



second language development. Studies carried out in the informal environment have shown that those who report more free reading achieve higher levels of competence in second languages. In the present study, university-level English as a Foreign Language (EFL) students in Hong Kong who participated in a popular literature class that emphasised reading for content and enjoyment, including some self-selected reading, made superior gains on measures of vocabulary and reading rate, when compared to students enrolled in a traditional academic skills class. Eighty-eight per cent of the literature students felt that what they learned from the course would help them in other university courses, but only 12% of the traditional academic skills students had this opinion about their class. These results are seen as consistent with previous studies showing that meaningful reading is an important source of literacy competence.

## Language testing

**01-142 Kenyon, Dorry M.** (Center for Applied Linguistics, Washington, USA; *Email*: dorry@cal.org) and **Tschirner, Erwin** (Universität Leipzig, Germany; *Email*: erwin.tschirner@t-online.de). The rating of direct and semi-direct oral proficiency interviews: comparing performance at lower proficiency levels. *The Modern Language Journal* (Malden, MA, USA), **84**, 1 (2000), 85–101.

As states and universities institute oral proficiency requirements with vast numbers of students to be tested, there is a need to investigate effective alternatives to the ACTFL (American Council on the Teaching of Foreign Languages) Oral Proficiency Interview (OPI) that allow group testing. This article reports on a study comparing student performances and test reliabilities for the German Speaking Test (GST) developed by the Center for Applied Linguistics, a semi-direct tape-mediated oral proficiency test, and the ACTFL OPI. Both the GST and a German OPI were administered as final oral exams to a randomly selected group of 20 students (out of a total of 59) enrolled in a fourth-semester German course at a large Midwestern university. The OPI levels of the students tested ranged from Novice High ( $n = 5$ ) and Intermediate Low ( $n = 9$ ) to Intermediate Mid ( $n = 6$ ). At these three levels, final ratings on the GST and the OPI agreed with each other perfectly in 90% of the cases. There were only two one-step disagreements, both involving students who were rated Novice High on the ACTFL OPI, but who received other ratings on the GST. Although the results indicated a high score equivalency between ACTFL proficiency ratings obtained on both tests, this study is seen as underscoring the pressing need for double ratings and arbitration procedures in high stakes testing situations.

**01-143 Krajnović, Marta Medved** (Filozofski fakultet, Zagreb U., Croatia). Učeničke strategije i testovi stranih jezika. [Learner strategies and foreign language tests.] *Strani Jezici* (Zagreb, Croatia), **28**, 3/4 (1999), 149–55.

This article is an attempt to connect two important aspects of foreign language learning and teaching: learner strategies and testing. It discusses the importance of test-taking strategies research in validating language tests and gaining a deeper insight into learners' knowledge. The author describes the most widely used research methodology – test-takers' verbal or written reports, and recent research in test-taking strategies used for different testing techniques (multiple-choice, cloze-tests, c-tests, oral interviews). The possible pedagogical implications of the research results are also pointed out.

**01-144 Krause, Wolf-Dieter and Sändig, Uta** (Universität Potsdam, Germany). Zur Effektivierung von Leistungskontrollen und Tests für die Überprüfung der Kommunikationskompetenz im Fremdsprachenunterricht (unter Berücksichtigung von ersten Ergebnissen des europäischen Kooperationsprojektes ECCELLENTT). [Increasing the effectiveness of performance controls and tests for checking communicative competence in foreign language teaching (taking into account the preliminary results of the European co-operative project ECCELLENTT).] *Fremdsprachen und Hochschule* (Bochum, Germany), **57** (1999), 88–116.

When the setting and development of tests is not merely concerned with linguistic knowledge, but is to be evaluated on the basis of language-communicative ability, the task becomes a complex one. This article deals with the aims of such tests and how they should be created accordingly, focusing in particular on task type and the corresponding test process in evaluating receptive and productive language activities in an FL context (here, German as a foreign language). Such considerations are set against the background of the preliminary results of the European co-operative project ECCELLENTT (Evaluation of Communicative Competence in European Language Learning Encompassing New Testing Technologies) which sets out seven criteria for the setting and evaluation of tests: objectivity, authenticity, reliability, validity, practicability, impact and involvement/interactiveness. Task types and corresponding tests are discussed in the light of each of these criteria. Two sample tests are included in the appendix.

**01-145 Laurier, Michel** (U. of Montreal, Canada; *Email*: laurierm@SCEDU.Umontreal.ca). Can computerised testing be authentic? *ReCALL* (Cambridge, UK), **12**, 1 (2000), 93–104.

The concept of authenticity first appeared with the development of the communicative approach. More

recently, in the field of educational measurement, authentic assessment methods have been proposed. Although adaptive testing seems to be the most important application of computers in language assessment, these tests are usually not authentic. Since many real world tasks are accomplished with computers, it is suggested here that these may be used for authentic direct testing. Computers may also be used in semi-direct testing as a way to enhance the context. Finally, in authentic assessment, computers may be used as a tool to process the data when the learners use them to organise their portfolio. It is concluded that, in using the computer, test developers can also create better authentic tests.

**01-146 Laviosa, Sara** (U. of Birmingham, UK). Didattizzare la traduzione per acculturare e comunicare. [Using translation as a teaching tool for education and communication.] *Rassegna Italiana di Linguistica Applicata* (Rome, Italy), **32**, 1 (2000), 117–36.

This article looks at the role of translation in Italian as a Foreign Language (IFL) exams currently administered in American and British universities. The data are based on questionnaire feedback from 38 universities, and indicate a preference for translation tasks at elementary level in the US, as compared to non-elementary use in the UK, with only a minority of respondents (under 20%) avoiding translation altogether. Contrastive awareness of lexical and grammatical features of the target language is therefore an essential component of IFL assessment in both countries. But effective translation also requires bi-cultural competence, which involves functional rather than formal equivalence between texts: for this reason, the author advocates the use of authentic parallel texts in English/Italian and views translation as a learning resource which combines linguistic and sociocultural skills, as conceptualised in Balboni's pyramid of communicative competence.

**01-147 Lutje Spelberg, Henk C.** (U. of Groningen, The Netherlands; *Email*: h.c.lutje.spelberg@ppsw.rug.nl), **de Boer, Paulien and van den Bos, Kees P.** Item type comparisons of language comprehension tests. *Language Testing* (London, UK), **17**, 3 (2000), 311–22.

Several test procedures are available to measure language comprehension. In this article two tests with different item types are compared. The tests are the Dutch Reynell test – in which the subject has to carry out verbal instructions, and the BELL test – in which the subject has to select which of four pictures best represents a given statement. The researchers studied the contribution of the type of item, and the linguistic features of the item to the prediction of item difficulty. Both tests were administered to 64 Dutch kindergarten children with an average age of 70.3 months. The simple correlation between type of item and item difficulty is relatively high (.40). However, regression analyses

indicate that type of item does not contribute significantly to the prediction of item difficulty. The linguistic features of the item, in contrast, do. The two tests cover successive age ranges. The results are taken as indicating that the BELL test can be used as a follow up test of the Reynell test for language comprehension.

**01-148 Oller, Jr., John W.** (U. of Louisiana at Lafayette, USA; *Email*: joller@louisiana.edu), **Kim, Kunok and Choe, Yongjae.** Testing verbal (language) and nonverbal abilities in language minorities: a socio-educational problem in historical perspective. *Language Testing* (London, UK), **17**, 3 (2000), 341–60.

This paper begins from the premise that there is a widespread socio-educational problem with language testing at its heart: speakers of minority languages are over-represented in classes for the learning disabled, disordered and educable mentally retarded, and under-represented in classes for the gifted. It is claimed that this imbalance is owed to mental measurement practices which involve language tests both directly and indirectly. The source of the problem is a general failure to appreciate the role of language proficiencies in psychological and educational testing. Also the relation between acquired (socially dependent) language proficiencies and so-called non-verbal abilities may be closer than commonly supposed. Among the questions addressed in the paper are the following: (a) to what extent it is possible to measure non-verbal abilities without invoking acquired language/dialect proficiencies; (b) whether it is possible to get across to test takers the instructions to non-verbal tasks without recourse to one or more particular languages or dialects; and (c) whether it is possible to make linguistically and culturally unbiased judgements about intellectual abilities (including abnormalities, disorders and giftedness) on the basis of 'non-verbal' tasks.

**01-149 Salaberry, Rafael** (Rice U., Texas, USA; *Email*: salaberry@rice.edu). Revising the revised format of the ACTFL Oral Proficiency Interview. *Language Testing* (London, UK), **17**, 3 (2000), 289–310.

Since the early 1980s proponents of proficiency examinations such as ACTFL (American Council on the Teaching of Foreign Languages) have been criticised for the low validity and reliability of tests such as the OPI (Oral Proficiency Interview). Despite these strong concerns, the most recent (1999) edition of the ACTFL Tester Training Manual does not reveal substantial changes from the previous manual published in 1986. While a complete elimination of proficiency tests such as the ACTFL-OPI may be neither feasible nor necessary, the present author considers that some practical changes may be appropriate. This article describes minor structural changes of the ACTFL-OPI framework which, it is suggested, while not eliminating the practical benefits of a proficiency test, would address



substantial concerns related to the validity and reliability of the instrument.

**01-150 Takala, Sauli and Kaftandjieva, Felianka** (U. of Jyväskylä, Finland; *Email*: sjtakala@cc.jyu.fi). Test fairness: a DIF analysis of an L2 vocabulary test. *Language Testing* (London, UK), **17**, 3 (2000), 323–40.

The purpose of the study reported here is to analyse gender-uniform differential item functioning (DIF) in a second language (L2) vocabulary test with the tools of item response theory (the separate calibration *t*-method), and to study potential gender impact on the test performance measured by different item composites. The English vocabulary test used was part of the Finnish Foreign Language Certificate Examination test battery, Intermediate Level, and was taken by 475 adults (182 males and 293 females). The results show that, despite the fact that there are test items with indications of DIF in favour of either females or males, the test as a whole is not gender-biased. In spite of this, it was demonstrated that some item composites are gender-biased. In view of item bank building and use, it means that some of the tests constructed on the basis of an item bank might be biased if the item bank contains items with indication for DIF. Although the results of some empirical research suggest that the requirements for items with DIF to be excluded from the final test version may on the whole be too restrictive, this study demonstrated that the traditional advice of excluding biased items gains new significance in the light of item bank building and use, since doing so will prevent possible biased item composites.

**01-151 Turner, Carolyn E.** (McGill U., Canada). Listening to the voices of rating scale developers: identifying salient features for second language performance assessment. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes* (Toronto, Ont.), **56**, 4 (2000), 555–84.

This article describes the process and discourse stances of a team of teachers involved in deriving a rating scale for writing ability. The research was carried out within a Ministry of Education of Quebec project whose objective was to develop empirically based rating scales for secondary-level English as a Second Language provincial exams. The study focused on instances during the process where actions of the participants and/or their use of the data sample (i.e., student writing samples) could be shown to influence the criteria for the rating scale and in turn the final ratings (i.e., areas where there was potential for variation within the two test method characteristics of scale development team and sample used). Through a qualitative analysis, it expands on earlier research (Turner & Upshur, 1999) which reports on the quantitative results of method characteristics in such empirically derived scales. This study provides a description of the nature of these test method characteristics.

## Teacher education

**01-152 Burgess, John** (U. of Manchester, UK; *Email*: john.burgess@man.ac.uk) and **Spencer, Sheila**. Phonology and pronunciation in integrated language teaching and teacher education. *System* (Oxford, UK), **28**, 2 (2000), 191–215.

This paper addresses the relationship between two fields: (1) teaching and learning pronunciation in a second or foreign language; and (2) the study of pronunciation-teaching and of phonology in the training and education of language teachers. It reports research conducted to inform the planning of the phonology component of an initial teacher-training course in TESOL (Teaching English to Speakers of Other Languages). The research involved two strands: a questionnaire, and an analysis of some current English as a Foreign Language textbooks, and was intended to provide a picture of current teacher attitudes towards, and pedagogic practice in, the teaching of pronunciation. The paper argues for a strongly integrated approach to the relationship between the two fields addressed, but for different priorities in those fields.

**01-153 Wildner, Siegrun** (U. of Northern Iowa, USA; *Email*: wildner@uni.edu). Technology integration into preservice foreign language teacher education programs. *CALICO Journal* (San Marcos, TX, USA), **17**, 2 (2000), 223–50.

Using national technology goals and standards, and institutional factors as a framework, this article describes the planning process and the initial phases of the implementation of a model of technology integration into the foreign language preservice teacher education program for K–12 teachers. It also provides suggestions for instructional goals and subject-related content areas for a curricular component on technology for foreign language teachers.

## Bilingual education/ bilingualism

**01-154 Dijkstra, Ton, de Bruijn, Ellen, Schriefers, Herbert and Ten Brinke, Sjoerd** (U. of Nijmegen, The Netherlands; *Email*: dijkstra@nici.kun.nl). More on interlingual homograph recognition: language intermixing versus explicitness of instruction. *Bilingualism: Language and Cognition* (Cambridge, UK), **3**, 1 (2000), 69–78.

This paper reports a study which contrasted the effect of instruction-induced expectancies and language