'Tone' and assessing writing: applying Bernstein and Halliday to the problem of implicit value assumptions in test constructs

Johanna Motteram
The University of Adelaide

The value assumptions of language test constructs are inherently difficult to identify. While language test scoring descriptors refer to concepts such as tone and appropriateness, the role of socially determined value assumptions in written language assessment has not been adequately modelled or discussed. This paper presents a framework to link the results of analysis of written test scripts with the value assumptions of the pedagogic discourse of the test.

Key words: Tone, value assumptions, pedagogic device, assessing writing

Introduction

This paper introduces a theoretical framework for investigating the "value assumptions of a construct theory" (Messick, 1980, p. 1022) through analysis of test scripts. The framework was selected to investigate differences between medium and high scoring responses to an IELTS General Module Written Task One (GMWT1) prompt. The responses were collected from prospective candidates under exam conditions and were each rated by two senior IELTS markers. The initial focus of the study was to investigate differences in 'tone', an element of the public version of the scoring criteria for the GMWT1, between high and medium scoring scripts. Tone is not present in the scoring criteria for any of the three other IELTS written tasks (Academic Module tasks one and two, General Module task two) and tone, referring to written language, is not a technical term in linguistics. This paper contends that in the context of the scoring criteria of this high stakes language test 'tone' can be associated with

the implicit value assumptions of the test. Messick (1980) recognises that value assumptions and ideology have inherent links. The framework presented in this paper connects language choices made in candidate scripts and the ideological beliefs inherent in the construct theory of the test through application of two complementary theories, Bernstein's Pedagogic Device and Halliday's Systemic Functional Linguistics. First, the paper provides a review of discussion of tone, value assumptions and their role in validity in the language testing literature, then justifies the use of, and describes an application of the framework. This paper is intended as an introduction to the theoretical framework of the study. Future publications will present the analytical work with reference to the test scripts collected for the study.

Tone, Value Assumptions and Validity

As mentioned, 'tone' with reference to writing is not a technically defined term in linguistics. The use of the term 'tone' with reference to written language can be traced to I. A. Richards (1893–1979), a moral philosopher and one of the first lecturers in the English School at Cambridge, UK (Day, 2008). Use of the term 'tone' has evolved with trends in literary criticism. Evidence for this can be found in glossary entries in a multi-edition dictionary of literary terms (Abrams, 1965, 1971, 1981, 1993, 1999). Over time the entry for 'tone' has changed to include reference to persona and voice. However, the dominant theme of the definition has been to relate tone to implicit expressions of moral and social values. According to Abrams, tone in writing refers to how "the way a person speaks subtly reveals his concept of the social level, intelligence, and sensitivity of his auditor, his personal relation to him, and the stance he adopts toward him" (Abrams, 1971, p. 123). This use of the term clearly relates to implicit value assumptions in writing. Thus, the borrowing of the term from literary criticism and the history and use of the term in specialised discourse shows that the term 'tone' in scoring criteria can be linked to the value assumptions of the construct theory.

Indeed, the problematic use of the term tone was identified by Shaw and Falvey (2008) when they used scare marks when they first introduced the term in their report on the review of the IELTS writing rating scales (Shaw & Falvey, 2008, p. 38). Further, Shaw and Falvey (2008, p. 68) make the link between response appropriacy and tone. The inherent link between questions of appropriateness and the implicit value assumptions of the social group in question is clearly made by Fairclough (1992). Shaw and Falvey (2008, p. 38) also see the inclusion of tone in the scoring criteria as covering concerns about appropriacy and consistency of register, which refers to the way the social and material context

of use of language shapes the language selections made in the text (Hasan, 1995).

The value assumptions and ideologies of a construct theory are explicitly referred to in Messick's (1989) formulation of validity as a unitary notion, as reproduced in Table 1 below. The validity matrix identifies 'value implications' which is explained as referring to "What social and cultural values and assumptions underlie test constructs and hence the sense we make of scores?" by McNamara and Roever (2006, p. 14), at the intersection of test interpretation and the consequential basis of tests (or the "overt social context of testing" [McNamara & Roever, 2006, p. 14]). Messick's matrix highlights the importance of addressing the social contexts of testing, the social and cultural value assumptions and ideologies inherent in tests, and the social consequences of the use of tests as elemental in test validation.

Current trends in language test validation are moving away from Messick's unitary validity and towards the development of validity arguments (Chapelle, 2008; Chapelle, Enright & Jamieson, 2010; Kane, 1992, 2012). A significant limitation of validity arguments is that the current schemata for proposition and refutation of validity arguments do not include explicit reference to the implicit value implications of tests, nor do they include as a mandatory step investigation of the social consequences of test use. There is a danger that with a move towards validity arguments the language testing community's research gaze may be diverted from the influence of social and cultural value assumptions in language testing. It is of great importance, on the assumption that language use and learning are essentially social (Halliday, 1978; Halliday & Matthiessen, 2004), that questions related to the overt social context of language testing remain in view.

Table 1. Messick's validity matrix (Facets of Validity)

	Test Interpretation	Test Use
Evidential basis	Construct Validity	Construct Validity + Relevance/Utility
Consequential basis	Value implications	Social consequences

Adapted from Messick, 1989, p. 20.

The social consequences cell of Messick's unitary validity matrix has been the focus of the language testing community's response to the social bottom line of the matrix. The social consequences of language tests are usually identified and discussed in the literature as washback (e.g. Bailey 1999; Green 2006) and impact (e.g. Hawkey, 2006; Saville, 2009; Wall & Horak, 2006). Washback research describes the effects of a test on teaching and learning, and is usually associated with "test preparation or teaching to the test" (Green, 2006, p. 5) and

refers to teachers and learners doing things "*they would not necessarily otherwise do*" (Alderson & Wall, 1993, p. 117) as a result of the test. Impact research investigates "the influences of language programmes and/or tests on stakeholders, beyond the immediate learning programme context" (Hawkey, 2006, p. 7). Washback and impact research is then interested in the effects these tests have on teaching and learning after they have been introduced into a language teaching and learning environment. This research is achieved through pre- and post-test introduction studies (Wall & Alderson, 1993), comparative studies between test related and non-test related classrooms (Alderson & Hamp-Lyons, 1996; Brown, 1988; Hayes & Read, 2004), investigation of test preparation classroom discourse (Mickan & Motteram, 2008) and surveys of teacher and student perception and experience of language test preparation (Hawkey, 2006).

The washback/impact paradigm is the dominant research approach to the social bottom line of Messick's validity matrix. This means that the questions raised in the value implications quadrant related to implicit social and cultural value assumptions in language tests are largely ignored. It is accepted that the purview of washback research is the consequences of test use, not Messick's value implications. However, the focus on washback and impact leads to further limitations. Throughout washback research there is a tendency to describe the effects of test influence rather than to explain the processes which lead to the effects. Alderson and Wall (1993, p. 127) recognised the need to go beyond description and to "*account for* what occurs" as a result of test influence. In addition, there is an inclination to locate the test at the centre of the research; this positioning of the test limits engagement with other factors in teaching and learning which may be equally significant to the questions of the research, a limitation which Alderson and Wall (1993, p. 127) also recognised. With respect to impact, due to the expense of large scale investigations of the sort required for impact research, it is most often conducted either internal to (Banerjee, 2012; Hawkey, 2006) or with funding from the testing bodies which own the test (Wall & Horak, 2006). This funding relationship determines the focus of the research which is usually related to the expansionist business plans of the test owning body. Thus the impact research is not usually couched in critical frameworks, and not concerned with social and cultural questions. An exception to this, Moore, Stroupe and Mahony (2012), is considered later in this paper.

Critical approaches to language testing

Critical language testing is a branch of the language testing literature which addresses questions related to the social and cultural value assumptions of tests. It overlaps in part with the washback and impact literature as many researchers of washback and impact locate their work, at least implicitly (Shih, 2008), in Critical language testing. A progenitor of the Critical Language Testing movement is Shohamy, and her book 'The Power of Tests' (2001) is a catalyst for much discussion. In the book Shohamy sets forward 15 principles of the Critical Language Testing movement. These principles locate critical language testing practice within "cultural, social, political, educational and ideological agendas" (Shohamy, 2001, p.131), and state broad areas of questioning in which Critical Language Testing is interested. Particularly significant for a discussion of implicit value assumptions of language tests are these three:

- Critical testing calls for a need to question the purposes and actual uses of tests. Are they intended to measure and assess knowledge or are they used to define and dictate knowledge?
- Critical testing perceives testing as being caught up in an array of questions concerning education and social systems. The notion of 'just a test' is an impossibility because it is impossible to separate language testing from the many contexts in which it operates.
- Critical testing challenges the knowledge on which tests are based. Is that knowledge limited to those in power who are eager to preserve and maintain their power and knowledge, or is it a democratic representation of the multiple groups in society? (Shohamy, 2001, p. 132)

These statements of interest provide a starting point and locate the search for a framework to discuss implicit value assumptions of language tests in Critical language testing. However, Shohamy's Critical Language Testing does not provide practical guidance on how to conduct research within a CLT frame, especially research into the social and cultural value assumptions of a particular test task.

Guidance for researchers interested in framing their questions within CLT can be found in Lynch (2001). Lynch first reviews the Critical Theory of the Frankfurt School and then the use of Critical Theory to develop a research paradigm within the educational research literature. He then proposes, with reference to Pennycook's (2001) Critical Applied Linguistics and Shohamy's (2001) Critical Language Testing, four characteristics to inform research which

represents a critical approach to applied linguistics research and particularly to language assessment related research. These four characteristics are:

- An interest in particular domains such as gender, class, ethnicity, and the ways in which language and language-related issues (like all human relations and activities) are interconnected with them;
- The notion that our research needs to consider paradigms beyond the dominant post positivist-influenced one;
- A concern for changing the human and social world, not just describing it, i.e., 'the transformative agenda', with the related and motivational concern for social justice and equality; and
- The requirement that critical applied linguistics be self-reflexive. (Lynch, 2001, p. 357).

A study which attempted to operationalise Lynch's characteristics of Critical Language Testing found that the characteristics do not lend themselves to a straightforward application (Moore et al., 2012). Employing the characteristics in the research design and analysis stages lead to a massive project. This leads to the awareness that a Critical Language Testing study needs to be closely focussed if it is to address substantial questions. In spite of the difficulties Moore et al. (2012) met when applying Lynch's recommendations, this pursuit of a framework is supported by his recommended research characteristics. In particular it is influenced by Lynch's observation that attention to social class, along with ethnicity and gender, is an important area for Critical Language Testing research, and that implementation of innovative research frameworks is required to extend Critical Language Testing research.

Additional theorising on possibilities for research into values and assumptions in language testing comes from McNamara (2001, 2006). McNamara suggests that an initial step in addressing Messick's questions about the social and cultural value assumptions of test constructs could be taken by applying the notion of performativity (Butler, 1990, 1993, as cited in McNamara, 2001) to the understanding of language test constructs. In its original state, performativity explains the obscuration of the social construction of gender through the support of a belief that gendered behaviour represents an 'inner self' which exists of its own volition. Performativity then suggests that this male or female 'inner self' is a fiction which is performed into being through socially conditioned choices in behaviour.

Following this discussion, McNamara links the existence of a measurable construct of language proficiency or language competence to the existence of testing. He asks: "But what if…the act of testing itself constructs the notion of

language proficiency? Are the constructs of language testing susceptible to a performative unveiling, to be revealed as social constructs serving social ideologies?" (McNamara, 2001, p. 339). McNamara (2001, p. 239) states that to research this possibility would require the investigation of the "social origins and derivations of test constructs". According to McNamara (2001) this could be accomplished through questioning of the epistemological beliefs, especially as these relate to individuality and language performance, underlying the construction of the test constructs, and critical historical research of the origins of constructs. In addition, policy analysis which investigates the political uses of concepts such as language proficiency, empirical analysis of test performances for evidence of the socially constructed nature of language in use, and "studies of the institutionality of language testing" (McNamara, 2001, p. 340) are possible areas for research.

Shohamy, Lynch and McNamara's discussions and recommendations inform the location of this paper firmly in a Critical language testing paradigm.

Justification of Adoption of Framework

The framework presented in this paper connects the results of linguistic analysis and broad notions such as social and cultural value assumptions. As has been argued, present discussion in the language testing literature does not adequately address the inherent social and cultural value assumptions of language test constructs and the washback and impact literature does not provide an accessible model to describe the processes which cause the influence of tests on teaching, learning and the wider society. Adopting Bernstein's framework in conjunction with Halliday's grammar is valuable as it broadens the focus of the research beyond the immediate boundaries of teaching and learning directly related to the test, to the larger social realm, while at the same time offering a conceptual structure which connects with systematic analysis of texts.

The dialogue between the theories began in the 1960s, when Bernstein, Halliday and Hasan were working in London (Christie, 1999). Bernstein was working on a sociological framework which would make sense of a problem he was observing. He found that young students from working class homes were unable to achieve in the English language sections of gate keeping examinations, even though they were achieving high results in the mathematics sections (Christie, 1999; Sadovnik, 1995). Success in both the language and mathematics sections of the examinations was necessary for access to academic stream secondary education and subsequent higher education. Bernstein's

hypothesis was that working class children spoke, and wrote, a different 'code' of English from the code required by the examinations. The two codes were labelled 'restricted' and 'elaborated' codes, with elaborated code being the preferred code of the examinations (Bernstein, 2003). It is important to note that Bernstein's conception of 'code' was not equivalent to dialect (different ways of saying the same thing according to regional or social boundaries [Halliday, 2009, p. 429]) but that speakers of different codes had intrinsically different ways of understanding and representing the world and their place in it.

Bernstein's sociology was unusual for its time as he saw language as intrinsic to sociology and to explanations of social life. Similarly, Halliday, as a linguist, identifies the social as being centrally important to linguistics. This is in contrast to other influential schools of thought in linguistics, such as Chomskian linguistics or traditional grammar approaches. These theories focus on "their own object of study, isolating it from all else" (Hasan, 1999, p. 13). These endotropic theories do not allow connection between language and social life. By contrast, Halliday's linguistic model firmly situates language use in social context, with each instance of language use (or text) modelled as a realisation of a particular context of situation (Halliday, 1991; Hasan, 1995). Essentially, it is this paradigm of language as situated in social context which allows for a dialogue between Halliday's linguistics and Bernstein's sociological model. This complementarity between the theories is described by Hasan (1999) as meta-dialogism. Meta-dialogism allows movement across the boundaries of the two theories such that the results of systemic grammatical analysis of texts can be used as support for arguments about the sociological rules proposed in Bernstein's theories (Christie, 2007; Hasan, 1996, 1999).

In the 1960s, Halliday supported Bernstein's reasoning and considered that his linguistic theory was sympathetic to Bernstein's hypotheses around the use of different codes of language by different social groups and the subsequent disadvantage of those who do not have sufficient control of the preferred code in institutional settings (Halliday, 1995). However, at that time, Halliday's tools for linguistic analysis were not sufficiently developed to identify the differences in code between the institutionally preferred texts and the texts produced by working class children (Christie, 2007). Some years later, Hasan and Cloran, in their work on the language used by mothers from households of Low and High Autonomy Professions (Cloran, 1999; Hasan, 1996), and Williams (1999) in his work on adult child joint book reading, were able to develop tools of analysis which identified significant differences between the codes used in those households. In the study, Hasan, Cloran, and others investigated spontaneous conversation between mothers and their three and a half year old children at home. Over 20,000 messages were semantically analysed and from this data

clear patterns of differential patterns of meaning were identified. These were, in Halliday's words:

> consistent orientations to different ways of meaning, which construed boys and girls as different social beings. And the same study – same analysis, same data, same program – showed up other differences, equally significant both ideologically and statistically, between mothers from the working class and mothers from the middle class. (Webster, 2009, p. 429)

The work on the language used in Low and High Autonomy Profession households identified elements of Bernstein's restricted and elaborated codes, or in more theoretically and analytically developed terms, 'coding orientation' (Cloran, 2000; Halliday, 1995; Hasan, 1999), 'semantic variation' (Halliday, 1995) or 'semiotic style' (Hasan, 1996).

The theoretical link that was made by Hasan and her colleagues between language as instance (the data they collected in the form of mother/child talk in various households) and Bernstein's theory of coding orientation involved the adoption of the complementary theories; Halliday's Systemic Functional Linguistics and Bernstein's pedagogic device. The remainder of this paper will introduce Bernstein's pedagogic device and explain the application of the theory to a single task in a high stakes test of English language.
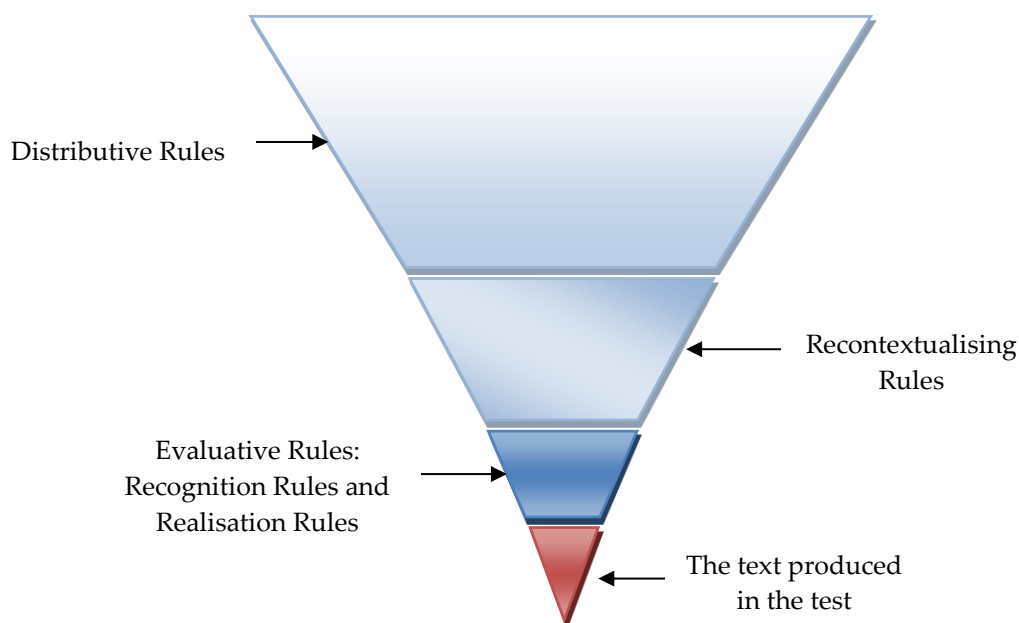
Bernstein's pedagogic device

Bernstein's pedagogic device is a theoretical model which explains the institutionally controlled link between social class determined distribution of material wealth and esoteric knowledge with the differently societally valued "habitual ways of being doing and saying" (Hasan, 1999, p. 15) of individuals (Bernstein, 2000). In the case of this study, the way of saying under investigation is a response to a language test prompt. While this may not represent a habitual text for the individuals who produced the responses for the study, the test prompt itself has been developed to characterise an authentic communicative task which is representative of habitual language use of the English speaking community (Cambridge English Language Assessment, n.d.).

Adopting Bernstein's framework provides a structure to understand the location of the IELTS test within a virtual institutional setting; the broad geographical reach of the test combined with the hardcopy and online publication of preparation textbooks, and the target of international reliability of written and spoken test results, has resulted in a global pedagogic discourse

related to the test. Bernstein's framework expresses through a series of hierarchical rules how that pedagogic discourse is produced. The rules explain how invisible societal forces regulate individuals' access to knowledge, how knowledge is selected to form disciplinary knowledge, and how that knowledge is evaluated. A simplified diagram of the rules can be found in figure one.

The diagram represents the rules as an inverted triangle. This emphasises the condensation of the implicit and explicit knowledge and skills which are required to produce a successful text in the test. The relationship between the rules is one of manifestation (Hasan, 1999). This means that the lower level represents how the level above is done; for example, the recontextualising rules represent the processes which enact the distributive rules, and the successful test text enacts the realisation rules. Because of this manifestation relationship, analysis can then move in the opposite direction. With sufficient data, trends in linguistic choices made in successful texts in a test can yield information about the recognition and realisation rules, recontextualising rules, and the distributive rules of the pedagogic discourse which coexists with the test.

Distributive Rules

Recontextualising Rules

Evaluative Rules: Recognition Rules and Realisation Rules

The text produced in the test

Figure 1. The system of rules of Bernstein's Pedagogic Device and the location of a test task with respect to the pedagogic device

The distributive rules

Bernstein identifies two types of knowledge in any society; the esoteric and the mundane (Bernstein, 2000). The distributive rules differentiate between the two

types of knowledge and determine who in the society will have access to the esoteric knowledge (with the assumption that the mundane will be available to all members of the society). The esoteric knowledge can be seen as the abstract and decontextualised knowledge of the society. In simple societies, access to esoteric knowledge is managed by the institution of religion. In modern societies, access to esoteric knowledge is largely managed by educational institutions. In the case of English language testing, the esoteric knowledge is managed by educational institutions, such as universities, and language test providers. The esoteric knowledge in language testing can be identified as both the knowledge of language as object of study and the knowledge of the technical aspects of language testing. Above the distributive rules are class relations which influence the distributive rules through the division of material wealth and esoteric knowledge.

The recontextualising rules

As mentioned earlier, the recontextualising rules are the social rules which do the work of the distributive Rules. Through the recontextualising rules the distribution of esoteric and mundane knowledge, to the groups which are permitted access, is achieved. The recontextualising rules thus create the pedagogic discourse. The recontextualising rules explain the processes through which knowledge is 'mpac re-packaged' and made available for learners. The construction of pedagogic discourse is explained as the movement of knowledge from its original context into a teaching and learning context. As the knowledge is relocated into the pedagogic discourse there is a process of recontextualisation. This process re-situates the knowledge as something to be learnt, rather than knowledge which exists. The process of recontextualising involves an opportunity to re-semiotise the knowledge by creating a new context to surround the knowledge. This new context has been likened to a space for ideology to play (Bernstein, 2000; Kell, 2000). In the case of English language teaching, learning, and testing this recontextualisation of language transforms language from a tool for intra- and inter-subject semiotic mediation into an object of study. This transformation is necessary for language testing as the process leads to the definition of aspects of language use into competencies which can then be identified and evaluated.

The evaluative rules

The evaluative rules are internal to individuals and consist of recognition rules and realisation rules. The recognition rules assist the individual to identify specific texts as being appropriate for different contexts, or to identify that different contexts require different texts. The realisation rules are the rules for the composition of the legitimate text of the context. The realisation rules will

vary with context and may include elements such as selection of appropriate topics of conversation, personal presentation, appropriate utterances or silences, and management of relationships with others in the context. In the case of IELTS general module written task one the legitimate text is restricted to words selected and written on a piece of paper.

Bernstein saw the evaluative rules as the condensation of all meaning of the pedagogic device. For this reason, in the diagram the test is represented as the tip of the inverted triangle, or the pointy end of the business. This image also recalls a funnel, which directs all fluid to the lowest point (John Knox, personal communication), in a similar manner to the way the pedagogic device condenses meaning in the evaluative rules.

Conclusion

Bernstein's framework for the construction of pedagogic discourse provides a structure for connecting the legitimate texts of the pedagogic discourse with the invisible social rules which determine what is included, and what is valued, in the pedagogic discourse. Those things which are included and valued are linked to "the value assumptions of the construct theory" and their "more subtle links to ideology" to which Messick (1980, p. 1022) referred. In the case of the IELTS general module written task one, the inclusion of the term 'tone' in the band descriptors is likely to be an oblique reference to these implicit value assumptions. The next stage of this project is an empirical study which employs the framework described in this paper so as to uncover the value assumptions of the test construct of the IELTS general module written task one in a systematic and justifiable manner. Comparison of the Systemic Functional Grammatical analysis of legitimate (and less legitimate) texts of the IELTS general module written task one will uncover those valued patterns of language choices, which are related to the ideologies of the test. These patterns will identify and describe elements of the test construct which are not explicitly classified in the public version of the band descriptors (Fulcher, 1996). This will in turn improve understanding of the role of hitherto neglected value assumptions in the assessment of writing.

Acknowledgements

reviewers for their criticism and comments on an earlier version of the paper. In addition, thanks to Dr Bill Winser for his significant assistance in revising the paper.

The Author

Johanna Motteram is a PhD candidate in the Discipline of Linguistics at The University of Adelaide. Her background is in English language teaching and Applied Linguistics. She is particularly interested in issues of test fairness.

References

Abrams, M. H. (1965). *A glossary of literary terms* (2nd ed.). New York: Holt, Rinehart and Winston Inc.

Abrams, M. H. (1971). *A glossary of literary terms* (3rd ed.). New York: Holt, Rinehart and Winston Inc.

Abrams, M. H. (1981). *A glossary of literary terms* (4th ed.). New York: Holt–Saunders International Editions.

Abrams, M. H. (1993). *A glossary of literary terms* (6th ed.). Fort Worth: Harcourt Brace College Publishers.

Abrams, M. H. (1999). *A glossary of literary terms* (7th ed.). Fort Worth: Harcourt Brace College Publishers.

Alderson, J. C., & Hamp–Lyons, L. (1996). TOEFL preparation courses: a study of washback. *Language Testing, 13*(3), 18.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115–129.

Banerjee, J. (2012). Introduction. *Research Notes, 50*, 2–3.

Bernstein, B. (2000). *Pedagogy, Symbolic Control and Identity:Theory, Research, Critique* (Revised Edition ed.): Rowman & Littlefield Publishers, Inc.

Bernstein, B. (2003). *Class, Codes and Control, Volume IV: The Structuring of Pedagogic Discourse* (Second ed. Vol. 4). London: Routledge.

Brown, J. D. H. (1988). An investigation into approaches to IELTS preparation, with particular focus on the Academic Writing compenent of the test. In S. Wood (Ed.), *IELTS Research Reports* (Vol. 1) (pp. 20–37). Sydney: ELICOS/IELTS Australia.

Cambridge English Language Assessment. (n.d.). *IELTS general training writing description*. Retrieved June 27, 2013, from https://www.teachers.cambridgeesol.org/ts/exams/academicandprofessional/ielts/generaltrainingwriting.

Chapelle, C. A. (2008). The TOEFL validity argument. In C. A. Chapelle, M. Enright & J. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 319–352). London: Routledge.

Chapelle, C. A., Enright, M., & Jamieson, J. (2010). Does an argument–based approach to validity make a difference? *Educational measurement: Issues and practice, 29*(1), 3–17.

Christie, F. (1999). Introduction. In F. Christie (Ed.), *Pedagogy and the shaping of consciousness* (pp. 1–9). London: Continuum.

Christie, F. (2007). Ongoing dialogue: functional linguistic and Bernsteinian sociological perspectives on education. In F. Christie & J. R. Martin (Eds.), *Language, knowledge and pedagogy* (pp. 3–13). London: Continuum.

Cloran, C. (1999). Contexts for learning. In F. Christie (Ed.), *Pedagogy and the Shaping of Consciousness* (pp. 31–65). London: Continuum.

Day, G. (2008). *Literary criticism: A new history*. Edinburgh: Edingburgh University Press Ltd.

Fairclough, N. (1992). The appropriacy of "appropriateness". In N. Fairclough, (Ed.), *Critical language awareness* (pp. 233–252). London: Routledge.

Fulcher, G. (1996). Does thick description lead to smart tests? A data–based approach to rating scale construction. *Language Testing, 13*(2), 208–238.

Green, A. (2006). *IELTS Washback in context: Preparation for academic writing in higher education* (1st ed.) (Vol. 25). Cambridge: Cambridge University Press.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.

Halliday, M. A. K. (1991). The notion of "context" in langauge education. In J. J. Webster, (Ed.), *Collected works of M. A. K. Halliday* (Vol. 9). London: Continuum.

Halliday, M. A. K. (1995). Language and the theory of codes. In A. R. Sadovnik, (Ed.), *Knowledge and pedagogy: The sociology of Basil Bernstein* (pp. 127–143). New Jersey: Ablex Publishing Corporation.

Halliday, M. A. K. (2009). Varieties and variation in language: Dialect, register, code. In J. Webster, (Ed.), *The essential Halliday* (pp. 427–450). London: Continuum.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). London: Hodder Education.

Hasan, R. (1995). The conception of context in text. In P. H. Fries & M. Gregory, (Eds.), *Discourse in society: Systemic functional perspectives* (pp. 183–283). Norwood, NJ: Ablex.

Hasan, R. (1996). The ontogenesis of ideology: an interpretation of mother–child talk. In C. Cloran, D. Butt & G. Williams, (Eds.), *Ways of saying, ways of meaning: Selected papers of Ruqaiya Hasan* (pp. 133–151). London: Cassell.

Hasan, R. (1999). Society, language and the mind: the meta–dialogism of Basil Bernstein's theory. In F. Christie, (Ed.), *Pedagogy and the shaping of consciousness* (pp. 10–30). London: Continuum.

Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000* (Vol. 24). Cambridge: Cambridge University Press.

Hayes, B. & Read, J. (2004). IELTS Test Preparation in New Zealand: Preparing Students for the IELTS Academic Module. In L. Cheng, Y. Watanabe & A. Curtis, (Eds.), *Washback in language testing: Research contexts and methods* (pp. 154–191). Lawrence Erlbaum and Associates.

Kane, M. T. (1992). An argument–based approach to validity. *Psychological Bulletin, 112*(3), 527–535.

Kane, M. T. (2012). Validating score interpretations and uses: Messick lecture, Language Testing Research Colloquium, Cambridge, April 2010. *Language Testing, 29*(1), 3–17.

Kell, C. (2000). Teaching letters – the recontextualisation of letter–writing practices in literacy classes for unschooled adults in South Africa. In D. Barton & N. Hall, (Eds.), *Letter writing as a social practice* (pp. 209–232). Amsterdam: John Benjamins Publishing Company.

McNamara, T. F. (2001). Language assessment as social practice: challenges for research. *Language Testing, 18*(4), 333–349.

McNamara, T. F. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly, 3*(1), 31–51.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*(11), 1012–1027.

Mickan, P. & Motteram, J. (2008). An ethnographic study of classroom instruction in an IELTS preparation program. In L. Taylor, (Ed.), *IELTS Research Reports* (Vol. 8). Canberra: IELTS Australia.

Moore, S. H., Stroupe, R. & Mahony, P. (2012). *Perceptions of IELTS in Cambodia: A case study of test impact in a small developing country*. IELTS Research Reports (Vol. 13). Manchester, UK/Melbourne, Australia: IDP: IELTS Australia and British Council.

Pennycook, A. (2001). *Critical applied linguistics: A critical introduction*. New Jersey: Lawrence Erlbaum.

Sadovnik, A. R. (1995). Bernstein's theory of pedagogic practice: A structuralist approach. In A. R. Sadovnik (Ed.), *Knowledge and pedagogy: The sociology of Basil Bernstein* (pp. 3–35). New Jersey: Ablex Publishing Corporation.

Shaw, S., & Falvey, P. (2008). *The IELTS writing assessment revision project: Towards a revised writing scale*. Cambridge: University of Cambridge ESOL Examinations.

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London: Longman/Pearson.

Wall, D. & Alderson, J. C. (1993). Examining washback: the Sri Lankan impact study. *Language Testing, 10*(1), 28.

Wall, D. & Horak, T. (2006). *The impact of changes in the TOEFL Examination on teaching and learning in Central and Eastern Europe: Phase 1*. Princeton: Educational Testing Services.

Webster, J. J. (Ed.). (2009). *The Essential Halliday*. London: Continuum.

Williams, G. (1999). The pedagogic device and the production of pedagogic discourse: a case example in early literacy education. In F. Christie, (Ed.), *Pedagogy and the shaping of consciousness* (pp. 88–122). London: Continuum.