



98-599 Wennerstrom, Ann (U. of Washington). Intonation as cohesion in academic discourse. *Studies in Second Language Acquisition* (New York), **20**, 1 (1998), 1-25.

This paper reports the results of a study of the intonation of 18 Mandarin Chinese speakers lecturing in English. As a basis for the study, it is proposed that intonation be considered a grammar of cohesion in English discourse: drawing from the intonational model of Pierrehumbert and Hirschbery (1990), it is argued that discrete morphemes of intonation correspond to the categories of cohesion in Halliday and Hasan's (1976) typology. The study investigated the hypothesis that the non-native speakers who were able to use the intonation system of English most effectively would score higher on a global language test. Using a Computerised Speech Lab to measure pitch, four aspects of intonation were averaged for each speaker: (a) the pitch difference between newly introduced content words and function words; (b) the use of high pitch at phrase boundaries to link related constituents; (c) the use of pitch to distinguish contrasting items from given items; and (d) the *paratone* or increase in pitch range at rhetorical junctures to signal topic shift. These four measures were chosen for their contribution to the cohesion of the lectures. Multiple regression analysis indicates that the fourth intonation variable, the *paratone*, was a significant predictor of these subjects' test scores. Examples are given of the other variables in context from both low- and high-scoring speakers. It is emphasised that intonation is not only a stylistic component of accent but also a meaning-bearing grammatical system.

98-600 Youngju, Han (Yongsan U. of International Affairs) **and Ellis, Rod**. Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research* (London), **2**, 1 (1998), 1-23.

This paper explores ways of measuring implicit and explicit second language knowledge and examines the relationship between these measures and measures of general language proficiency. Scores were obtained from a timed oral production test, a timed grammaticality judgement test (administered twice), a delayed grammaticality judgement test and an interview designed to tap metalingual knowledge, all of which focused on learners' knowledge of verb complementation in English. A factor analysis revealed a two-factor solution, reflecting a clear distinction between those measures that incorporated a time constraint (hypothesised to reflect implicit knowledge) and those that did not (hypothesised to tap explicit knowledge). Both factors were found to correlate with scores on the Secondary Level English Proficiency Test. However, only one measure of explicit knowledge (the Delayed Grammaticality Judgement test) was found to be significantly related to scores on the Test of English as a Foreign Language. The significance of these results for language teaching and testing is considered.

Language testing

98-601 Brindley, Geoff (Macquarie U., Sydney). Assessing listening abilities. *Annual Review of Applied Linguistics* (New York), **18**, (1998), 171-91.

Assessment of listening ability has received relatively limited coverage in the language testing literature. This paper begins by presenting current views of second language listening. Whilst drawing attention to the inadequacies of current models of listening comprehension, it identifies two common points of consensus in the literature on the nature of listening processes: the assumption by test developers that there are identifiable, hierarchical listening skills; and the move towards a more interactive model of the listening process. It then reviews various issues and challenges in the assessment of listening: assessing higher level skills and listening in oral interaction, the confounding of skills (which threatens the validity of listening tests), and dealing with authenticity (purpose, task, text). Aspects of task construction for assessment include factors affecting test performance, practical issues in listening test construction, and item formats. Some consideration is given to potential applications of new technology (video and computers), and some avenues for further research are suggested. There is both an annotated and extensive unannotated bibliography.

98-602 Brindley, Geoff (Macquarie U., Sydney). Outcomes-based assessment and reporting in language learning programmes: a review of the issues. *Language Testing* (London), **15**, 1 (1998), 45-85.

In recent years educational authorities in many countries have introduced outcomes-based assessment and reporting systems in the form of national standards, frameworks and benchmarks of various kinds which are used both for purposes of system accountability and for assessing individual progress and achievement in language learning. However, in some cases the introduction of these systems has proved problematic, owing to a number of political, technical and practical factors. These include the difficulty of combining formative assessment with summative reporting, the differing information requirements of different audiences, concerns about the validity and reliability of outcome statements, and the lack of appropriate resources to support implementation. This paper suggests that such problems may be able to be alleviated by closer consultation between policy-makers, administrators and practitioners, by undertaking further research into the validity and consistency of outcome statements, and by strengthening the links between assessment and reporting. A major investment in teacher professional development is deemed necessary if teachers are to be responsible for carrying out their own assessments. Ongoing research needs to be conducted into the effects of outcomes-based assessment and reporting on student learning.

98-603 Egbert, Maria and Maxim, Hiram (U. of Texas at Austin). Incorporating critical thinking and authenticity into Business German testing. *The Modern Language Journal* (Madison, WI), **82**, 1 (1998), 19–32.

For some time, foreign language proficiency theoreticians and practitioners as well as executives in international business have identified critical thinking as an essential skill. Despite this demand, the leading worldwide examinations for assessing oral and written communicative competence in Business German, the *Prüfung Wirtschaftdeutsch International* [International Business German Examination] (PWD) and the *Zertifikat Deutsch für den Beruf* [German Certificate for Professional Purposes] (ZDfB), test only students' information retrieval skills and fail to test their critical thinking abilities. This article proposes to integrate critical thinking and problem solving into the exams and to contextualise the tests' tasks in a more authentic business setting without compromising the PWD's and the ZDfB's content. These proposals should then be researched using qualitative as well as statistical methods in order to develop a stronger test. For each component of these exams, the article offers ways to include an assessment of the students' ability to reflect on, analyse, and critically evaluate business information. Parallels are drawn with the ACTFL (American Council on the Teaching of Foreign Languages) Proficiency Guidelines.

98-604 Glisan, Eileen W. and Foltz, David A. (Indiana U. of Pennsylvania). Assessing students' oral proficiency in an outcome-based curriculum: student performance and teacher intuitions. *The Modern Language Journal* (Madison, WI), **82**, 1 (1998), 1–18.

This article describes the results of a study conducted with two school districts in order to investigate (a) students OPI (Oral Proficiency Interview) ratings at the end of a typical four-year sequence of study, (b) the degree to which teachers could accurately predict their students' proficiency ratings, and (c) the relationship between classroom achievement and oral proficiency ratings. Findings indicate that (a) students at a given level of foreign language study demonstrate a range of OPI ratings; (b) a typical four-year sequence, for many students, is not sufficient time to attain Intermediate-Low oral proficiency; (c) teachers need some type of OPI training in order to predict accurately their students' OPI ratings using the ACTFL (American Council on the Teaching of Foreign Languages) Proficiency Guidelines; (d) in a proficiency-oriented curriculum there is a relationship between classroom achievement and students' OPI ratings; and (e) teachers of proficiency-oriented classes have some accurate intuitions regarding their students' oral progress and OPI ratings. The study is considered to provide valuable information for states and school districts as they assess student oral performance in an outcome-based or standards-driven curriculum.

98-605 Kuo, Jane and Jiang, Xixiang (American Grad. School of Internat. Management). Assessing the assessments: the OPI and the SOPI. *Foreign Language Annals* (New York), **30**, 4 (1997), 503–12.

This paper examines the two forms of oral proficiency testing currently in use in most language teaching programmes across the United States: the Oral Proficiency Interview (OPI) and the Simulated Oral Proficiency Interview (SOPI). The authors, who have extensive experience in conducting the OPI and in training SOPI raters, attempt to provide an objective assessment of the two tests by comparatively examining the test administration, response elicitation, and rating procedures. The paper also provides a concise but general overview of the two test instruments regarding their theoretical underpinnings, administrative support from their test developers, and the test structures and contents. While recognising that the two tests are both valid and reliable oral proficiency measurements, the authors have examined the characteristics that distinguish them as two different tests suitable for different purposes in different situations. Teachers are encouraged initially to identify the specific testing needs of their own institutions, and then to align the features of the two tests with these before deciding on a particular form of the test.

98-606 McNamara, Tim (U. of Melbourne). 'Interaction' in second language performance assessment: whose performance? *Applied Linguistics* (Oxford), **18**, 4 (1997), 446–66.

This paper addresses the notion of *interaction* within performance-based language assessment. Discussions of interaction in second language performance assessment have generally been loosely psychological in orientation, forming part of attempts to model the nature of communicative ability within the individual. But in investigating the validity of performance assessments involving interactions between individuals (for example, between candidate and interlocutor in speaking assessments), it is suggested that the intrinsically social nature of performance needs to be recognised. The author asks what the consequences might be for language testing research if it added a social perspective on the nature of interaction. The paper explores the necessity for such a re-orientation, and suggests areas that would feature in a consequent research agenda.

98-607 McNamara, Tim (U. of Melbourne). Policy and social considerations in language assessment. *Annual Review of Applied Linguistics* (New York), **18** (1998), 304–19.

There is increasing recognition of language testing as a social, value-laden and intrinsically political activity, and this paper summarises recent developments in the field. In Australia and North America, language assessment has played a significant role in government policy in the fields of education, vocational training and – notoriously – immigration. Evidence from a range of countries

shows the need for accountability on the part of language testers, particularly with regard to minority language students who are frequently disadvantaged by cultural bias and test constructs. Two professional codes of ethics – those of the International Language Testing Association (ILTA) and the Association of Language Testers in Europe (ALTE) – are examined. Both focus on practice, neglecting the possibility of a broader sociopolitical critique of testing. The emergence of critical language testing, whereby tests are contextualised with reference to their social contexts, is identified as an area of future research. Overall, it is anticipated that the impact of technology in testing will attract increasing attention, as will issues connected with the growing prominence of English as an International Language. Such issues are becoming ever more relevant with heightened awareness of the socially constructed nature of test performance and test score interpretation.

98-608 Ross, Steven (Kwansei Gakuin U.). Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing* (London), **15**, 1 (1998), 1–20.

Self-assessment has been widely used in language testing research, but has produced variable results. In many quarters it is considered a viable alternative to formal second language (L2) assessment for placement and criterion-referenced interpretations, although variation in self-assessment validity coefficients suggests potential difficulty in accurate interpretation. This article first summarises the research literature with the use of a formal meta-analysis conducted on 60 correlations reported in the L2 testing literature. These are the basis for estimates of median effect sizes for L2 speaking, listening, reading and writing tests. The second phase of the study is an empirical analysis of the validity of a self-assessment instrument. 236 'just-instructed' English as a Foreign Language learners completed self-assessment of functional English skills derived from instructional materials and from general proficiency criteria; their teachers also provided assessments of each of the learners. The criterion variable was an achievement test written to assess mastery of the just-completed course materials. Contrastive multiple regression analyses revealed differential validities for self-assessment compared to teacher assessment depending on the extent of learners' experience with the language skill self-assessed.

98-609 Shameem, Nikhat (U. of Auckland). Validating self-reported language proficiency by testing performance in an immigrant community: the Wellington Indo-Fijians. *Language Testing* (London), **15**, 1 (1998), 86–108.

The Wellington Indo-Fijians are recent immigrants to New Zealand, having arrived in the country after the 1987 Fiji military coups. A performance test was developed and implemented to validate self-reported first language proficiency of 35 teenagers in this immigrant community. The main drawback of self-report studies is

the likelihood of gathering inaccurate data, particularly if the first language, like Fiji Hindi in this community, is preliterate and perceived by the community members as a lower-status, less useful language than English. The performance test consisted of an oral interview, a listening comprehension test and a vocabulary test. The results of the performance test correlated strongly with the self-report data, thereby demonstrating the validity of the self-report scale. Significant differences between oral performances and self-reports as well as general trends in data suggested, however, that the respondents were often reporting their oral Fiji Hindi ability at a level higher than their judged level of performance.

98-610 Talbot, Dennis (City U. of Hong Kong). Testing an L2 reading strategy programme. *RELC Journal* (Singapore), **28**, 2 (1997), 128–41.

This paper discusses some of the difficulties associated with using standardised reading tests in assessing students' use of second language (L2) reading strategies for expository text in short-term intervention programmes. These tests normally attempt to capture gross distinctions, and are deemed unsuitable for showing differences over a relatively short intervention study. The author instead suggests the use of alternative approaches, including identifying omitted structural material, unscrambling texts, and identifying and correcting illogical texts. It is claimed that such testing approaches help to illustrate transfer across tasks with the same underlying structure but different surface formats.

98-611 Turner, Jean (Monterey Inst. of Internat. Studies). Assessing speaking. *Annual Review of Applied Linguistics* (New York), **18**, (1998), 192–207.

This review of oral proficiency assessment briefly outlines the nature of the early research in interview-format oral proficiency testing, before reporting on new directions in the investigation of the construct validity of interview-format tests and other types of oral skills tests through analyses of examinee, interviewer and rater performance. Discussion focuses in particular on whether examinee performance in interview-format proficiency tests resembles non-interview discourse – i.e. are such tests a measure of conversational skill, as is often assumed? – and on factors that affect examinee and rater performance. The paper also reviews research into the development of empirically-derived rating scales. The final section discusses the findings of several studies reporting on the development of oral skills tests for specific environments (particularly with regard to oral testing of children, and appropriate tests to measure communicative competence), and efforts directed towards integration of teaching and assessment.

98-612 van Elmpt, Marjolein and Loonen, Pieter (Rijksuniversiteit Groningen). Open vragen: antwoorden in de vreemde taal? [Open questions: answers in the foreign language?] *Toegepaste Taalwetenschap in Artikelen* (Amsterdam), **58** (1998), 149–54.

The new examinations for text comprehension at upper secondary level in the Netherlands will include open questions, besides the traditional multiple-choice ones. It has been provisionally suggested that these open questions should be answered in the first language (Dutch) rather than the second (L2). This paper questions whether the L2 is a serious handicap. In the research project reported, 67 Dutch students at pre-examination level were asked to take the same text comprehension test for English, 35 of them with questions and answers in Dutch, and 32 with questions and answers in English. The scores turned out to be roughly equal for the two groups. A similar (but much more limited) project for German showed considerable differences between the Dutch and the German groups.

98-613 Yi'an, Wu (Beijing Foreign Studies U.). What do tests of listening comprehension test? A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing* (London), **15**, 1 (1998), 21–44.

A test-taker's performance in a listening comprehension test is seen to be a function of at least two variables: his or her listening comprehension ability and the test method. To know how this trait-method unit works, it is necessary to look into the test-taking process of test-takers. Accordingly, employing an immediate retrospective verbal report procedure, a study was conducted among 10 Chinese EFL (English as a Foreign Language) university student test-takers. The test format investigated was multiple-choice. The results identified the participants' listening processes leading to comprehension and comprehension breakdowns, and captured the effect of the multiple-choice format on their test performance. The immediate retrospection research methodology as applied to the test of listening comprehension was explored.

Teacher education

98-614 Gray, John (Internat. House, Barcelona). The language learner as teacher: the use of interactive diaries in teacher training. *ELT Journal* (Oxford), **52**, 1 (1998), 29–37.

This article describes the piloting of interactive learner/trainee diaries as part of a short pre-service teacher training course. The outcome of the experiment suggests that the formal inclusion of adult learners in the training process provides trainees with access to essential information about the learning process, learner criteria for effective teaching, and the importance of

affective factors in ensuring a mutually rewarding teaching encounter. It is suggested that, by listening to learners' voices, trainees can begin to see their learners as educators with a valuable contribution to make to their development as teachers.

98-615 Ilyushina, Maria (British Council, St Petersburg). Perceived needs and wants in INSET provision – a study of the professional needs of teachers and the attitudes of trainers to INSET provision in St Petersburg, Russia. *English Language Teacher Education and Development (ELTED)* (U. of Warwick / U. of Birmingham), **3**, 1 (1997), 1–35.

This article reports on a study carried out into aspects of the needs of a selected group of recipients of teacher education in the St Petersburg area. First there is discussion of the place of English in Russian education, the status of teachers in Russian society and the current provision of pre-service and inservice (INSET) courses in the area. The author then focuses on the results of a questionnaire that formed part of the qualitative and quantitative research methods used in her survey. The questionnaire was designed to find out teachers' expectations of INSET: especially outcomes, the balance of theory and practice, content, methodology, course type, input modes and role allocation, and teacher trainers' expectations of INSET in the same areas as above. The results of the study point to the complexity of teachers' needs and also to a serious mismatch in perspective between teachers and trainers. The author feels that the teachers saw training predominantly in connection with transmission and accumulation of knowledge, especially procedural knowledge, whereas trainers, whilst being aware of this, felt that INSET should also deal with interpersonal and conceptual knowledge or change. The author also sees a mismatch between the declared views and the actual behaviour of teacher trainers, who, while theoretically being 'pro-reflection', worked predominantly in lecture mode.

98-616 Lynch, Tony (IALS, U. of Edinburgh). Classroom transcripts and 'noticing' in teacher education. *Edinburgh Working Papers in Applied Linguistics* (Edinburgh), **9** (1998), 42–51.

The focus of this paper is on raising teachers' awareness of what language learners do in and with the target language, rather than on the language produced by teachers in managing the classroom process. The author's interest in the potential of transcripts arises from an ongoing study of native/non-native talk in IALS (Institute for Applied Language Studies) speaking classes. The background to the study is briefly explained; the focus is on possible differences in student interactions involving the classroom assistant and the class tutor. Two extracts from transcripts of a particular type of group work are then presented: the data comprise audio-recordings of group work in two English for Academic Purposes classes. Finally, suggestions are made for ways of using classroom transcripts in pre-service and in-