

6

Fairness and Justice in Language Assessment: Principles and Public Reasoning

Antony John Kunnan

National Institute of Education, Singapore

antony.kunnan@nie.edu.sg

This paper is based on the talk given at the 2013 CELC Conference at NUS, Singapore

Introduction

Fairness has been debated for the last 20 years; there is agreement regarding the concept, but conceptions of fairness have differed. The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) have a section titled “Fairness in Testing”; Codes of ethics and practice (ILTA, ETS, ALTE, etc.). Scholars have written about this as well: from the point of view of situated ethics: Kunnan and Davidson 2000); how to investigate fairness: Xi, Kane, Kunnan, Davies (in *Language Testing*, 27) and the overlap and separation between these two concepts.

Different conceptions of Fairness

The concept of fairness has been interpretation to mean: “unbiased,” “equal opportunity,” “equitable treatment,” “similar outcomes in terms of scores.” Questions have been raised regarding the scope: Does fairness include validity or does validity include fairness? Or, are they two separate entities? And, does any action need to be taken: Should we do anything about it? (Kane, Xi, Davies from *Language Testing*, 27).

The concept of justice has been interpreted as follows: Distributive justice - institutions providing benefits are distributed to a society in a just manner; Retributive or corrective justice - punishments are just; and Compensatory justice - fair compensation for injuries.

Purpose of paper

The purpose of this paper is to provide principled bases for fairness and justice as applied to the institution of assessment:

Fairness as *relating to persons* - how assessments ought to be fair to test takers

Justice as *relating to institutions* - how institutions ought to be just to test takers

Preliminary questions

Here are a few preliminary questions that can be raised to begin the discussion:

- 1 Does every test taker have the right to a fair assessment? Is this rule inviolable? Are rights of test takers to a fair assessment universal or only applicable in states that provide equal rights?
- 2 Is it adequate that most test takers are assessed fairly while a few are not? Would it be appropriate to use a cost-benefit analysis to evaluate whether assessments should be improved or not? And, if harm is done to test takers, does such harm need to be compensated?
- 3 Would the rights of test takers to a fair assessment be supported in authoritarian states that do not provide for equal rights? Would institutions in such states feel less compelled to provide a fair assessment?
- 4 Should an assessment be beneficial to the society in which it is used?
- 5 Should assessment developers and users be required to offer public justification or reasoning?

In order to offer clarity to the discussions of the abstract concepts and principles, a few real-world vignettes are provided.

Vignette 1: Pre-testing

Imagine a new staff member joined a large language assessment organization that develops

assessments for high-stakes contexts. After she had worked at the organization for three months, she began to be concerned about many of the practices: first, did not pre-test their test tasks; instead, they used the un-pretested tasks in a real administration and did not delete the scores from those tasks when they computed the scores for the test takers. The staff member approached her supervisor. He was at first disinterested in the staff member's concern but later admitted that pre-testing tasks would cost too much, and, if they conducted a pre-test, the assessment would also cost more to the test taker.

Reflection

Did the staff member do the right thing in bringing to the attention of her supervisor the lack of pre-testing? Is pre-testing of tasks for high-stakes assessments a requirement? Is the head's lack of understanding of the situation acceptable? Are his reasons for not conducting any pre-testing justified? Would people of any persuasion (teachers, test takers, business leaders, etc.) be able defend such practice? Is there a violation of an accepted Code of Ethics and Practice? Could this practice lead to an unfair assessment?

Vignette 2: Checks for bias

...The staff member also found that the organization did not conduct any review to examine whether the assessment was fair to all test takers in terms of content, dialect, test delivery, or test performance. She brought this matter up with her supervisor too. The supervisor said that while these are important matters, they did not have staff with expertise to conduct such investigations. He also reminded her that, once again, these studies would also cost the organization a lot of money and the final result would be that the assessment would cost more to the test taker.

Reflection

Did the staff member do the right thing by bringing to the attention of her supervisor the lack of any investigations regarding fairness? Are such investigations required in an assessment that is a high-stakes assessment? Are the supervisor's reasons for not conducting these investigations defensible? Is there a violation of an accepted Code of Ethics and Practice? Would this be an example of an unfair assessment?

examples include abortion rights; taxes; affirmative action; torture of terror suspects; and differential pricing. In terms of assessment, quality of assessment v. cost-benefit analysis and Fairness of assessments v. maximizing utility are useful examples I will return to later.

Arguments from Philosophy

What are the main perspectives and who are the main proponents? From the Utilitarian point of view: outcomes-based philosophers like Bentham and Mill and from the Deontological point of view and duty-based philosophers like Kant, Rawls and Sen.

The main idea behind Utilitarianism is that it promotes the notion of the greatest happiness of the greatest number of people; majority rules. The utility principle trumps moral values such as individual rights. Another aspect is consequentialist thinking in which outcomes of an event are used as tools to evaluate an institution. In terms of assessment, decisions about an assessment would be made solely on the basis of utility and consequences of an assessment. This could also lead the view that minority members need not be protected.

Rawls' ideas and arguments are presented in "A Theory of Justice" (1971) and "Justice as Fairness" (2001) and Sen's non-parochialist approach for global justice in "The Idea of Justice" (2009).

Rawls' Fairness principles are expected to work for a "well-ordered society" with a "fair system of social cooperation" and "citizens who are free and equal persons." He puts forward two principles:

- 1st Principle: Each person has the same claim to a fully adequate scheme of equal basic rights and liberties.
- 2nd Principle: (a) Equal Opportunity principle: offices and positions open to all under conditions of fair equality of opportunity;
(b) Difference principle: economic arrangements in which the least-advantaged members of society are better off when primary goods are unequally distributed than when primary goods are equally distributed.

Sen's views are somewhat similar to Rawls': he argues that his idea of justice would work in a democratic state (in the sense of "government by discussion" with political and personal freedoms) with free and equal persons (who are capable of challenging injustice) that would be able to safeguard principles of fairness through public reasoning. He also takes a non-parochial, global perspective view: each country may have parochial

Ethical dilemmas

Ethical dilemmas are not new to us as we grapple with these all the time. Here are a few: right and wrong; fair and unfair; equality and inequality; individual rights and the common good; just and unjust. More practical

practices that need to come under examination and scrutiny from others with distant judgments (as impartial spectators).

Public reasoning: critical component in advancing justice.

One of Sen's key points is that there should be public reasoning in the defense of institutions. Therefore, it would seem obvious that assessments as public institutions should be subjected to investigations of fairness and justice through public reasoning.

Based on the standards and codes of ethics and practice movements and the literature from philosophy, I put forward two principles and sub-principles that are relevant for assessment:

Principle of Fairness

An assessment ought to be fair to all test takers; there is a presumption of treating every test taker with equal respect.

Sub-principle 1: An assessment ought to provide adequate opportunity to learn the knowledge or abilities that are to be assessed.

Sub-principle 2: An assessment ought to be consistent and meaningful in terms of its test-score interpretation.

Sub-principle 3: An assessment ought to be free of bias against any test taker groups, in particular by assessing construct-irrelevant matters.

Sub-principle 4: An assessment ought to use appropriate access, administration and standard setting so that decision-making is equitable to test taker groups.

Principle of Justice

An assessment institution ought to bring about benefits to society and advance justice through public reasoning.

Sub-principle 1: An assessment institution ought to bring benefits to society by making a positive social impact.

Sub-principle 2: An assessment institution ought to advance justice through public reasoning of their assessment.

Here are a few vignettes that will help clarify some of the practical issues in assessment from the perspective of fairness and justice.

Vignette 3: Defective tasks

Imagine there were two forms of a paper and pencil assessment - Forms A and B. The forms were from

a high-stakes university admissions assessment. It was known previously through pre-testing that there were a few defective tasks in the two forms: 10 tasks in Form A and 5 tasks in Form B out of a total of 100 tasks in each. But the administrator went ahead and used both forms as a cost-benefit analysis conducted earlier showed that only 10% of the test takers who took Form A and 5% of the test takers who took Form B were misclassified as failed due to the defective tasks. She felt these figures were within the usual margin of error. The administrator also wrote in her report that the costs of replacing the defective tasks would be much higher than errors in misclassification.

Reflection

If we consider the two philosophical persuasions, each may support or criticize the actions of the administrator. From the Utilitarian view: cost-benefit analysis provided the basis for the administrator's decision, and that such decisions have to be made to run a business; the administrator should have preferred Form B to Form A as it had utility. From the deontologist view: the administrator did not act morally as she did not uphold the rights of all test takers to a fair assessment by using defective assessments.

So, what should the administrator have done? Which of these perspectives appeals to us? What is the right thing to do?

Vignette 4: Assessment quality

Imagine a high-stakes high-school exit examination that is conducted by the Ministry of Education for many years. After the examination, some students got together and exchanged thoughts on the examination. They concluded that some of the tasks and topics that were new to them. When the results were announced, it turned out these students had received low scores. The students could do nothing else (as there was no appeal process) but their parents complained to the Ministry of Education that something was wrong with the examination. Ministry officials said that there could not be nothing wrong as their examinations were written by expert teachers who had been doing this for many years. When pressed to show that the examination was an appropriate assessment, Ministry officials defended their examination by saying there were no prior complaints and therefore no prior analyses of the examination were conducted.

Reflection

Did the MoE have the motivation and expertise to provide the best possible examination? Possible problems: Did some of the students not

have the opportunity to learn all the material? Did the MoE regulations not have any provision for appeal? Why were there no analyses of the examination tasks conducted although they were written by experts? Why were there no research studies that examined the quality of the assessments and the test performance? Was the examination providing a beneficial service to the community?

Did the MoE owe the students public justification? Did the MoE act responsibly? Why was there lack of provision for basic rights to the students?

Vignette 5: Public reasoning

...Imagine that the parents of the students who received low scores protested the Ministry's approach and demanded that they provide a public justification of the assessment. The Ministry replied with a firm NO as it had never responded to such a request before and did not consider it necessary to do so.

Reflection

Questions for Ministry officials: From the Deontologist view: Test takers were denied basic freedoms such as the basic right to be treated with respect and dignity, to have fair assessments and assessment practice. The Ministry did not have public reasoning (through public forums) would be the only way to ensure that the assessment is fair and the institution is just.

Vignette 6: Differential pricing

Imagine a testing organization had differential pricing for some aspects of testing. Higher prices for the following: Better assessments (reliability, fairness, valid), Detailed diagnostic feedback (individual not generic), Better graders (more experienced; not harsh), Front row seating (for listening assessments with speakers in front of the room), First in line for an interview, Faster turn-around time for results, Better heating, air-conditioner, seats, Multi-modal upgrades: color, video, better animation, No experimental section

Reflection

For which aspects are we willing to pay more? For which aspects are we not willing to pay more? Why? Is our thinking utilitarian or outcomes-based? Or is it deontologist or duty-based? Does the testing organization have an obligation to provide quality assessments? Ought these obligations required across borders? Across parochial lines?

Here is a real-world example from the Netherlands:

The Netherlands case

Immigrants to the Netherlands have to pass three stages of testing: admission to the country, civic integration after arrival, and naturalization to citizenship. The Law on Integration Abroad passed in 2006 requires immigrants to take a computerized phone test of the Dutch language called the *Toets Gesproken Nederlands* (using Versant's computer-scoring technology); and, a test of knowledge of Dutch politics, work, history, etc. This requirement is first in the modern world for family unification. It has been criticized on grounds of human rights: as discriminatory because it violates the right to marry and start a family.

Reflection on The Netherlands case

Is this institution's requirement that a spouse demonstrate a certain level of language ability prior to travelling and living in the receiving country a violation of human rights? Is it appropriate to relax this law for citizens of certain countries? How is this policy beneficial to the community? Is this an example of an unjust institution? What could language assessment experts or professionals do about this case?

Conclusion

A principled foundational basis for fairness and justice in language assessment drawing on the work of Rawls and Sen. Two principles, the Principle of Fairness and Principle of Justice were proposed. The principles help to help design and establish fair assessments and just institutions and to remove manifest unfairness and injustice that exists today

So, in conclusion, the two basic questions: How do we set up fair assessments and just institutions? How do we remove unfairness and injustice? One way is by using principles of fairness and justice to design, develop, administer, and analyze assessments so that assessments are beneficial to society. And, by using public reasoning to justify the benefits of assessment.

References

1. Kunnan, A. J. (2014). Fairness and justice in language assessment. In A. J. Kunnan (Ed.) *The Companion to Language Assessment* (pp. 1098-1114). Boston, MA: Wiley-Blackwell.
2. Kunnan, A. J. & Davidson, F. (2004). Situated ethics in language assessment. In D. Douglas (Ed.), *English language tests and testing practice* (pp. 115-132). Washington, D.C.: NAFSA.
3. *Language Testing*, 27(1) (2010). Test Fairness. Special Issue.
4. Rawls, J. (1999). *The law of peoples*. Cambridge, MA: Harvard University Press.
5. Sandel, M. (2009). *Justice: What's the right thing to do?* London, England: Penguin.
6. Sen, A. (2009). *The idea of justice*. London, England: Penguin Books.