Editorial

The theme of the previous special issue of *Prospect* 18,1 continues in the first article of this volume. The special issue looked at the unexpected outcomes that sometimes result from policy decisions, and how they can affect people and situations in unintended ways. This issue brought together several papers that examined the main theme from a variety of perspectives. In the first article of this volume, the topic is again addressed, this time in the context of secondary school policies in Britain. Leung and Creese examine the real impact of an educational policy which, while designed to embrace all school children, including those from ethnic and linguistic minorities in mainstream classes, does not necessarily do so. In investigating this topic, Leung and Creese focus on two specific areas of school policy – English as an additional language, and multiculturalism. The authors’ interest in this article is in examining the ways in which the discourses of the policy makers impact, and are interpreted by, the teachers in the classroom, and how this is reflected in their views of the children they teach. The authors argue that policy will be interactively interpreted by teachers through their own local communities of practice. In other words, the ways in which policy is interpreted by teachers will be influenced by the ideological situations in which they find themselves. The data used to explore these issues is drawn from two sets of semi-structured interviews with practicing teachers. These interviews, quoted extensively throughout the article, provide fascinating insights into the different ways in which policy may be interpreted by teachers both at a personal level, and at a professional level, and speak to the importance of professional development activities which explore the implications of policy at a local level.

The second article, by Woodward-Kron, also uses qualitative research methodologies, but this time to investigate the kinds of problems that arise for students who are required to write critical evaluations of journal articles for assessment in their undergraduate years. To investigate this question, 14 essays were selected for analysis that ranged across all pass level grades. The essays in the study were classified as falling into the Evaluative Account genre, and in this context, they were initially analysed for schematic structure. A more detailed analysis of a subset of the essays was then subjected to Appraisal theory analysis. The analyses revealed that only half of the essays were successful in terms of the schematic structure analysis, and many of the students did not appropriately critically analyse the articles. The difficulty the students appear to have with this type of assignment raises issues for the ways in which such assignments should be explained to the students, and demonstrates the importance of undertaking such detailed analyses of the discourse.
Issues in relation to writing and literacy are also the subject of the next article, by Gunn. It reports an action research project in which the author worked with a group of low literacy Horn of Africa women over a period of three terms. The group was formed after consultation with community members, and two bilingual aids were assigned to the class for one and a half hours a week, out of a total of the six hours weekly that the class met. The focus of the class was on developing the literacy skills of the participants, but the philosophy underlying the teaching approach – that of pursuing an opportunity, rather than making up a deficit – provided for a very positive force in the classroom. The paper documents the type and range of activities which were used in the classroom, outlining and discussing the approach adopted. We see, as well, the reactions of the learners as a result of interviews conducted with them, which reflect not only their positive attitude toward the class, but also their aspirations and hopes for the future. In addition, the teacher’s own reflective journal provides insight into the nature of the process for her.

The final paper, by Tercanlioglu, uses more quantitative methodologies in investigating the ways in which preservice language teachers themselves learn a language. The study uses survey techniques to explore the manner in which preservice teachers in Turkey approach learning a language. This is done with a view to enhancing our understanding of how these teachers think about themselves and their activities because of the ways in which their own orientations toward learning may affect their success in their future careers. The data is drawn from preservice teachers who are native speakers of Turkish learning English, and both differences across gender, and year level, are investigated. The value of the study lies in its detailed exploration of the variety of approaches to learning the teachers adopt, which identify the importance of educators recognising the range of preferences in learning style that their students many adopt.

This volume concludes with three book reviews. Cathie Elder reviews *The power of tests* by Elana Shohamy, which debates the ways in which language tests have been used, and in some cases abused. The second review, Kieran O’Loughlin’s *The equivalence of direct and semi-direct speaking tests*, is a detailed and measured account of a range of issues related to the testing of speaking skills, and explores the ways in which face-to-face interviews versus tape-based speaking elicitations are, or are not, equivalent.

Gaby Duigu reviews *The good grammar book* by Michael Swan and Catherine Walter, a pedagogical grammar which she recommends – it is an attractive book in terms of layout, its exercises are varied and it is generally well thought out.

GILLIAN WIGGLESWORTH
Book reviews

The power of tests: A critical approach to language testing

and

The equivalence of direct and semi-direct speaking tests

Reviewed by Catherine Elder

The power of tests is a volume dedicated to the study of what the author calls ‘use-oriented’ testing, or in other words, the decisions underlying the introduction of tests, the practices surrounding their delivery and the effects these tests have on the various stakeholders who either participate in some aspect of the testing enterprise or who are affected directly or indirectly by its results. She sets use-oriented testing in opposition to what she terms ‘traditional testing’ which she describes as a being solely focused on ‘the creation of tests which accurately measure knowledge of those tested’.

Traditional testing, she states:

... is not interested in the motives for introducing tests, in the intentions and rationale for using tests or in the examination of whether intentions were fulfilled. It is not interested in the steps taken in preparation for tests or in how test takers feel about tests. It is especially not interested in the consequences of tests and their effects on those who failed or succeeded in them. It also overlooks how the test affected knowledge, learning patterns and habits. Traditional testing views tests as isolated events, detached from people, society, motives, intentions, uses, impacts, effects and consequences. (p 4)

In her introduction to the volume, Shohamy tells us that the turning point on the road to her particular Damascus was the moment when she became aware of the gulf between the group into which she was socialised (that is, the traditional language testers concerned only with the ‘what’ of language testing) and the policy makers concerned with the ‘how’, that is, with using tests as a means of promoting particular political and educational agendas. It is these agendas, and the way tests are put to their service, which now intrigue her and she dwells particularly on the power of tests when placed in the hands of their users to divide and rule, to brand and disempower, and to constrain or distort what teachers teach and learners learn.

The volume begins with personal anecdotes from test takers regarding the powerful role of tests in their lives and goes on to sketch case studies of tests introduced into the school education system in Israel, and elsewhere,
which the author has researched over the years. The case studies include a test of Hebrew reading comprehension introduced in an attempt to monitor national achievements, a test of Arabic designed to raise the status of this language amongst native speakers of Hebrew and a test of EFL designed to bring about a greater emphasis on spoken English in the school curriculum. Shohamy documents not only the intentions underlying these state-mandated testing projects but also the complex effects they had on the educational settings into which they were introduced – effects which were often at odds with what was intended. Reference is also made to other examples of tests from around the world, which are categorised according to the particular political and bureaucratic agendas they serve. There follows an attempt to draw together and theorise the processes she has described in terms of Bourdieu’s notion of symbolic power, whereby tests are used to perpetuate existing structural inequities and dominated parties are co-opted into accepting and supporting decisions made by those in authority.

The final part of the volume presents an agenda for change involving, on the one hand, the application of a set of principles to guide an alternative testing paradigm under the rubric ‘critical testing’ and, on the other, a series of examples of more enlightened collaborative approaches to assessment involving dialogue and discussion between the various stakeholders involved in the assessment process. In addition to her own suggestions, Shohamy offers a compendium of existing codes of practice (including one recently formulated by the International Language Testers Association) and summarises recent debates on the responsibilities of language testers (here defined as anyone involved in the enterprise) and on the rights of test takers.

As Candlin notes in his introduction to the volume, a monograph focusing specifically on the uses and politics of language testing was long overdue. In the end, however, I felt the author did not do full justice to her topic. The picture painted in the first half of the volume is uniformly grim – a tale of missed opportunities, with no instances of test takers with positive experiences and no circumstances in which educational settings appear to have benefited in any way from the introduction of a state-mandated testing project. A more nuanced discussion of the issues would, in the end, have been more convincing. Shohamy makes few concessions to the efforts of those working in the so-called ‘traditional’ language testing paradigm to reflect on the purpose of their tests, to make explicit the nature of their constructs, to detect and control, in as far as possible, for test bias and to optimise the predictive power of their instruments. While these efforts may well have dominated the language testing agenda in recent decades at the expense of critical discussions of test use, they should nevertheless be considered integral, because they too have
to do with the inferences which are drawn from test scores and hence the uses which are made of them. They are an important aspect of the ‘Do it with care’ (p 61) advice offered by Shohamy towards the end of the volume, and should surely have been acknowledged, especially by one who herself has participated actively in such efforts over the years. What testers have failed to do effectively, and here I am in complete accord with Shohamy, is to communicate the importance and also the limitations of what they do to the policy makers – and this is one reason for the unfortunate rift between the ‘traditional’ and ‘use-oriented’ paradigms.

Shohamy’s proposals for reform at the end of the volume are imaginative but often utopian and curiously naïve. While the use of collaborative dialogic assessments has obvious potential in teaching and learning contexts, what of the other testing purposes? Would collaborative assessment of the kind the author proposes be appropriate in all situations (for example, to negotiate the content and standards to be applied in the context of, say, a high stakes language proficiency test for air traffic controllers)? Solutions such as ‘a test taker should be granted the possibility of being assessed by an alternative method other than the traditional “test-only” system. Such evidence can be used as counter evidence against decisions based on tests only’ (p 158) are offered. This sounds good, but which alternative methods would be acceptable, and in which situations? Who would decide? How would these very different methods be evaluated in the event of a dispute over, say, which of two students should accepted for a popular course with a quota on the number of enrolments? And to what extent would such a permissive approach be compatible with the feasibility standards mentioned on page 152?

In brief, Shohamy’s discussion of alternative ways of doing assessment fails to acknowledge the importance of assessment purpose and social context in decisions about assessment design. In addition, references to recent important work in the alternative assessment paradigm are disappointingly scant and one is left pretty much in the dark as to how the proposed new approaches to assessment would be validated.

In sum, the book is written so as to be accessible to a lay audience and offers a fluent and often stirring discussion of language test uses and abuses in different contexts, as well as some important and salutary reminders about the roles and responsibilities of testers and the powerful consequences of what they do. It is a valuable adjunct to the library of any test maker or user but it does not tell the whole story. Those seeking a more balanced and considered discussion of both the craft and conduct of language assessment of either the traditional or alternative variety must look elsewhere.

One place to look is at a recent volume in the series on language testing.
put out by UCLES/CUP. While O’Loughlin’s monograph bears the far less inspiring title *The equivalence of direct and semi-direct speaking tests*, it gives the reader an excellent insider perspective on what goes into the development and validation of a test of speaking ability. A direct test, as O’Loughlin explains, is one involving some form of face-to-face interview whereas a semi-direct test of speaking is one which, usually for efficiency reasons, does away with the interviewer in favour of a more cost-effective ‘non-human’ alternative such as a tape recorder or a printed booklet with a series of prompts which serve to elicit a speech sample from the test-takers. (An indirect test, in case you are curious, may not require the candidate to speak at all, but instead, for example, to tick a box on an answer sheet to indicate a word is pronounced differently from others in a given set). Why, you may ask, would we care about the equivalence of direct and semi-direct tests of speaking? The answer is indeed important because tests are powerful (as Shohamy is at pains to point out) and because in Australia in the 1990s the seemingly trivial issue of how a test was delivered, whether with a ‘live’ interlocutor or by speaking into a tape, might well have been critical in decisions made about whether or not an overseas applicant could migrate to the country. The context for O’Loughlin’s study is the access: test (Australian Assessment of Communicative English Skills) introduced in response to a testing policy by the Commonwealth Department of Immigration and Multicultural and Indigenous Affairs to ascertain the vocational English language competence of skilled immigrants before admitting them to entry to Australia. A four-skill test (with reading, writing, listening and speaking sub-components) was devised, and because one-on-one interviewing is expensive and unwieldy, it was decided to develop a parallel tape-based version of the test, which could be administered in a language laboratory. The ‘live’ format was to be used only at those offshore administration centres where language laboratory equipment was unavailable. O’Loughlin’s purpose in exploring the equivalence of these alternative methods of testing is to determine whether they do, in fact, measure the same kind of ability and whether this ability is measured with equal precision in each mode. Lack of equivalence would be a threat to construct validity (in the sense that the meaning of scores assigned to test takers might differ according to the method of testing), and also for test fairness in the event that some candidates’ performance was jeopardised by the fact of speaking to a tape rather than to a person (or vice versa).

This researcher’s approach to determining equivalence is exemplary in its thoroughness. Different forms of the test were administered in direct and semi-direct mode to a large sample of candidates and multiple methods of analysis were undertaken, starting with an examination of the test specifications.
of the live and taped version which were designed to match one another as closely as possible, and then proceeding to number quantitative analyses of score patterns to determine whether the two versions could be mapped on to a single unidimensional scale (and could therefore be regarded as statistically equivalent). The findings of the statistical analyses were somewhat ambiguous (that is they varied according to version and to the type of analysis undertaken) but nevertheless cast some doubt on the equivalence of the two testing modalities. O’Loughlin presses on to investigate the quality of the language produced by a sub-set of candidates, using both quantitative and qualitative discourse measures, one of which was devised expressly for this study. Both analyses revealed important differences in the quality of speech produced in the two modes, although one method was more powerful in detecting such differences. What O’Loughlin concludes is that the two versions were drawing on different skills overall, with the ‘live’ version producing a more natural, interactive type performance from the candidates. This may seem blindingly obvious, but note that in an attempt to maximise reliability and comparability of performance across versions, interviewers’ prompts in the live version were tightly scripted. Most of the speaking tasks elicited ‘one-way’ monologic speech, requiring minimal input from the interlocutor, who was trained to intervene as little as possible once the candidate had started speaking. Differences in candidate behaviour on the direct and semi-direct modes were therefore unexpected.

The next stage in the investigation was to look at testing processes in order to gain further insight into performance differences which had emerged across versions. This involved attending meetings and interviewing all stakeholders, including test development committee members, the test administrator, the raters, the test interlocutors and the candidates. Two test candidates were chosen for in-depth case studies and were observed during the test-taking encounter. One of these received startlingly different ratings according to the modality of testing and a number of reasons are offered as to why this might be the case, one being the candidate’s greater sense of ease in the tape-based mode, which could well be a matter of individual personality.

The conclusion which O’Loughlin arrives at, via this multifaceted validation process, is that spoken interaction involving more than one party is jointly constructed and hence fundamentally different in its character from communication with a machine. In some sense this is reassuring. No matter how much we attempt to constrain interviewers via training, they cannot refrain from giving feedback. Humans are relational beings and the discourse samples on the live version were littered with small reactive tokens such as hmm, yes, right which made a measurable difference to the character of the test
taker’s response. Attempts to get away with the more cost-effective, indirect, solutions to testing come at a price, resulting in somewhat crude approximations of human interaction. Of course this may be something that testers have to put up with for the sake of reliability and practicality, which are paramount in a high stakes selection test like this one, but O’Loughlin cautions against using the direct and semi-direct forms of the test interchangeably. The issue is not of course confined to access, and has implications for many other testing contexts, all the more so now that testers are experimenting with computer-based delivery and scoring.

The important message which this book conveys is the seriousness of this particular language tester’s attempts to address issues of construct validity and test fairness, his patient and intelligent engagement with multiple methods of analysis (all of which are meticulously described in this volume) and his unwillingness to settle for easy answers. To be sure, his account qualifies as ‘traditional’ in Shohamy’s terms because it focuses above all on the test properties, dealing only cursorily with the political agenda that the test served. The policy makers’ agenda in this instance, was to stem the unprecedented flow of skilled immigrants in a climate of economic recession and to reduce the cost of English language training in the adult migrant sector which had previously been offered gratis on the basis of need. Hawthorne’s (1997) test-external commentary on this situation is cited approvingly in Shohamy’s volume: ‘… where the measurement of language proficiency is clearly a pretext for achieving some broad political purpose, construct validation procedures, which are concerned with what the test purports to measure, may be an insufficient means of ensuring a test’s ethicality’ (p 58).

This point is certainly worth making, but construct validation, while insufficient, remains necessary. Even those ideologically opposed to the notion of gatekeeping, would surely concede that a test, if it must be used for this purpose, should be more than just a random measure and be designed and administered in such a way as to produce meaningful and accurate information about relevant skills. Making sure that this is so is an exacting process, and the efforts of language testers to achieve this and to make this process public as lucidly as O’Loughlin has done here should not be taken lightly.

REFERENCE
The good grammar book: A grammar practice book for elementary to lower-intermediate students of English (with answers)
Reviewed by Gabi Duigu

What constitutes a good grammar book? In this case, the answer from a publisher’s point of view is boldly presented in the form of a colourful booklet, as well as a glossy promotional sheet provided with the review copy of this book. They ask the question: ‘What makes you recommend one grammar book over another?’ The answers come in the shape of sound-bites, and in speech bubbles in the case of the glossy sheet: ‘big chunks are less easy to digest than little pieces’, ‘active practice is better for the brain than mechanical repetition’, ‘themed questions are good’, and ‘the colours of the rainbow are more pleasing to the eye than clouds of grey’.

Pedagogical grammars, like toothpaste, all have basically the same ingredients, and so marketing gurus now appear to be installed among the publishers of ESL books to ensure that their product reaches the very extensive and lucrative college market for course books. Appearance, brand recognition and catchy advertising are all essential in this enterprise. Here brand recognition is supplied by the venerable name of the publisher itself, but also by the use of authors who are well established, even if they also write for the opposition. (Swan and Walter are well known for their Cambridge series of course books).

The appearance of *The good grammar* book is indeed attractive: trendy colours for the cover, good clear fonts and good use of colour within the pages of the text. Illustrations are provided, but relatively discreetly, so that one does not feel that one is paying mostly for pictures rather than information, as is often the case with lower level textbooks.

I believe that the pedagogical suitability of a grammar can be assessed according to something like the following criteria:

**A  Suitability for students by level of English, as determined by:**

1 Grammar content:  
   a) choice of items  
   b) sequencing  
   c) clarity of explanations

2 Exercises:  
   a) for understanding  
   b) for practice/drift

**B  Suitability for students in terms of style of presentation, as determined by:**

1 Topical content of exercises:  
   a) interest-value/relevance to learners  
   b) cultural appropriacy
2 Realisation:  
   a) level of vocabulary employed for instructions and exercises  
   b) variety and type of exercises  
   c) quality of illustrations and other graphics  
   d) general layout.

How does The good grammar book score by these criteria?  
The choice of items is obviously constrained by the nature of English and of learners' language needs in general, and so one finds, as with other such books, that the early units deal with the verbs to be and to have followed by sections on the present, future, past and perfect tenses in that order. The verb system is also dealt with in the following four sections, (modals, passives, questions and negatives, infinitives and -ing forms, and special structures with verbs, in that order) before the articles and determiners are tackled. Interestingly, personal pronouns only make their appearance in Unit 13, followed by nouns and other parts of speech.

   While the content is as one would expect it to be, the sequencing is not critical in this case, since the items are presented in such a fashion that teachers and students can easily move around the book as the need arises. There is no deliberate build-up from one unit to the next, although the sentences do become rather longer and more complex as the units progress. Indeed, learners are advised in the introduction not to go through the book from beginning to end, but to select those sections which are necessary for them. The grammatical items are divided up in such a way as to cover either one complete page, or two facing pages and no more. At the top of each page, or two-page spread, the item being dealt with is presented in large, colour-coded, italic print, with sample sentences in some cases for ease of comprehension. Each section is preceded by a grammar summary in which the explanations are kept short and simple, with cross-references for more complex issues, and by a pre-test which helps the learners to decide whether the section is necessary for them. Additional explanations are given, as briefly as possible, in coloured boxes interspersed among the exercises on the following pages of each section. The explanations also draw attention to common errors, for example, We usually put to with infinitives, and, I want to go home.

   In pedagogical terms, therefore, the book seems eminently suitable, as one would indeed expect from such experienced writers. However, the style of presentation is often equally important when choosing a grammar book. In this case the British, as opposed to USA, provenance of the book is clear, and I must confess to a distinct preference for the former. Place names are
mostly British, but Canada and China and so on also get a look-in. Humour is introduced by the use of quotations (for example, the verb *to be* is accompanied by this quotation from Seamus Heaney: *Is there life before death?*) and cartoons, but without being intrusive. However, clearly the imagined users are young adults studying in Britain. References are to young adult preoccupations and occupations: socialising, travel, work and so on. Cultural references are also very Eurocentric and the people depicted in the illustrations and cartoons are mostly Caucasian with English names. Nevertheless, the sentences and short texts used in the exercises are rather more interesting than those found in other grammar books that remain determinedly neutral and bland (for example, *Because Jim was hot he stayed under the shade of a tree but Susan went back to work.* Who wants to say or write sentences like that?) They do reflect what people actually want to say (for example, *I hate getting up in the winter before the sun is up.*). There is more value in texts which are relatively authentic within one cultural sphere than those which attempt to be all things to all people by avoiding any specific context.

An online discussion group for ESL materials writers that I subscribe to recently dealt with two relevant issues: whether grammar exercises were better if they were contextualised, or used stand-alone sentences, and how to make low-level materials contain a spark of interest. Participants on the whole preferred short contextualised texts for grammar, and all agreed that low-level texts should employ humour, and should reflect a world with change and conflict, since this is what people normally want to talk about. *The good grammar* book makes little attempt to contextualise, and only occasionally has links between the sentences in the exercises. This does not seem to be a serious drawback, however, since it scores well in terms of the second criterion for maintaining interest. Actions, change, general knowledge quizzes and common social situations are all well represented.

A common fault in low-level textbooks is that the instructions and explanations utilise a vocabulary and sentence structure well above the linguistic level of the items being taught. Swan and Walter have done extremely well in giving short, succinct, but also clear and not over-simplified information and instructions. For example, on the page *by with passives* we find the following:

> We only use *by* … if it is really necessary. (80% of passive sentences are made without ‘by’…).

This is followed by an exercise for which the instruction is:

> Cross out the expression *in italics* if you feel it gives no useful information (that is, expressions beginning with *by*).
The exercises are varied, the layout is attractive and functional and the illustrations successful.

The good grammar book is indeed a good grammar book (publisher’s pun clearly intended). For the gain of one review copy I therefore do indeed recommend it, irritated though I may initially have been by the publicity material supplied.
Notes on contributors

**Angela Creese** is a lecturer in bilingualism in education and TEFL at the School of Education, University of Birmingham.

**Gabi Duigu** has been teaching ESL for nearly 40 years, in Africa, Asia and the Pacific, and most recently at the University of New South Wales Institute of Languages. She has published two books for IELTS preparation.

**Catherine Elder** is Associate Professor in the Department of Applied Language Studies and Linguistics at the University of Auckland. Her teaching and research interests span the areas of language assessment, language program evaluation, bilingualism and second language literacy. She is author of the *Dictionary of language testing* (Cambridge University Press, 1999) and coeditor of *Experimenting with uncertainty: Essays in honour of Alan Davies* (CUP, 2001) and the *Handbook of applied linguistics* (Blackwell, forthcoming).

**Margaret Gunn** is currently a lecturer at English Language Services, Adelaide Institute of TAFE in South Australia. She has worked in the Adult Migrant English Program in Victoria, Tasmania and South Australia. Her previous experience of preliterate learners was with the Hmong of Laos in the AMEP in Tasmania. Other research and teaching interests currently include second language acquisition, curriculum design, settlement support and community development. She has recently produced a series of readers and workbooks for new adult learners of English.

**Constant Leung** teaches at King's College, London. He has written and published widely on additional/second language education and language assessment issues both nationally and internationally. His most recent publications include *English as a second language in the mainstream: Teaching, learning and identity* (with B Mohan and C Davison, Longman 2001) and *Language and additional/second language issues for school education* (2002, NALDIC).

**Dr Leyla Tercanlioglu** is an Assistant Professor at Atatürk University, Erzurum, Turkey. Her main academic interests are foreign language teacher education, reading, L1 and L2 reading instruction.

**Robyn Woodward-Kron** has recently completed her doctoral studies in the Faculty of Education, University of Wollongong, Australia. Her research interests include academic literacy, learning through writing and genre. She has co-published with Elizabeth Thompson a CD-ROM writing guide for students, *Academic writing: A language based approach*. 