



they typically concentrate on the decoding of words in an attempt to achieve total comprehension. The paper offers a description of the proficient reading process and the typical L2 learner reading process, and suggests a reading programme which aims to help L2 learners to become successful readers by encouraging them to follow from the beginning a reading process as close as possible to the one they use in their L1 as a result of establishing experiential reading as their L2 norm and sufficient understanding as their reading goal.

**01-615 Waggot, Valerie** (U. of Edinburgh, Scotland, UK; *Email: valerie@srv0.apl.ed.ac.uk*). Features, cobwebs or clines: towards a possible model of lexical retrieval in bilingual readers. *Edinburgh Working Papers in Applied Linguistics* (Edinburgh, Scotland), **10** (2000), 72–79.

This article reports an investigation of how 120 teenage bilingual Chinese/English readers comprehended a verb and a noun in a Chinese text. The data are parallel translations made by examination candidates, all in the British educational system. A brief justification of the use of the data is given. This is followed by an overview of approaches to the theory of meaning and the mental lexicon. The data consist of a range of interpretations of *moni* ('to imitate') and *zawen* ('literary essay'). Evidence from the data and support from the literature lead to the conclusion that the comprehension of the meaning of lexical items is influenced to a great extent not only by recovery of meaning from the text but also by personal experience. Able readers are more likely to provide a 'majority' or near dictionary equivalent, which may be regarded as central on a continuum, while less able readers veer towards either end of the continuum.

**01-616 Walz, Joel** (U. of Georgia, USA). Reading hypertext: lower-level processes. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes* (Toronto, Ont.), **57**, 3 (2001), 475–94.

The World Wide Web presents an amazing array of authentic documents in a form called 'hypertext' that language learners can use not only to find out about target language cultures but also to develop reading skills. However, as authentic documents, they are more difficult to read than texts traditionally associated with language instruction. This article analyses such texts in light of reading research and proposes solutions to two potential areas of difficulty. The first is the need to read documents from a computer screen. Since successful reading involves familiarity and habits, activities that add to students' reading experiences are described. The second relates to the varieties of the target language found on Web sites, which teachers can approach through lower-level reading processes including learning basic vocabulary related to the Web, searching native language versions of bilingual sites, using a translation site, pronouncing words aloud, and guessing contextually. These techniques can help students over-

come obstacles to reading comprehension. The content of Web sites will be the subject of a future article on higher-level processes.

## Language testing

**01-617 Berry, Vivien and Lewkowicz, Jo** (The U. of Hong Kong; *Emails: vberry@hkucc.hku.hk, jolewkow@hkucc.hku.hk*). Exit-tests: Is there an alternative? *Hong Kong Journal of Applied Linguistics* (Hong Kong), **5**, 1 (2000), 19–49.

This paper examines the major issues relating to the introduction of an assessment of English language proficiency for students prior to graduation from tertiary education. It looks at the on-going debate relating to the introduction of an exit language test and considers possible alternatives to formal standardised tests for reporting on language proficiency. It then describes a small-scale pilot study, carried out at the University of Hong Kong, which set out to discover students' views on the introduction of such an exit mechanism. The paper goes on to caution against introducing any mechanism of assessment without full consultation with the primary stakeholders of the test, that is, the students and employers who are eager to see language standards of graduating students improved. It concludes by suggesting how a valid and reliable reporting mechanism can be developed for students in Hong Kong, drawing on the current work being done on the assessment of Core Competencies and on the European Portfolio Project.

**01-618 Kenyon, Dorry M. and Malabonga, Valerie** (Center for Applied Linguistics, Washington, DC, USA; *Emails: dorry@cal.org; valerie@cal.org*). Comparing examinee attitudes toward computer-assisted and other oral proficiency assessments. *Language Learning and Technology* (<http://llt.msu.edu/>), **5**, 2 (2001), 60–83.

This article reports the results of a study of examinee attitudinal reactions to taking different formats of oral proficiency assessments across three languages: Spanish, Arabic, and Chinese. All 55 students in the study were administered both the tape-mediated Simulated Oral Proficiency Interview (SOPI) and a new Computerized Oral Proficiency Instrument (COPI). In addition, the 24 students participating in the Spanish study were administered the face-to-face American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview (OPI). Participants were graduate and undergraduate students taking language courses at their universities. The order of test administration was based on self- and teacher-assessed proficiency levels and experience in learning the target language. After each test, the learners completed a Likert questionnaire on six aspects of their attitudes towards and perceptions of that test. Finally, they were requested to compare the two technology-mediated tests directly on those six

aspects. This article presents the examinees' responses on those questionnaires, focusing primarily on differences between the two technology-mediated tests. It was found that the adaptive nature of the COPI allowed the difficulty level of the assessment task to be matched more appropriately to the proficiency level of the examinee. This led examinees, particularly those at the lower proficiency levels, to feel the COPI was less difficult than the SOPI. In most other aspects, the two tests were rated similarly. For the Spanish students, the two technology-mediated tests were rated similarly to the OPI, though the OPI appeared to them to be a better measure of real-life speaking skills.

**01-619 Kim, Mikyung** (U. of California, Los Angeles, USA; *Email*: kimmi@ucla.edu). Detecting DIF across the different language groups in a speaking test. *Language Testing* (London, UK), **18**, 1 (2001), 89–114.

The investigation of differential item functioning (DIF) is crucial in language proficiency tests in which test-takers with diverse backgrounds are involved, because DIF items pose a considerable threat to the validity of tests. To date, DIF analysis in language testing has been conducted mainly for multiple-choice items. However, examining DIF with polytomous response items such as in writing and speaking tests should also be taken into consideration for validating tests. This study aims to investigate DIF across two different broad language groupings, Asian and European, in a speaking test in which the test-takers' responses are rated polytomously. Data in this study were collected from 1038 nonnative speakers of English from France, Hong Kong, Japan, Spain, Switzerland and Thailand who took the SPEAK test in 1988 (cf. Educational Testing Service, 1985). The methods used for DIF analysis were the likelihood ratio test and the logistic regression procedure. The primary scoring categories of interest in this study were 'grammar', 'pronunciation' and 'fluency'. The results showed that 'grammar' and 'pronunciation' functioned differentially across the two groups. A content analysis of the DIF items suggested that the types and the numbers of scoring scales might influence the test validity. The study provides methodological information on differences between two approaches to DIF analysis and suggestions for future research.

**01-620 Li, Xiajou and Wang, Yi** (Guangzhou Foreign Languages U.; *Email*: lixiajou@gdufs.edu.cn). Testing oral English on a mass scale: Is it feasible? – The oral component of the MET in China. *Hong Kong Journal of Applied Linguistics* (Hong Kong), **5**, 1 (2000), 160–86.

The Matriculation English Test (MET) is the English test required nationwide for enrolment at universities in China. This paper reports on the development of the MET oral subtest (METOS). The work on the METOS was subject to many constraints, including the huge size of the candidature and the rigid limitations on human

and time resources, which are particularly acute in China. Having undergone repeated piloting and moderation, the METOS format has attained validity on the strength of its being a message-based test of the interactive and contextualised use of spoken English. It has developed a number of unique features, such as a balance between control and spontaneity in the output, a union of the analytical and the holistic approach in the rating, and a combination of the single-examiner method in the testing and the double-marking method in the scoring. It is argued that the METOS is not just a test that suits the conditions and meets the needs of China. In the field of language testing it can claim to be a successful innovation in mass-scale oral testing.

**01-621 Lumley, Tom and Stoneman, Bernadette** (Hong Kong Poly. U.; *Emails*: egluml@polyu.edu.hk, ecstonem@polyu.edu.hk). Conflicting perspectives on the role of test preparation in relation to learning? *Hong Kong Journal of Applied Linguistics* (Hong Kong), **5**, 1 (2000), 50–80.

This paper reports on the development of a Learning Package for a newly introduced test in Hong Kong, the Graduating Students' Language Proficiency Assessment (GSLPA)-English. This Learning Package, which was designed for self-access use or for classroom teaching, aims to go beyond merely providing practice materials in the form of samples of test tasks. Instead it is intended to encourage students to take some level of responsibility for their learning, and to view the test not merely as a barrier to be passed but as an opportunity to improve their proficiency in English. This paper examines reactions of students and teachers to the Learning Package during the piloting of the materials. Preliminary feedback suggests that, while teachers are both supportive of the package and consider that the materials developed have excellent potential for use as teaching materials, students have a more complex set of reactions. These seem to be related partly to their level of English proficiency (cf. Lewkowicz, 2000), but have clear messages about their expectations of test preparation materials, as well as their attitudes to learning.

**01-622 Lynch, Tony** (U. of Edinburgh, Scotland, UK; *Email*: A.J.Lynch@ed.ac.uk). An evaluation of the revised Test of English at Matriculation at the University of Edinburgh. *Edinburgh Working Papers in Applied Linguistics* (Edinburgh, Scotland), **10** (2000), 61–71.

This paper reports a second-cycle validation study of the Test of English at Matriculation (TEAM), following revisions made to the test in 1993. Candidates' scores from four academic sessions (1993–1997) were used to assess the relationship between performance on TEAM at the beginning of their degree course and their eventual academic outcome. Results suggest that (1) the predictive capacity of the revised test is similar to that of other tests of English for Academic Purposes and (2)



performance in the Listening section remains the strongest single linguistic predictor of candidates' success in taught Master's degrees. The overall correlation is slightly lower than for the first version of TEAM: possible reasons are discussed in the paper.

**01-623 Marius, Tamas and Perrin, Geoffrey**

(Bundesprachenamt, Lehrerschulung, Hürth, Germany). Examinations and student priorities. *Fremdsprachen und Hochschule* (Bochum, Germany), **60** (2000), 32–41.

The last decade has witnessed a flurry of research interest in the topic of washback. By far the greater part of this, however, has focused on the influence of examinations on teacher practice. The present study seeks to redress to some extent the relative neglect of the learner by looking at the way in which examination requirements impinge on student attitudes and expectations during a course of language instruction. The data are drawn from student interviews, student diaries, and day-to-day observation (one of the authors was also a teacher in the class chosen for the case study at the German Government Language Centre). The investigation leads to two suggestions: firstly, that exams may be a less pervasive presence than is sometimes assumed; and, secondly, that at least some of the difficult issues of research ethics relating to investigations in examination classes can be resolved if certain conditions are met.

**01-624 McKay, Penny and Ferguson, Rebecca**

(Queensland U. of Technology, Brisbane, Australia; Email: pa.mckay@qut.edu.au). English language standards for schools in Australia and China: finding the 'third place'. *Hong Kong Journal of Applied Linguistics* (Hong Kong), **5**, 1 (2000), 108–27.

The present authors explore the possibilities for the application of principles which informed the construction of a set of ESL (English as a Second Language) standards for schools in Australia to the development of English Language standards for an EFL (English as a Foreign Language) context. The specific EFL context under consideration is mainland China. Drawing on Kramsch's (1993) notion of the 'third place', and recognising the presence of cultural values and perspectives in all educational materials, the authors offer principles gained from experience in the construction of the NLLIA (National Languages and Literacy Institute in Australia) *ESL Bandscales*, for consideration in the development of EFL standards for the Chinese context. They recognise the need for professionals from mainland China to take up the principles, rather than the Australian materials, to develop EFL standards for the Chinese educational and cultural context.

**01-625 Oller, Jr., John W.** (U. of Louisiana at Lafayette, USA; Email: joller@louisiana.edu), **Kim, Kunok, Choe, Yongjae and Jarvis, Lorna**

**Hernandez.** Testing relations between language (verbal) and nonverbal abilities in children and adults

acquiring a nonprimary language. *Language Testing* (London, UK), **18**, 1 (2001), 33–54.

General sign theory per Oller *et al.* (2000) predicts that, to the extent that valid measurements are possible, nonverbal abilities should correlate positively with primary language abilities (Hypothesis 1). Further, nonverbal abilities of persons in the early stages of acquiring a nonprimary language should correlate significantly more positively with proficiencies in their primary language than in their nonprimary language (Hypothesis 2a); but as persons approach parity between their primary and any nonprimary language, correlations between nonverbal scores and proficiencies in the two languages should both be positive and not significantly different (Hypothesis 2b). The Cattell-Horn theory predicts that nonverbal abilities should correlate equally with primary and nonprimary skills throughout the course of development. Gardner's theory predicts nonsignificant correlations. This article reports the examination of hypotheses 1, 2a and 2b in within-subjects, repeated measures designs. Study 1 examines 50 children acquiring English as a Foreign Language (EFL) in a bilingual school in Mexico, and Studies 2 and 3 examine 165 and 203 Korean adults learning EFL. Hypotheses 1, 2a and 2b are confirmed ( $p < .01$ ). The Gardner and Cattell-Horn predictions are disconfirmed.

**01-626 Pelly, Christine Parimala and Allison, Desmond**

(National U. of Singapore; Emails: pelly@pacific.net.sg, elldma@nus.edu.sg). Investigating the views of teachers on assessment of English language learning in the Singapore education system. *Hong Kong Journal of Applied Linguistics* (Hong Kong), **5**, 1 (2000), 81–106.

This study sets out to explore and characterise an important insider perspective, that of primary school teachers, on assessment in English language in Singapore and the impact of assessment on their experiences of teaching. A review of related literature focuses on the importance of research into teachers' views on curriculum and assessment issues, in addition to – and in the light of – their own classroom experiences. The present exploratory study elicits teachers' views on assessment in the form of questionnaire responses ( $n = 58$ ) and semi-structured in-depth interviews with four participants. Findings indicate, at least typically for this sample, that teachers are strongly aware of the prevailing examination culture; that they are not against formal tests in principle, but see the need to supplement them with other kinds of assessment; and that they are markedly divided and uncertain in their views of the efficacy of current tests. Besides pointing to the need for further research, the discussion brings out implications for teacher education, the role of assessment in the school and the wider culture, and participant involvement in initiatives for educational change.

**01-627 Read, John** (Victoria U. of Wellington; Email: john.read@vuw.ac.nz) **and Chapelle,**

**Carol A.** A framework for second language vocabulary assessment. *Language Testing* (London, UK), **18**, 1 (2001), 1–32.

Vocabulary tests are used for a wide range of instructional and research purposes, but a comprehensive basis for evaluating the current instruments or developing new lexical measures for the future is needed. This article presents a framework that takes as its starting point an analysis of test purpose and then shows how purpose can be systematically related to test design. The link between the two is based on three considerations which derive from Messick's (1989) validation theory: construct definition, performance summary and reporting, and test presentation. The components of the framework are illustrated throughout by reference to eight well-known vocabulary measures; for each one there is a description of its design and an analysis of its purpose. It is argued that the way forward for vocabulary assessment is to take account of test purposes in the design and validation of tests, as well as considering an interactionist approach to construct definition. This means that a vocabulary test should require learners to perform tasks under contextual constraints that are relevant to the inferences to be made about their lexical ability.

**01-628 Roever, Carsten** (U. of Hawai'i at Manoa; Email: roever@hawaii.edu). Web-based language testing. *Learning and Technology* (<http://lt.msu.edu/>), **5**, 2 (2001), 84–94.

This article describes what a Web-based language test (WBT) is, how WBTs differ from traditional computer-based tests, and what uses WBTs have in language testing. After a brief review of computer-based testing, WBTs are defined and categorised as low-tech or high tech. Since low-tech tests are the more feasible, they constitute the focus of the paper. Next, item types for low-tech WBTs are described, and validation concerns that are specific to WBTs are discussed. After a brief overview of the marriage of computer-adaptive and Web-based tests, the general advantages as well as design and implementation issues of WBTs are considered before examining the role that testing consequences play in deciding whether a WBT is an appropriate assessment instrument. It is argued that WBTs are most appropriate in low-stakes testing situations; but with proper supervision, they can also be used in medium-stakes situations, although they are not generally recommended for high-stakes situations. Some possible areas for future research are suggested.

**01-629 Sawaki, Yasuyo** (U. of California, LA, USA; Email: ysawaki@ucla.edu). Comparability of conventional and computerised tests of reading in a second language. *Language Learning and Technology* (<http://lt.msu.edu/>), **5**, 2 (2001), 38–59.

Computerisation of second language (L2) reading tests has been of interest among language assessment researchers for the past 15 years, but few empirical studies have evaluated the equivalence of the construct

being measured in computerised and conventional L2 reading tests and the generalisability of computerised reading test results to other reading conditions. In order to address various issues surrounding the effect of mode of presentation on L2 reading test performance, the present study reviews the literature in cognitive ability testing in educational and psychological measurement and the non-assessment literature in ergonomics, education, psychology, and L1 reading research. Generalisation of the findings to computerised L2 assessment was found to be difficult: the nature of the abilities measured in the assessment literature does not necessarily involve language data; mode of presentation studies in the non-assessment literature involving L2 readers are scarce; and there are limitations in the research methodologies used. However, the literature raises important issues to be considered in future studies of mode of presentation in language assessment.

**01-630 Schmitt, Norbert** (U. of Nottingham, UK; Email: norbert.schmitt@nottingham.ac.uk), **Schmitt, Diane and Clapham, Caroline**. Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* (London, UK), **18**, 1 (2001), 55–88.

The Vocabulary Levels Test, originally designed by Paul Nation, has been widely used in language assessment and vocabulary research. Despite this widespread use, however, the test has never been properly validated. This article reports on a study which uses a range of analysis techniques to present validity evidence, and to explore the equivalence of two revised and expanded versions of the Vocabulary Levels Test.

**01-631 Shohamy, Elana** (Tel Aviv U., Israel; Email: elana@post.tau.ac.il). Using language tests for upgrading knowledge: The phenomenon, source and consequences. *Hong Kong Journal of Applied Linguistics* (Hong Kong), **5**, 1 (2000), 1–18.

This paper is divided into two sections. The first part looks at the way in which many high stakes tests have been introduced for reasons of political expediency rather than educational benefit. It examines the consequences of using tests in this way and suggests that when tests are used as disciplinary tools their introduction leads to a narrowing of the curriculum rather than an improvement in learning. The second part of the paper examines the introduction in Hong Kong of the Language Proficiency Assessment for Teachers in the light of the earlier discussion and suggests that a number of questions need to be addressed if the assessment is not to have the reverse effect to that intended by the Education Department.

## Teacher education

**01-632 Chamberlin, Carla R.** (Penn State Abington Coll., Pennsylvania, USA). TESL degree