

language and learning believe that the latter kind of understanding is crucial to children's education. The nature of dialogue and its role in second language

teaching are explored, and implications for classroom practice are suggested.

97-75 Zimmermann, Günther (Technical U. of Braunschweig, Germany). Die Elaborationsstrategien guter und schlechter Textverarbeiter. [The elaboration strategies of good and poor processors of text.] *Zeitschrift für Fremdsprachenforschung* (Bochum, Germany), **7**, 1 (1996), 43-59.

Four subjects were asked to read some pages from a grammar book in preparation for a test, and to think aloud as they read. The think-aloud protocols were analysed for the frequency and quality of the elaborations (rephrasing, expanding etc.) which they used in explaining the grammar to themselves. Quality was assessed on a 3-level scale (Schnolz), Level 1 being simple memorising and surface processing, Level 3 involving understanding and manipulating underlying concepts. Two of the subjects, both from Class 8 (age approx. 13-14) of a *Hauptschule* (non-academic secondary school), were considered 'poor processors'; the others, one from

Class 9 (age approx. 15-16) of a *Gymnasium* (grammar-school) and the other with four years at university, were considered 'good processors', and it was hypothesised that the second two would produce more and better elaborations. This was confirmed far beyond expectations: the first two produced one and five elaborations respectively, all at Levels 1 or 2, the others produced 63 and 64, many at Level 3. Examples are quoted, illustrating misunderstandings by the first pair, sophisticated strategies from the second. It is recognised that the findings are not generalisable, and a further study is planned.

97-76 Zimmerman, Hansmartin and Werlen, Iwar (U. of Bern). Das Projekt "Zweitsprachunterricht im obligatorischen Schulsystem": Konzept, empirische Daten und erste Ergebnisse. [The project 'Second language teaching in the compulsory school system': Concepts, empirical data and first results.] *Bulletin Suisse de Linguistique Appliquée* (Neuchâtel, Switzerland), **64** (1996), 35-59.

The project 'Second Language Teaching in the Compulsory School System' investigates the processes of language acquisition and teaching in language classrooms of state schools in the Canton of Berne. Only too often have there been complaints from politicians, journalists and education specialists in the media and elsewhere that Swiss school children appear to have trouble in attaining a satisfactory level of knowledge of a second national language (L2). According to the principle of territoriality, the Canton of Berne is officially bilingual with a majority of its population living in the German-speaking area and a minority in the French-speaking area. For this reason, and in

conformity with the national language policy, the first L2 must be either French or German. Research has shown that the majority of students are neither strongly motivated to learn nor able to use the L2 competently even after several years' training. The project starts from the hypothesis that in the language classroom as an institutional setting, characteristic forms of discourse are generated in the interaction between teacher and students which result in the creation of typical linguistic patterns. A second hypothesis maintains that typical patterns of discourse which operate in the L2 classroom may have been conditioned in the first language classroom earlier on in the child's school career.

Language testing

97-77 Alderson, J. Charles (Lancaster U.) and **Hamp-Lyons, Liz** (Hong Kong Poly. U.). TOEFL preparation courses: a study of washback. *Language Testing* (London), **13**, 3 (1996), 280-97.

'Washback' is the influence that writers on language testing, syllabus design and language teaching believe a test will have on the teaching that precedes it. Much has been written about the influence of

testing on teaching. To date, however, little empirical evidence is available to support the assertions of either positive or negative washback. The English proficiency test TOEFL (Test of



English as a Foreign Language) is among the best-known examinations in the field of English language teaching. The study reported in this article set out to investigate common claims that the TOEFL exerts an undesirable influence on language teaching. The data consist of interviews with teachers and students, and observations made of two sorts of classes: 'normal' language proficiency classes, and parallel classes intended for students preparing to take the TOEFL. The authors observed both TOEFL

preparation classes and non-TOEFL preparation classes by the same teachers in order to be able to separate washback from the TOEFL from any possible effect of individual teacher style. They suggest as a result that simple forms of washback hypotheses are too naive: influences on what happens in class are much more complex than unexamined beliefs about washback allow, and more complex hypotheses about washback are needed.

97-78 Bailey, Kathleen M. (Monterey Inst. of International Studies). Working for washback: a review of the washback concept in language testing. *Language Testing* (London), **13**, 3 (1996), 257-79.

This article is a literature review which seeks to answer four questions: (1) What is washback? (2) How does washback work? (3) How can positive washback be promoted? (4) How can washback be investigated? Building on the 'Washback Hypothesis' proposed by Alderson and Wall, and

suggestions from Hughes, the article proposes a model that identifies participants, processes and products which may influence, or be influenced by, washback. Strategies for investigating washback are also discussed.

97-79 Brown, Annie and Iwashita, Noriko (U. of Melbourne). Language background and item difficulty: the development of a computer-adaptive test of Japanese. *System* (Oxford), **24**, 2 (1996), 199-206.

The use of new statistical procedures such as Item Response Theory has greatly facilitated the development of computer-adaptive tests, where the adaptiveness is based on measures of item difficulty resulting from the performance of trial candidates. However, studies into the acquisition of second language (L2) grammar by learners with different first languages (L1s) indicate that the learners' L1 strongly influences their acquisition of grammar in the L2. Thus, it would be expected that grammar test items would present different levels of difficulty to candidates from different language backgrounds. Where a computer-adaptive grammar test is to be used with such candidates it is, therefore, questionable whether set item difficulty measures can validly be used for all types of candidate. The study investigates the performance of students from

different language backgrounds, using data from a computer-adaptive Japanese grammar test developed as a placement tool. The trial pen and paper test consisted of 225 multiple choice items. Fourteen hundred students in Australia, China and Japan (all of whom had studied Japanese for 150-500 hrs) each completed 50 items. In this study, data are presented from native speakers of English and Chinese. Item difficulties were found to be quite different for the three groups of candidates. The authors suggest that this has implications for the validity of use of computer-adaptive tests, in that where actual candidates are from a different background from that of the trial population not only does the test fail to measure such candidates efficiently, but their measures of ability are also affected.

97-80 Elder, Catherine (U. of Melbourne). The effect of language background on 'foreign' language test performance: the case of Chinese, Italian, and Modern Greek. *Language Learning* (Cambridge, MA), **46**, 2 (1996), 233-82.

This article considers the validity of applying common assessment instruments and scales to assess the language skills of Australian school-age language learners from different first language (L1) backgrounds. The author used data gathered from standardised national tests of reading and listening proficiency in three heritage languages: Italian,

Greek, and Chinese. The tests, known as the Australian Language Certificates, are administered annually to a large population of language learners at Year 8 or Year 9. Mann Whitney and Mantel-Haenszel producers were applied to data from the 1993 administration to investigate score differences and differential item functioning (DIF) between

those with and without a home background in each of the three target languages. Content analyses of DIF items attempted to identify possible sources of group difference. Findings reveal a strong relationship between home exposure to the language and level of performance on the listening and, to a lesser extent, the reading component of the tests. There is also evidence of differential item functioning, although this effect is stronger for

Chinese and Italian than for Greek. The content analysis suggests that the DIF effect may be due to 'true' differences in language ability, rather than to test bias. Nevertheless, the use of common instruments and reporting procedures for first and second language learners may be inappropriate and demotivating and, in the later years of schooling, could lead to serious misuse of test scores.

97-81 Harrison, Colin (Nottingham U.). Balancing the priorities of the classroom teacher with the imperatives of High Stakes assessment of reading: an English perspective. *Interchange* (Dordrecht, the Netherlands), **27**, 3/4 (1996), 349-60.

The question underpinning this paper, and the international symposium which preceded it, is framed thus: 'In what ways can we learn from each another how to improve reading assessment, at the personal, classroom, local and national levels?'. The author first presents some background information on the New Paradigms in Reading Assessment seminar. An account is then offered of language arts assessment (the subject termed 'English') in the United Kingdom, which he suggests goes some way to explaining why in England there has been a

particularly urgent need to learn from the experiences of other countries; and this is followed by examples of new approaches elsewhere, particularly in the United States, Australia and Scotland. Finally, the author puts forward some principles which might underpin new approaches within the English system, together with an indication of how some of these are being put into practice in classrooms in England, as pilot work on an international collaborative study.

97-82 Kempe, Vera (Toledo U.) and **MacWhinney, Brian** (Carnegie Mellon U.). The crosslinguistic assessment of foreign language vocabulary learning. *Applied Psycholinguistics* (New York), **17**, 2 (1996), 149-83.

This study examines a task which it is claimed can be applied in a uniform fashion across different languages to compare levels of vocabulary development in foreign language learning. Experiment 1 tested native speakers of Russian and German and demonstrated the basic comparability of the subjects' judgements for both words and nonwords. The results for Russian showed an influence of word length, which can be understood in terms of the Orthographical Depth Hypothesis. Experiment 2 applied the same task to learners of Russian and German and found that learners of Russian had achieved a lower level of vocabulary control than learners of German at comparable language exposure levels. This disadvantage for

Russian can be attributed to the novelty of the Cyrillic graphemic system, which restricts the accessibility of written language input at early stages. There was a nonlinear increase over time in word sensitivity, which can be attributed to the increasing contribution of lexical plausibility factors at later stages of learning. Moreover, the lexical decision task appeared to be sensitive to inhibitory effects of concurrently studied languages, as well as to decay due to the lack of regular exposure. It is claimed that, taken together, these results indicate that the lexical decision task can be a useful tool for the assessment and crosslinguistic comparison of lexical development in foreign language learning.

97-83 Liddicoat, A. (Australian National U.) The Language Profile: oral interaction. *Babel* (Victoria, Australia) **31**, 2 (1996) 4-7, 35.

In the fields of second language (L2) teaching and testing, listening and speaking are generally identified as discrete skills. This paper examines the Australian schools' Language Profile oral interaction component, which recognises listening and speaking

as interdependent skills. It is based on a 'conduit' model of communication, and assesses students according to their ability to communicate in the L2. The aim is for learners not only to be able to produce utterances in the L2, but to participate



successfully in spontaneous conversational exchanges. The progression from receptive to productive skills as learners become more proficient mirrors actual language development. Proficient L2 speakers are expected to employ the features of native speaker conversation such as shortened messages, non-verbal signals, and a wide range of

speech acts. In order to extend classroom interaction beyond the traditional teacher-led model, learners need to practise a broad range of communicative tasks. As a result, they develop knowledge and understanding of linguistically and culturally appropriate language use.

97–84 Messick, Samuel (Educational Testing Service). Validity and washback in language testing. *Language Testing* (London), **13**, 3 (1996), 241–56.

Washback, a concept prominent in applied linguistics, refers to the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning. Some proponents have even maintained that a test's validity should be appraised by the degree to which it manifests positive or negative washback, a notion akin to the proposal of 'systemic

validity' in the educational measurement literature. This article examines the concept of washback as an instance of the consequential aspect of construct validity, linking positive washback to so-called authentic and direct assessments and, more basically, to the need to minimise construct under-representation and construct-irrelevant difficulty in the test.

97–85 Padilla, Armado M. (Stanford U.) **and others.** Development and implementation of student portfolios in foreign language programs. *Foreign Language Annals* (New York), **29**, 3 (1996), 429–38.

Issues related to portfolio development and use in assessing language learning in foreign language education are discussed. Among the issues are 'audiences' and 'purposes' of the portfolio. With teachers involved in an evaluation project of less commonly taught languages (Japanese, Chinese, Korean, and Russian) representing different educational levels (elementary and secondary), the authors collaborated in the design and implementation of student portfolios to examine

growth in foreign language proficiency. The contents of portfolios were analysed to determine their usability as a means of assessing foreign language learning. Analysis of the portfolios showed that many factors should be considered by teachers in deciding on contents and objectives of the portfolio. The article concludes with a series of recommendations for foreign language educators interested in using portfolios to document their students' language learning progress.

97–86 Shohamy, Elana and others (Tel Aviv U.). Test impact revisited: Washback effect over time. *Language Testing* (London), **13**, 3 (1996), 298–317.

In countries with centralised educational systems national tests are used as primary devices through which changes in the educational system are introduced. This article examines the impact over time of two national tests which have been in operation for a number of years: one in Arabic as a second language (ASL) and one in English as a foreign language (EFL), through questionnaires, interviews and document analysis from a sample of teachers, students and language inspectors. Results showed different washback patterns for the two tests. Slight modifications in the ASL test created no effect in classroom activities, test preparation, or the status and prestige of the subject tested, yet the inspectors expressed satisfaction and wished to continue the

administration of the test as they feared that without the tests proficiency levels would drop. Slight modifications in the EFL test, on the other hand, created major impact in terms of teaching activities, time devoted for test preparation, production of new teaching material, etc. Although negative attitudes were expressed by teachers regarding the quality of the test, they would like it to continue. Inspectors believe that the test creates a meaningful change and is powerful enough to trigger changes without a need to provide training and a new curriculum. The study shows that washback varies over time, owing to many factors such as the status of the language and the uses of the test.

97–87 Wall, Dianne (Lancaster U.). Introducing new tests into traditional systems: insights from general education and from innovation theory. *Language Testing* (London), **13**, 3 (1996), 334–54.

Educators are concerned about the effects that tests may have on teaching, and want to know what they should do to ensure that these effects are beneficial. Some English language teaching specialists have offered advice on how to produce positive washback, calling for attention to test design and the communication between testers and teachers. Research in general education has pointed out the need for feedback from testers to many key players in the educational system, and for adequate

resourcing and training. Innovation theory provides further insights into why attempts to introduce change in the classroom are often not as effective as their designers hoped they would be. This article reviews several key concepts in educational innovation, showing how these concepts are manifested in a case study in washback and outlining how they are being applied in recent test development projects.

Teacher education

97–88 Burgess, John (U. of Manchester) and **Carter, Iain G.** (Inter. School of Helsinki). Common codes for mainstream ESL support across the curriculum. *System* (Oxford), **24**, 2 (1996), 211–22.

This paper reports a five-day pre-term inservice training (INSET) course for the staff of an international school in Colombo, Sri Lanka. The first section gives a description of the school and its perceived need for INSET on cross-curricular English language support for bilingual students developing their knowledge of, and skills in, English. The second section describes the course, whose contents the authors suggest to be relevant to

all contexts of English-medium school education, and indicates related curriculum action. The principal argument is that all teachers involved in the support of English as a second language (ESL) learners – i.e. both mainstream subject teachers and language support specialists – can best communicate with each other and facilitate their students' learning through the medium of 'common codes'. These common codes are defined and exemplified.

97–89 Kaufman, Dorit and Grennon Brooks, Jacqueline (New York State U.). Interdisciplinary collaboration in teacher education: a constructivist approach. *TESOL Quarterly* (Washington, DC), **30**, 2 (1996), 231–51.

The authors assert that teacher education programmes must begin to foster in beginning teachers of all disciplines new images of collaboration, involvement and inquiry: images of classroom environments where students of all cultures engage in interdisciplinary activities and construct knowledge rooted in personal experiences. They cite the high numbers of language minority students who score below the national norms in mathematics and science to underscore the negative ramifications of uncoordinated instruction on their academic success. The article then describes the

evolution of a collaborative initiative involving graduate and undergraduate students in two teacher education programmes. The collaboration was motivated by constructivist approaches, and integrates language pedagogy and science instruction. It is based on the premise that, if teachers are to collaborate in schools and create enhanced interdisciplinary classroom environments that better foster students' linguistic and academic growth, they must experience such pedagogy in university teacher education programmes.

97–90 Neophytou Eleftheria East, Patricia and Chan Sui-Mee, Michelle (N. London U. and United Medical and Dental Schools, London). Enhancing the educational experience of bilingual students on initial teacher education. *Multicultural Teaching* (Stoke on Trent, Staffs.), **15**, 1 (1996), 29–33.

At the School of Teaching Studies 46 different languages are spoken by students. Two to four hour

language learning workshops were set up for bilingual students in order to develop the students'