

Language Testing

<http://ltj.sagepub.com>

Rethinking assessment from a critical perspective

Brian K. Lynch

Language Testing 2001; 18; 351

DOI: 10.1177/026553220101800403

The online version of this article can be found at:
<http://ltj.sagepub.com/cgi/content/abstract/18/4/351>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Language Testing* can be found at:

Email Alerts: <http://ltj.sagepub.com/cgi/alerts>

Subscriptions: <http://ltj.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.co.uk/journalsPermissions.nav>

Citations <http://ltj.sagepub.com/cgi/content/refs/18/4/351>

Rethinking assessment from a critical perspective

Brian K. Lynch *Portland State University*

This article examines language assessment from a critical perspective, defining critical in a manner similar to Pennycook (1999; 2001). I argue that alternative assessment, as distinct from testing, offers a partial response to the challenges presented by a critical perspective on language assessment. Shohamy's (1997; 1999; 2001) critical language testing (CLT) is discussed as an adequate response to the critical challenge. Ultimately, I argue that important ethical questions, along with other issues of validity, will be articulated differently from a critical perspective than they are in the more traditional approach to language assessment.

I Introduction

A rethinking of assessment from a critical perspective requires first of all that we agree on the meaning of critical. As I use it here, 'critical' has strong ties to the philosophical and social scientific tradition of Critical Theory. However, following Pennycook (1999; 2001), I adopt a critical perspective towards applied linguistics that differs from this tradition in significant ways. In order to appreciate this position fully, I first briefly summarize Critical Theory as well as its realization in the educational research literature. Next, I contrast it with a critical approach to applied linguistics as espoused by Pennycook. Finally, this perspective on applied linguistics is used to rethink language assessment.

II Critical Theory

Critical Theory is identified with the work of the Institute for Social Research, affiliated with the University of Frankfurt, usually referred to as the Frankfurt School. Inspired by the German workers' movement of the 1920s, Critical Theory has since undergone a few significant transformations. Rasmussen (1996: 11) gives an accessible summary of those transformations, and offers the following genealogy and generic definition for critical theory:

Address for correspondence: Brian K. Lynch, Portland State University, Department of Applied Linguistics, PO Box 751, Portland, Oregon 97202-0751, USA; email: bklynch@pdx.edu

Critical theory is a metaphor for a certain kind of theoretical orientation which owes its origin to Kant, Hegel, and Marx, its systematization to Horkheimer and his associates at the Institute for Social Research in Frankfurt, and its development to successors, particularly to the group led by Jurgen Habermas, who have sustained it under various redefinitions to the present day . . . Critical theory is a tool of reason which, when properly located in an historical group, can transform the world.

In general, then, critical theory strives to link reason with transformation, thought with emancipation. It has historically distinguished itself by attempting to address issues of political oppression.

With the rise of Nazism in Germany in the 1930s, and the experience of the Stalinist era in Russia, members of the Institute for Social Research moved from Frankfurt (some, like Horkheimer, Adorno and Marcuse ending up in Southern California) and began to distance themselves from the Marxist focus of earlier Critical Theory. There was a general disillusionment with the ideal of reason leading to a more enlightened world. This disillusionment led Horkheimer and Adorno (1972) to a critique of what was termed 'instrumental reason', implying that reason had become a tool of social control. That is, the emancipatory promise of the enlightenment, the potential for reason to provide a peaceful and just society, was seen to have failed. They termed this the 'dialectic of enlightenment', signalling that the modern project of enlightenment carried the seeds of its own destruction; it illuminates and destroys at the same time. The dilemma for critical theorists, then, was how to recover the emancipatory potential of reason. Horkheimer and Adorno ultimately retained their belief in reason's ability to function critically in order to challenge the oppressive orders of the day and lead to human emancipation. However, they rejected the notion that reason could ground itself in one particular framework or perspective, and in doing so foreshadowed postmodernism (Benhabib, 1996; Rasmussen, 1996).

Adorno, influenced by Walter Benjamin, finally embraced art as the embodiment of a non-instrumental reason that could be used to guide critical theory.

The modernity of art lies in its mimetic relation to a petrified and alienated reality. This, and not the denial of that mute reality, is what makes art speak. . . art internalizes the repressing principle, i.e., the unredeemed condition of the world, instead of merely airing futile protests against it. Art identifies and expresses the condition, thus anticipating its overcoming. (Adorno, 1984: 31 and 26, cited in Rasmussen, 1996. 29 and 30)

The primary development of Critical Theory, however, has come from Habermas (1971; 1987a; 1987b), who proposed a theory of communication that seeks to provide a non-instrumental theory of reason, but one which remains tied to modernity. He is critical of the work of Horkheimer and Adorno (and, by implication, that of

Weber), claiming that it represents an oversimplification and misstatement of the modern, Western perspective on reason and rationality (Rasmussen, 1996: 35). By making a distinction between strategic (presumably motivated by interests) and communicative (non-interested, non-instrumental) action, Habermas's theory of communicative action argues that rationality in a non-instrumental sense can be retained, and the emancipatory potential of critical theory can be recovered. Note that this is a different use of strategic and communicative than appears in the applied linguistics literature. Habermas's theory assumes a 'communicative community which is by nature predisposed to refrain from instrumental forms of domination'; that is, it assumes that 'the original form of discourse is emancipatory' (Rasmussen, 1996: 36). Basically, Habermas has moved critical theory from its earlier structural analysis of history (Marxist) to an analysis based on the philosophy of language.

Habermas (1971) acknowledges the interested nature of knowledge, of course, and this remains a crucial tenet of Critical Theory. However, some feel that recent, postmodern thought retains the duality of knowledge and interest, at the expense of any knowledge claims whatsoever. 'The stark opposition between knowledge and interest, which traditionally meant ignoring the latter to better pursue the former and recently seems to mean the converse, also has to be superseded by critical theory' (McCarthy, 1996: 346). Critical Theory attempts to retain the possibility of reasoned knowledge claims, rather than assuming that the only alternative to pretending that knowledge can be objective, separated from interests, is to give up on knowledge altogether and celebrate those interests.

In order to be non-instrumental, Habermas's approach to knowledge claims relies on group consensus, but does not use consensus as the means to establish validity. Instead, consensus is a by-product of verified knowledge claims. 'Habermas's discourse theory of validity is not meant to *define* either truth or moral rightness but to offer an account of what is involved in "redeeming" or *justifying* truth and rightness claims' (McCarthy, 1996: 359). McCarthy thus argues that Habermas is not taking an objectivist stance towards reality, i.e., that his theory is not ontological, but epistemological. 'Habermas's claim is that the idea of uncoerced, reasoned agreement is a pragmatic presupposition of certain types of discursive practice which are central to modern forms of life and to which we have no viable alternatives' (p. 365).

Another aspect of Critical Theory that remains essential for current theorists is the possibility of the universal. For critical theorists, like McCarthy (1996), no history of any sort (including Foucault's genealogy) can be written without an appeal to some sort of universal

truth. Furthermore, the rejection of a universal, normative ethics in favour of each individual's life being taken as a work of art, as articulated by Foucault (1997), is an:

inadequate ethical-political response to a world in which misery and injustice are rampant. . . We also have to investigate the social, economic, political, and cultural conditions that perpetuate misery and injustice and render the chances of making one's life into a work of art very different at different societal locations. (McCarthy, 1996: : 355–56).

An 'acceptable pluralism' that respects difference 'requires an overarching framework of justice, so that one group's well-being does not come at the expense of another's' (p. 359).

III Critical Theory closer to home: the educational research literature

In the educational research literature, Critical Theory has been realized in ways that are generally referred to as a research paradigm (Guba, 1990; 1994). Kincheloe and McLaren (1994: 139–40) summarize this critical paradigm as one which uses its work for the purposes of social and cultural criticism, believing that:

- all thought is fundamentally mediated by power relations that are social and historically situated;
- facts can never be isolated from the domain of values or removed from some form of ideological inscription;
- the relationship between concept and object and between signifier and signified is never stable or fixed and is often mediated by the social relations of capitalist production and consumption;
- certain groups in any society are privileged over others . . . and the oppression that characterizes contemporary societies is most forcefully reproduced when subordinates accept their social status as natural, necessary or inevitable;
- oppression has many faces and that focusing on only one at the expense of others (e.g., class oppression vs. racism) often elides the interconnections among them; and
- mainstream research practices are generally, although most often unwittingly, implicated in the reproduction of systems of class, race, and gender oppression.

This application of Critical Theory to educational research and practice differs from the preceding discussion in a manner that may give us pause. It may seem as if there is a political agenda going on here (especially with the neo-Marxist references to capitalist production and consumption) that is as oppressive as some of the forces it seeks to confront. We may agree with the general political, social

and cultural sentiment and goals but find this perspective 'utopian' in light of our previous professional experience. The questions that these characteristics cause us to pose, let alone the answers we may receive, may convince some of us that a critical perspective is not helpful or relevant to our work.

I am purposefully including myself in the 'us' of the previous paragraph. Although my sympathies and intellectual preferences line up well with the critical perspective, I am not a critical applied linguistics 'insider' at this point: to date, the majority of my own work would be categorized as 'mainstream' or postpositivist. (I use the term 'post positivist' to refer to the current, modified version of positivism that is the dominant research paradigm in applied linguistics; for a full explanation of the assumptions of this paradigm, see Phillips, 1990.) I would not want to misrepresent myself as an expert on this perspective, and I hope to avoid sounding evangelistic about it. Instead, this article is a portrait of how I see the critical perspective affecting and challenging language assessment. While a critical approach to language assessment is only one of many possible perspectives that can be adopted, this perspective needs to be explicitly argued for since it is at odds with the dominant research paradigm within which most of us were trained and whose assumptions continue to guide most of our work. Before moving to the 'rethinking' of language assessment, however, I present below a summary of Pennycook's current view of a critical approach to applied linguistics. I am aware that Pennycook is not the only voice of critical applied linguistics (see, for example, Tollefson, 1989; 1991; 1995; Fairclough, 1989; 1995; Phillipson, 1992; Auerbach, 1993; 1995; Pierce (Norton), 1995; Wodak, 1996; Morgan, 1998) but I believe that his most recent writing helps to provide a critical perspective that, while departing from certain aspects of Critical Theory, offers a valuable framework for our field in general and for language assessment in particular.

IV Critical applied linguistics

From his earlier works to the present, Pennycook has avoided the modernist aspect of Critical Theory: he has argued against replacing one rationalist critique with another. In his 1989 TESOL Quarterly article, 'The concept of method, interested knowledge, and the politics of language teaching', he outlined a similar position to that of other critical applied linguists (e.g., Fairclough, 1989; 1995) on the role of ideology in our research and understanding of language teaching and learning. However, that article also made reference to theoretical perspectives other than Critical Theory, such as feminism, Third World writers and postmodernists. In 1990 he defined 'critical applied

linguistics' as a linking of our professional practice with social, political and cultural concerns, including the exploration of 'the ways in which our work supports the increasingly sophisticated forms of physical, social and above all ideological coercion' (p. 9). This definition still fits reasonably well within a Critical Theory perspective.

Most recently, as guest editor of a special issue of *TESOL Quarterly* (1999, vol. 33, no. 3), Pennycook has summarized the characteristics that he feels define a critical approach:

- having an interest in particular domains such as gender, class, sexuality, race, ethnicity, culture, identity, politics, ideology and discourse; along with a resistance to normative responses to questions relevant to those domains;
- embracing a transformative pedagogy (and, by implication, transformative research practices); and
- taking a self-reflexive stance on critical theory.

In terms of the first characteristic, an interest in particular domains, Pennycook avoids the criticism that this approach follows a narrowly defined leftist political line but insists instead on a plurality of possible ways of responding to the central issues of inequality and oppression. Here he refers to later work of Foucault (1980: 190) as a guide for creating 'new schemas of politicization' (cited in Pennycook, 1999: 334). This is important because the emphasis on politics in critical applied linguistics is often confused with where an individual lines up on the political spectrum. Those of us who may consider ourselves to be leftists, or 'progressive' in our social and political beliefs, may feel that we are being unfairly judged as 'the enemy' if our research practices and beliefs do not line up with the critical perspective. (Of course there are plenty of examples of people in our field combining radical political stances with traditional, conservative research perspectives. Whether this is intellectually consistent and desirable or not is the subject for another article.) The 'avoidance of normative responses' also makes a related call for research paradigms (postmodernism, feminism, postcolonialism, constructivism) that offer alternatives to the dominant, positive-influenced approaches to research and practice.

In arguing for the 'transformative' agenda, Pennycook comes the closest to a version of critical applied linguistics that matches Critical Theory. At the same time he acknowledges the problems with critical pedagogy (e.g., its tendency to take a patronizing, dogmatic position in relation to people and the social, political and cultural issues they are struggling with). He also talks about transformation not so much in the sense of reason triumphing over false consciousness, but as 'ways of engaging with difference not merely in terms of inclusivity

and issues but also at the level of desire' (1999: 341). Again, he seems to be drawing on postmodern thought more than Critical Theory and its application in critical pedagogy for his approach to issues of oppression and injustice.

Finally, a critical applied linguistics is seen to require a constantly self-reflexive stance, explicitly in relation to Critical Theory. By self-reflexive Pennycook means an approach in which terms such as 'oppression', 'inequality', 'ideology' and 'empowerment' are examined for the ways in which they reveal particular understandings of the world. I would add that these terms might or might not have the shared meanings that are often assumed by those who use them. This sense of self-reflexiveness is also tied to the avoidance of critical pedagogy's problems and the essentially modernist aspects of Critical Theory. Pennycook, following others such as Dean (1994) feels that Critical Theory replaces one modernist rationalism with another.

This view of critical theory touches on the concerns I have raised about a form of modernist-emancipatory politics that seeks to make people more aware of the truth of their condition: There is a problem if critical theory only offers a rationalist account of social conditions that is supposed to supplant a possibly rationalist account (an understanding obscured by ideology). A critical approach that claims only to emancipate people through a greater awareness of their conditions is both arrogant and doomed to failure. As the discussion of engagement suggests, a more plausible way forward is through a critical engagement with people's wishes, desires, and histories, that is, a way of thinking that pushes one constantly to question rather than to pontificate. (Pennycook, 1999: 343)

Drawing upon the most recent work of Pennycook (1999; 2001), then, I would offer the following characteristics for a critical approach to applied linguistics:

- 1) 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;
- 2) the notion that our research needs to consider paradigms beyond the dominant, postpositivist-influenced one;
- 3) 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
- 4) the requirement that critical applied linguistics be self-reflexive.

This position differs from that of Critical Theory in terms of the second and fourth characteristics. The research paradigm issue differentiates this articulation of a critical approach from Critical Theory in its resistance to constructing a grand theory of rationality and consciousness as the basis for social transformation. It is here that the

critical perspective as defined by Pennycook overlaps with postmodernism (see also Lather, 1991) and responds to the critique by Guba (1990) that Critical Theory is, at the ontological level at least, grounded in the same critical realist philosophy that defines the current version of positivism (i.e., postpositivism). The self-reflexive characteristic acts to ensure that the approach does not result in merely replacing one research paradigm with another. This is a key element of Pennycook's critical approach in that it allows for some theoretical grounding without committing to a fixed theoretical framework. There are advantages and disadvantages to this strategy, of course. Fixed theoretical frameworks allow researchers access to a common language and agreed-upon assumptions that can assist in the refinement of our understanding. The disadvantages come when adherence to the fixed frameworks prevents us from being open to new perspectives that could deepen our understanding.

V Assessment and testing

In order to discuss the implications of a critical perspective for language assessment, the term assessment needs to be defined. It has tended to be used as either a synonym for testing, a synonym for evaluation, or has signalled a broader collection of measurement techniques. As I use the term here, assessment will refer to the systematic gathering of information for the purposes of making decisions or judgements about individuals. Its relationship to testing and measurement is depicted in Figure 1. Assessment, in this conceptualization, is the superordinate term for a range of procedures that includes measurement and testing but is not restricted to these forms. That is, at times the systematic information we gather in order to make decisions about individuals comes from tests or other measurement procedures. At other times, however, we gather systematic information in a non-quantitative procedure, and we use that information to make decisions about individuals without quantifying it. An example of non-measurement, non-testing assessment would be portfolio assessment, when results are reported in the form of a qualitative profile, rather than a set of scores.

1 Alternative assessment

Given the definition of assessment that allows for non-measurement approaches to the gathering of information for decision making about individuals, the term 'alternative assessment' needs to be clarified for the purposes of this discussion. Figure 1 is not meant to imply that all characteristics of all forms of assessment are the same as those of

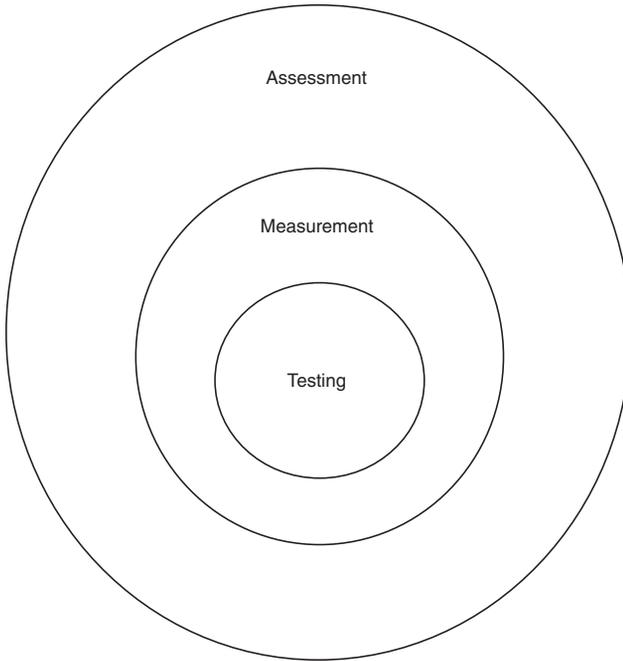


Figure 1 Assessment, measurement and testing

testing. What unites measurement and testing under the assessment umbrella is the notion of the systematic use of information for making decisions and judgements about individuals. This is not to say that the ways in which those decisions are made – what will count as reliable and valid information – will be the same for all forms of assessment. The outer circle in Figure 1 – non-measurement/non-testing forms of assessment – is meant to signal forms of assessment that have a different basis for determining the validity of decisions and judgements.

If all forms of assessment were united by the same requirements for reliability and validity, then I would agree with Brown and Hudson (1998) who prefer the term ‘alternatives in assessment’ (which they credit to John Norris; see Norris *et al.*, 1998). They make the important criticism that proponents of what has been called ‘alternative assessment’, such as Huerta-Macias (1995), must provide evidence for the validity of inferences made from these assessment procedures, rather than claiming that they are somehow inherently valid and trustworthy. However, most proponents of alternative assessment would not agree with, and have never argued for, its

exemption from questions of validity (see, for example, Wolf *et al.*, 1991; Shepard, 1993; Moss, 1994; 1996; Birenbaum, 1996). What these researchers do argue for is a recognition that there are different validity requirements for different approaches to assessment (when these approaches are viewed as more than different techniques).

In order to understand the meaning of alternative assessment and its potential to contribute, along with testing, to our ability to make informed decisions and judgements about individual language ability, we need to allow for different research paradigms. In this sense, alternative assessment is meant to describe something more than just procedures and methods. It acknowledges needs for assessment that fall outside of the traditional testing approach and its research paradigm. Alternative assessment signals a research paradigm, a 'culture', that differs from traditional, testing culture.

The cultural metaphor for alternative assessment helps us to understand how it might respond to the questions posed by a critical approach to applied linguistics. As described by Wolf *et al.* (1991) and Birenbaum (1996), assessment culture can be characterized by the following propositions:

- Teaching and assessment practices should be considered as integral.
- Students should be active participants in the process of developing assessment procedures, including the criteria and standards by which performances are judged.
- Both the process and the product of the assessment tasks should be evaluated.
- Reporting of assessment results should usually be in the form of a qualitative profile rather than a single score or other quantification.

At first glance, these propositions may seem amenable to traditional testing culture as well. However, within the testing perspective they would be seen as desirable properties if, and only if, the postpositivist, psychometric requirements with respect to reliability and validity – defined by traditional psychometric characteristics such as inter-rater reliability, objectivity and construct generalizability – can be met. When attempting to make use of the alternative assessment paradigm, we are not seeking measurement and tests as the means for providing the evidence we need to make decisions. And since, within this paradigm, we are not conceiving of that which we want to assess as an independently existing entity to be measured, the traditional criteria of reliability, generalizability and objectivity will not be the ones we need to make judgements about validity (see Moss, 1994; 1996). This is not to say that measurement and tests should be thrown out – I

am not arguing for the alternative assessment paradigm to replace testing – but that when we choose to use this approach, we need to judge its validity with criteria that are appropriate to its underlying paradigmatic assumptions.

Returning to the relationships between assessment, measurement and testing depicted in Figure 1, assessment would be all activity – not just ‘alternative assessment’ – involved in making decisions and interpretations about individuals, in this case about their language ability. Language testing would be identified with the test producing varieties of measurement that exist for assessment purposes. Alternative assessment would then be identified with a portion of the circle outside of measurement and testing, i.e., that portion which views what we are trying to assess from a non-postpositivist perspective or paradigm, and making use of non-quantitative techniques for data collection and analysis. The use of qualitative methods would not be simply a methodological choice, but an epistemological and ontological one, entailing a view of language as something that is created and exists in the act of our using, inquiring and interpreting, not as an independent, objective entity waiting to be discovered and measured. This still allows for a portion of the assessment circle outside of measurement and testing that remains within the postpositivist paradigm, where qualitative data and analysis might be used as a methodological strategy only (and therefore qualify as ‘alternatives in assessment’ rather than ‘alternative assessment’).

In part, I am arguing that alternative assessment can respond better to the implications of a critical perspective than can traditional testing or assessment procedures carried out as methodological alternatives to tests. The reason for attempting this response is one of intellectual curiosity rather than an ideological commitment to Critical Theory or critical applied linguistics. I believe that the critical perspective may have important elements to offer our research and practice in language assessment, not as a replacement for current testing practices, but as an additional approach to making decisions and judgements about individual language ability. If alternative assessment, as a response to the critical perspective, is to be judged fairly, the differences with testing need to be acknowledged and represented in our judgements.

Any attempt to characterize the paradigm of testing, and then contrast it with the paradigm of alternative assessment, within the confines of one section of a journal article runs the risk of constructing a ‘straw man’ argument. All paradigms are complex and, as our research evolves, the paradigm boundaries are beginning to blur (see Denzin and Lincoln, 2000). Even with this paradigm blurring, or seeing paradigms as a continuum of beliefs and assumptions, there are certain characteristics that remain influential in our research thinking.

My goal here is not to argue that the paradigmatic assumptions that I see underlying testing are false or insufficient in comparison with those of the alternative assessment paradigm, but that they are different. Testing, as a measurement-driven enterprise, is wedded to the current postpositivist research paradigm. It is centrally concerned with measuring, however imperfectly, traits and abilities. Underlying that research and practice are the assumptions that reality – in our case the reality of language and language use – exists independently of our attempts to understand it; that it, is an objective entity that can be measured with the proper tools and procedures. Alternative assessment, as an alternative paradigm, takes the view that language ability and use can best be understood as realms of social life that do not exist independently of our attempts to know them. Judgements or decisions about language ability and use cannot, therefore, be accomplished as a measurement task: there is no ‘true score’ waiting to be approximated.

It is within this discussion of the differences in research paradigms that the potential for responding to the critical perspective seems to me to be most obvious. Specifically, the difference in paradigms leads to a difference in the way we conceptualize validity and its criteria, and these criteria can be examined in relation to the characteristics of a critical perspective for applied linguistics. Before turning to a discussion of what the validity criteria for alternative assessment might look like, I summarize below the critical perspective on language testing (and assessment) as articulated by Elana Shohamy.

2 Critical language testing

Just as the work of Alastair Pennycook has been definitive for the critical applied linguistics being used in this article, Elana Shohamy has provided the most definitive response to date for a critical approach to language assessment. Her 1997 American Association of Applied Linguistics plenary address, ‘Critical language testing and beyond’, effectively defined the critical perspective within the field of language testing. Note that this term uses testing instead of assessment; it may be more appropriate from the perspective examined here to label it critical language *assessment*. Even in this article, Shohamy referred to ‘some models of assessment . . . ’ with ‘not testing!’ in parentheses, before proposing nine features of critical language testing, acknowledging the influence of Kramsch (1993), Pennycook (1994) and Moss (1996). In a 1999 article (based on sections of her now published book: Shohamy, 2001), she elaborated those features as 15 ‘principles that underlie critical language testing’ (p. 10). Table 1 groups these principles under the characteristics of a critical applied

Table 1 Shohamy's critical language testing principles and the critical perspective

Critical perspective characteristic 1: an interest in particular domains such as gender, class, ethnicity, and the ways that language and language-related issues (like all human relations and activities) are interconnected with them.

Shohamy's CLT Principles:

- 1) Critical language testing (CLT) is not neutral, but is shaped by cultural, social, political, educational and ideological agendas.
- 3) CLT views test-takers as political subjects within a political context.
- 4) CLT views tests as tools within a context of social and ideological struggle.
- *5) CLT asks questions about which and whose agendas tests serve.
- *6) CLT claims that testers need to understand the tests they create within a larger vision of society and its use of those tests.
- *7) CLT examines tests in terms of their measurement and assessment of knowledge vs. their definition and dictation of knowledge.
- *8) CLT questions the nature of knowledge that tests are based upon: whose knowledge? Independent 'truth' or negotiated and challengeable?
- 9) CLT examines the influence and involvement of the range of stakeholders in a testing context.
- 10) CLT perceives the embeddedness of tests within social and educational systems.

Critical Perspective Characteristic 2: the notion that our research needs to consider paradigms beyond the dominant, postpositivist-influenced one.

Shohamy's CLT Principles:

- *7) CLT examines tests in terms of their measurement and assessment of knowledge versus their definition and dictation of knowledge.
- *8) CLT questions the nature of knowledge that tests are based upon: Whose knowledge? Independent 'truth' or negotiated and challengeable?
- 11) CLT admits to the limited knowledge of any tester and the need for multiple sources of knowledge.
- 12) CLT challenges the dominant psychometric traditions and considers 'interpretive' approaches to assessment that allow for different meanings and interpretations rather than a single absolute truth.
- 13) CLT considers the meaning of test scores within this interpretive framework, allowing for the possibility of discussion and negotiation across multiple interpretations.
- 15) CLT challenges the primacy of the 'test' as assessment instrument and considers multiple procedures for interpreting the knowledge of individuals.

Critical perspective characteristic 3: a concern for changing the human and social world, not just describing it: the 'transformative agenda', with the related and motivational concern for social justice and equality.

Shohamy's CLT Principles:

- 2) CLT encourages an active, critical response from test-takers.
- *5) CLT asks questions about which and whose agendas tests serve.
- *6) CLT claims that testers need to understand the tests they create within a larger vision of society: What vision do the tests create? What vision and purposes are they used for?
- 14) CLT challenges the knowledge that tests are based upon and advocates a democratic representation of the multiple groups of society.

Critical perspective characteristic 4: the requirement that critical applied linguistics be self-reflexive.

Shohamy's CLT Principles:

- *5) CLT asks questions about which and whose agendas tests serve.
- *8) CLT questions the nature of knowledge that tests are based upon: Whose knowledge? Independent 'truth' or negotiated and challengeable?
- *13) CLT considers the meaning of test scores within this interpretive framework, allowing for the possibility of discussion and negotiation across multiple interpretations.

Notes: Numbering refers to the order in which Shohamy (2001) presents the principles. Asterisks (*) indicate that the principle appears under more than one characteristic.

linguistics presented earlier. The numbering refers to the order in which Shohamy presents her principles, with asterisks indicating that the principle appears under more than one characteristic.

As can be seen, Shohamy's critical language testing (CLT) responds to all of the characteristics of a critical approach. In particular, it is explicit and outspoken about the need for alternative research paradigms. The response to the requirement for self-reflexivity does not distinguish itself from the call for alternative paradigms (the principles overlap), but the questioning of 'agendas' (Principle 5) in combination with these appeals to interpretive research paradigms seems to articulate a recognition of the importance of a self-reflexive stance.

VI Ethics and validity from a critical perspective

As mentioned previously in the discussion of alternative assessment, if there is a need to recognize research paradigms beyond the dominant, postpositivist approach, then we need to rethink our approach to validity as well. My colleague Peter Shaw and I have attempted to begin some thinking in this direction in relation to portfolio assessment (Lynch and Shaw, 1998). It is important to reiterate that assessment techniques and procedures are not, in themselves, exemplary of alternative assessment or a critical perspective. It is the assumptions of the research and practice within which they are embedded that determine their critical potential or alternative paradigm character. These assumptions are reflected in the ways in which assessment procedures like portfolios are structured, used and interpreted (for more on this issue, see Lynch, 1997; 2001). The validity framework we developed integrates validity with ethical considerations, especially in terms of consciously addressing the power relations that are at play in the assessment context. I present the derivation and definition for each of the framework's categories in the following subsections.

1 Fairness

Although present in most validity typologies, fairness is approached in different ways, depending on the research paradigm being adopted. From the alternative assessment perspective used in the present framework, reaching a 'fair' consensus on the meaning of a student's portfolio (as opposed to a statistically significant inter-rater correlation, a criterion of fairness in the traditional validity/reliability frameworks) involves a consideration of the following questions:

- Are the perspectives of all affected participants in the portfolio assessment process being taken into account?

- Is the assessment structured such that it maximizes ethical behaviour in the sense provided by Foucault (1982); that is, so that the relations of power are ‘mobile’, ‘reversible’, ‘reciprocal’?

2 *Ontological authenticity*

This category of ethics and validity comes from the work of Guba and Lincoln (1989); specifically, their ‘authenticity criteria’. For Guba and Lincoln, ontological authenticity means being able to access and use information from research (or assessment) in a meaningful way. Here, this notion is focused on being able to establish a meaningful identity for oneself, which draws upon Foucault’s (1990; 1997) ‘care of self’ and ‘practices of the self’: the active practice of constituting an identity or, more accurately, identities for oneself. In this sense, ontological authenticity, as an ethics and validity criterion, asks: Do the participants in the assessment process establish a meaningful identity, a sense of who they are?

It is important to remember that this category, like the others, includes all participants and stakeholders in the assessment process: the teachers, the assessors, the administrators, and parents/community members. In Lynch and Shaw (1998) we suggest that a better term might be ‘ontological creativity’, since the use of Foucault’s thought places an emphasis on the creative, active construction of identity, rather than on authenticity *per se* (authenticity in the Sartrean sense of being ‘true to one’s true self’).

3 *Cross-referential authenticity*

This category also draws on Guba and Lincoln (1989), specifically their criterion of ‘educative authenticity’. As used here, it is extended to examine the understanding of the identities that others have constituted for themselves as a result of the assessment process. It asks the question: Are the participants in the assessment process able to gain an improved understanding of the perspectives outside their own group (e.g., do students understand teachers better; does the teacher of a different class understand a colleague and/or her students better)?

4 *Impact/consequential validity*

Here the term ‘impact’ is adapted from Bachman and Palmer’s (1996) test ‘usefulness’ framework. It corresponds, also, to Messick’s (1989) ‘consequential validity’ (the value implications and social consequences of test interpretation and use), and to Guba and Lincoln’s

(1989) ‘catalytic authenticity’ (the degree to which something happens as a result of the research or assessment process). In our framework, accordingly, ethical and valid assessment is taken to include an examination of the outcomes, intended and unintended, of the assessment process. The question to be examined is: what specifically is done as a result of the assessment (e.g., is a change made in the curriculum; does a teacher-in-development alter some aspect of his or her teaching style)?

As pointed out in Lynch and Shaw (1998), the link between values and consequences (highlighted by Messick, 1989) becomes clear in a consideration of this category. Consequences, or impact – the things that occur as a result of assessment – once identified need to be examined for their value. Are they good or bad? Remaining consistent with the other categories of this framework, determining good from bad would need to include the multiple perspectives that make up the assessment setting. It would be a negotiated consensus of some sort.

5 Evolved power relations

The construction of this category represents a combination of educational research and philosophy, once again. Guba and Lincoln’s (1989) ‘tactical authenticity’ (the degree to which participants are empowered to carry out the changes that are made possible through the research or assessment process) is combined with Foucault’s (1982) notion of power relations, as central to a determination of ethics. The overlap between this category and the others in the framework is clear, and underscores their non-discrete character. For example, a determination of ontological authenticity will clearly involve an examination of power relations. If power relations change, this would be an example of impact. The focus of this category, continuous with others as it may be, derives from Foucault’s highlighting the importance of free and ethical power relations. It asks:

- Do the participants change the way in which they relate to each other and to themselves (e.g., do the students assume and obtain more responsibility in the curriculum; do the teachers gain control over assessment policies previously established by others)?
- Do these changes become fixed, or are they established as reversible, mobile relations of power?

Once we have discovered the degree to which power relations have changed, as with impact/consequential validity, we need to evaluate the change as positive or negative. In the second question above, however, there is a basis for evaluation: power relations that establish themselves as reversible and mobile are seen as ‘good’.

How well does this framework respond to the challenges of a critical perspective? By emphasizing a recognition of multiple perspectives and a negotiation of consensus on assessment outcomes, most of the concerns of a critical approach are addressed. Certainly, it responds to the requirement of considering research paradigms beyond the dominant, postpositivist-influenced one. A self-reflexivity is built into the categories of ontological authenticity and cross-referential authenticity. Impact/consequential validity and evolved relations of power recognize the interconnections between language and social, cultural and political questions; and by making change an area of concern for ethics and validity, these categories link to the critical requirement for transformation.

This framework may, however, be criticized from three standpoints. First, from the postmodern-influenced version of critical being adopted here, the motivation and need for such a framework may be questioned. Are we not simply replacing one *a priori* set of categories (i.e., internal and external validity) with another? The answer to this question is yes, but the replacement categories are meant to be less fixed and more open to reinterpretation on an assessment case-by-case basis than in traditional validity typologies. This leads to the second critique: does this not place a potentially impossible burden on achieving local consensus to make assessment decisions? The answer to this question is 'possibly'; however, without the use of multiple perspectives and negotiated consensus, assessment cannot meet the critical challenges of alternative paradigm use and self-reflexivity. Finally, there may be a lack of explicit attention to the critical requirement of having an interest in particular domains such as gender, class, ethnicity and race.

VII Conclusion

The issue of domains of interest, and the related issue of research paradigms, which characterize the critical approach of Pennycook, makes an important difference with Critical Theory. In particular, these characteristics are associated with a resistance to constructing a grand theory of rationality and consciousness as the basis for social transformation, which the Habermasian version of Critical Theory attempts. The self-reflexive characteristic of Pennycook's critical approach acts to ensure that it does not result in merely replacing one research paradigm with another. The danger in this position is that it may result in a lack of theoretical coherence for research and practice. Theory is given an ambiguous role: the need to ground our research and practice in some sort of theory is acknowledged, but there is a suspicion of theory turning into grand narrative. Pennycook's solution

to this is a 'constant skepticism, a constant questioning of the normative assumptions of applied linguistics' (2001: 8). That is, theory becomes an ever-changing framework, rather than a fixed point of reference. For some this will result in a lack of coherence, while for others it will allow the best chance at a thoughtful engagement with the complex and crucial issues of our time and with the variety of ideas being constructed to account for those issues.

Although the focus of this article is on the implications for ethics and validity in language assessment, the critical perspective raises the related questions of what critical language assessment procedures would look like; what would they do? As a preliminary answer, I would suggest that issues of gender, sexuality, class, ethnicity, race, ideology, identity and culture would need to be engaged with as a part of the assessment process. The process would be evaluated as valid to the extent that it led to changes in the lives of people in the assessment context, if not the larger social and cultural context, that involve struggling against and overcoming oppression, inequality and injustice.

In critical language assessment such high-sounding goals might be realized as a procedure that assesses someone's ability to identify homophobic assumptions in language interaction and to produce language that exposes these assumptions and allows for the participants to interact in a way that respects the sexual identities of all. This would be an example where the domains of a critical perspective are the subject matter for the assessment. This same critical perspective might be applied to more traditional subject matter as well, designing the assessment procedure such that alternative research paradigms and their associated validity frameworks are drawn upon; aiming to establish an assessment context in which normally marginalized voices are given a more active role and expression, i.e., a context in which traditional power relations are recognized and made more reversible and flexible. It would also aim for transformation, i.e., for participating individuals to change their negative attitude and behaviour towards persons with different sexual orientations to theirs (and this would overlap with the evaluation of the validity of the procedure, using alternative criteria such as cross-referential authenticity and evolved power relations). Finally, such a critical assessment would need to be self-reflexive: the basis for arriving at the assessment results, and the procedure itself, would have to be constantly questioned and any move towards a mandatory and normalized attitude and behaviour resisted. This, of course, renders problematic our traditional reliance on notions of objectivity and right-and-wrong answers for making assessments.

There are other implications for language assessment from a critical

perspective that need to be addressed. In addition to the openness to other research paradigms and explicit examination of fairness already discussed, McNamara (1998) suggests the following:

- an awareness of issues raised by the existence of ‘world Englishes’;
- a reconsideration of the social impact of technology in test delivery.

Other important features of a critical language assessment include the need to reconsider the expert status of language testers and more democratic approaches to assessment. Assessment, rather than testing, may facilitate collective action such as the formation of advocacy groups as suggested by Shohamy (1999). There is also work relevant to testing as well as assessment designed to make testing, as well as assessment, more open to the active participation of stakeholders other than testers (see Davidson and Lynch, 2002). The need for a reconsideration of the ethical issues raised by testing is currently being taken up by the International Language Testing Association, in the form of a code of ethics and professional practice, which will include Shohamy’s call for a recognition of test-takers’ rights.

The process of writing this article – i.e., of thinking through the issues, realizing how many there are to consider and reflecting on the complex nature of their interactions – has been exciting, yet frustrating. The challenges of the critical perspective have, at times, made me think that language assessment of any sort is incompatible with such a perspective. However, I am glad that critical applied linguists like Alastair Pennycook and critical language testers like Elana Shohamy have asked us to take a step back and put our assessment research and practice into the broader social, political and cultural picture. It is a difficult project, but to limit our concerns to technical sophistication and efficiency, focusing only on the goals of an objective and neutral language assessment, would surely be a mistake.

VIII References

- Adorno, T.W.** 1984: *Aesthetic theory*. London: Routledge and Kegan Paul.
- Auerbach, E.** 1993: Reexamining English only in the ESL classroom. *TESOL Quarterly* 27, 9–32.
- 1995: The politics of the ESL classroom: issues of power in pedagogical choices. In Tollefson, J., editor, *Power and inequality in language education*. New York: Cambridge University Press.
- Bachman, L.F.** 1990: *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F.** and **Palmer, A.S.** 1996: *Language testing in practice*. Oxford: Oxford University Press.

- Benhabib, S.** 1996: Critical theory and postmodernism: on the interplay of ethics, aesthetics, and utopia in critical theory. In Rasmussen, D.M., editor, *The handbook of critical theory*. Oxford: Blackwell, 327–39.
- Birenbaum, M.** 1996: Assessment 2000: towards a pluralistic approach to assessment. In Birenbaum, M. and Dochy, F.J.R.C., editors, *Alternatives in assessment of achievements, learning processes, and prior knowledge*. Dordrecht, Netherlands: Kluwer, 3–29.
- Brown, J.D. and Hudson, T.D.** 1998: The alternatives in language assessment. *TESOL Quarterly* 32, 653–75.
- Davidson, F. and Lynch, B.K.** 2002: *Testcraft: a teacher's guide to writing and using language test specifications*. New Haven: Yale University Press.
- Dean, M.** 1994: *Critical and effective histories: Foucault's methods and historical sociology*. London: Routledge.
- Denzin, N.K. and Lincoln, Y.S.** 2000: *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Fairclough, N.** 1989: *Language and power*. London: Longman.
- 1995: *Critical discourse analysis*. London: Longman.
- Foucault, M.** 1980: *Power/knowledge: selected interviews and other writings, 1972–1977*. New York: Pantheon Books.
- 1982: The subject and power. In H. L. Dreyfus and P. Rabinow, editors, *Michel Foucault: beyond structuralism and hermeneutics*, pp. 208–26. Note: pp. 208–16 written in English by Michel Foucault; pp. 216–26 translated from the French by Leslie Sawyer. Brighton: Harvester Press.
- 1990: *The history of sexuality: Volume 3: the care of the self*. Translated by R. Hurley. New York: Vintage Books/Random House. Original work published 1976.
- 1997: The ethics of the concern for self as a practice of freedom. In Rabinow, P., editor, *Michel Foucault, Ethics: Subjectivity and truth (The essential works of Michel Foucault, 1954–1984, Vol. 1)*. London: Allen Lane/Penguin, 281–302.
- Guba, E.G.** 1990: The alternative paradigm dialog. In Guba, E.G., editor, *The paradigm dialog*. Newbury Park, CA: Sage, 17–27.
- Guba, E.G. and Lincoln, Y.S.** 1989: *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Habermas, J.** 1971: *Knowledge and human interests*. Boston, MA: Beacon Press.
- 1987a: *The theory of communicative action*. Boston, MA: Beacon Press.
- 1987b: *The philosophical discourse of modernity*. Boston, MA: MIT Press.
- Horkheimer, M. and Adorno, T.W.** 1972: *Dialectic of enlightenment*. New York: Herder and Herder.
- Huerta-Macias, A.** 1995: Alternative assessment: Responses to commonly asked questions. *TESOL Journal* 5, 8–11.
- Kincheloe, J.L. and McLaren, P.L.** 1994: Rethinking critical theory and

- qualitative research. In Denzin, N.L. and Lincoln, Y.S., editors, *Handbook of qualitative research*. Thousand Oaks, CA: Sage, 138–57.
- Kramsch, C.** 1993: *Context and culture in language teaching*. Oxford: Oxford University Press.
- Lather, P.** 1991: *Getting smart: feminist research and pedagogy with/in the postmodern*. New York: Routledge.
- Lynch, B.K.** 1997: In search of the ethical test. *Language Testing*, 14, 315–27.
- 2001: The ethical potential of alternative assessment. In Elder, C., Brown, A., Grove, E., Hill, K., Iwashita, N., Lumley, T., McNamara T. and O’Loughlin, K., editors, *Experimenting with uncertainty: essays in honour of Alan Davies*. University of Cambridge Local Examinations Syndicate Studies in Language Testing Vol. Cambridge: Cambridge University Press, 218–29.
- Lynch, B.K. and Shaw, P.** 1998: Portfolios, power and ethics. Paper presented at TESOL 98, Seattle, WA.
- McCarthy, T.** 1996: Critical theory and postmodernism: a response to David Hoy. In Rasmussen, D.M., editor, *The handbook of critical theory*. Oxford: Blackwell Publishers, 340–68.
- McNamara, T.F.** 1998: Policy and social considerations in language assessment. *Annual Review of Applied Linguistics* 18, 304–19.
- Messick, S.** 1989: Validity. In Linn, R.L., editor, *Educational measurement*. 3rd edn. New York: Macmillan, 13–103.
- Morgan, B.** 1998: *The ESL classroom: teaching, critical practice, and community development*. Toronto: University of Toronto Press.
- Moss, P.A.** 1994: Can there be validity without reliability? *Educational Researcher* 23, 5–12.
- 1996: Enlarging the dialogue in educational measurement: voices from interpretive research traditions. *Educational Researcher* 25, 20–28.
- Norris, J.M., Brown, J.D., Hudson, T. and Yoshioka, J.** 1998: *Designing second language performance assessments*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai’i Press.
- Peirce (Norton), B.N.** 1995: Social identity, investment, and language learning. *TESOL Quarterly* 29: 9–31.
- Pennycook, A.** 1989: The concept of method, interested knowledge, and the politics of language teaching. *TESOL Quarterly* 23, 589–618.
- 1990: Towards a critical applied linguistics for the 1990s. *Issues in Applied Linguistics* 1, 8–28.
- 1994: *The cultural politics of English as an international language*. London: Longman.
- 1999: Introduction: critical approaches to TESOL. *TESOL Quarterly* 33, 329–48.
- 2001: *Critical applied linguistics: a critical introduction*. Mahwah, NJ: Lawrence Erlbaum.
- Phillips, D.C.** 1990: Postpositivistic science: myths and realities. In Guba, E.G., editor, *The paradigm dialog*. Newbury Park, CA: Sage, 31–45.
- Phillipson, R.** 1992: *Linguistic imperialism*. Oxford: Oxford University Press.

- Rasmussen, D.M.** 1996: Critical theory and philosophy. In Rasmussen, D.M., editor, *The handbook of critical theory*, Oxford: Blackwell, 11–38.
- Shepard, L.A.** 1993: The place of testing reform in educational reform: a reply to Cizek. *Educational Researcher*, 22, 10–13.
- Shohamy, E.** 1997: Critical language testing and beyond. Plenary talk presented at the American Association of Applied Linguistics (AAAL) Meeting, Orlando, Florida. Published in: *Studies in Educational Evaluation*, 1998, Volume 24, 331–45.
- 1999: Critical language testing, responsibilities of testers and rights of test takers. Paper presented at the AERA Convention, Montreal.
- 2001: *The power of tests: a critical perspective on the uses of language tests*. London: Longman.
- Tollefson, J.** 1989: *Alien winds: the reeducation of America's Indochinese*. New York: Praeger.
- 1991: *Planning language, planning inequality : language policy in the community*. London: Longman.
- editor, 1995: *Power and inequality in language education*. Cambridge: Cambridge University Press.
- Wodak, R.** 1996: *Disorders of discourse*. London: Longman.
- Wolf, D., Bixby, J., Glenn, J., III, and Gardener, H.** 1991: To use their minds well: investigating new forms of student assessment. *Review of Research in Education*, 17, 31–74.