

English-speaking students in Computer Science developed into a student newspaper published on the World Wide Web. The students were involved at every stage in the analysis of their writing and their presentation. The results of an analysis of their work at surface level (grammar, vocabulary, sentence structures), meaning level and presentation level are discussed. It is suggested that this type of activity involves the students much more than traditional writing and leads to a different approach to and style of writing. Advice that the students found helpful, both in their writing and in composing their web pages, is given. The author found the project time-consuming but very rewarding, and argues that it is important to make use of a technology which is becoming more and more common.

**01-445 Sasaki, Miyuki** (Nagoya Gakuin U., Japan). Toward an empirical model of EFL writing processes: an exploratory study. *Journal of Second Language Writing* (Norwood, NJ, USA), **9**, 3 (2000), 259–91.

The study reported here investigated English as a Foreign Language (EFL) learners' writing processes using multiple data sources including their written texts, videotaped pausing behaviours while writing, stimulated recall protocols, and analytic scores given to the written texts. Methodologically, the study adopted a research scheme that has been successfully used for building models of Japanese first language writing. Three paired groups of Japanese EFL writers (experts vs. novices, more- vs. less-skilled student writers, novices before and after six months of instruction) were compared in terms of writing fluency, quality/complexity of their written texts, their pausing behaviours while writing, and their strategy use. The results revealed that (a) before starting to write, the experts spent a longer time planning a detailed overall organisation, whereas the novices spent a shorter time, making a less global plan; (b) once the experts had made their global plan, they did not stop and think as frequently as the novices; (c) second language proficiency appeared to explain part of the difference in strategy use between the experts and novices; and (d) after six months of instruction, novices had begun to use some of the expert writers' strategies. It was also speculated that the experts' global planning was a manifestation of writing expertise that cannot be acquired over a short period of time.

**01-446 Tarnopolsky, Oleg** (Dnepropetrovsk State Technical U. of Railway Transport, Dnepropetrovsk, Ukraine). Writing English as a foreign language: a report from Ukraine. *Journal of Second Language Writing* (Norwood, NJ, USA), **9**, 3 (2000), 209–26.

This report investigates teaching writing in English in Ukraine. The past and present situations in teaching writing and the reasons for avoiding teaching commu-

nicative writing skills in English courses in that country are considered. The results of Ukrainian English as a Foreign Language (EFL) students' needs analysis are presented, indicating the necessity of introducing writing into EFL courses. The process-genre approach is postulated as a foundation for elaborating an effective writing course for Ukraine, and the first version of the course based on such an approach is analysed. Causes of the failure of this course are reported. It is demonstrated that a successful EFL writing course has to be not only communicative but also state-of-the-art. To motivate students, it also has to involve them from the beginning level in activities, making writing itself fun. The second (successful) version of the course, with a great part of learning organised as writing for fun, is presented, and its advantages are shown.

**01-447 Warden, Clyde A.** (Chaoyang U. of Technology, Taiwan, Republic of China; *Email*: warden@mail.cyut.edu.tw). EFL business writing behaviors in differing feedback environments. *Language Learning* (Malden, MA, USA), **50**, 4 (2000), 573–616.

This paper raises the question of the steps English as a Foreign Language (EFL) students go through when completing business-oriented English documents, and whether these steps can be differentially influenced by the type of feedback provided. To understand students' reaction to feedback, path analysis was used to generate models of student writing activity. Three groups received differing combinations of feedback and redrafting opportunity over four assignments. A total of 563 first drafts from 141 students were analysed, resulting in 5,630 data points for modeling. Results suggest that Taiwan business English writing students may implement proof-reading and self-directed redrafting behaviours aimed towards improving their writing when objective sentence-level error feedback is supplied with no opportunity for further correction. The addition of redrafting opportunity resulted in generally unproductive writing strategies but also lower objective error rates. Students receiving general teacher-based feedback exhibited redrafting behaviours and a tendency to copy directly from reference material.

## Language testing

**01-448 Alderson, J. Charles** (Lancaster U., UK; *Email*: c.alderon@lancaster.ac.uk). Technology in testing: the present and the future. *System* (Oxford, UK), **28**, 4 (2000), 593–603.

As developments in information technology have moved apace, and both hardware and software have become more powerful and cheaper, the long-prophe-

sied use of information technology for language testing is finally coming about. The Test of English as a Foreign Language (TOEFL) is mounted on computer; CD ROM-based versions of University of Cambridge Local Examinations Syndicate tests are available; and the Internet is beginning to be used to deliver language tests. This paper reviews the advantages and disadvantages of computer-based language tests, explores in detail developments in Internet-based testing using the examples of TOEFL and DIALANG – an innovative on-line suite of diagnostic tests and self-assessment procedures in 14 European languages – and outlines a research agenda for the next decade.

**01-449 Billmyer, Kristine and Varghese, Manka** (U. of Pennsylvania, USA). Investigating instrument-based pragmatic variability: effects of enhancing discourse completion tests. *Applied Linguistics* (Oxford, UK), **21**, 4 (2000), 517–52.

A significant dilemma in sociolinguistic research concerns the methods used to collect data, the validity of different types of data and 'their adequacy to approximate the authentic performance of linguistic action' (Kasper & Dahl 1991: 215). Manes and Wolfson (1981) claimed the most authentic data in sociolinguistic research to be spontaneous speech gathered by ethnographic observation. Difficulties in relying on this method are well-documented and have led to the wide use of an elicitation procedure known as the discourse completion test (DCT). Justified criticisms have been levelled at the DCT, some labelling it an instrument that limits the capturing of authentic communication, and others making it look almost obsolete. However, there are to date no other sociolinguistic data collection instruments that have as many administrative advantages as the DCT, making it, practically speaking, a resource pragmatics testing and teaching will continue to rely on. Furthermore, a better understanding of communication in such a constructed context may lead to a better understanding of authentic communication. This study investigates the effect of systematic modification to DCT situational prompts used to elicit requests on the responses of native and non-native speakers of English. Modification included enhancing the prompts by adding information on a number of social and contextual variables considered relevant to this study. Results indicate that enhancement did not affect the request strategy or amount of internal modification. However, enhancing situational prompts did produce significantly longer, more elaborated requests in both groups. These findings point to the importance of external modification of speech act production and the need for an instrument that can satisfactorily account for both variation and context. The study has implications for both teaching and testing in interlanguage pragmatics.

**01-450 Bogaards, Paul** (Leiden U., The Netherlands). Testing L2 vocabulary knowledge at a high level: the case of the *Euralex French Tests*.

*Applied Linguistics* (Oxford, UK), **21**, 4 (2000), 490–516.

Testing second language (L2) vocabulary knowledge should take into account all aspects of what is implied by lexical knowledge: knowledge of form, meaning, morphology, syntax, collocations, and use in discourse. Different types of tests will be necessary to tap all these aspects. Some general aspects of vocabulary testing are briefly discussed in this context. In the second part of the paper, a new type of test which aims at testing very high levels of L2 vocabulary knowledge is analysed and investigated. Although the validity of the test as it stands is questionable, it is shown that it is possible to improve on both its reliability and its validity, and to adapt it to the level of vocabulary knowledge which is common with advanced university students.

**01-451 Chalhoub-Deville, Micheline** (U. of Iowa, USA; *Email*: m-chalhoub-deville@uiowa.edu) **and Turner, Carolyn E.** (McGill U., Canada; *Email*: cx9x@musica.mcgill.ca). What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System* (Oxford, UK), **28**, 4 (2000), 523–39.

The article is intended to familiarise test-users with the issues they need to consider when employing assessments for screening/admission purposes. It examines the purpose, content, and scoring methods of three English as a second language admission tests – the Cambridge certificate exams, International English Language Testing System, and Test of English as a Foreign Language-computer-based test – and discusses reliability and validity considerations salient to each instrument. The validity and reliability discussion is guided by the (1999) 'Standards for Educational and Psychological Testing'. Since the assessment scores obtained are used to help make critical decisions affecting test-takers' lives, it is critical that they provide high quality information. Developers of such large-scale tests have the responsibility to: construct instruments that meet professional standards; continue to investigate the properties of their instruments and the ensuing scores; and make test manuals, user guides and research documents available to the public. Test-users also have a responsibility: they need to be cognisant of the properties of the instruments they employ and ensure appropriate interpretation and use of test scores provided; and they need to carry out local investigations to make sure that their admission requirements are based on an informed analysis of their academic programmes and the language ability score profiles necessary to succeed in these programmes.

**01-452 Clapham, Caroline** (Lancaster U., UK; *Email*: caroline.clapham@lancs.ac.uk). Assessment for academic purposes: where next? *System* (Oxford, UK), **28**, 4 (2000), 511–21.

This paper starts by discussing research into the effect of background knowledge on English for Academic Purposes (EAP) tests, and discusses EAP tests in which the content of at least some of the test components is related to students' fields of academic study. Research has demonstrated that students do not necessarily do better if they are given tests in their own academic subject areas; and, because of the difficulties inherent in test-equating, such tests may not be testing the students fairly. The paper suggests, therefore, that for international EAP tests, English for Specific Academic Purposes testing be abandoned. The paper then looks to the future, and suggests that, instead of EAP proficiency tests, there should be aptitude tests to find out whether first and second language (L1/L2) students would be capable of rapidly acquiring the requisite academic discourse practices once they had embarked on their academic courses. Such tests for L2 students should include a test of specific grammatical skills, so that receiving institutions can be sure that students have the requisite linguistic infrastructure needed to carry out academic work in English.

**01-453 Cobb, Tom** (Université du Québec à Montréal, Canada). One size fits all? Francophone learners and English vocabulary tests. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes* (Toronto, Ont.), **57**, 2 (2000), 295-324.

This paper raises the question of which need is greater, the need for standard measures of vocabulary knowledge, or the need for vocabulary measures tailored to learners' first languages. The question is explored using placement test data from more than 1,000 Francophone students entering English language courses at a francophone university in Montreal in 1997 and 1998. The test consisted of several measures including a standard vocabulary size test, Nation's (1990) Levels Test. The study shows that a standard vocabulary measure can miss important information about learners' knowledge. It also suggests that an inter-language sensitive measure can be a better predictor of broader language proficiency, and concludes that different tests may be needed for different stages of second language development.

**01-454 Davidson, Fred** (U. of Illinois at Urbana-Champaign, USA; *Email*: fgd@uiuc.edu). The language tester's statistical toolbox. *System* (Oxford, UK), **28**, 4 (2000), 605-17.

Using analogy to carpentry, this article describes various statistical analysis tools in language testing, chiefly Classical Test Theory and Item Response Theory. Computer software for statistical analysis is briefly reviewed and divided into three tiers: commonly available (e.g., Excel); statistical packages (e.g., SPSS); and speciality software (e.g., FACETS). Simplicity is stressed: often the most challenging aspect of statistics

in language testing is interpretation, not the actual computational analysis. Regardless of the complexity of the analysis, it is considered unwise for statistics to drive language test development. Statistical determinism is seen as a dangerous epistemological force.

**01-455 Fulcher, Glenn** (U. of Surrey, UK; *Email*: g.fulcher@surrey.ac.uk). The 'communicative' legacy in language testing. *System* (Oxford, UK), **28**, 4 (2000), 483-97.

This article looks at the phenomenon of 'communicative' language testing as it emerged in the late 1970s and early 1980s as a reaction against tests constructed of multiple choice items and the perceived over-emphasis on reliability. Lado in particular became a target for communicative testers. It is argued that many of the concerns of the communicative movement had already been addressed outside the UK and that Lado was done an injustice. Nevertheless, the jargon of the communicative testing movement, however imprecise it may have been, has impacted upon the ways in which language testers approach problems today. The legacy of the communicative movement is traced from its first formulation, through present conundrums, to tomorrow's research questions.

**01-456 Hamp-Lyons, Liz** (Hong Kong Polytechnic U.; *Email*: eglhl@polyu.edu.hk). Social, professional and individual responsibility in language testing. *System* (Oxford, UK), **28**, 4 (2000), 579-91.

This paper explores some of the developments in the philosophy and epistemology of language testing with particular reference to the notion of 'responsibility'. This term is used to include the ethical practices of the individual language testing practitioner; the professional standards to which language testing as professional practice should be held; and the accountability to society and by society which should drive all those involved in the industry of creating, administering, consuming, preparing people for, and critiquing tests. The paper argues that, as the millennium changes, language testing is at a point of change too. Contending with complexities on multiple levels, the future is still unclear, but it is clear that it will take us in some directions quite different from the past.

**01-457 Larson, Jerry W.** (Brigham Young U., USA; *Email*: Jerry\_Larson@byu.edu). Testing oral language skills via the computer. *CALICO Journal* (San Marcos, TX, USA), **18**, 1 (2000), 53-66.

Although most foreign and second language teachers today stress the development of oral skills in their teaching, it is very difficult for them to find time to assess these skills. This article discusses the importance of testing language students' oral skills and describes

computer software which has been developed to assist in this important task. Information about various techniques that can be used to score oral achievement test performance is also presented.

**01-458 Mattos, Andréa Machado de Almeida** (Federal U. of Minas Gerais, Brazil; *Email:* amam@horizontes.net). A Vygotskian approach to evaluation in foreign language learning contexts. *ELT Journal* (Oxford, UK), **54**, 4 (2000), 335–45.

This paper reports on the results of a small-scale research project applying Vygotskian concepts to the evaluation of English as a Foreign Language (EFL), and suggesting that language assessment should be adapted to meet more 'real-life' learning situations. The study, which involved two pairs of EFL learners working on a cooperative picture-story task, is based on the Vygotskian concepts of the zone of proximal development, scaffolding, private speech, and task and activity. The research design draws on work by Donato (1994) and Coughlan and Duff (1994), and applies their ideas to suggest a sociocultural approach to the evaluation of EFL learners, which is seen as providing a better setting for language learning and teaching.

**01-459 McKee, Gerard, Malvern, David and Richards, Brian** (U. of Reading, UK; *Email:* b.j.richards@reading.ac.uk). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing* (Oxford, UK), **15**, 3 (2000), 323–27.

This paper describes software (*vocd*) that implements a solution to problems encountered in quantifying vocabulary diversity. Researchers in various fields of linguistic enquiry have calculated vocabulary diversity using the ratio of different words (Types) to total words (Tokens) – the Type-Token Ratio (TTR) – or measures derived from it. Such measures are flawed, however, because the values obtained are related to the number of words in the sample. The paper shows how the relationship between TTR and sample size can be described by a new mathematical model, which in turn leads to an innovative method of measuring vocabulary diversity. The software automates measurement from transcripts prepared in a widely used computer-readable set of conventions: the CHAT format of the CHILDES project. Options in *vocd* are described to show how the user can determine which linguistic items will count as valid types and tokens in the analysis. The new measure is calculated by, first, randomly sampling words from the transcript to produce a curve of the TTR against Tokens for the empirical data. Then the software finds the best fit between this empirical curve and theoretical curves calculated from the model by adjusting the value of a parameter. The parameter, *D*, is shown to be a valid and reliable measure of vocabulary diversity without the problems of sample size found with previous methods.

**01-460 Moser, Urs** (Universität Zürich, Switzerland). Evaluationen in den 'Fremdsprachen': Chancen und Grenzen. Methodische Zugänge, Konzeptionen und Aufwand. [Evaluation in 'foreign languages': opportunities and limits. Methodological approaches, concepts and expenditure.] *Babylonia* (Comano, Switzerland), **1** (2000), 10–14.

Evaluations based on measuring competence can be carried out with many different intentions. Self-evaluations (i.e., evaluated by the teacher or the school in question) are mainly used to check the competence of individual students, classes or even schools. Third-party evaluations on the other hand often foreground educational policy issues and problems of instruction. In theory, self-evaluations, third-party evaluations and educational policy and didactically-oriented evaluations may be combined. However, the greater the number of interests involved, the greater the effort and cost. This article looks at just such questions from the point of view of Switzerland, offering as it does cultural diversity, diversity of educational policy, a differentiated description of foreign languages related to the four mother tongues. Three areas are well suited to study in the Swiss situation: comparison of foreign language competencies between the linguistic regions; comparison of the curricular frameworks between cantons; and the analysis and comparison of different methodological approaches to teaching. The author looks at each of these three areas and outlines some of the opportunities and limits for evaluations imposed by the Swiss situation.

**01-461 North, Brian** (Eurocentres Foundation, Zurich, Switzerland; *Email:* bnorth@eurocentres.com). Linking language assessments: an example in a low stakes context. *System* (Oxford, UK), **28**, 4 (2000), 555–77.

Linking language assessments is becoming a matter of greater concern with the advent of educational frameworks used to orient curricula and profile attainment. Generally speaking these projects do not command the resources or know-how for fully fledged comparability projects, but the extent to which they often appear to rely solely upon expert judgement is surprising. The measurement literature recognises five classic ways of linking separate assessments with a decreasing degree of rigour: (1) equating; (2) calibrating; (3) statistical moderation; (4) predicting (or benchmarking); and (5) social moderation (or standards-oriented assessment). This article outlines practical ways in which the principles of these techniques have been employed on a modest scale to solve linking problems in a low stakes context: Eurocentres. It then goes on to argue that social moderation is potentially the strongest form of linking because it implies the development and validation of a common view of a construct. Finally, the article introduces the Common European Framework recently published by the Council of Europe, which consists of a descriptive scheme and validated proficiency scales, but

suggests that putting a common framework into practice will be accomplished more easily if the ideas behind the classic linking methodologies are employed in the process.

**01-462 Shohamy, Elana** (Tel Aviv U., Israel; Email: elana@post.tau.ac.il). The relationship between language testing and second language acquisition, revisited. *System* (Oxford, UK), **28**, 4 (2000), 541-53.

This paper examines the relationship between and the relevance of second language acquisition (SLA) and language testing (LT). It examines the interfaces of the two fields based on articles published in recent issues of the journals *Language Testing* and *Studies in Second Language Acquisition*. The examination is based on (a) three dimensions of potential contributions of LT to SLA, i.e., (1) defining the construct of language ability; (2) applying LT findings to test SLA hypotheses; and (3) providing SLA researchers with quality criteria for tests and tasks; and (b) three dimensions of potential contribution of SLA to LT, i.e., (1) identifying language components for elicitation and criteria assessment; (2) proposing tasks for assessing language; and (3) informing language testers about differences and accommodating these differences. The relevance of LT to SLA is examined based on written interviews with leading scholars in SLA who were asked about the relevance of LT to their work. The results indicate very limited interfaces between the two fields as well as limited relevance of LT to SLA. It is suggested that the findings point to the potential need of LT to broaden its focus and scope by addressing broader views of language learning and language processing.

**01-463 Spolsky, Bernard** (Bar-Ilan U., Ramat-Gan, Israel; Email: spolsb@mail.buu.ac.il). Language testing in *The Modern Language Journal*. *The Modern Language Journal* (Malden, MA, USA), **84**, 4 (2000), 536-52.

The *Modern Language Journal* published on average two articles a year dealing with language tests in its first 80 years. This probably reflects the actual (if not the desirable) level of interest in language testing of the language teaching profession. For the early years, before more specialised journals appeared, it gives an excellent picture of the history of the field in the United States. Later, the coverage became spottier, but there continued to appear a number of important articles, especially on topics like prognosis and aptitude tests, the cloze test, oral testing, and the controversy over the ACTFL [American Council on the Teaching of Foreign Languages] Proficiency Guidelines. As a whole, the articles show a valuable concern with the use rather than the form of language tests.

**01-464 Wall, Dianne** (Lancaster U., UK; Email: d.wall@lancaster.ac.uk). The impact of high-stakes

testing on teaching and learning: can this be predicted or controlled? *System* (Oxford, UK), **28**, 4 (2000), 499-509.

One of the issues which attracted the attention of language testers in the 1990s was the impact of high-stakes tests on the classroom: what kind of influence did such tests have on teaching and learning and what could educators do to ensure that this was beneficial rather than harmful? Some progress was made in defining notions such as 'impact' and 'washback', and a number of studies appeared which analysed the relationship between tests and teachers' and learners' attitudes and behaviour. There was a growing awareness of the importance of factors other than test design in determining whether tests would have the impact that was desired. These factors also appear in the literature of educational innovation, and it is to this field that some testers turned for guidance on whether test impact could be predicted or controlled. This paper summarises what language testers have learned about test impact in the last decade and discusses what one model of educational innovation has revealed about how tests interact with other factors in the testing situation. It concludes with a set of recommendations about the steps future test developers might take in order to assess the amount of risk involved in attempting to create change through testing.

## Teacher education

**01-465 Akyel, Ayşe** (Boğaziçi U., Istanbul, Turkey; Email: akyel@boun.edu.tr). Collaboration to explore teaching: a case study report. *TESL Canada Journal / La Revue TESL du Canada* (Burnaby, B.C.), **18**, 1 (2000), 58-74.

The study presented here is based on a pilot project focused on English as a Foreign Language (EFL) teachers' collaborative reflection on individual teaching practice. The purpose of the study was to analyse the collaborative efforts of two EFL teachers to explore teaching with a view to solving problems in their respective classes. In so doing the study also attempted to investigate whether this process of collaborative reflection would lead to a change in these EFL teachers' instructional practices and attitudes toward teaching. The results indicate that, during the process of collaboration, the participants formulated their own agenda for developing strategies to cope with their problems and for a possible change in their classroom practice. Moreover, the dialogue between the teachers seemed to affect their perception of an ideal teacher, although there was no change in their beliefs concerning to the extent to which they themselves had these characteristics of an ideal teacher.

**01-466 Bayliss, Doreen and Vignola, Marie-Josée** (U. of Ottawa, Canada). Assessing language