Thailand) and assessment practices, a situation certainly shared by many other economies around the world, and one that adversely affects English language education. Assessing knowledge in a more integrative and direct fashion has considerable associated costs, which is why more efficient and psychometrically reliable multiple-choice tests are often selected.

The argument could be made that these more "efficient" and cost-effective tests are good *indirect* measures of oral ability. However, they have very poor face validity in that regard. This trend of misaligned curriculum and assessment is very discouraging for students and teachers who, rather than embrace 21st century curriculum and standards or respond to the particular interests and needs of their own students, must teach to the standardized test. That is, the test leads to negative "washback" in teaching (Cheng, Watanabe, & Curtis, 2004) and is therefore not conducive to best practices in language education. Even if tests seem to indirectly measure a particular skill like speaking and writing, if those skills are not visible to potential test-takers or to teachers, they are unlikely to devote sufficient attention to their development. The tests' construct validity in the light of standards and curriculum developed with other explicit objectives is then easily challenged. It was largely in response to such concerns that the US-based Educational Testing Service (ETS) recently concluded its extensive redevelopment of the TOEFL exam after many years of research at ETS and consultation with the professional community of scholars and language educators. As a result, the Internet-based TOEFL now includes both speaking and writing components, whereas the Test of Written English was optional before and there was no test of speaking for general test-takers; other changes were also made. An expected consequence of that test reform will be a concomitant increase in attention paid to those skills in schools, in test-preparation centers, in related language teaching/learning materials, and in the consciousness of learners, teachers, and parents about valued competencies and skills—in other words, positive washback effects are expected.

IV.   Exemplary Standards "Frameworks": Language Learning Proficiency Scales for S/FL Learner Profiles (e.g., Common European Framework)

The EDNET report by Chen et al. (2008) provides a commendable analysis of the following four well known and generally well respected standards for English and other L2 learning developed in different regions of the world:

- USA (ACTFL) – originally college-level, oral[2]
- Europe (Common European Framework of Reference, CEFR) – broadest appeal
- Canada (Canadian Language Benchmarks) – adult workplace
- Australia (International Second Language Proficiency Rating) – adult primarily

Another standards documents not included in the report, which has a shorter history of development and implementation in any case and less related testing research, include the international organization of Teachers of English to Speakers of Other Languages' (TESOL's) "ESL Standards for Pre-K-12 Students.[3]" These standards have a great deal in common with the four standards documents reviewed in terms of their underlying principles of language learning and language pedagogy, stressing language for communication, language for academic learning, and pragmatic or functional aspects of language use.

---

[2] See Svender & Duncan's (1998) guidelines for ACTFL use with k-12 learners.
[3] Available at: http://www.tesol.org/s_tesol/seccss.asp?CID=95&DID=1565.

The four standards documents listed above all benefited from a long period of incubation, considerable revision, expert consultation and research (from the testing community, language educators, and policy-makers), and many years of implementation. Not surprisingly, there was also a good degree of cross-fertilization among them, as many of the same expert consultants worked on them at different points since the standards were expected to reflect the state of the art internationally and not just nationally. Furthermore, all have much to offer APEC standards/practices, especially the CEFR (Buck, 2007; Byrnes, 2007; Chen et al., 2008). Below I elaborate on the CEFR specifically, which has much to offer APEC economies concerned with adopting or referencing a common metric of language proficiency should consider carefully.

1.   Some advantages of CEFR

CEFR has had wide internationally impact and implementation and serves as an excellent model or reference point for APEC economies, although their local contexts are naturally quite different from those of European Union economies. CEFR has also spawned important new trends in assessment, such as the European Language Portfolio, giving students more agency in recording and reflecting on their own functional abilities and experiences with the languages in their repertoire. It encourages formative and summative self assessment, multilingual "biographies" and identities, and dossiers, all in the spirit of cultivating a "plurilingual" citizenry.

Excellent recent position papers on CEFR appeared in the *Modern Language Journal*, 2007 (Alderson, 2007; Byrnes, 2007; Little, 2007; North, 2007), pointing out both its strengths and limitations. In general, the strengths far outweigh any limitations. CEFR has three main levels of proficiency (A, B, C, with C the highest) and then proficiency distinctions within each level. It is generally lauded for being teacher-friendly and intuitive, using non-technical language that is easily accessible to non-specialists trying to implement it. It has been adopted by all countries in Europe and others far beyond Europe, such as New Zealand. The Council of Europe, which sponsored its development, wanted to facilitate the "mutual recognition of language qualifications in Europe," (http://www.coe.int/t/dg4/linguistic/CADRE_EN.asp), and it has gone a long way toward doing precisely that. In addition, CEFR has demonstrated a positive potential impact on teaching and curriculum, as well as on preservice and inservice teacher education--and not just on assessment. It also has had a positive impact on stated learning outcomes. For example, in France, students are expected to attain "B1" standing (as "independent users") in their first L2 and A2 level (as "basic users") in their second L2. University graduates are expected to have reached a C2 level ("mastery", or near-native ability), the highest in the CEFR, in their L2.

Experts reviewing the CEFR also note that it has a favourable influence on classroom assessment, it is functional and task-oriented, and can also be applied to language learning for a variety of purposes: learning language for work, study, social activity or tourism, and so on. Finally, the CEFR's very positive orientation is often cited as an appealing aspect of its use for assessment, stressing what learners *can* do, rather than what they *cannot* do. It therefore is more motivating and encouraging for students than assessment criteria framed in terms of deficiencies or error types or other inadequacies. For example, as the table below, adapted from the Association of Language Teachers of Europe (http://www.alte.org), illustrates, at level C2-5, a student "*can* advise on or talk about complex or sensitive issues, understand colloquial references and deal confidently with hostile questions." In writing, students "*can* write letters on any subject and full notes of meetings or seminars with good expression and

accuracy". At the lowest level, A1-Breakthrough, on the other hand, students "*can* understand basic instructions" or "complete basic forms." At B1-2, about half way between the other two extremes and representing an intermediate level, students "*can* express opinions on abstract/cultural matters in a limited way or offer advice within a known area" and "*can* write letter or make notes on familiar or predictable matters."

Examples of "CAN-DO" Levels from CEFL
(http://www.alte.org/can_do/general.cfm)

Levels	Listening/speaker	Reading	Writing
C2 – Level 5	CAN advise on or talk about complex or sensitive issues, understanding colloquial references and dealing confidently with hostile questions.	CAN understand documents, correspondence and reports, including the finer points of complex texts.	CAN write letters on any subject and full notes of meetings or seminars with good expression and accuracy.
B1 – Level 2	CAN express opinions on abstract/cultural matters in a limited way or offer advice within a known area, and understand instructions or public announcements.	CAN understand routine information and articles, and the general meaning of non-routine information within a familiar area.	CAN write letters or make notes on familiar or predictable matters.
A1 – Breakthrough level	CAN understand basic instructions or take part in a basic factual conversation on a predictable topic.	CAN understand basic notices, instructions or information.	CAN complete basic forms, and write notes including times, dates and places.

2. *Some limitations of CEFR*

Despite these many attractive features of CEFR, the European context, as noted earlier, is certainly not the same as APEC's, with respect to the range and types of languages represented, the mobility of students and teachers, the official policies espousing multilingualism and immigration, and then the economic, political, and other relationships across regional economies. At present, CEFR levels are not anchored to any specific language (but have been translated into 23 European languages), therefore issues of transferability, or comparability of levels across languages must be explored to a greater extent. Within Europe, for example, many languages have familial links and learning other languages within the same language family is generally considered less time-consuming than learning typologically unrelated languages (e.g., see an oft-cited study by Liskin-Gasparro, 1982, summarized by Hadley, 2001, that supports this assertion). APEC obviously also represents a geographically much vaster area than Europe, in terms of potential mobility for educational purposes.

More daunting perhaps, is that, in practice, it is often difficult to get raters of tasks on tests to agree on the specific levels of speech or writing that they are assessing or targeting, especially across countries and distinct languages. For example, it is difficult to determine whether a particular task for either testing or teaching purposes is a B1 or a B2 task and similarly it can be difficult to assess whether students' performance is B1 or B2 level (Marianne Nikolov, personal communication, October, 2007, with respect to the adoption of CEFR and inter-rater training in Hungary; see Alderson, 2007).

Another critique of CEFR is that, although it was based on extensive L2 testing research and consultation with L2 teachers, it has not really been validated by parallel second language acquisition developmental data, for example monitoring how students progress from one level to another, if indeed that is how they progress. The levels make great sense intuitively but a stronger interface between testing research and second language acquisition research would further strength them. Alderson (2007) therefore suggests that the test data need to be verified with test corpus data. Alderson and Little (2007) point out that the CEFR has to date had more impact on the field of testing such as the Association of Language Testers of Europe (ALTE), and especially private companies' testing interests, than on official high school matriculation testing, curriculum design, materials, and pedagogy.

Other limitations of the CEFR are the following:

(1) It has been used primarily with young adults. With the introduction of foreign language teaching (and assessment) at earlier grade levels CEFR tasks or competencies likely need to be adapted somewhat.

(2) For content-specific learning (called "language of schooling" in Europe) rather than general-proficiency language teaching and learning, additional modifications might be necessary.

(3) Although it accounts for second-language pragmatics (appropriateness of language use), CEFR doesn't directly and explicitly take into account cultural or literary knowledge.

V. Other Issues Related to Assessment and Standards

1. Assessing language learners across APEC economies

The previous section highlighted the strengths and limitations of CEFR for potential adaptation in and across APEC economies. Certainly, it has numerous strengths. In considering the matter of adopting or adapting such instruments in APEC, a tension must be acknowledged between the desire to establish comparisons in learning outcomes (or standards) across economies/languages by using well-field-tested instruments, on the one hand, and the need for local autonomy, responsiveness to local contexts, and a sense of agency and ownership of policy/standards/practices on the part of local experts/teachers, on the other hand. Furthermore, borrowing curriculum or assessment instruments developed in a very different educational and geopolitical context does require a full understanding of how and why particular instruments were developed in the first place and how best to use or adapt them.

Within APEC economies presently, according to the 2007 EDNET survey, there are many approaches to testing: from local classroom-based and national standardized instruments to international standardized tests such as those developed by the University of Cambridge, UK. In general, it appears that most APEC language tests are locally developed, but ensuring that tests reflect curriculum contexts/levels and objectives well has been an ongoing concern.

One advantage of using an internationally standardized examination system is that it facilitates comparisons of results across contexts and helps establish the readiness of learners to study abroad or in second-language immersion programs, for example. However, again the