Second Language Testing: Interfaces between Pedagogy and Assessment. An Introduction

It should go without saying that language assessment[1] and language pedagogy are clearly connected –i.e., that there are clear interfaces between the two. In fact, Davies (1991) claims that language testing, both the British as well as the North American tradition, "grew out of language pedagogy needs":

> Language testing as practiced today, grew out of the special demands on language assessment of large-scale, on-demand intensive adult language courses in World War 2, which had a similar influence on parallel developments in language teaching materials and methodology (Davies, 1991: 136).

Clearly both disciplines are inter-related and affect each other. New theories of teaching and learning may lead to changes in testing practices (Spolsky, 1995) while language tests, especially high-stakes tests, which have a direct impact on students' immediate futures, may affect teaching and learning (Cheng, 1997; Clapham, 2000; Wall, 1996). One fundamental interface of these two fields is the issue of methodological selection. If a second language is taught, say, following the methodological guidelines of the communicative approach (cf. Breen & Candlin, 1980; Brumfit & Johnson, 1979; Canale & Swain, 1980; Johnson, 1982; Johnson & Morrow, 1981; Littlewood, 1981; Richards, 2006; Savignon, 1983, 1991, 2002; Widdowson, 1978), it should then be assessed accordingly by using test tasks that allow for the measurement of the communicative competence of the test-taker. Otherwise, it would not be possible to make inferences about the test-takers' abilities on the basis of the scores yielded by the test —i.e., the assessment carried out would lack construct validity (cf. Bachman,

---

[1] Although many authors have used the term 'assessment' as both a general umbrella term which includes all methods of testing, and as a term to distinguish 'alternative assessment' or more informal testing techniques from testing, in this introduction both terms are used interchangeably. (Clapham 2000; Gipps, 1994)

1990; Bachman & Palmer, 1996). While there is some evidence pointing to language testing researchers drawing on work in language education research, it appears to be little communication and collaboration between both these fields (Rea-Dickins, 2007). That is why some authors call for increased efforts to align classroom language teaching with testing approaches:

> For testing and assessment to be valid (testing what is supposed to be tested) and reliable (getting consistent results), the instructor must make the connection between the way she/he teaches and the way she/he tests. It is just common sense, but there are also pedagogically sound reasons for finding ways to connect one's teaching and testing. These reasons include key factors for the treatment of students: equity, fairness, and transparency (Hancock, 2006: 12).

When considering the interface of methodological selection, especial emphasis should be placed on the arrival of the communicative movement in the late 1970s and early 1980s, which highlighted both communicative language teaching methods and language testing. Thus, as communicative language teaching approaches evolved, language testing researchers showed a growing interest in promoting communicative language testing, whose advent was christened 'The Promised Land' by Morrow (1979). It goes without saying that the wide acceptance of the communicative approach has had a pervasive influence on teaching, learning and testing researchers worldwide who have consequently developed new ways of approaching language tests (Fulcher, 2000). In fact, communicative language teaching and testing approaches have continued to be encouraged over recent decades (Bachman, 1990; Bachman & Palmer, 1996; Canale & Swain, 1980, 1981). Furthermore, as Fulcher (2000: 493) explains, "it would be difficult to market a new large-scale test that did not claim to be 'communicative' –whatever the term may mean for different users". In this regard, one of the most influential publications of the last decade in the field of language pedagogy and language testing in Europe has been *The Common European Framework for Language: learning, teaching and assessment* (Council of Europe, 2001; CEFR) (Figueras *et al.,* 2005). The CEFR describes a model of language use referred to as the "action-oriented approach" which highlights the development of communicative language competences based on learners' "communicative needs". With regard to testing methods, the CEFR places an emphasis on performance testing in an attempt "to describe language proficiency through a group of scales composed of ascending level descriptors couched in terms of outcomes" (Weir, 2005: 281). The main aim of the CEFR is summarized below:

> The Common European Framework provides a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe. It describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively. The description also covers the cultural context in which language is set. The Framework defines levels of proficiency which allow learners' progress to be measured at each stage of learning and on a life-long basis. (Council of Europe 2001:1)

However, if there is a fundamental principle in language assessment that has a clear interface with second language teaching and learning, that is washback. It has been sufficiently proved now that language testing has an influence on teaching and learning (Alderson & Wall, 1993; Bailey, 1996, 1999; Burrows, 2004; Cheng, 2005; Cheng & Curtis, 2004; Messick, 1996; Shih, 2007; Shohamy *et al.,* 1996; Spratt, 2005; Wall, 2000, 2005, 2011, 2012; Watanabe, 2004). In other words, and in answer to the question posed a couple of decades ago by Alderson and Wall (1993), washback does exist, although it seems to be an extremely complex phenomenon (Rea-Dickins, 2007). Although traditionally language testing has been mainly concerned with issues of test design rather than with the actual consequences of the use of tests within educational systems (Green, this volume; Linn, 1997; Messick, 1996; Stobart, 2003), there has been a growing concern about consequential validity, which includes evidence and rationales for evaluating the intended and unintended consequences of score interpretation and the use of tests (Messick, 1996), as well as their impact on both the larger macro level (i.e. educational system, textbook publishers and wider society) and the micro level (i.e. teachers and students' attitudes, curriculum, classroom materials, etc.) (Shohamy 1997, 1998, 2001; Taylor 2005, 2006; Wall 1997). In fact, many authors point to test use and associated consequences as the major factors influencing teaching and learning (Gipps, 1994; Shohamy, 1992). In this regard, and despite the fact that washback has traditionally been associated with rather more damaging than beneficial effects (Chapman & Snyder, 2000; Hughes, 1989; Spolsky, 1996), over recent decades, tests, especially high-stakes tests, have been started to be exploited to promote pedagogical reform and improve educational practices (Spolsky, 1996; Weir, 1990). Furthermore, communicative language testing has placed special emphasis on the promotion of positive washback in order to enhance communicative teaching and therefore support desired instructional practices and learning (Green, 2007, this volume). Underlying this latter approach there is the assumption that tests should aim at developing and improving student performance and not just merely evaluate it (Wiggins, 1998). This can only be achieved where testing is linked to teaching and learning: "test should serve as a nexus to connect classroom based instruction practices and work in tandem" (Adair-Hauck *et al.,* 2006: 365). Otherwise, research suggests that tests may have a negative impact on the development of communicative skills since most of the class time is devoted to the teaching of skills featured in the test neglecting untested skills and materials (Amengual 2009, 2010; Amengual & Méndez García, 2012; Hughes, 1989; Laborda, 2010; Laborda & Martín Monje, this volume; Spolsky, 1996). In spite of the potential benefits attributed to the manipulation of high-stakes tests to deliberately produce beneficial washback, available evidence suggests that language tests seem to affect certain areas of teaching (i.e. content, curriculum) but have only limited apparent success on others areas such as teaching methods (Alderson & Wall, 1993; Cheng 2005; Luxia, 2007; Shohamy *et al.,* 1996). Therefore, recent research has made it clear that washback is not deterministic and that tests cannot, of themselves, reform instructional practices and produce beneficial

washback effects (Green, 2007). Thus, if effective improvements in language education are to be realized, testing should be connected to teaching and be used in support of teaching and learning processes (Gipps, 1994; Green, 2007; Shohamy, 2001). Also, the increasing interest in the ethics of language testing (Shohamy, 1997, 1998, 2001) may contribute to enhance instructional programs and develop standards of good testing practice:

> As language testing continues to grow and develop into the twenty-first century, concerns for such issues as public accountability, fairness to test takers, the uses of test results, and technical qualities of tests (validity, reliability, construct definition, scale definition, score interpretation) must be clearly and explicitly stated and monitored by the international language testing community. (Douglas, 1995: 176)

According to Hancock (2006: 12), "foreign or second language teaching and instruction must link testing in deliberate ways". As has been acknowledged, there seems to be a great potential for fruitful cooperation between language testing researchers and language teaching researchers. In this regard, many authors advocate a greater involvement of tests developers, administrators and language teachers (Clapham, 2000; Standsfield, 2008). We trust that the collection of papers in this volume will help to provide a better insight into testing and the ways in can be integrated into classroom teaching. It is our belief that this mutual cooperation will undoubtedly contribute to a more "fair and just society" (Standsfield, 2008: 323).

AN OVERVIEW OF THE VOLUME

In their article, **Glenn Fulcher and Agneta Svalberg** discuss the concepts of criterion- and norm-referenced (CR and NR) assessment with a double purpose: a) to reject claims that criterion-referenced testing is *a dead enterprise* (on the grounds that what has so far passed as CR testing is in fact not so) and, most importantly, b) to scrutinize the concept of CR testing so that a more accurate definition of what it actually involves can be provided. The authors argue that the establishment of cut-scores with reference to descriptive models such as the CEFR is illusory, based on the corruption of the term "criterion" and, above all, it is not a sufficient condition for a test to be considered CB. True CB tests are those that insist on (i.e., provide an accurate description of) actual behavior in a specified domain beyond the test. Because further applied linguistic description of domain specific communication is necessary, the authors suggest that experimental research using CR tests, group difference studies with groups who are known to differ on the criterion on test-external grounds, and/or intervention studies should be carried out, as CR tests should be more sensitive to instruction than NR tests. The authors conclude by reminding that, even though CB tests will very likely continue to be used for high-stakes decision making, it should never be forgotten that language testing is a social science (in spite of all the statistics and measurement theory involved).

Irina Argüelles Álvarez contributes an empirical study on the rationale and validation process of a multiple-choice grammar test at the *Universidad Politécnica de Madrid* (UPM) for the accreditation of a B2 level, in accordance with the *Common European Framework of Reference for Languages* (CEFR). The researcher advocates for the introduction of an indirect grammar test to assess proficiency in tertiary settings on the grounds of practicality. Argüelles presents multiple-choice tests, not as substitutes for direct tests, but rather as adequate alternatives for increasing reliability and maintaining good levels of criterion-related validity by placing students above or below the minimum B2 proficiency level, which was established to regulate students' access to a new compulsory subject implemented at the UPM. The results obtained from the large-scale application of the test to 924 students enrolled in different degree courses based on final test version item analysis will lead to conclusions about students' eventual acquisition sequences of grammar. Apart from being aware of some of their students' grammatical difficulties, these latter results may help teachers to introduce grammar points accordingly integrating language testing completely into the teaching learning process.

In this paper, **Anthony Green** continues his examination of the issue of washback in language assessment (Green 2006a, 2006b, 2007). He presents us with a review of the evolution of washback studies in second language testing in the quarter century. Particularly, he describes the most important strands in the research into this fundamental principle of language assessment and the development of models of washback. But the part of his contribution that is more directly related to the general theme of this volume (i.e., the interfaces between language testing and language teaching) is the one in which he offers and agenda for test developers who wish to build washback into their programs. Basically, he recommends that greater attention be paid to test design features and to the outcomes of learning and that other aspects such as learner motivation and cultural background be also explored, given that all of them can contribute to foster certain reactions to tests (read 'positive reactions') and prevent other reactions (read, 'negative reactions'). He concludes that washback should not be regarded so much as the *influence* of tests on teaching and learning but as the *interaction* between tests, teaching, and learning, and that research into washback should not only focus on teacher perceptions and practices but also on the roles of other participants such as course leaders, policy makers, textbook writers and, perhaps the most significant of all, learners.

The paper by **Mª Luisa Roca-Varela and Ignacio M. Palacios** examines some of the most long established and widely recognised tests and their main features with particular attention to their oral modules. More specifically, the authors focus on the weighting of the oral part, the evaluation criteria of the oral skills and their relationship with the guidelines provided by *The Common European Framework of Reference* (CEFR), the selection of tasks used for the assessment and finally the scoring procedure employed for the oral part in these tests. The results highlight the relevance of the CERF and its impact on oral tests, which

emphasizes today's increased effort to align language test with more communicative approaches to language teaching. Although the tests are believed to be valid instruments for the assessment of oral performance, attention is drawn to some serious limitations of these tests which may end up having a negative washback effect on preparatory courses for these exams. The authors conclude by highlighting the need to perform a continuous evaluation of such tests with particular attention to their spoken oral modules to find out whether they meet the requirements of a good language test.

Jesús García Laborda and Elena Martín-Monje take up the issue of washback previously explored by Green (this volume) and analyse the consequences of test use by examining the current English section of the University Entrance Examination (PAU), a high-stakes public tests, and its foreseeably substitution for the High School Leaving Diploma (HSLD). The authors justify the introduction of the HSLD in terms of a re-design of the tasks emphasising the qualities of usefulness proposed by Bachman and Palmer (1996), and taking into account the distinctive features of language tests (i.e. construct validity, reliability, impact and practicality, VRIP) in order to improve test validity and reliability and achieve beneficial washback. The authors' interest is mainly motivated by two main reasons: the poor results of Spanish students on national and international surveys (i.e. the European Survey on Language Competence (2012) Eurostat press Release (2013)), and the psychometrics of the current test. The researchers also highlight the need to adapt the forthcoming HSLD to the recommendations of the *Common European Framework of Reference* (CEFR) integrating in this way language testing with more communicative teaching approaches. Some further suggestions are also provided for computer-based delivery through the use of cutting-edge technology such as smartphones or even iPads and tablet PCs.

Finally, **John Read and Janet Von Randow**'s contribution describes a project carried out at the University of Auckland, in New Zealand, which is a clear example of what is understood in this volume as interface between second language testing and second language pedagogy. While the norm in most universities is testing for entry purposes –i.e., prior to admission–, the University of Auckland has designed and is implementing since 2002 post-entry language assessment (PELA) whose main goal is to identify, from all the students that have already been admitted, those "who are likely to have significant academic language needs, in order to guide or direct them to appropriate learning resources on campus." These learning resources are gathered in the Diagnostic English Language Needs Assessment (DELNA), an academic language enrichment program available for all the students who need to it –those for whom English is an Additional Language, that is, non-native speakers, but also native speakers whose academic skills are deficient or below standard for successful performance in their degree subjects. The authors discuss the background that led to the development of PELA in their university and describe its current operation, with a focus on how the feedback from the academic staff involved and the students who have completed it has guided the changes that have taken place in the assessment instrument since its initial

implementation back in 2002, mainly in terms of the composition and the administration of the assessment as well as the delivery of effective English language enhancement.

To conclude, we would not like to end without expressing our appreciation and thanks to the General Editor of IJES, Aquilino Sánchez, and to the Editorial Assistant, Raquel Criado, for their help and assistance. We would also like to extend our gratitude to the colleagues who very kindly assisted us in the review process, and to the authors for their invaluable cooperation.

LOURDES CEREZO
MARIAN AMENGUAL

MURCIA, OCTOBER 2013

REFERENCES

Adair-Hauck, B., E. W.Glisan, K. Koda, E. Swender B. & P. Sandrock (2006). The Integrated Performance Assessment (IPA): Connecting Assessment to Instruction and Learning. *Foreign Language Annals*, *39,* 359-382.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129.

Amengual, M. (2009). Does the English test in the Spanish University Entrance Examination influence the teaching of English? *English Studies*, *90*(5), 582-598.

Amengual, M. (2010). Exploring the washback effects of a high-stakes English test on the teaching of English in Spanish upper secondary schools. *Revista Alicantina de Estudios Ingleses, 23,* 149-170.

Amengual, M. & Méndez García, C. (2012). Implementing the oral English task in the Spanish University Admission Examination: An international perspective of the language. *Revista de Educación, 357,* 105-127.

Bachman, L.F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L.F. & Palmer, A.S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests.* Oxford: Oxford University Press.

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing, 13*(3), 257-279.

Breen, M.P. & Candlin, C.N. (1980). Essentials of the communicative curriculum. *Applied Linguistics, 1*(2), 89-112.

Brumfit, C.J. & Johnson, K. (Eds.). (1979). *The Communicative Approach to Language Teaching.* Oxford: Oxford University Press.

Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp. 113-128). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Canale, M. & Swain, M. (1981).A theoretical framework for communicative competence. In A. S. Palmer, P.J.M. Groot & G.A. Trosper (Eds.), *The Construct Validation of Tests of Communicative Competence* (pp. 31-36). Washington, DC: TESOL.

Chapman, D.W. & Snyder, C. W. (2000). Can High stakes National Testing Improve Instruction: Re-examining Conventional Wisdom. *International Journal of Educational Development, 20,* 457-474.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, *11*(1), 38-54.

Cheng, L. (2005). *Changing Language Teaching through Language Testing: A Washback Study. Studies in language testing, 21*. Cambridge: Cambridge University Press.

Cheng, L. & Curtis, A. (2004). Washback or backwash: a review of the impact of testing on teaching and learning. In L. Cheng & Y. Watanabe (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp. 3-178). Mahwah, NJ: Lawrence Erlbaum Associates.

Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics, 20,* 147-161.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge: Cambridge University Press.

Davies, A, (1991). Language Testing in the 1990s. In Ch. Alderson & B. North (Eds.), *Language Testing in the 1990s: The Communicative Legacy* (pp. 136-141). London: Macmillan Publishers limited.

Douglas, D. (1995). Developments in language testing. *Annual Review of Applied Linguistics, 15*, 167-187.

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. (2000). The 'communicative' legacy in language testing. *System, 28,* 483-497.

García Laborda, J. (2010). ¿Necesitan las universidades españolas una prueba de acceso informatizada? El caso de la definición del constructo y la previsión del efecto en la enseñanza de idiomas para extranjeros. *Revista de Orientación y Psicopedagogía, 21*(1), 71-80.

García Laborda, J. & Martín-Monje, L. (this volume). Item and test construct definition for the new Spanish High School Leaving Diploma: A proposal. *International Journal of English Studies*, *13*(2),

Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment.* London: The Falmer Press.

Green, A. (2006a). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing, 11*(2), 113-134.

Green, A. (2006b). Watching for washback: observing the influence of the IELTS academic writing test in the classroom. *Language Assessment Quarterly, 3*(4), 333-367.

Green, A. (2007). Washback to learning outcomes: a comparative study of IELTS Preparation and University pre-sessional language courses. *Assessment in Education, 14*(1), 75-97.

Green, A. (this volume). Washback in language assessment. *International Journal of English Studies*, *13*(2),

Hancock, Ch. (2006). Language teaching and language testing; a way and ways to make appropriate connections. *Acción Pedagógica, 15,* 6-13.

Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Johnson, K. (1982). Communicative syllabus design and methodology. Oxford: Oxford University Press.

Johnson, K. and Morrow, K. (1981). *Communication in the Classroom. Applications and Methods for a Communicative Approach*. Harlow: Longman.

Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement: Issues and Practice*, *16*(2), 14-16.

Littlewood, J. (1981). *Communicative Language Teaching*. New York, NY: Cambridge University Press.

Luxia, Q. (2007). Is Testing an Efficient Agent for Pedagogical Change? Examining the Intended Washback of the Writing Task in a High-stakes English Test in China. *Assessment in Education*, *14*(1), 51-74.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.

Morrow, K. (1979). Communicative language testing: revolution or evolution. In C.K. Brumfit, K. Johnson (Eds.), *The Communicative Approach to Language Teaching* (pp. 143-159). Oxford: Oxford University Press.

Rea-Dickins, P. (2007).Editorial. Washback from language tests on teaching, learning and policy: evidence from diverse settings. *Assessment in Education*, *14*(1), 1-7.

Richards, J.C. (2006). *Communicative Language Teaching Today*. New York, NY: Cambridge University Press.

Savignon, S. (1983). *Communicative Competence: Theory and Classroom Practice*. Reading, MA: Addison-Wesley.

Savignon, S. (1991). Communicative language teaching. State of the art. *TESOL Quarterly, 25*(2), 261-277.

Savignon, S. (Ed) (2002). *Interpreting Communicative Language Teaching. Contexts and Concerns in Teacher Education*. New Haven, CT: University of Yale Press.

Shih, C.-M. (2007). A new washback model of students' learning. *Canadian Modern Language Review, 64*(1), 135-162.

Shohamy, E (1992). Beyond proficiency testing: a diagnostic feedback testing model for assessing foreign language learning, *The Modern Language Journal*, *76*(4), 513-521.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing, 13(3)*, 298-317.

Shohamy, E. (1997). Testing methods, testing consequences; are they ethical? Are they fair? *Language Testing*, *14*(3), 340-349.

Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, *24*(4), 331-345.

Shohamy, E. (2001). *The Power of Tests: A Critical Perspective on the Uses of Language Tests.* Harlow: Longman.

Spolsky, B. (1995). *Measured Words*. Oxford: Oxford University Press.

Spolsky, B. (1996). The examination-classroom backwash cycle: some historical cases. In D. Nunan, R. Berry & V. Berry (Eds.), *Bringing about Change in Language Education* (pp. 55-66), Hong-Kong: The university of Hong-Kong, Department of Curriculum Studies.

Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research, 9,* 5-29.

Standsfield, Ch. W. (2008). Where we have been and where we should go. *Language Testing*, *25*(8), 311-326.

Stobart, G. (2003). Editorial. The impact of assessment: intended and unintended consequences. *Assessment in Education*, *10*(2), 139-140.

Taylor, L. (2005). Washback and impact, *ELT Journal*, *59*(2), 54-155.

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

Wall, D. (1996). Introducing new tests into traditional systems:insights from general education and from innovation theory. *Language Testing*, *13*(3), 334-354.

Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Language Testing and Assessment* (pp. 291-302). Amsterdam: Kluver Academic Publishers.

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System, 28*, 499-509.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching. Studies in Language Testing 22.* Cambridge: Cambridge University Press.

Wall, D. (2011). *The impact of changes in the TOEFL on teaching in a sample of countries in Europe*: phase 3, the role of the coursebook, phase 4, describing change. TOEFL iBT Reports series. Princeton, NJ: Educational Testing Service.

Wall, D. (2012). Washback. In G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing* (pp. 79-92). New York, NY: Routledge.

Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp. 129-146). Mahwah, NJ: Lawrence Erlbaum Associates.

Weir, C. (1990). *Communicative Language Testing*. Hemel Hempstead: Prentice Hall.

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

Widdowson, H. (1978). *Teaching Language as Communication.* Oxford: Oxford University Press.

Wiggins, G.P. (1988). *Educative Assessment*. San Francisco, California: Jossey-Bass Publishers.