C-ESOL organizes the test development and validation process into five stages: initial planning and consultation, development, validation, implementation, and operation (Falvey & Shaw, 2006). Weir (2005b) has created a socio-cognitive framework for prioritizing and conducting crucial validation activities that enable test developers to build compelling validity arguments for tests. His framework contains five elements (context validity, theory-based validity, scoring validity, consequential validity, and criterion-related validity) and considers three dimensions (test taker characteristics, task response, and score). Weir's framework reflects current trends in the design and validation activities associated with large-scale, high-stakes EFL tests and it has informed the activities of C-ESOL test developers.

ETS operates an active program of research and development that supports its EFL tests and the results are published in a series of monographs and technical papers that are available on the publisher's Web site. The results of many of these papers were integrated into a recently published case study of the development of the iBT (Chapelle, Enright, & Jamieson (2008). The volume presents one of the most comprehensive descriptions of the evidence and validity argument for a high-stakes EFL test currently available. In the book, project participants articulate a framework for the project and summarize the validation activities that informed the design of the test and support the interpretations and use of iBT scores. One key aspect of the project was the construction of an interpretive argument for the new TOEFL and it was based on recent developments in validation theory and current standards of educational measurement.

VI.   Conclusion

Language testing is increasingly acknowledged to be not only a form of educational practice but a form of social and political practice as well (McNamara, 2008; Shohamy, 2001). Given the broad impact of tests on individuals and society, language education policymakers, testing specialists, and test users are obliged to strive to minimize the negative consequences of using high-stakes tests of L2 ability and to maximize the positive consequences. This is more likely to occur in a context in which test development and use are viewed as a shared responsibility and where the highest professional standards and best practices occur. In this paper, I have reviewed some of the recent developments and current standards that are being applied to the design and use of large-scale, high-stakes tests of English language ability.

References

ALTE/Association of Language Testers in Europe (2001). *Code of practice*. Retrieved 11, June, 2008, from http://www.alte.org
American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, D.C.: AERA.
Bachman, L. F. (1988). Problems in examining the validity of the ACTFL Oral Proficiency Interview. *Studies in Second Language Acquisition, 10,* 149-164.
Bachman, L. F. (1990). *Fundamental considerations in language testing*. New York: Longman.
Bachman, L. F. (2002). Some reflections on task-based language performance assessments. *Language Testing, 19*(4), 453-476.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Canale, M. (1983). On some dimensions of language proficiency. In J. Oller (Ed.), *Issues in language testing research* (pp. 333-342). Rowley, MA: Newbury House.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Chalhoub-Deville, M. (1997). Theoretical models, assessment frameworks, and test construction. *Language Testing, 14*(1), 3-22.

Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing, 20*(4), 369-383.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York: Routledge.

Chapelle, C. A., Grabe, W., & Berns, M. (1997). *Communicative language proficiency: Definitions and implications for TOEFL 2000* (TOEFL Monograph Rep. No.10). Princeton, NJ: Educational Testing Service.

Chapelle, C. A., Jamieson, J. M., & Hegelheimer, V. (2003). Validation of a Web-based ESL test. *Language Testing, 20*(4), 409-439.

Chen, H., Sinclair, P., Huang, S-y, & Eyerman, L. (January, 2008). APEC EDNET project seminar on language standards and their assessment: Background research paper. Symposium conducted at the APEC EDNET meeting, Xi'an, China.

Cheng, L. (2008). The key to success: English language testing in China. *Language Testing, 25*(1), 15-37.

Cheng, L., & Qi, L. (2006). Description and examination of the National Matriculation English Test. *Language Assessment Quarterly, 3*(1), 53-70.

Choi, I-c. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing, 25*(1), 39-62.

Crystal, D. (1997). *English as a global language*. Cambridge: Cambridge University Press.

Davies, A., Hamp-Lyons L., & Kemp, C. (2003). Whose norms? International proficiency tests in English. *World Englishes*, *22*(4), 571-584.

Douglas, D., & Hegelheimer, V. (2007). Assessing language using computer technology. *Annual Review of Applied Linguistics 27*, 115-132.

Duff, P. A. (January, 2008). APEC second/foreign language standards and their assessment: Trends, opportunities, and implications. Symposium conducted at the APEC EDNET meeting, Xi'an, China.

Eckes, T., Ellis, M., Kalnberzina, V., Pižorn, K., Springer, C., Szollás, K., & Tsagari, C. (2005). Progress and problems in reforming language examinations in Europe: Cameos from the Baltic States, Greece, Hungary, Poland, Slovenia, France and Germany. *Language Testing, 22*(3), 355-377.

Educational Testing Service. (2002). *The ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2004). *English language competency descriptors*. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2007). *TOEIC examine handbook: Listening and reading*. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2007). *TOEIC speaking and writing sample tests*. Princeton, NJ: Educational Testing Service.

Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics 23*, 282-301.

Falvey, P., & Shaw, S. (2006). IELTS writing: Revising assessment criteria and scales (Phase 5) (*Research Notes 23*). Cambridge: Cambridge ESOL.

Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly, 1*(4), 253-266.

Gottlieb, N. (2008). Japan: Language policy and planning in transition. *Current Issues in Language Planning, 9*(1), 1-68.

Hamp-Lyons, L., & Davies, A. (2008). The English of English tests: Bias revisited. *World Englishes, 27*(1), 26-39.

Jenkins, J. (2006a). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly, 40*(1), 157-181.

Jenkins, J. (2006b). The spread of EIL: A testing time for testers. *ELT Journal, 60*(1), 42-50.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, D.C.: AERA.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice, 21*(2), 31-41.

Kaplan, R., & Baldauf, R. (2005). Language-in-education policy and planning. In E. Hinkle (Ed.), *Handbook of research in second language teaching and learning* (pp. 1013-1034). Mahwah, NJ: Erlbaum Associates, Publishers.

Kunnan, A. J. (2008). Large scale language assessments. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education (2nd ed.)*, (Vol. 7, Language testing and assessment, pp. 135-155). New York: Springer.

Norris, J. (2002). Interpretations, intended uses and designs in task-based language assessment. *Language Testing*, *19*(4), 337-346.

Norris, J., Brown, J. D., Hudson, T., & Bonk, W. (2002). Examinee abilities and task difficulty in task-based second language performance assessment. *Language Testing, 19*(4), 395-418.

McNamara, T. (2008). The socio-political and power dimensions of tests. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education (2nd ed.)*, (Vol. 7, Language testing and assessment, pp. 415-427). New York: Springer.

Prapphal, K. (2008). Issues and trends in language testing and assessment in Thailand. *Language Testing, 25*(1), 127-143.

Purpura, J. E. (2008). Assessing communicative language ability: Models and their components. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education (2nd ed.)*, (Vol. 7, Language testing and assessment, pp. 53-68). New York: Springer.

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing, 22*(2), 142-173.

Qian, D. (2008). English language assessment in Hong Kong: A survey of practices, developments and issues. *Language Testing, 25*(1), 85-110.

Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing, 18*(1), 1-32.

Ross, S. (2008). Language testing in Asia: Evolution, innovation, and policy challenges. *Language Testing, 25*(1), 5-13.

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London: Pearson.

Stoynoff, S. J. (2009). Recent developments in language assessment and the case of four large-scale tests of ESOL ability. *Language Teaching*, *42*(1), 1-40.

Stoynoff, S. J. (2007). Assessing communicative competence: From theory to practice. In J. Lui (Ed.), *English language teaching in China* (pp. 127-149). London: Continuum.

Tannenbaum, R., & Wylie, E. C. (2005). *Mapping English language proficiency scores onto the Common European Framework* (TOEFL Research Report Rep. No. 80). Princeton, NJ: Educational Testing Service.

Taylor, L. (2006). The changing landscape of English: Implications for language assessment. *ELT Journal, 60*(1), 51-60.

Weir, C. J. (2005a). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing, 22*(3), 281-300.

Weir, C. J. (2005b). *Language testing and validation: An evidence-based approach*. Basingstoke, UK: Palgrave.

Wigglesworth, G. (2008). Task and performance based assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education (2nd ed*.), (Vol. 7, Language testing and assessment, pp. 111-122). New York: Springer.

Zenisky, A., & Sireci, S. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education, 15*(4), 337-362.

Zheng, Y., & Cheng, L. (2008). College English Test in China. *Language Testing, 25*(3), 408-417.