2011), who established that NES seem to have
had very little impact on teachers’ perceptions or their actual teaching. It might there-
fore be debatable whether the NES will have a positive effect on foreign language
teaching in terms of rich, communicative learning environments and eventually in
terms of Bildung in the Humboldtian tradition.
European Language Portfolio

A (potentially) powerful aspect of washback of the CEFR is represented by the European Language Portfolio (ELP). Again, the acceptance of the ELP with regular teachers is considerably lower than with scholars or educational authorities, given the fact that the use of the ELP is recommended in curricula (e.g. Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen 2007). The ELP, and more specifically its language biography part, offers a great potential for alternative assessment, particularly self-assessment. The can-do descriptors place an emphasis on what a learner can already do in the foreign language, even on an elementary level. This positive approach can be rather motivating for learners – for once their deficiencies are not at the centre of attention. Also, learner self-assessment can form an addition (and represent a counterbalance) to teacher assessment, which still dominates the foreign language classroom. Self-assessment can be used for diagnostic purposes and thus might represent the first step towards differentiated language learning, again contributing to a standards-based approach, albeit by incorporating learner-centredness. Small-scale research among teachers of English in Germany has shown, however, that teachers were not all that convinced of the advantages of the ELP and partly only used it when its use was made compulsory at their school (Tsagari and Vogt 2011; for more research on the use of ELPs in upper secondary cf. Flächer 2011). Materials developers have included portfolio-style documents with either the dossier function or as checklists, often asking learners to rate their progress after a unit of learning in the textbook. Although this does not equal the systematic work that could be done with an ELP, it could be a first step towards including more innovative forms of assessment in more "traditional" foreign language classrooms.

4. Conclusion: Assessment in Post-CEFR Times

The outcome-oriented approach to learning and teaching foreign languages has brought about new forms of assessment and a tendency towards more large-scale and standardized assessment in the German context. In operationalizing this approach, the CEFR has been misused as a normative document, for NES in particular but also for central exams. Although it has a relatively solid empirical basis (North 2000), the CEFR is not without flaws. More importantly, it is not "the truth" on language learning and the authors have never claimed it was. It only partly lends itself to certain assessment purposes, e.g. observation of a known learner over time in a formative assessment context. Being a frame of reference only (albeit a good one), it must not be misused for certain assessment purposes, e.g. high stakes exams, which it is not suitable for.

The ELP as a document closely linked to the CEFR has positive potential for alternative forms of assessment that could complement traditional teacher assessment. Although self-assessment and peer-assessment could contribute to more learner-centred foreign language learning, they seem not to have been embraced by foreign language teachers on a larger scale. The reasons for this might be manifold; one of them could be the open question of how teacher assessment and alternative forms of assessment can be reconciled.

In conclusion, the dangers of a reduced pragmatism in focusing on testable linguistic aspects in a communicative foreign language classroom should be given due
attention. Communicative language learning comprises further competences which ought to be assessed, too. This might be an item on researchers' to-do-lists, particularly if problematic washback is to be avoided. If we truly want to establish a communicative foreign language classroom, we must strive to find appropriate instruments for testing and assessment, particularly for classroom-based assessment.

**Works Cited**


Tsagari, Dina, and Karin Vogt. *Classroom-Based Language Assessment – Teachers’ Perspectives*. Poster presented at the 8th Annual Conference of the European Asso-


