Annual Review of Applied Linguistics (2002) **22**, 221–242. Printed in the USA.
Copyright © 2002 Cambridge University Press 0267-1905/01 $9.50

12. DISCOURSE AND ASSESSMENT

Tim McNamara, Kathryn Hill, and Lynette May

Many contemporary currents in applied linguistics have favored discourse studies within assessment; there have been calls for cross-fertilization with other areas within applied linguistics, critiques of the positivist tradition within language testing research, and the growing impact of Conversation Analysis (CA) and sociocultural theory. This chapter focuses on the resulting increase in discourse-based studies of oral proficiency assessment techniques. These studies initially focused on the traditional oral proficiency interview but have since been extended to new test formats, including paired and group interaction. We discuss the research carried out on a number of factors in the assessment setting, including the roles of interlocutor, candidate, and rater, and the impact of tasks, task performance conditions, and rating criteria. Recent research has also concentrated more specifically on the assessment of pragmatic competence and on the applications of technology within the assessment of spoken language, including the comparability of semidirect and direct methods for such assessment and the use of computer corpora.

The Growth of Interest in Discourse Studies in Assessment Research

One of the encouraging features of research in language assessment in the last decade has been the impact of ideas and insights from outside the traditional ambit of language testing. This is no more so than in the area of discourse, particularly spoken discourse. If structural linguistics was the source of the views on language of the formative period of postwar language testing, best represented in the work of Lado (1961), then discourse analysis has taken its place in the assessment of oral language. As the communicative movement has found performance assessment to be its natural accompaniment, a feature first pointed out by Morrow (1979), so interest in the assessment of spoken language has grown. This interest first lighted on the practical, policy-driven tradition of oral language assessment represented by the Oral Proficiency Interview or OPI (Clark & Clifford, 1988). But discourse analysis was soon found to offer discomforting insights into the construct of language proficiency implicit within the OPI, and

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

more generally into the construct of proficiency as a projection of individual competence.

A number of intellectual currents in applied linguistics have favored discourse studies within assessment. The first is a general call for a greater openness to the possibility of cross-fertilization of research in language testing from research in other areas of applied linguistics, including second language acquisition (Bachman & Cohen, 1998; Young, 1995) and, specifically, discourse analysis (Jacoby & McNamara, 1999; McNamara, 1997; Shohamy, 1998; Young & He, 1998). The challenge to the positivist epistemology of applied linguistics generally (Pennycook, 2001) and of language testing in particular (Hamp-Lyons & Lynch, 1998; Lynch, 2001; Lynch & Hamp-Lyons, 1999; Shohamy, 2001) has favored non-quantitative research methodologies including discourse analysis and other interpretative methods (Halvari & Tarnanen, 1996). The increasing influence on applied linguistics of Conversation Analysis, triggered by the intellectual synergy between conversation analysts and applied linguists within the applied linguistics program at UCLA, has been a significant factor (Egbert, 1998; He, 1998; Kim & Suh, 1998; Lazaraton, 1992, 1997, in press; Schegloff, Koshik, Jacoby, & Olsher, this volume; Riggenbach, 1998). The current strong interest in Vygotskyan sociocultural theory and Activity Theory has also been an increasing influence on studies of joint activity in assessment settings, which has tended toward a close examination of the discourse of participants (Coughlan & Duff, 1994; Johnson, 2001; Swain, 2001; Young, 2001). A feature of all these developments has been their critical edge, challenging and disturbing the assumptions underlying current practice.

Discourse Analysis and the Validation of Oral Tests

In this section, research is reported involving various kinds of discourse analysis in the service of the validation of a range of oral tests. We begin where the work on discourse analysis in oral assessment began, with critical work on the familiar Oral Proficiency Interview. The role of discourse analysis in the definition of the construct of speaking in oral tests in general is then considered, including a short section on tests of pragmatics; then we discuss work on the impact of various facets of the assessment setting: interlocutor, task, performance conditions, test takers, and raters. Occasional reference will be made where relevant to the scant research that exists applying discourse analysis techniques to the validation of tests of other skills, especially writing.

The Oral Proficiency Interview (OPI)

Discourse analysis has been an effective and illuminating tool for the analysis of aspects of validity of Language Proficiency Interviews (LPIs), the most well known of which is the ACTFL Oral Proficiency Interview (OPI). Van Lier

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

(1989), in a now classic paper, noted a lack of research pertaining to the kind of speech event an OPI actually constituted, and concluded, "a comparison of OPIs and basic features of conversation reveals significant differences between these two modes of social interaction" (p. 505). Relying on techniques of Conversation Analysis, Lazaraton (1992) found that "the most that can be said about the question 'interview or conversation?' is that the encounters share features with conversations, but they are still characteristically instances of interviews, and interviews of a distinctive kind, for the participants" (p. 383). An important collection of papers on the topic of the OPI from a variety of discourse analytic perspectives appeared in the late 1990s (Young & He, 1998). In their introduction to this volume, He & Young (1998) provide a theoretical introduction to the issues in the volume. They stress the interactional nature of the event, which in some ways masks its institutional character, as participants (perhaps most significantly, test designers and assessors) may be led to think of it as a conversation, or perhaps an interview, without fully appreciating its institutional and culturally specific character. They suggest that a range of possible discourse analytic perspectives may be required to illuminate its nature. Egbert (1998), using CA techniques, compares OPI interactions involving American learners of German and an extensive corpus of naturally occurring conversation in German, and concluded that the OPI differed from conversation in its handling of the organization and means of repair. Moder and Halleck (1998), on the other hand, while conceding that the OPI is not an informal conversation because of its lack of interviewee topic nomination and control, nevertheless find no difference between native and nonnative speaker interviewees in their interpretation of the framing of the interviews. They suggest that the OPI can be viewed as an authentic instance of talk-in-interaction, a point which perhaps was not in dispute. Johnson (2000, 2001) is perhaps the most strident critic of the OPI. She concluded from a study of 35 OPIs conducted over the telephone that the OPI tests speaking ability in the context of a specific type of interview, rather than in the real-life context of a conversation, as claimed by its proponents. Similarly, Johnson and Tyler (1998), who analyzed an interview found on a training tape for OPI assessors, claimed that the OPI handles several important features of everyday conversation in specific ways that are untypical of everyday conversation.

It is not clear what we are to conclude from these findings. Clearly the OPI, as with any other test, is not 'natural'; it is constructed and contrived. The issue is whether the interpretations that we draw from it about ability to perform in nontest situations are sustainable. For this we need evidence. The contribution of the work on the OPI is to problematize the claims for validity it makes. This is in contrast to many of the proponents of the OPI over the years who have assumed it is automatically valid because of its surface resemblance to face-to-face conversational interaction, its 'directness.' The most interesting of the detailed studies reported below address the validity of interpretations of scores from oral assessment tasks, and make practical suggestions about changes—in interviewer

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

behavior, in rating scale development, in interlocutor and assessor training, in candidate preparation—that will improve the relative validity of test scores.

Meanwhile, those wishing to stay with the OPI have also used insights from discourse analysis of test performance to suggest modifications to the wording of its rating scales or accompanying materials. For example, Liskin-Gasparro (1996) analyzed the use of communication (lexical repair) strategies in the OPI by learners of Spanish, and concluded that the statements in the ACTFL Guidelines about communication strategies should be expanded to include strategies other than circumlocution, particularly to assist in defining differences between levels. In an attempt to address long-standing concerns about the validity and reliability of the OPI, Salaberry (2000) proposes a number of changes, including to the development and relative weighting of assessment criteria, based on a more explicit description of the criterion.

Definition of the Construct of Speaking in Oral Tests

Test development begins with the development of a view of the construct being measured. This will draw on current debates about the nature of oral test performance, which in turn are informed by insights from discourse analysis or other fields which themselves use discourse analysis. The revision of major tests such as the Test of Spoken English (TSE) (Douglas & Smith, 1997) and the Test of English as a Foreign Language (TOEFL) (Butler, Eignor, Jones, McNamara, & Suomi, 2000; Chapelle, Grabe, & Berns, 1997; Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 2000) has involved the articulation of relevant constructs, and insights from discourse analysis are a significant influence on the frameworks on which the revisions are based. The field of second language acquisition offers relevant discourse-based viewpoints. For example, Selinker (2001) points to the gap between speaking performance in testing and nontesting contexts to identify a possible 'test taking domain,' which presumably would constitute a problem for the validity of oral tests. Upshur and Turner (1999) argue that the combined insights of second language acquisition and language testing fall short of providing an adequate account of systematic effects in second language performance testing. They propose the need for a new paradigm, one that would consider the effect of test method on discourse, the effect of discourse on ratings and of task on assessment criteria. More generally, McNamara (1997) shows how developments in discourse analysis, particularly Conversation Analysis, provide fundamental challenges to the notion of oral test performance as being essentially a psycholinguistic phenomenon, and instead require us to face the interactive, social and thereby jointly constructed character of the performance. He cites the work of Coughlan and Duff (1994), discussed here, which problematizes the notion of oral test task, this time from the point of view of Activity Theory.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Interlocutor Effects

A greater understanding of the interactive nature of discourse has led to a detailed consideration of the role of the interlocutor in speaking tests and its impact on test validity. This line of research, building on a crucial early paper of van Lier (1989), has been influenced most clearly by Conversation Analysis in the work of graduates of the program at UCLA, particularly Lazaraton (1992, 1996, 1997, in press, this volume) and Ross (1992, 1995, 1996, Berwick & Ross, 1996, Ross & Berwick, 1992). This latter work has focused on the notion of accommodation between interviewer and candidate, including differences among interviewers such as cultural background and level of experience. Lazaraton (1996) conducted detailed discourse analysis of interviews from the Cambridge Assessment of Spoken English (CASE) and identified eight types of linguistic and interactional support offered to candidates by examiners, then speculated on the possible effects of such support on proficiency ratings. Lazaraton (1997) considers the significance of the reception by the interlocutor of self-deprecating remarks by the candidate in a context where the assessment is used to channel students into elective classes, and uses the CA notion of 'preference sequences' to identify the institutional character of the talk. Conversation Analysis has also been influential as a research tool in Australian research on this subject. Filipi (1994) looked at expansion sequences in a high stakes Italian language test for school leavers; Cafarella (1997) used CA techniques to reveal a range of interviewer support strategies in the same test. Brown and Hill (1998) were able to demonstrate that variability in interviewer style does, in fact, impact ratings and used discourse analysis to identify language behaviors which appeared to distinguish the 'easier' from the 'more difficult' interviewers. In an extension of this study Brown (in press) used a discourse analysis of the interaction in two interviews involving the same candidate with two different interviewers in order to investigate the interviewer's role in the construction of candidate proficiency. An analysis of raters' comments about the two test performances indicated that interviewing style led to different impressions of the candidate's ability. Both studies concluded that more emphasis needed to be placed on interviewer training.

One response to the issue of interviewer variability has been to prescribe the behavior of interlocutors more closely, by providing 'scripts' or suggesting a wording for individual prompts. This has been tried in the Cambridge Assessment of Spoken English (CASE). Lazaraton (1996) examined the adherence of interlocutors to such scripts, and found considerable variation among them; she recommends more attention to interlocutor training as part of assessor training (in those cases where the interviewer and assessor roles are combined in the one person).

Mohan (1998) explores a related phenomenon in the use of an interview as part of the assessment of international teaching assistants at a Canadian university.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Lumley and Brown (1996) identified interlocutor behaviors that hindered or helped candidate performance in medically related role-plays in the speaking subtest of the Australian Occupational English Test for immigrant health professionals. Katona (1998) investigated differences in interactions with unknown and familiar interlocutors, an important consideration when, for example, teachers are required to conduct interviews with members of their own classes.

The impact of several different categorical differences among interviewers in relation to candidates has been explored using discourse analytic techniques. Among the most important of these is the impact of gender. Berry (1997) found that personality factors (extraversion/introversion) and gender influenced the discourse produced in dyads of varying composition in terms of these factors in paired oral tasks (see below). However, O'Loughlin (in press) found no significant effect for gender on either discourse or scores, suggesting that multiple factors in the test setting were influencing candidate discourse. Another factor is native speaker status: Richards and Malvern (2000) found that what had previously been discovered about the extent of native speaker interviewer accommodation in oral interviews (e.g., in the work of Ross, cited above) held true in certain respects for nonnative speaking interviewers, although less delicately.

A particularly interesting and seriously underresearched situation arises in the case of a relatively innovative format for oral tests, the paired or group oral. In this type of assessment, candidates interact with each other rather than with a native speaker interlocutor and are observed by one or more assessors. Ikeda (1998) draws on Vygotskian theory to characterize the interaction in paired learner interviews in a Japanese classroom setting. Iwashita (1998) carried out a small-scale study of a paired oral assessment in Japanese, with candidates matched with higher proficiency and lower proficiency candidates, respectively. She found that being matched with a higher proficiency candidate generally led to the production of more talk, but that this did not necessarily have an impact on scores. Egyud and Glover (2001) provide brief examples of candidate discourse in a paired oral test in a secondary school setting in Hungary. The paired oral format is used extensively in the University of Cambridge Local Examinations Syndicate (UCLES) suite of examinations, but to date little of their research on the validity of this format has been published (UCLES, 2000). The research is said to include studies of the discourse generated in the paired interaction. An extension of the paired oral format is the group oral. Among the relatively few studies of this format are Fulcher (1996) and particularly Berry (1995, 1998). Berry (1995) shows that differences between learners in their approach to the task forming the basis of the group discussion influences the discourse they produce, and that raters are affected by these differences, although in ways that raise issues about the validity of the format. Berry (1998) examined the discourse of extravert and introvert students in a group oral, and found that the discourse varied according to this factor. She also discovered that the accommodative features of interaction of

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

native speaker interlocutors found in previous studies were present in the discourse of these learners interacting with each other.

Research on Tasks

Kormos (1999) uses discourse analysis in a comparison of interview and guided role play formats in oral assessments. She finds the role-play to have more of the features of casual conversation than the interview, with greater candidate control of topic introduction, topic maintenance, and topic shift. Riggenbach (1998) uses discourse analysis to suggest the advantages of tasks in which learners choose native speaking interlocutors outside the classroom and record their conversation in an informal setting. However, Spence-Brown (2001) uses interviews to reconstruct the process through which the tape recordings of such conversations are carried out as part of assessment in a university language course, and reveals the artifice and strategic sophistication underlying such apparently artless conversation. Her study will be seen as important for establishing the authenticity of tasks as a research field, with discourse analysis as an important tool in the investigation. In a study of EFL learners in Brazil, Ejzenberg (1995) reports an effect for task variation (cued and uncued dialogues and monologues) on the nature of the discourse produced and resulting fluency measures. Lazaraton and Wagner (1996) used discourse analysis to validate tasks within the Test of Spoken English (TSE) in terms of the functions they were intended to elicit, and found considerable support for the proposed functions in test discourse. Wigglesworth (1997) studied the impact on test discourse of differing types of tasks, particularly those containing an information gap. Tasks with and without an information gap were found to differ significantly in the type of discourse produced and the degree of interactivity.

Task Performance Conditions

An important line of research on tasks used in oral assessments stems from the work of Skehan (1998) on a cognitive approach to the characterization of task demands. This work, which originated in research on task-based learning, proposes that task structure and the conditions under which tasks are carried out can impact task difficulty. Research on this approach in a nontesting context has involved careful analysis of discourse produced with different task types and under varying performance conditions. Skehan (1998) has argued that the approach may be useful in constructing an interpretable scale of performance on oral tasks, based on characterizing performance in terms of task demands. In order to investigate the potential of the approach to help understand the impact of task demands under test conditions, Wigglesworth (1997) investigated the effect of planning time on oral test discourse, and found some impact on discourse measures of complexity and fluency, although not for subjects at all proficiency levels. Iwashita, McNamara, and Elder (2001) in a large scale study of this issue as part of the

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

development of a speaking test for New TOEFL found little support for the impact of the proposed dimensions of task difficulty in discourse on monologic narrative tasks elicited under test conditions. This finding may have implications for the development of the theory of task demands as a whole, and supports the call for increased collaboration between second language acquisition and language testing researchers.

Test Taker Characteristics

The view of oral test performance as interactive, so central to much current work, means that it is difficult to consider the impact of test taker characteristics in isolation from those of interlocutors. For example, the impact of gender in the work of O'Loughlin (in press), cited above, involves consideration of the gender of both participants in an oral assessment. Similarly, investigations of the impact of the cultural and linguistic background of testtakers only makes sense in terms of the match with or divergence from the background of their interlocutors. Some research has begun in this area, although studies comparing matched and unmatched interactions are yet to be done. Morton (1988) demonstrated evidence of cultural and linguistic background in test-taker discourse on a monologic narrative task, although there was no impact on test scores. The cross-cultural character of the encounter in an oral interaction assessment has been investigated by Berwick and Ross (1996), who analyzed the discourse of interactions between Japanese learners of ESL interacting with an American male, and Canadian learners of Japanese as a foreign language interacting with a Japanese male. They found differences in the two cross-cultural settings, and suggest that the issue of second language pragmatic competence as a potential feature of the assessment is raised by the analysis. Ross (1998) emphasizes that the expectations of extended (rather than minimal) candidate responses and of self-disclosure in ESL oral interviews with Japanese speaking subjects runs counter to Japanese cultural norms. Young and Halleck (1998) explore the impact of cultural norms in relation to interviews involving American-Korean, Mexican-American, and Japanese-American dyads. In a study of American students taking an oral exam in Korean, Davies (1998) argues that "the complex cross-cultural context of the face-to-face interaction creates competing role expectations for the participants" (p. 271). Kim and Suh (1998) examine the impact of speaker status on topic development in oral assessments in Korean. He (1998) similarly raises the issue of culture-specific conversational styles as one explanation for the performance of native speakers of Chinese in an English oral proficiency interview used with international teaching assistants.

Rating Criteria

Research on oral assessment has come increasingly to understand the attributes and behavior of raters as affecting test scores, and the way in which

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

rating scales and rating criteria, and their interpretation by raters, act as de facto test constructs in such assessments. Discourse analysis has proved to be a useful tool in investigating rater behavior and in assisting in the development of empirically-based rating scales.

A number of studies have used think-aloud methods used to study rater cognition in the assessment of second language oral performance (Brown, 2000; Brown, McNamara, Iwashita, & O'Hagan, 2001; Meiron, 1998; Pollitt & Murray, 1996). For example, Brown (2000) used protocol analysis to investigate what IELTS raters were paying attention to in the assessment of the IELTS oral interview, and its impact on the scores they gave. Such information can be used in rater training, or in revising or developing the wording of rating scales, and extensions of this work are now informing the development of rating scales for rating performance on cognitively complex speaking tasks as part of the proposed New TOEFL. Further studies on rater cognition in relation to writing scales are relevant here. Lumley (2000) used protocol analysis of raters scoring a test of written English for adult immigrants to Australia to develop insight into the rating process and the use of a rating scale. He demonstrated the fundamentally institutional character of the process, most clearly revealed when the usually automatic process of allocation of scripts to score categories was disturbed in the context of resolving some uncertainty in score allocation because of some particular feature of the script in question. Using retrospective written reports, introspective verbal reports, and group interviews with trained and untrained raters, Milanovic, Saville, and Shen (1996) were able to identify four distinct approaches to marking written compositions as well as a number of linguistic elements that influenced ratings. Turner (2000) explores discourse data from teachers engaged in a process of creating a rating scale for writing ability in Québec, and demonstrates how the actions of the participants and features of sample scripts influenced the criteria included in the scale.

Jacoby and McNamara (1999) draw attention to the significant gap between linguistically based assessment criteria typically used in performance assessments and criteria that are actually used by participants in the target use setting. These latter are often unconsciously held and applied, but can be revealed by close analysis of discourse in real-world settings where judgments of performance occur naturally, for example in oral feedback on performance. Jacoby (1998) refers to such locally motivated assessment occasions as 'indigenous assessment.' Jacoby and McNamara suggest they can be used as the basis for deriving relevant criteria for performance assessments, and give examples from research on feedback in a physics research team involving native and nonnative speaking participants. Interestingly, no clear distinction is made between the criteria used in feedback to the two groups, and linguistic features of performance are rarely commented upon and do not appear to be salient. Douglas (2001)

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

suggests that this issue of assessment criteria has problematic implications for the construct validity of many specific purpose language assessments.

Another approach to rating scale development involves close analysis of test taker discourse. Critical of the a priori, non-empirical basis of rating scales, Fulcher (1996) proposes a data-based approach to rating scale development, using discourse from oral performances on the ELTS test to construct a fluency rating scale.

Indirect Methods and the Investigation of Sociopragmatic Competence

Another site in which the insights of discourse analysis are having an impact of tests, and this time in tests used in research as well as in assessments of individual proficiency, is in the area of the assessment of interlanguage pragmatics. Investigation of interlanguage pragmatics has been conducted primarily with the aid of the Discourse Completion Task (DCT), which is usually presented in written form. A given social situation is presented, and the subject asked to choose what they would say in it. A thorough introduction to the format of these tests is provided in Hudson, Detmer, and Brown (1995). There has been much discussion of this format. Johnston, Kasper, and Ross (1998) demonstrated the sensitivity of data from questionnaires used to elicit subject responses to the format of the questionnaire. Billmyer and Varghese (2000) modified the format of the DCT to achieve different results. Hinkel (1997) cautions about the use of DCT for the investigation of sociopragmatic issues in Chinese. Yuan (2001) goes further: In an investigation of the validity of the written DCT in comparison with an oral version of the DCT, oral interviews, and field notes of actual usage, the written DCT was shown to be the least valid of the approaches.

Bisshop (1996) questions the validity of the DCT from another perspective: She elicited open-ended responses to stem items on apology from native speakers of English and nonnative speakers from Asian first language backgrounds. Then, after removing surface errors of syntax and lexical choice so that the native or nonnative speaker status of the respondents would not be immediately apparent, she presented a combined, randomized list of the responses to a further group of native speaker subjects and asked them to judge their acceptability and to offer reasons for their judgments. While the native speaker responses were reliably judged to be better than the nonnative speaker ones, the reasons given for why they were unsatisfactory entirely failed to support the rationale for the content of the DCT distracter items in the literature. Johnston, Kasper, and Ross (1998) claim their results highlight the need for different types of data and further validation studies. Elsewhere, in a review of different methods of data collection, Kasper (1998) concludes that differential, sequential, or triangulated combinations of data collection methods are usually called for.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Discourse Analysis and Classroom Assessment

Unfortunately, there has been a marked absence of studies using discourse analysis in the context of classroom assessment, a situation that matches the neglect of classroom assessment research in the field more generally. The notable exception, however, is the ongoing work of Rea-Dickins (Gardner & Rea-Dickins, 2001; Rea-Dickins, 2001; Rea-Dickins & Gardner, 2000) on formative assessment in the primary classroom in the context of the implementation of assessment requirements under the National Curriculum in England and Wales. In this work, careful transcription and analysis of classroom discourse is carried out to identify modes and functions of assessment in such classrooms in the immediate context of lessons. This work is exemplary in its extension of the concept of assessment beyond formal assessment occasions and its reconceptualization of the fact of assessment in primary school classrooms.

Discourse Analysis and Technology

This section explores the growing use of technology, both in the delivery of speaking tests, most commonly in semidirect formats such as the Simulated Oral Proficiency Interview or SOPI, and in the analysis of discourse that is used in the development of test material.

Semi-direct Tests of Speaking

The Simulated Oral Proficiency Interview (SOPI) has emerged in the last decade as a cheaper and more flexible alternative to the OPI. The equivalence of the two, and more generally the validity of the SOPI, relying as it does on monologic discourse in the absence of a 'live' interlocutor, has been an issue since Shohamy's seminal paper (Shohamy, 1994), which criticized the SOPI largely on discourse grounds. O'Loughlin (1995, 2001) has studied the issue intensively in the context of a different test, the *Access* test, used to award points for English language proficiency for certain categories of intending immigrants to Australia, where immigration places are allocated on a competitive basis. The test was offered in two formats, face-to-face and tape-mediated, and the insights of the Shohamy study were used at the design stage in an attempt to ensure their equivalence. Nevertheless, using a range of qualitative and quantitative studies, including analyses of lexical density of candidate discourse, O'Loughlin demonstrates that the two formats are not equivalent, testifying to the indelible impact on discourse of the existence of face-to-face interaction, no matter how muted that interaction is planned to be. O'Loughlin's finding that discourse elicited in semidirect formats is more formal, and more like written language than discourse elicited in face-to-face interaction, is supported by the findings of Koike (1998) on a Spanish SOPI. She found more organized and less conversation-like

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

discourse in this format. Malone (2000) used counts of discourse features to validate levels of performance on an English SOPI based on the ACTFL Guidelines, with mixed results. Luoma (1997) carried out a study similar to that of O'Loughlin in many ways, again using a range of methods, including analysis of test taker discourse, to establish similarities and differences between the direct and semidirect formats of an English test in Finland.

Testing Pragmatics Online

The widespread use of multimedia in the second language classroom has been accompanied by increasing interest in computer-based approaches to language test delivery. Roever (2001a, b) reports on a web-based test of ESL pragmalinguistic knowledge, operationalized in the test as knowledge of implicature, situational routines, and specified speech acts. Both qualitative and quantitative data from ESL/EFL students and a native speaker comparison group were used during the validation process.

The Use of Corpora

The use of corpora, of both spoken and written English, is an increasing feature of the design of language tests, particularly in academic purpose contexts. Biber (2001) describes a corpus of 2.5 million words of spoken and written academic discourse used in the development of the new TOEFL. At the University of Michigan, the MICASE (Michigan Corpus of Academic Spoken English) project constitutes a shared corpus resource for the development of listening and other tests of academic spoken English. The growing interest in the field is attested by the convening of a major symposium at the 2001 Language Testing Research Colloquium entitled *Using language corpora in language testing*. Rapid developments in Natural Language Processing (NLP) offer great promise for automatizing the assessment of spoken language: These developments rely on advances in the application of discourse analysis within the field of computational linguistics. Burstein, Kaplan, Rohen-Wolff, Zuckerman, and Lu (1999) summarize the potential of computer-based speech technology for the development of an assessment of spoken language within the new TOEFL. Applications of such technologies have already been made in some areas of written assessment, using automatic scoring, known as 'e-rater', as a second rater in the scoring of essays at ETS (Burstein, Marcu, Andreyev, & Chodorow, in press).

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

ANNOTATED BIBLIOGRAPHY

Brown, A., & Hill, K. (1998). Interviewer style and candidate performance in the IELTS oral interview. In S. Wood (Ed.), *IELTS Research Reports, 1*, 1–19.

This study explored the consequences of variation in interviewer style, within the context of IELTS interviews. By separating the role of interviewer and rater and using multiple ratings to remove rater effects, they were able to demonstrate that interviewer style can have a significant impact on candidate performance and hence on ratings of candidate ability. Through an analysis of candidate and interviewer discourse, Brown and Hill were able to identify distinct interviewer styles, which could be categorized as "easy" or "difficult." Both this and a subsequent study (Brown, in press) highlighted the need for more comprehensive rater training, and a deeper understanding of the co-constructed nature of interaction within the interview by the test developers.

Coughlan, P., & Duff, P. A. (1994). Same task, different activities: Analysis of an SLA task from an activity theory perspective. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian Approaches to Second Language Research* (pp. 173–193). Norwood, NJ: Ablex.

Though not strictly in the field of language testing, this thought provoking article uses Vygotsky's sociocultural theory of learning to question the belief that tasks can be regarded as a constants in second language acquisition research. The social construction of tasks is explored through discourse analysis, leading to the conclusion that the same task can be conceptualised and realized quite differently by different people. This has important implications for language testing.

Jacoby, S., & McNamara, T. (1999). Locating competence. *English for Specific Purposes*, *18*(3), 213–241.

This paper argues that the criteria used in the assessment of performance in specific purpose contexts should reflect criteria that are actually used in the setting in question. It proposes discourse analysis of naturally occurring instances of feedback on performance in such settings—termed 'indigenous assessment'—as a means of making these implicit criteria available for use in the design of formal assessments.

Johnson, M. (2001). *The art of non-conversation: A re-examination of the validity of the Oral Proficiency Interview*. New Haven, CT: Yale University Press.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Discourse analysis is a key tool for the investigation of the construct validity of the Oral Proficiency Interview (OPI) in Johnson's study. Johnson explores the nature of the speech event that constitutes the OPI, and proposes a prototypical model of the OPI communicative event. Her analysis is grounded in Vygotsky's sociocultural theory.

Lazaraton, A. (in press). *A qualitative approach to the validation of oral language tests. Studies in Language Testing 14.* Cambridge: Cambridge University Press/University of Cambridge Local Examination Syndicate.

Lazaraton's research over the last decade represents perhaps the most consistent application of discourse analysis to problems of oral language testing. This new book represents the most complete account of her work to date.

McNamara, T. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics*, *18,* 444–466.

This paper makes the case for the unsettling implications for the validity of oral assessments of views of performance implicit in current work in discourse analysis, particularly Conversation Analysis. The idea that test performance is 'co-constructed' challenges the focus of testing theory on the individual cognitive abilities central to current notions of language proficiency.

O'Loughlin, K. (2001). *The equivalence of direct and semi-direct speaking tests. Studies in Language Testing 13.* Cambridge: Cambridge University Press/University of Cambridge Local Examination Syndicate.

This study of two closely parallel versions of an oral test, one version delivered face-to-face and the other delivered on tape, demonstrates the lack of equivalence of the two formats. This has implications for the validity of semidirect tests, and attests to the co-constructed nature of test performance.

Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing, 11* (2), 99–123.

This classic article used discourse analysis as one means through which to explore aspects of the validity of semidirect and direct tests of oral proficiency. Her evidence strongly favors the direct format over the semi-direct one.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Young, R., & He, A.W. (Eds.). (1998). *Talking and testing: Discourse approaches to the assessment of oral proficiency*. Amsterdam: John Benjamins.

In this collection of fourteen articles, aspects of the validity of Language Proficiency Interviews (LPIs) are examined through discourse analysis. The focus of many of the studies is the extent to which LPIs exhibit the features of natural conversation. The findings of most studies in this volume add to the growing awareness within the field that there may be a specific oral test discourse, giving rise to concerns about validity.

OTHER REFERENCES

Bachman, L., & Cohen, A. (Eds.). (1998). *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press.

Berry, V. (1995). *A qualitative analysis of factors affecting learner performances in group oral tests*. Paper presented at the 17th Language Testing Research Colloquium, Long Beach, CA.

Berry, V. (1997, March). *Gender and personality as factors of interlocutor variability in oral performance tests*. Paper presented at 19th Language Testing Research Colloquium, Orlando, FL.

Berry, V. (1998, March). *Personality and oral test score variability*. Paper presented at TESOL '98 Conference, Seattle, WA.

Berwick, R., & Ross, S. (1996). Cross-cultural pragmatics in oral proficiency interview strategies. In M. Milanovic, & N. Saville (Eds.), *Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem* (pp. 34–54). Cambridge: Cambridge University Press and University of Cambridge Local Examinations Syndicate.

Biber, D. (2001, February). Academic discourse: Corpus-based perspectives. Plenary paper presented at the American Association for Applied Linguistics Annual Conference, St Louis, MO.

Billmyer, K., & Varghese, M. (2000). Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. *Applied Linguistics, 21,* 517–552.

Bisshop, C. (1996). *'I am apologize. . .': Asian speakers of English saying sorry in the Australian context*. Unpublished master's thesis, University of Melbourne, Australia.

Brown, A. (2000). An investigation of rater's orientation in awarding scores in the IELTS Interview. In R. Tulloch (Ed.), *IELTS Research Reports, 3,* 30–49.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Brown, A. (in press). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*.

Brown, A., McNamara, T., Iwashita, N., & O'Hagan, S. (2001). *Investigating raters' orientations in specific-purpose task-based oral assessment*. Final report to TOEFL 2000 Oversight Committee, Educational Testing Service, Princeton, NJ.

Burstein, J., Marcu, D., Andreyev, S., & Chodorow, M. (in press). Towards automatic classification of discourse elements in essays. *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, Toulouse, France.

Burstein, J., Kaplan, R. B., Rohen-Wolff, S., Zuckerman, L. E., & Lu, C. (1999). *A review of computer-based speech technology for TOEFL 2000. Monograph Series 13*, Princeton, NJ: Educational Testing Service.

Butler, F., Eignor, D., Jones, S., McNamara, T., & Suomi, B. (2000). TOEFL 2000 Speaking Framework: A Working Paper. *Monograph Series 20*, Princeton, NJ: Educational Testing Service.

Cafarella, C. (1997). Assessor accommodation in the V.C.E.Italian Oral Test. *Australian Review of Applied Linguistics, 20* (1), 21–41.

Chapelle, C., Grabe, W., & Berns, M. (1997). Communicative language proficiency: Definition and implications for TOEFL 2000. *Monograph Series 10,* Princeton, NJ: Educational Testing Service.

Clark, J. L., & Clifford, R. T. (1988). The FSI/ILR/ACTFL Proficiency scales and testing techniques: Development, current status, and needed research. *Studies in Second Language Acquisition, 10,* 129–147.

Davies. C. E. (1998). Maintaining American face in the Korean Oral Exam: Reflections on the power of cross-cultural context. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 271–296). Amsterdam: John Benjamins.

Douglas, D. (2001). Language for Specific Purposes assessment criteria: Where do they come from? *Language Testing, 18,* 171–185.

Douglas, D., & Smith, (1997). Theoretical underpinnings of the Test of Spoken English Revision Project. *Monograph Series 9*, Princeton, NJ: Educational Testing Service.

Egbert, M. M. (1998). Miscommunication in language proficiency interviews of first-year German Students: A comparison with natural conversation. In R. Young, & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 147–169). Amsterdam: John Benjamins.

Egyud, G., & Glover, P. (2001). Oral testing in pairs: A secondary school perspective. *ELT Journal 55*(1), 70–76.

Ejzenberg, R. (1995, March). *The role of task structure in oral fluency assessment.* Paper presented at TESOL '94 Conference and at the 17th Annual Meeting of the American Association for Applied Linguistics, Long Beach, CA.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Filipi, A. (1994). Interaction in an Italian oral test: The role of some expansion sequences. *Australian Review of Applied Linguistics, 11,* 119–136.

Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing, 13,* 23–49.

Gardner, S., & Rea-Dickins, P. (2001). Conglomeration or chameleon? Teachers' representations of language in the assessment of learners with English as an additional language. *Language Awareness, 10,* 2–3.

Halvari, A., & Tarnanen, M. (1996). Qualitative procedures for comparability—experiences from a multi-language, multi-lingual testing system. In A. Huhta, V. Kohonen, L. Kurki-Snonio, & S. Luoma (Eds.), *Current developments and alternatives in language assessment: Proceedings of Language Testing Research Colloquium 1996* (pp. 127–136). Jyväskylä: University of Jyväskylä.

Hamp-Lyons, L., & Lynch, B. K. (1998). Perspectives on validity: A historical analysis of language testing conference abstracts. In A. J. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th Language Testing Research Colloquium, Long Beach* (pp. 253–276). Mahwah, NJ: Lawrence Erlbaum.

He, A.W. (1998). Answering questions in LPIs: A case study. In R. Young, & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 101–116). Amsterdam: John Benjamins.

Hinkel, E. (1997). Appropriateness of advice: DCT and multiple choice data. *Applied Linguistics, 18* (1), 1–26.

Hudson, T., Detmer, E., & Brown, J. D. (1995). *Developing prototypic measures of cross-cultural pragmatics*. Technical Report #7, Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa, Honolulu, HI: University of Hawai'i Press.

Ikeda, K. (1998). The paired learner interview: A preliminary investigation applying Vygotskian insights. *Language, Culture and Curriculum, 11,* 71–96.

Iwashita, N. (1998). The validity of the paired interview in oral performance assessment. *Melbourne Papers in Language Testing, 5* (2), 51–65.

Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning, 51,* 401–436.

Jacoby, S. (1998). *Science as performance: Socializing scientific discourse through conference talk rehearsals*. Unpublished doctoral dissertation, University of California, Los Angeles.

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). TOEFL 2000 Framework: A Working Paper. *Monograph Series 16,* Princeton, NJ: Educational Testing Service.

Johnson, M. (2000). Interaction in the Oral Proficiency Interview: Problems of validity. *Pragmatics, 10*(2), 215–231.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Johnson, M., & Tyler, A. (1998). Re-analyzing the OPI: How much does it look like natural conversation? In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 27–51). Amsterdam: John Benjamins.

Johnston, B., Kasper, G., & Ross, S. (1998). Effect of rejoinders in production questionnaires. *Applied Linguistics, 19,* 157–182.

Kasper, G. (1998). Datenerhebungsverfahren in der Lernersprachenpragmatik [Data collection methods in interlanguage pragmatics]. *Zeitschrift fur Fremdsprachenforschung*, *9* (1), 85–118.

Katona, L. (1998). Meaning negotiation in the Hungarian Oral Proficiency Examination of English. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 239–267). Amsterdam: John Benjamins.

Kim, K., & Suh, K. (1998). Confirmation sequences as interactional resources in Korean Language Proficiency Interviews. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 297–332). Amsterdam: John Benjamins.

Koike, D. A. (1998). What happens when there's no one to talk to? Spanish Foreign Language discourse in Simulated Oral Proficiency Interviews. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 69–98). Amsterdam: John Benjamins.

Kormos, J. (1999). Simulating conversations in oral-proficiency assessment: a conversation analysis of role plays and non-scripted interviews in language exams. *Language Testing*, *16,* 163–188.

Lado, R. (1961). *Language testing*. London: Longman.

Lazaraton, A. (1992). *A conversation analysis of structure and interaction in the language interview*. Unpublished doctoral dissertation, University of California, Los Angeles.

Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing*, *13,* 151–172.

Lazaraton, A. (1997). Preference organization in Oral Proficiency Interviews: The case of language ability assessments. *Research on Language and Social Interaction, 30,* 53–72.

Lazaraton, A. (this volume). Quantitative and qualitative approaches to discourse analysis.

Lazaraton, A., & Wagner, (1996). The revised test of spoken English: Discourse analysis of native speaker and nonnative speaker data. *Monograph Series 7*, Princeton, NJ: Educational Testing Service.

Liskin-Gasparro, J. E. (1996). Circumlocution, communication strategies, and the ACTFL proficiency guidelines: An analysis of student discourse. *Foreign Language Annals, 29,* 317–330.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Lumley, T. (2000). *The process of the assessment of writing performance: The rater's perspective*. Unpublished doctoral dissertation, The University of Melbourne, Australia.

Lumley, T., & Brown, A. (1996). Specific-purpose language performance tests: Task and interaction. *Australian Review of Applied Linguistics, 13*, 105–136.

Luoma, S. (1997). *Comparability of a tape-mediated and a face-to-face test of speaking: A triangulation study*. Unpublished Licentiate thesis, Centre for Applied Language Studies, University of Jyväskylä, Jyväskylä, Finland.

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18,* 351–372.

Lynch, B. K., & Hamp-Lyons, L. (1999). Perspectives on research paradigms and validity: Tales from the Language Testing Research Colloquium. *Melbourne Papers in Language Testing, 8*, 57–93.

Malone, M. E. (2000). *The development of the English Speaking Test: An investigation of reliability and validity*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor, MI.

Meiron, B. (1998). *Rating oral proficiency tests: A triangulated study of rater thought processes*. Unpublished master's thesis, University of California, Los Angeles.

MICASE project (Michigan Corpus of Academic Spoken English). http://directory.umich.edu/dirsvcs-bin/search?search_text=micase

Milanovic, M., Saville, N., & Shen, S. (1996). A study of the decision-making behaviour of composition markers. In M. Milanovic & N. Saville (Eds.), *Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem* (pp. 92–114). Cambridge: Cambridge University Press and University of Cambridge Local Examinations Syndicate.

Moder, C., & Halleck, G. (1998.) Framing the language proficiency interview as a speech event: Native and non-native speakers' questions. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 117–146). Amsterdam: John Benjamins.

Mohan, B. (1998). Knowledge structures in Oral Proficiency Interviews for international teaching assistants. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 173–204). Amsterdam: John Benjamins.

Morrow, K. (1979). Communicative language testing: Revolution or evolution? In C.J. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp.143–157). Oxford: Oxford University Press.

Morton, J. (1988). A cross-cultural study of second language narrative discourse on an oral proficiency test. *Prospect, 13*(2), 20–35.

O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language Testing, 12,* 217–237.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

O'Loughlin, K. (in press). The impact of gender in oral proficiency testing. *Language Testing, 19, 2.*

Pennycook, A. (2001). *Critical applied linguistics: A critical introduction.* London: Routledge.

Pollitt, A., & Murray, N. L. (1996). What raters *really* pay attention to. In M. Milanovic & N. Saville (Eds.), *Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem* (pp. 74–91). Cambridge, UK: Cambridge University Press and University of Cambridge Local Examinations Syndicate.

Rea-Dickins, P. (2001). Mirror, mirror, on the wall: Identifying processes of classroom assessment. *Language Testing, 18,* 429–462.

Rea-Dickins, P., & Gardner, S. (2000). Snares or silver bullets: Disentangling the construct of formative assessment. *Language Testing, 17,* 215–241.

Richards, B. J., & Malvern, D. D. (2000). Accommodation in oral interviews between foreign language learners and teachers who are not native speakers. *Studia Linguistica, 54,* 260–271.

Riggenbach, H. (1998). Evaluating learner interactional skills: Conversation at the micro level. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 53–67). Amsterdam: John Benjamins.

Roever, C. (2001a, February). Validation of a web-based test of second language pragmatics. Paper presented at the American Association of Applied Linguistics Conference, St. Louis, MO.

Roever, C. (2001b). *A web-based test of interlanguage pragmalinguistic knowledge: Speech acts, routines, implicatures.* Unpublished doctoral dissertation, University of Hawai'i at Manoa.

Ross, S. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition, 14,* 159–76.

Ross, S. (1995). Aspects of communicative accommodation in oral proficiency interview discourse. Unpublished doctoral dissertation, University of Hawai'i at Manoa.

Ross, S. (1996). Formulae and inter-interviewer variation in oral proficiency interviewer discourse. *Prospect, 11,* 3–16.

Ross, S. (1998). Divergent frame interpretations in language proficiency interview interaction. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 333–353). Amsterdam: John Benjamins.

Ross, S., & Berwick, R. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition, 14,* 159–76.

Salaberry, R. (2000). Revising the revised format of the ACTFL oral proficiency interview. *Language Testing, 17,* 289–310.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Schegloff, E., Koshik, I., Jacoby, S., & Olsher, D. (this volume). Conversation analysis and applied linguistics.

Selinker, L. (2001, February). Speaking as performance within a discourse domain. Presentation at a joint symposium of Language Testing Research Colloquium and American Association for Applied Linguistics Annual Conference, *The Brahmin and the Elephant: Defining and assessing Speaking Ability*. St Louis, MO.

Shohamy, E. (1998). How can language testing and SLA benefit from each other? The case of discourse analysis. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp.156–176). Cambridge: Cambridge University Press.

Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing, 18,* 373–391.

Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics, 18,* 268–286.

Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing, 18,* 463–481.

Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn form test scores. *Language Testing, 18,* 275–302.

Turner, C. E. (2000). Listening to the voices of rating scale developers: Identifying salient features for second language performance assessment. *Canadian Modern Language Review/Revue Canadienne des langues vivantes, 56,* 555–584.

University of Cambridge Local Examinations Syndicate (UCLES) 2000: Investigating the paired speaking test format. UCLES research Notes (August 2000). URL: http://www.cambridge-efl.org/rs_notes2_7.cfm

Upshur, J. A., & Turner, C. E. (1999). Systematic effects in the rating of second-language speaking ability: Test method and learner discourse. *Language Testing, 16,* 82–111.

van Lier, L. (1989). Reeling, writhing, drawling, stretching and fainting in coils: Oral proficiency interviews as conversations. *TESOL Quarterly, 23,* 480–508.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14,* 85–106.

Young, R. (1995). Discontinuous interlanguage development and its implications for oral proficiency rating scales. *Applied Language Learning, 6* (1–2), 13–26.

Young, R. (2001, February). The role of speaking in discursive practice. Presentation at *The Brahmin and the Elephant: Defining and Assessing Speaking Ability,* a joint symposium of Language Testing Research Colloquium and American Association for Applied Linguistics Annual Conference, St Louis, MO.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press

Young, R., & Halleck, G. (1998). "Let them eat cake!" or how to avoid losing your head in cross-cultural conversations. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 355–382). Amsterdam: John Benjamins.

Yuan, Y. (2001). An inquiry into empirical pragmatics data-gathering methods: Written DCTs, oral DCTs, field notes, and natural conversations. *Journal of Pragmatics*, *33,* 271–292.

https://doi.org/10.1017/S0267190502000120 Published online by Cambridge University Press