

# Language Testing

<http://ltj.sagepub.com>

---

## **Introduction: the limits of ethics in language testing**

Alan Davies

*Language Testing* 1997; 14; 235

DOI: 10.1177/026553229701400301

The online version of this article can be found at:  
<http://ltj.sagepub.com/cgi/content/abstract/14/3/235>

---

Published by:

 SAGE Publications

<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.com/journalsPermissions.nav>

# Introduction: the limits of ethics in language testing

**Alan Davies** *University of Edinburgh and University of Melbourne*

The articles in this special issue were written for a symposium on the ethics of language testing held at the triennial congress of the Association Internationale de Linguistique Appliquée in August 1996 in Finland. This special issue addresses the role of ethics (and the limits of that role) in professional activities such as language testing. The nine articles that follow can be divided into four sections. The two articles in the first section (Spolsky and Hawthorne) consider language testing as a means of political control. The two in the second section (Elder, and Norton and Starfield) are concerned essentially with the definition of the test construct – Elder considering whether the construct has the same meaning for different groups and Norton and Starfield questioning whether it is form or content that is being assessed. The two articles in the third section (Hamp-Lyons and Rea-Dickins) consider the effects of language tests on the various stakeholders who are involved. In the fourth section, the three articles (Shohamy, Lynch and Davies) offer criteria for promoting ethicality in language testing, Lynch by an approach from principle (deontological), Davies through the professionalizing of the activity (teleological), and Shohamy through attention both to method and to consequence.

## I Introduction

At the triennial congress of the Association International de Linguistique Appliquée (AILA) held in Jyväskylä, Finland, in August 1996, the writer convened a symposium on the theme of 'Good conduct in language testing: ethical concerns'. The motivation behind the symposium was the growing feeling among scholars in language testing that challenges as to the morality of language testing were increasing both within and outside the field. Attendance at the symposium indicated wide-ranging interest in the topic; as indeed does the selection of the theme for this special issue. The nine articles in this special issue are based on a number of the presentations at Jyväskylä, with the addition of the article by Pauline Rea-Dickins who was scheduled to present at Jyväskylä but had to withdraw.

The special issue explores ethicality in language testing. Should testing specialists be responsible for decisions beyond test construction? Who decides what is valid? Does professionalism conflict with 1) public and 2) individual morality? Other issues explored are

relationships with various stakeholders, conflict between fairness and face validity, the relationship between bias and fairness, washback, and the politics of gatekeeping.

The purpose of this special issue is not to arrive at any agreed conclusion on an ethical position in language testing or even on the place of ethics in the field. Rather it is to raise some of the issues that confront language testers, to consider these in a range of language testing situations and for different language tests. While the growing professionalization of language testing is perceived as a strength and a major contribution towards a growing sense of ethicality, the increase in commercial and market forces, as well as the widespread use of language assessment as an instrument in government policy, may pressure language testers into dangerous (and unethical) conduct.

The remainder of this introduction attempts an explanation of the role of ethics (and the limits of that role). The eight articles that follow can be divided into four sections: language testing as a means of political control, procedures for detecting bias, the effects of language tests on the various stakeholders and criteria for promoting ethicality in language testing.

## **II Group ethics**

In the articles that follow there is an underlying concern with group ethics, the ethics of the profession of language testing.

O'Hear (1985: 277) points out that 'participation in such groups is characterized by the recognition on the part of all of reciprocal duties, rights, loyalties and deserts, which are not a matter of individual choice or preference'. In order to consider what is meant by 'reciprocal duties, rights, loyalties and deserts' it will be useful to examine the question of normative ethics.

Normative ethics makes a distinction between two types of reason for following norms, that is acting out an 'ought'. The first is deontological, the second teleological, the difference being that while teleological theories appeal to value considerations, deontological theories do not.

We can caricature the deontological and the teleological approaches thus:

- Deontological: do it because it is inherently right.
- Teleological: do it because it will bring the best results.

While deontological theories emphasize the importance of fairness as an absolute, thereby allowing for interventions such as positive discrimination in favour of minorities in order to ensure fairness (the 'level playing field'), teleological theories are primarily concerned

with validation in terms of outcome, and therefore not of fairness as an absolute but of securing the best result. Clearly there is uncertainty as to what the best result might be and certainly deontologists might claim that their fairness will produce the best results. Teleologists on the other hand might respond that fairness is chimerical anyway and rather than intervene on behalf of a chimerical fairness it is better to allow fairness to find its own level. For example, in a scholarship selection competition the deontological approach might be to weight results by adjusting for handicapped starts. Such interference, teleologists might say, introduces irrelevant issues into the selection process. In the name of a spurious fairness (spurious because you can never be fair to everyone since group adjustments take no account of variation within groups) we end up selecting the wrong people, not the best but the most deserving according to certain criteria. The dispute is between the idealistic and the realistic, what Stewart (1996) refers to as the general good and the individual good.

Rawls (1967: 221) points out that these are two (equal) principles of justice:

... first each person engaged in an institution or affected by it has an equal right to the most extensive liberty compatible with a like liberty for all; and second, inequalities as defined by the institutional structure or fostered by it are arbitrary unless it is reasonable to expect that they will work out to everyone's advantage and provided that the positions and offices to which they attach or from which they may be gained are open to all.

And so one of the chief roles for ethics is to balance these two principles, the individual and the social. This requires thought and imagination as much as law-making, offering '... a way of conceptualising difference which renders it compatible with equality, but also, and crucially, does not simply increase social differentiation' (Mendus, 1992: 414).

Given that all testing endeavours to discriminate among individuals, the articles that follow all grapple with this issue by attempting to reconcile the two principles, the individual and the social. Their concern is how social justice can be maintained while at the same time individual differences are fully recognized and taken account of.

The danger, of course, is that in our attempts to be fair we end up by destroying completely the social, making all morality individual and therefore never achieving fairness anyway. Indeed, Osborne (1992: 181), discussing the influence on philosophy of poststructuralism and feminism, suggests that we may have already got to that point where 'Perhaps philosophy has been reduced to personal ethics or the search for small forms of valid knowledge'.

This is a counsel of despair, as is the concern that since all morality

is contextually and culturally relative there is no general morality. Jackson (1996) shows a way of avoiding such a solipsist trap.

### III Ethics and morality

Jackson (1996: 1) equates ethics and morality: 'Practical ethics (or practical morality) – I will use the terms interchangeably'. Webster's *Ninth new collegiate dictionary of the English language* (1994 edition) makes a subtler distinction, thus:

MORAL and ETHICAL are both concerned with rightness or wrongness of actions and conduct, but MORAL is more often applied to the practice or actions of individuals, often specifically in sexual relations, ETHICAL more often to theoretical or general questions of rightness, fairness or equity.

Moral may indeed be more often applied to the practice or actions of individuals but in practice the terms ethics/ethicality and morals/morality appear to be used interchangeably.

Discussing codes of practice Jackson (1996) points out that morality is never absolute. For example, codes of health and safety require appropriate protection of employees. At the same time, in all such cases there is a clause (either implicit or explicit) which limits employers' responsibility to 'within reason'. Otherwise, their duty would be impossible to fulfil. What this means in language testing is that it is necessary that language testers ('members') are responsible 'within reason' for the output of their professional work and not for all possible uses and consequences of the tests they produce.

Without the recognition of the 'within reason' limitation, we are likely to exaggerate the demands of morality and to assume wrongly that you cannot get on in business or carry out your profession unless you are prepared to cast aside or compromise principles. This is not so. Morality as typically encoded in codes of practice constrains responsibility within reasonable limits.

### IV This special issue

The nine articles that follow can be divided into four sections. The two articles in the first section (Spolsky, Hawthorne) consider language testing as a means of political control. The two in the second section (Elder and Norton and Starfield) are concerned essentially with the definition of the test construct – Elder considering whether the construct has the same meaning for different groups and Norton and Starfield questioning whether it is form or content that is being assessed. The two articles in the third section (Hamp-Lyons and Rea-Dickins) consider the effects of language tests on the various stakeholders who are involved. In the fourth section, the three articles

(Lynch, Davies and Shohamy) offer criteria for promoting ethicality in language testing, Lynch by an approach from principle (deontological), Davies through the professionalizing of the activity (teleological), and Shohamy through attention both to method and to consequence.

In section 1, Spolsky points out that the gatekeeping use of language tests is not new and that their misuse was recognized a hundred years ago. Spolsky urges the need for scepticism with regard to test results, particularly in the middle bands of the scale.

Hawthorne approaches the topic from a background in immigration research. She describes the origin and development of two recent English language tests for immigrants to Australia, arguing compellingly that their purpose is largely if not entirely political and that they represent a cynical manipulation of language testers' professional skills.

In section 2, Elder examines what language testers actually do internally (in terms of test construction and analyses) in order to respond to ethical demands. Elder uses data from her study of three foreign language examinations (Italian, Greek and Chinese) for school leavers in Australia to argue for 'the central role of judgement in determining whether or not there is bias in a given test'. Since this must mean that bias analysis, however sophisticated, is still influenced by judgements of what is fair, we cannot escape from ethical queries about those judgements.

Norton and Starfield report their study of students in a South African university who think they are being assessed on their academic knowledge but who are in fact being assessed on the basis of their language accuracy. Norton and Starfield argue the need for an explicit departmental policy specifying the extent to which proficiency in written English or other language is assessed in academic subjects. The authors provide detailed recommendations which, if implemented, would help both teachers and students. They acknowledge that the fundamental obstacle to real facilitation in this regard is our lack of progress in separating the language and the content, the medium and the message, the great unresolved dilemma of all work on languages for specific purposes.

In section 3, Hamp-Lyons considers language testing from the strong Messick position on consequential validity. Hamp-Lyons notes that the often proposed solution of alternative testing must itself be regarded sceptically, illustrating her argument from her own research on portfolio assessment. Her conclusion modifies the extreme position from which she began: language testers cannot, she acknowledges, be held accountable for all possible consequences, but they are accountable for 'those consequences of which we are aware'.

Rea-Dickins distinguishes two groups of stakeholders, those who hold power and make decisions and those affected by those decisions. She points out that language testers have never been autonomous but always subject to control by the first group of stakeholders. Rea-Dickins argues that fairness (and therefore ethicality) requires the involvement in the assessment process of all stakeholders. But she is sceptic enough to recognize that encouraging greater participation by stakeholders, such as teachers and parents, does not necessarily lead to their real empowerment. To avoid being seduced by a cynical exercise in face validity, testers must interact with other stakeholder groups in order better to 'understand the assessment culture in which [they] operate'.

In section 4, Lynch evaluates a standardized basic skills test (the Learning Assessment Project or LAP test) used in the Australian State of Victoria against frameworks for determining equity of educational opportunity and finds the LAP wanting. Lynch advocates alternative forms of assessment for which new validity criteria (such as the authenticity criteria used in qualitative research) would be needed.

Davies proposes that the very process of becoming more professional, of deliberately creating a profession, is itself ethical. He argues that such a process serves to protect both the professional and the client: in the case of language testing, the tester and other stakeholders. Davies makes a case for the responsibilities laid down by the profession to be indicative both of the extent and of the limits (the 'within reason' clause) of the professional's obligations. Given the lack of professional sanctions in language testing, as in other social science professions, Davies concludes that the various processes of professionalizing themselves create the implicit sanctions of the profession's ethical milieu.

Shohamy, citing tests of Arabic as a second language and of English as a second language, argues that methods which are not fair to all candidates are unethical. Equally, uses of language tests to exercise control rather than to provide information are unethical. Language testers need to exercise constant vigilance.

## V References

- Jackson, J.** 1996: *An introduction to business ethics*. Oxford: Blackwell.
- Mendus, S.** 1992: Losing the faith: feminism and democracy. In Stewart, R.M., editor, *Readings in social and political philosophy*, New York: Oxford University Press, 409–17.
- O'Hear, A.** 1985: *What philosophy is*. Harmondsworth: Penguin Books.
- Osborne, R.** 1992: *Philosophy for beginners*. New York: Writers & Readers Publishing, Inc.
- Rawls, J.** 1967: Distributive justice. In Laslett, P. and Runciman, W.G.,

editors, *Philosophy, politics and society* (3rd series), Oxford: Blackwell (reprinted under the same title in Stewart, R.M., editor, *Readings in social and political philosophy*, New York: Oxford University Press, 219–34.

**Stewart, R.M.**, editor, 1996: *Readings in social and political philosophy*. New York: Oxford University Press.