

99-476 Scholler, D. (Freie Universität Berlin, Germany). *Se débrouiller à l'écrit: ein Trainingsprogramm zur Verbesserung der Textproduktion*. [Getting by in writing: a training programme for improving text production.] *Fremdsprachenunterricht* (Berlin, Germany), **1** (1998), 22–31.

This article presents a graded series of classroom exercises [appended in full] for German students of French preparing the *Abitur*, designed to develop their skills in recognising and deploying the means by which written texts achieve coherence. The author considers this to be a neglected area of language pedagogy, attributable to the emphasis in the 1970s on spoken communicative competence, and also to theoretical uncertainties over the term 'coherence'. Building on the work of Charolles (his four rules of text coherence) and Bieber, the author develops a teaching unit based on the use of *connectors* – here given a wide definition, but divided into *logical connectors* and *connectors of orientation* (with further subdivisions). The exercises themselves range from identifying the connectors in a quasi-authentic text, through binary-choice and gap-filling exercises, to a free text-production exercise in which students play the part of a detective presenting a report of an investigation.

99-477 Tickoo, Asha (Hong Kong Baptist U.). Monitoring the structure of meaning in beginners' ESL prose. *TESL Canada Journal / La Revue TESL du Canada* (Burnaby, B.C.), **16**, 1 (1998), 1–18.

The prose of beginners' English as a Second Language (ESL) is characteristically both over-segmented and inappropriately segmented with respect to information structure, posing a unique type of challenge to readers attempting to process such texts. This article examines whether this unconventional form of information segmentation results in violation of certain information-packaging constraints that ordinarily ensure that readers can easily identify the macrostructure of the text [references given]. Comparison of 20 descriptive essays written by ESL beginners with the prose of skilled writers suggests that four constraints on information packaging are needed to facilitate the identification of macrostructure. Evidence is presented of the absence of these constraints in beginners' prose, and it is demonstrated that revising a sample essay to incorporate these constraints results in easier recognition of its macrostructure. These findings suggest that the fragmentation problem in beginners' ESL prose may be overcome if learners are made aware of the notion of macrostructure and taught to conform to the four organisational constraints that ensure its accessibility.

99-478 Villamil, Olga S. and de Guerrero, Maria C. M. (Inter American U. of Puerto Rico, USA). Assessing the impact of peer revision on L2 writing. *Applied Linguistics* (Oxford, UK), **19**, 4 (1998), 491–514.

The study reported here sought to investigate the impact of peer revision on writers' final drafts in two rhetorical modes, narration and persuasion, among 14 Spanish-speaking ESL (English as a Second Language) college students. The two questions addressed were how revisions made in peer sessions were incorporated by writers in their final versions; and how trouble-sources were revised according to different language aspects (content, organisation, vocabulary, grammar and mechanics). An analysis of audiotaped interactions, first drafts and final drafts revealed that 74% of revisions made in the peer sessions were incorporated. These revisions suggest a pattern of behaviour conducive to self-regulation among writers. Results also show that students focused equally on grammar and content when revising in the narrative mode and predominantly on grammar in the persuasive mode; organisation was the least attended aspect in either. Only 7% of false repairs were found overall. The study suggests that peer assistance can help L2 (second language) intermediate learners realise their potential for effective revision to the extent their linguistic abilities permit. The authors believe that peer revision should be seen as an important complementary source of feedback in the ESL classroom.

Language testing

99-479 Bae, Jungok and Bachman, Lyle F. (U. of California, LA, USA). A latent variable approach to listening and reading: testing factorial invariance across two groups of children in the Korean/English Two-Way Immersion Programme. *Language Testing* (London, UK), **15**, 3 (1998), 380–414.

The study reported here investigated the factorial distinctness of two receptive language skills, reading and listening, and the equivalence of factor structure across two groups using simultaneous multigroup co-variance structure analyses. Participants were two groups of students (Korean-American and non-Korean-American) from grades two, three and four, enrolled in the Korean/English Two-Way Immersion Programme in the Los Angeles Unified School District, all learning Korean as a primary/foreign language. The analyses were based on data from tests of listening and reading in Korean. The results indicate the following: (1) the two receptive skills are factorially separable; (2) a two-factor model with listening and reading factors applies across the two groups of learners; (3) the correlation between the listening and reading factors was high and the same across the two groups; (4) the variation in levels of listening and reading proficiency differed across the groups; (5) with the exception of one listening test task, the degree to which the listening and reading tasks measured listening and reading ability was the same across the two groups; and (6) the test task type that had the highest factor loadings for both groups was one which presented test takers with a set of passages (listening or reading), each of which was followed by



comprehension questions. The study also makes a methodological contribution in that it investigated the nature of the two receptive language skills in a latent variable framework, using simultaneous analyses of two groups of young children, and also demonstrated a way to detect measurement invariance as a critical prerequisite to achieve validity of inferences based on measures. [See also abstracts 99-486 + 99-491]

99-480 Child, James R. (Nat. Cryptologic School, Fort Lowe, MD, USA). Language skill levels, textual modes, and the rating process. *Foreign Language Annals* (New York, USA), **31**, 3 (1998), 381-91.

This paper first considers the organisation of the wording of the Interagency Language Roundtable (ILR) descriptions of four language skills across five levels, principally as they relate to second language learners. It then summarises another paper by the author in which he posits four textual levels that may be reasonably assumed to cover the communicative needs of speakers of most, if not all, languages. Finally, the author considers the complex role of the rater in assessing proficiency and performance skills, and suggests an approach to rating whereby texts typical of each level can serve as models in assigning skill levels.

99-481 Coombe, Christine, Kinney, Jon and Canning, Christine (United Arab Emirates U.).

Issues in the evaluation of academic listening tests. *Language Testing Update* (London, UK), **24** (1998), 32-45.

The authors attempt to provide a series of key criteria by which English for Academic Purposes practitioners can evaluate the validity and effectiveness of the tests of academic listening they currently use. They summarise the differences between general and academic listening in terms of micro-skill taxonomies, and highlight the significance of cultural factors and background knowledge as well as the implications of using more 'direct' tests which mimic real-life, authentic communicative situations – although here increased validity is potentially compromised by a decrease in statistical reliability (which is where 'indirect', discrete-point testing approaches have the advantage). It is felt that academic listening tests should aim to embody the 'cornerstones' of good testing practice (reliability, validity, practicality and positive washback), and display authenticity of text, task and medium, not least as academic listening is a proactive, 'top down' and holistic process rather than a 'bottom up', atomistic deciphering of acoustic signals. Test designers should select culturally appropriate topics, relevant to students' actual listening needs and local experience, and be sensitive to 'skill contamination', whereby the listening task itself (and a student's performance) is only accessible/demonstrable via competence in reading and writing. Whatever skill integration is used in academic listening tests, this should ultimately be consistent with the target environment.

99-482 Dunkel, Patricia A. (Georgia State U., USA). Considerations in developing or using

second/foreign language proficiency computer-adaptive tests. *Language Learning and Technology* (<http://polyglot.cal.msu.edu/llt>), **2**, 2 (1999), 77-93.

This article describes what a computer-adaptive test (CAT) is, examines its roots, and points out some of the challenges this innovative approach to assessment presents. A number of issues involving CATs are then discussed in terms of (a) the basic principles of assessment embodied in the CAT, (b) the special psychometric and technical issues peculiar to the CAT as opposed to traditional or paper-and-pencil tests, (c) the hardware and software used in the CAT, and (d) the administration of the CAT. Each of these issues is discussed in terms of the questions that CAT score users should ask (and have answered) and that developers must consider when designing second language (L2) CATs. It is suggested that the questions should highlight the need for published CAT tables of specifications, models, and blueprints; should help design evaluative criteria for assessing the reliability, validity, and utility of L2 CATs; may help developers fashion more valid, reliable, and authentic L2 CATs; and might familiarise CAT score users with some of the complexities involved in interpreting CAT scores, as opposed to more traditional paper-and-pencil scores.

99-483 Fulcher, G. (U. of Surrey, UK).

Widdowson's model of communicative competence and the testing of reading: an exploratory study. *System* (Oxford, UK), **26**, 3 (1998), 281-302.

This paper considers the validity of Widdowson's discourse model of communicative competence and performance (1983, *Learning Purpose and Language Use*, Oxford University Press) as the basis for developing tests of reading. The article describes the basic structure of the model, and attempts to add to the original description in such a way that it is possible to design an operational test of the model components. The purpose of the operational test is only to investigate the adequacy of the model as a theoretical basis for test design, but successful model validation is potentially very important for the interpretation of test scores. The results suggest that there is some empirical support for the position that one level of the model is valid for the stated purpose.

99-484 Gille, Erna and Mets, Jan (Cito, Arnhem, The Netherlands). Het ontwikkelen van een programma van toetsing en afsluiting. [Developing a final evaluation programme for integrated literature teaching.] *Levende Talen* (Amsterdam, The Netherlands), **534/535** (1998), 539-43.

This paper deals with the evaluation of literature teaching in Dutch secondary education. Due to recent educational innovations, literature will be partly examined by means of portfolios. The objectives for these portfolios have been formulated in such a way that formal, traditional testing is not possible. The portfolios will

contain checklists for recording whether students have adequately carried out certain tasks, and actual pieces of work produced. In addition, a formal test may be used for evaluating certain aspects of the literary skills, such as evaluative judgements and justifications of personal preferences. First and foremost, however, the objectives of the literature domain should be translated into a complete trajectory of activities that each student should go through. This outline, which should contain both the activities to be carried out and the criteria by means of which they will be evaluated, is officially called the 'final evaluation programme'. The authors describe the steps that should be taken in order to develop such a programme. The article is in fact a summary of the full guidelines to be published by Cito – the Dutch National Institute for Educational Measurement – later this year.

99-485 Kieweg, Werner (Munich U., Germany). Allgemeine Gütekriterien für Lernzielkontrollen. [Common standards for the control of learning aims.] *Der Fremdsprachliche Unterricht Englisch* (Stuttgart, Germany), **37**, 1 (1999), 4–11.

This article makes a plea for common standards in examining English as a Foreign Language. It is considered that within schools there is often little or no discussion of test formats, and error types tend to be equally weighted, along with other inconsistencies in assessment. The author proposes correct weighting of passive, productive and interactive knowledge. He discusses the advantages and disadvantages of three types of testing: discrete-point testing, exemplified in cloze tests; integrative testing, with a focus on the picture story; and pragmatic testing in the form of a letter of elicitation. Examples are given, with the focus on testing various past tense forms in English. The article then highlights factors which need to be taken into account when combining different test formats. It is suggested that, in order to achieve transparency for students, a common standard of testing needs to fulfil the following criteria: the various control processes need to be known; the tasks and questions need to be clearly and unambiguously formulated; the marks which can be achieved must be clearly indicated; and examiners' guidelines on these should be provided (i.e. how the marks are broken down).

99-486 Kunnan, Antony John (California State U., USA). An introduction to structural equation modelling for language assessment research. *Language Testing* (London, UK), **15**, 3 (1998), 295–332.

This article is the first in a special themed issue of *Language Testing* which introduces a relatively new methodology to the field of language assessment, i.e. structural equation modelling (SEM). The article provides an introduction to SEM for language assessment research in five sections. In Section I, the general objectives of SEM applications relevant to language assessment are presented. Section II offers a brief overview of

SEM which considers the methodology and the statistical assumptions about data that have to be met. In Section III, the commonly used steps and concepts in SEM are presented. Section IV discusses SEM application matters with example models. The final section addresses recent critical discussions and some directions for future SEM applications in language assessment research. [See also abstracts 99-479 + 99-491]

99-487 Liddicoat, Anthony J. (Australian Nat. U.). Trialling the languages profile in the A.C.T.. *Babel (AFMLTA)* (Victoria, Australia), **33**, 2 (1998), 14–38.

This article highlights concerns regarding the Australian Capital Territory (ACT) component of the 1994/95 National Professional Development Programme, whereby 'languages profiles' with reference to annotated student work samples for foreign languages (FLs) Chinese, French, German, Indonesian, Italian, Japanese and Spanish provide a set of proficiency guidelines for the evaluation of students' oral and written performance. The choice of appropriate proficiency tasks is considered problematic, as performance can be influenced by 'prior work' (i.e. preparing/rehearsing language) and by an assessment situation being 'routine' (resulting in formulaic language) or else demanding spontaneous and pro-active language use. Difficulties are also perceived in the definitions provided within the descriptors at each level (e.g. between the concept of 'several linked sentences' versus a 'paragraph'), and in the difference in cognitive load between the development of FL literacy in a character-based language as opposed to a roman script. It is felt that this difference could affect the expectation of likely achievement in each language in the Kindergarten to Year 10 period, and it is suggested that empirically-based descriptions of the achievements of actual learners should inform the ACT profiles. The profiles are also considered deficient in failing to identify possible levels of 'sociocultural' knowledge, which is felt to be part of successful FL learning.

99-488 Lowe, Jr., Pardee (Nat. Cryptologic School, Fort Lowe, MD, USA). Keeping the optic constant: a framework of principles for writing and specifying the AEI definitions of language abilities. *Foreign Language Annals* (New York, USA), **31**, 3 (1998), 358–80.

This article examines the need to keep constant the optic, i.e. the framework within which the AEI definitions, both the ILR Language Skill Level Descriptions and the ACTFL Guidelines, are revised and more fully specified. (In the acronym AEI, the 'A' refers to ACTFL – American Council on the Teaching of Foreign Languages; the 'E' to ETS – Educational Testing Service; and the 'I' to the ILR – the U.S. Government's Interagency Language Roundtable: these organisations have contributed historically to the development and elaboration of definitions of foreign language proficiency, which are periodically revised.) Although the definitions



have been used for a variety of purposes, such as curriculum development, the present discussion focuses on their use in describing and assessing foreign language user abilities. The author raises issues concerning the delimiting of the AEI ranges, parsimony, and more fully specifying the criterion-referenced nature of the definitions, as well as issues concerning their use for assessing the abilities of target language natives and of heritage speakers. Seventeen principles are presented and their utility examined for keeping the optic constant.

99-489 Morton, Janne (U. of Melbourne, Australia). A cross-cultural study of second language narrative discourse on an oral proficiency test. *Prospect* (Macquarie U., Sydney, Australia), **13**, 2 (1998), 20–35.

This article investigates cross-cultural factors in oral proficiency testing. The author proposes that the analysis of discourse produced on speaking tests can provide useful information about the validity of such tests. Recent studies have shown that variables such as interviewer, task, topic, and gender of candidates all contribute to candidate discourse variability [references are given]. The study reported here set out to investigate whether another variable – the cultural-linguistic background of test takers – had a systematic effect on the discourse produced on a single narrative task and on the way this discourse was rated. The data were collected from a test which contributed to the decision-making processes regarding the eligibility of certain categories of prospective immigrants to Australia. A Rasch multi-faceted analysis of the scores revealed no significant bias according to the background of the test takers. However, the analysis of the language produced was of value in understanding how test takers from different cultural-linguistic backgrounds approached the task and how performances were rated.

99-490 Pollard, John (Saudi Development and Training, Saudi Arabia). Research and development – a complex relationship. *Language Testing Update* (London, UK), **24** (1998), 46–59.

This article describes the development in Saudi Arabia of the ‘Five Star Test’, a computer-resourced proficiency test, and points out how valid tests are not only triggered ‘top down’ from theory and research, but also via commercial situations involving various stakeholders – not all of them sympathetic to the complexities of testing issues – and ‘real world’ constraints. The negotiations, concessions, improvisations and informed risk-taking attendant on such contexts do not, it is felt, negate the effectiveness or theoretical integrity of what may be developed. The history of the test’s development is outlined, and examples provided [appendices] of the basic format, which involves transactional oral sessions between native speaker rater/interlocutor and non-native speaking testee at a computer, the testee’s responses invisibly scored in a database. It is claimed that the communicative nature of this interaction –

upheld by an independent validity study carried out by Sheffield Hallam University – is in line with current testing theory favouring integrative and authentic testing scenarios. The author concludes that there is a role for ‘research-in-development’, where practical exigencies can develop good testing practice – in concert, of course, with the designer’s awareness of the theoretical background.

99-491 Purpura, James E. (Columbia U., New York, USA). Investigating the effects of strategy use and second language test performance with high- and low-ability test takers: a structural equation modelling approach. *Language Testing* (London, UK), **15**, 3 (1998), 333–79.

Implicit in the research on strategy use is the notion that high- and low-ability language learners utilise strategies differently, and that these differences are related to differential performance; however, to date, no study has tested these assumptions statistically. This article reports a study using structural equation modelling to examine the relationships between strategy use and second language (L2) test performance with high- and low-ability test takers. Strategy questionnaires and a language test were given to 1,382 test takers, and separate baseline models of strategy use and L2 test performance were established for each group. These models were then estimated simultaneously with cross-group equality constraints imposed. The similarities and differences in this multi-group model provided the means for evaluating the viability of the parameter estimates across ability levels. The results showed that the metacognitive strategy use (MSU) and L2 test performance models produced almost identical factorial structures for each group, while the cognitive strategy use (CSU) models were somewhat different. Then, while some evidence of cross-group equivalence was observed, the cross-group tests of invariance could not be uniformly supported in the data; the effect of MSU on CSU was found to be variant across the groups, as was the effect of the retrieval processes on lexico-grammatical ability, suggesting that the variation in lexico-grammatical ability and reading ability stems from the test takers’ use of the retrieval processes in concert with the metacognitive strategies. [See also abstracts 99-479 + 99-486]

99-492 Tsui, Amy B. M. (U. of Hong Kong) and **Fullilove, John**. Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics* (Oxford, UK), **19**, 4 (1998), 432–51.

Some of the studies in reading and listening comprehension which have investigated the processing skills used by skilled and less-skilled readers/listeners have observed that the former are those better able to engage in top-down processing; others have found similarly in favour of bottom-up processing. The study reported here investigates which kind of processing skill is more important in discriminating the perfor-

mance of second language (L2) learners on listening test items in large-scale public examinations in Hong Kong over a seven-year period. Two variables were investigated: the schema type of the aural text and the question type. Two types of schema were identified: 'non-matching', where the schema activated by the initial linguistic input is *not* congruent with the subsequent linguistic input; and 'matching', where it is. In the first case, candidates need to process the incoming linguistic cues rapidly and accurately, and revise their developing schema accordingly to get the correct answer; in the second case, they could rely on top-down processing. Two types of question were also identified. 'Global' questions required candidates to understand the text as a whole and draw conclusions or inferences; hence, they need to process all the key linguistic cues rapidly and accurately. 'Local' questions require candidates only to pick out specific details; hence they could get the correct answer even when failing to process some of the cues. When the mean criterion scores for the correct answers were compared for schema type and question type using a series of paired t-tests, results showed that 'non-matching' items yield significantly higher scores than 'matching' items regardless of question type, suggesting bottom-up processing to be more important than top-down processing in discriminating the listening performance of L2 learners on test items.

Teacher education

99-493 Borg, Simon (U. of Malta). Data-based teacher development. *ELT Journal* (Oxford, UK), **52**, 4 (1998), 273-81.

This article describes how data from English Language Teaching (ELT) classroom research can be exploited in teacher development activities. The term 'data' here refers to descriptions of ELT lessons and interviews in which teachers talk about their work. The article begins by outlining the contribution data-based activities can make to teacher development, and presents a practical example to illustrate the principles underlying their design. A case is made for using such activities to facilitate teachers' growth as reflective practitioners, and to promote a more productive view of the relationship between research and teacher development in ELT.

99-494 Bucher-Poteaux, Nicole (IUFM d'Alsace, Université Louis Pasteur, Strasbourg, France). Inspection et évaluation. [Inspection and evaluation.] *Les Langues Modernes* (Paris, France), **3** (1998), 42-46.

This paper argues against the current situation in France where, it is claimed, teachers and inspectors exist in a state of constant misunderstanding. It suggests that there is a tendency towards a pedagogical norm created by the Inspectorate and other powerful institutions which is holding back advances towards a more

enlightened professional world. This world would include: an inspectorate composed of people coming from different perspectives, dealing with the problems of evaluation from complementary angles; a formative notion of teacher evaluation with observation of practice seen in a developmental light; an evaluation focused on the person as much as on the content and in which the teacher's own reflections would be respected; training which included peer observation and teamwork; encouragement of self-evaluation; and, finally, training for the whole educational community based on questioning and enquiry rather than on certitude.

99-495 Burton, Jill (U. of South Australia). Professionalism in language teaching. *Prospect* (Macquarie U., Sydney, Australia), **13**, 3 (1998), 24-34.

Professionalism is defined as teachers coming to an understanding of what teaching is for themselves, a process which requires personal reflection, discussion, collaboration and a sense of community. This process will result in the development of a discourse of teaching which is needed to unify all participants in education (teachers, managers, administrators), particularly as the aims of these participants may not always coincide. The article moves on to describe the Australian Adult Migrant English Program (AMEP) as an example of how professionalism can develop. It traces the transformation of a programme, which in the 1970s employed temporary teachers to work in rather isolated circumstances, into a large-scale, national, project-based programme committed to professional development through collaboration by the mid-1980s. It is suggested that overlap between management goals and teaching goals and the terms of reference employed to articulate these resulted in sustainable working practices which could withstand cuts in funding and thus threats to professionalism. It is concluded that encouraging teachers to reflect on their actions and values and promoting collaboration between managers and teachers in setting shared goals for an institution will result in the development of professionalism in education.

99-496 Caspari, D. (DFG-Stipendiatin, Gießen, Germany). Subjektive Theorien von Fremdsprachenlehrern/-innen - für Studierende ein relevantes Thema? Überlegungen zum Gegenstand und seiner methodischen Umsetzung im Rahmen eines fachdidaktischen Hauptseminars. [The subjective theories of foreign language teachers - a relevant topic for trainees? Some thoughts on the subject and its systematic application in a series of teacher-training seminars.] *Fremdsprachen Lehren und Lernen* (Tübingen, Germany), **27** (1998), 122-45

This article gives an account of a university course which the author devised (and delivered) for trainee teachers of French in Germany. The course was aimed at allowing students to explore the professional self-