

Fairness and Justice in Language Assessment

Antony John Kunnan

Nanyang Technological University, Singapore

Introduction

The concept of *fairness*, as related to assessment and assessment practice, has been debated regularly since the late 1980s by researchers and practitioners. In the field of educational assessment, the concept of fairness was first introduced in employment-related testing of the General Aptitude Test battery (Hartigan and Wigdor, 1989). In language assessment, the term was first discussed at the Language Testing Research Colloquium in Finland when Kunnan (1997) presented a case for a fairness research agenda. Fairness also soon made its way into the influential *Standards for Educational and Psychological Testing* (APA, AERA, & NCME, 1999) with a section titled “Fairness in Testing” and subsections on fairness in testing and test use, the rights and responsibilities of test takers, testing individuals of diverse linguistic backgrounds, and testing individuals with disabilities. Codes of ethics and practice based on this pioneering work have now been established by many assessment agencies such as the International Language Testing Association, Educational Testing Service, Princeton, University of Cambridge, and the Association of Language Testers of Europe, among others. Since 2000, there have been frequent publications on fairness as a concept (Kunnan, 2000, 2004, 2008; Walters, 2012), situated ethics (Kunnan & Davidson, 2004), how to investigate fairness (Kunnan, 2010; Xi, 2010), differential item functioning as a method for detecting biased items (Ferne & Rupp, 2007), accommodations for test takers with disabilities (see Chapter 67, Accommodations in the Assessment of English Language Learners), the intersection of fairness and justice (McNamara & Ryan, 2011), and legal matters related to fairness (see Chapter 92, Language Testing in the Dock).

Disagreements, however, have regularly surfaced in these debates. The first is to do with the interpretation of the term. Depending on the researcher’s

perspective, fairness has meant “absence of bias,” “equal opportunity,” “equitable treatment,” “similar outcomes in terms of scores,” and so on. Additionally, the scope of the term has been contested (Kane, 2010; Xi, 2010): Does fairness include validity or does validity include fairness? Or are they two separate entities? This has led many researchers to set aside the concept of fair assessments as inferior to validation efforts. Davies (2010) even asked whether it was worth paying attention to. Finally, no foundational principles that drive the concept have been articulated; as a result, critics have argued that fairness studies are incoherent (e.g., Bachman, 2005). As a result, fairness is often invoked with ad hoc and post hoc investigations after assessments are written and launched but it is not part of the assessment design, development, administration, and standard-setting processes. A good example is the class of DIF/bias studies that examine test performance, often with no a priori hypotheses.¹

The term *justice*, on the other hand, is rarely mentioned in the assessment literature, although the idea of justice has been discussed in writings from Plato to recent work by John Rawls and Amartya Sen. Once again, the term is difficult to define. A few interpretations include “distributive justice,” which refers to institutions providing benefits that are distributed to a society in a just manner, “retributive or corrective justice,” which refers to whether punishments are just, and “compensatory justice,” which refers to fair compensation for injuries. In language assessment, Kunnan has tied the two concepts of fairness and justice together (Kunnan, 2004, 2008) and McNamara and Ryan (2011) have wrestled with the two concepts to offer separation and clarity to the concepts.

This chapter attempts to provide principled bases for fairness and justice as applied to the institution of assessment. It does this by applying the *idea of fairness as relating to persons—how assessments ought to be fair to test takers*—and the *idea of justice as relating to institutions—how institutions ought to be just to test takers*.

Preliminaries

Despite the disagreements regarding the concept of fairness, the very concept of the public examination (or assessment) includes notions of fairness and justice. This can be seen as the main goal of such examinations, which is to bring about a level playing field in awarding benefits through a process that assesses desired abilities; not to award benefits based on privilege and patronage. This was the main goal of the civil service selection process in China centuries ago, and in France, Germany, the UK, and colonial India in the late 18th and early 19th centuries. In more recent times, related concepts of equality, equal protection, and equal representation have become part of public discourse in most parts of the world, although such discussions have not always resulted in the active promotion of fairness, equality, and civil rights. Thus, in many countries fairness in schooling and employment has been advanced through fair assessments and just institutions.

The more practical aspects of fairness are noticeable. For example, anonymizing examination responses by removing test takers’ personal information so that test takers cannot be identified; the use of topics in test materials that are familiar to test takers; investigations regarding biases toward test

takers from different language, gender, age, and ethnic backgrounds and biases of raters and ratings; checks regarding whether test takers have had the opportunity to acquire the knowledge or skills prior to the assessment, and the use of appropriate accommodations for test takers with disabilities. Therefore, in many contexts, the practical aspects of fairness are not new. With this background, a few preliminary questions can be articulated:

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?

The chapter continues with hypothetical vignettes and real-world examples and reflections on these scenarios. These vignettes exclude concrete realistic details so that we can focus on a limited number of issues. Arguments from normative ethics regarding fairness and justice are then discussed in order to provide appropriate background for the proposed principle of fairness and principle of justice. The chapter concludes with how these principles can be used to advance fairness and justice in language assessment.

Vignette 1: Pretesting of assessment tasks

Imagine a new staff member has joined a large professional language assessment organization (university or commercial) 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 agency’s practices. She took note of them: First, they did not pretest or trial their test tasks; instead, they used the non-pretested tasks in a real administration and did not delete the scores from those tasks when they computed the scores for the test takers. In other words, the test takers received scores that included tasks that were not pretested. The staff member approached her supervisor who was head of assessment development. He was at first disinterested in the staff member’s concern but later admitted that pretesting tasks would cost too much money for the organization, and, if they conducted a pretest, the assessment would also cost the test taker much more.

The main questions here are: Did the staff member do the right thing in bringing to the attention of her supervisor the lack of pretesting? Is pretesting of tasks for high stakes assessments a requirement? Is the head of assessment development’s

lack of understanding of the situation acceptable? Are his reasons for not conducting any pretesting justified? Would people of any persuasion (teachers, test takers, business leaders, etc.) be able to defend such a practice? Is there a violation of an accepted code of ethics and practice? Would this be an example of an unfair assessment?

Vignette 2: Checks for biased assessment items/tasks

Continuing with Vignette 1 . . . The staff member also found that the organization did not conduct any review or investigation 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, the organization did not have staff with expertise to conduct such investigations. He also reminded her that, once again, these investigations would cost the organization a lot of money and the final result would be that the assessment would cost the test taker more.

Once again, the main questions here are: 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?

Arguments From Philosophy

One way to understand the concept of fairness and justice is to step back from its current theory and practice in assessment, and to examine how the concept is used in normative ethics (an important branch of philosophy). In this field, there have been numerous attempts at debating the concept of moral reasoning from normative ethical theories. Ethical dilemmas that we face include right and wrong, fair and unfair, equality and inequality, just and unjust, and individual rights and common good. The main theoretical perspectives and proponents in philosophy are: utilitarianism (Bentham, Mill) and social contract/deontology (Kant, Rawls, Sen).²

Utilitarianism

The dominant Western philosophical doctrine for many centuries until Rawls's work appeared was that of *utilitarianism* advanced by Bentham and Mill.³ Its thinking is that the highest principle of morality is to maximize utility and to balance pleasure over pain. It promotes the notion of the greatest happiness of the greatest number of people. As a result, the utility principle trumps individual rights. Related to this, the most important aspect of utilitarianism is *consequentialist* thinking in which outcomes of an event are used as tools to evaluate an institution. Thus, implementing utilitarianism in the assessment context would

mean that decisions about an assessment would be made solely on the basis of utility and consequences.

Rawls's *Justice as Fairness*

Rawls's (1971, 2001) theory and arguments of justice and fairness have been the basis for wide discussions in moral philosophy/reasoning from the 1970s onwards. His main focus is on inequalities in citizens' life prospects. In formulating a theory and principle of fairness and justice, he argues that *fairness is prior to justice* but foundational and central to justice, and that *fairness relates to persons* and *justice relates to choice over institutions*. In this chapter, these two terms will be used accordingly. To quote from Sen's (2009) summary of Rawls's work:

In the Rawlsian theory of "justice as fairness", *the idea of fairness relates to persons (how to be fair between them)* whereas the Rawlsian principles of *justice apply to the choice of institutions (how to identify "just institutions")*. The former leads to the latter in Rawls's analysis. (p. 72, emphasis added)

The case for justice as fairness that Rawls makes is a moral philosophy for a "well-ordered society" that has a "fair system of social cooperation" and has "citizens who are free and equal persons." Rawls's intention here is that, for his theory to work, society has to be well ordered—in other words, democratic—with a representative government and, as Sen puts it, with a "government by discussion" and not just with elections and balloting. Further, Rawls argues that members of such a society should accept the concept of justice and be free and equal to have the moral capacity to do this.

Rawls presents a procedural plan for how a just institution can provide social arrangements that promote justice. He introduces three inter-related concepts: the hypothetical thought experiment which he called "the original position and the veil of ignorance," "public justification," and "reflective equilibrium." In Rawls's original position and veil of ignorance, members of a society are not allowed to know their social position in society, any of their backgrounds (race, ethnicity, or gender), or their endowments (capabilities, talents). Therefore, as members do not know anything about themselves, they will not be able to gamble to become beneficiaries of any benefits. This setup, Rawls argues, would give members a way to derive principles of justice without any biases or prejudices as they know any decisions they make could affect them.

The principles of justice Rawls posits that would emerge from such a procedure would have unanimous agreement, although he amended this later (2001, p. 32) to accommodate the idea of "overlapping consensus," as he recognized the limits of agreement on justice in pluralistic democracies where conflicting religious, philosophical, and moral doctrines may make unanimity unlikely. Finally, for justice as fairness to work, Rawls contends, public justification is a necessary part of the process. He argues that it is necessary to justify political judgments to fellow citizens so that public consensus can be reached. He also suggests the use of the methodology of "reflective equilibrium" to help in the public justification process. In this methodology, initial ideas, beliefs, or theories are subjected to reason,

reflection, and revision until the ideas, beliefs, or theories reach a state of equilibrium in public justification.⁴

Rawls offers two principles of justice as way of guidance in the design of just institutions. The *first principle of justice* states:

Each person has the same indefeasible claim to a fully adequate scheme of equal basic rights and liberties, which scheme is compatible with the same scheme for all. (2001, p. 42)

This principle includes five sets of basic liberties: liberty of conscience and freedom of thought, freedom of association, equal political liberties, the rights and liberties that protect the rights and liberties of the person, and rights and liberties covered by the law.

The *second principle of justice* has two parts: The first part, the *equal opportunity principle*, is familiar as examinations in general provide fair equality of opportunity. The institution of examination by assessing abilities opens up opportunities that would otherwise be available only in terms of heredity, nobility, and social position by birth. The second part, the *difference principle*, refers to economic opportunities in which the least advantaged members of society are better off when primary goods are unequally distributed (between least advantaged and more advantaged members) than when primary goods are equally distributed between the two groups.

In summary, in Rawls's theory of justice as fairness, the focus is on developing ideal just institutions by identifying what just institutions would look like. This, Rawls proposes, can be achieved in a well-ordered society with free and equal citizens interested in social cooperation to bring about justice as fairness. By using the original position and the veil of ignorance, principles of justice can be publicly justified by the process of reflective equilibrium.⁵ When this is done, principles of justice would emerge as guidance to build a just institution.

Sen's Idea of Justice

Sen advances Rawls's thinking significantly with some major ideas. First, he contends that Rawls's theory of justice as fairness is primarily aimed at the ideal of establishing just institutions (which Sen terms "transcendental institutionalism") and that it does not have any mechanism to evaluate human transgressions that bring about unjust societies through public reasoning.⁶ Sen (2009) contends that Rawls's approach is *arrangement-focused* (justice conceptualized in terms of organizational arrangements like institutions, regulations, behavioral rules, and the active presence of these would indicate that justice is being done). This is in contrast to Sen's view of justice as *realization-focused* understanding of justice (examining what emerges in society, the kind of lives people can lead, given the institutions and rules, but also actual behavior that would inescapably affect human lives).

Second, Sen (2009) invokes Adam Smith's thought experiment of the "impartial spectator." This device, Sen argues, could be used "when judging one's own conduct, 'to examine it as we imagine an impartial spectator would examine it'" (p. 124). Sen argues that this approach of the impartial spectator has a major advantage over Rawls's original position with a veil of ignorance in arriving at principles of justice. Although both approaches attempt to remove the vested interests and goals of individuals, spectators or disinterested people from other societies can participate in deliberations in Smith's approach whereas outsiders are restricted in the Rawlsian approach. Sen argues that Smith's open impartiality—including the voice of the people who do not belong to the focal group—is superior to Rawls's closed impartiality—restricting the voice of the people to focal group members—as it provides the opportunity for cross-societal and crossborder deliberations.

In a related point, the nonparochial, global perspective view is a central part of Sen's thesis. He is concerned about the parochial nature of nations when it comes to the service of justice for two reasons: First, what happens in a particular country in terms of how its institutions operate can have huge consequences on the rest of the world, and, second, each country or society may have parochial practices that need to come under examination and scrutiny from others with distant judgments who are impartial spectators.⁷ Further, the global reach of justice is necessary in a world where globalization is taking place in other areas: trade, commerce, business, travel, technology, and so on.⁸ Similarly, Sen also criticizes the view of Asian government leaders from Singapore, Malaysia, and China regarding "(East) Asian values." Their leaders argue that the denial of political and personal freedoms and suppression of media freedoms in exchange for economic growth are part of "Asian values," different from those of the West. This is defective reasoning, Sen argues, as Asian countries have a tradition of democratic values and principles as well.⁹ Recall Martin Luther King's warning: "Injustice anywhere is a threat to justice everywhere."¹⁰

Third, Sen argues that public reasoning is a critical component in advancing justice. His requirements are similar to those of Rawls: in this case, a well-ordered society—in other words, 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 justification and reasoning. Such states would have in place transparent mechanisms for the fair selection and use of assessments, public reasoning of the assessment in use (in public forums), and regulations and laws that have adequate provisions for appeals and redress. An authoritarian regime, on the other hand, with few or no political and personal freedoms, will be less compelled to need or allow public justification and reasoning of principles of fairness. The lack of such reasoning along with inadequate accompanying regulations and laws for appeals and redress would make it difficult for such institutions to be just.

In summary, Sen argues that the focus of justice must be on the advancement of the cause of justice through the methodology of public reasoning. His methodology for doing this is through the distant judgment of the impartial spectator in his open partiality mode with outsiders and the world examining and scrutinizing

the practices of just institutions. He also indicates the need for a global reach of justice.

Applying Fairness and Justice

General Issues

Drawing on insights from Rawls and Sen on fairness and justice, we can now consider how their ideas and arguments can be applied to language assessment. First, individual rights and inequalities in test takers' life prospects have to be the central focus of the application. The main idea is that assessments ought to be fair and assessment institutions ought to be just to all test takers. Second, Rawls's idea of public justification and Sen's public reasoning have to be part of this application.

Two Principles of Fairness and Justice

Based on Rawls and Sen, two general principles and subprinciples of fairness and justice are proposed:¹¹

Principle 1—*the principle of fairness*: An assessment *ought* to be fair to all test takers, that is, there is a presumption of treating every test taker with equal respect.

Subprinciple 1: An assessment *ought* to provide adequate opportunity to acquire the knowledge, abilities, or skills to be assessed for all test takers.

Subprinciple 2: An assessment *ought* to be consistent and meaningful in terms of its test score interpretation for all test takers.

Subprinciple 3: An assessment *ought* to be free of bias against all test takers, in particular by avoiding the assessment of construct-irrelevant matters.

Subprinciple 4: An assessment *ought* to use appropriate access, administration, and standard-setting procedures so that decision making is equitable for all test takers.

Principle 2—*the principle of justice*: An assessment institution *ought* to be just and bring about benefits in society and advance justice through public reasoning.

Subprinciple 1: An assessment institution *ought* to bring benefits to society by making a positive social impact.

Subprinciple 2: An assessment institution *ought* to advance justice through public reasoning of their assessment.

A few remarks regarding the principles are necessary here. To begin with, the first principle, the principle of fairness, is prior to the second, the principle of justice, because if the first principle is not satisfied, then the second principle cannot be satisfied. In other words, if the presumption that treating every test taker with equal respect in an assessment is not satisfied, then the assessment will not succeed in being beneficial to society and bring justice to society. In terms of the relationship between the general principles and the subprinciples, the respective subprinciples provide the framings for the two general principles, and

therefore the subprinciples have to be individually satisfied in order for the general principle to be satisfied.

Second, the principles and subprinciples are written as obligations (obligatory actions signaled with the use of *ought*) and not as categorical or unconditional imperatives, but the assumption is that there will be universal application. As argued earlier, justice should be nonparochial and impartial beyond one's society as everyone should be treated in the same manner. This is particularly true of current globalized assessment institutions that operate in numerous countries. It does not seem defensible to propose otherwise, despite the objection of being imperialist as to how there could be different approaches to fairness and justice.

The Principle of Fairness: Grounds and Objections It is necessary to explicitly make a general case for this principle by articulating the grounds and rejecting some common objections. First, the principle states that an assessment ought to be fair to all test takers, which includes a presumption of treating every test taker with equal respect. This emphasis on the test taker rather than an assessment or its scores should be sufficient to reject the argument that validity of an assessment (or valid score interpretations or validity arguments) guarantees that all test takers will be treated with equal respect. The focus of validity concerns has either been on the assessment itself or at most on various aspects of assessment practice; the focus has never been on the individual test taker.

Second, the subprinciples provide guidance for detailed investigations of assessments so that a number of grounds for the compliance or noncompliance with the principle can be arrived at. Researchers could conduct investigations relevant to the subprinciples to build arguments regarding the general principle of fairness. The subprinciples focus on the test takers' opportunity to learn, the meaningfulness of the assessment to the test taker, and whether the assessment is free of bias and standard setting has been conducted in an equitable manner. These matters are relevant to the individual test taker and affect the test taker positively or adversely depending on the qualities of the assessment. Thus, they are essential components of the general first principle.

The Principle of Justice: Grounds and Objections As mentioned earlier, this principle follows the first principle but it is a necessary component. First, the principle states that an assessment ought to bring about benefits to society and that such institutions promote just institutions. The overall benefit to society should be the primary motivation to build any assessment in society; that is, if the motivation to build an assessment is not to resolve some difficulties and bring about benefits to society, one could conclude that there may not be a need for the assessment. This is particularly the case if an assessment is likely to cause adverse effects on test takers and society.

Second, the institution of assessment is not any different from institutions like banks or universities. But assessment institutions have a higher responsibility in society than the Department for Beautiful Gardens as assessment institutions are responsible for awarding benefits to test takers that can alter their life prospects. If such institutions bring benefits and just institutions, then the principle of justice is satisfied.

Finally, while it is possible that assessment institutions may have different ways of defending their assessments, it is essential that there is public reasoning of assessments. This can be offered through public forums such as conferences or research reports available to the public.

Challenges From Vignettes

Let us now consider vignettes from different assessment contexts that illustrate some real-world challenges. These challenges need to be resolved by assessment developers, administrators, score users such as administrators, teachers, and test takers, and decision makers. Some of the vignettes are related to the principle of fairness while others are related to the principle of justice.

Vignette 3: Defective tasks

Imagine there were two forms of a paper and pencil assessment—Forms A and B. The forms belonged to a high stakes university admissions assessment. It was known previously through pretesting 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 cost of replacing the defective tasks (designing and writing new tasks, pretesting them, assembling them into the forms, and printing the new assessments) would be much higher than that of errors in classification, although she did not assign a monetary value to the misclassified test takers’ lost opportunities due to the errors.

If we consider the different philosophical persuasions, each may take a view that supports or criticizes the actions of the administrator. The utilitarian could take the view that the cost–benefit analysis provided the basis for the administrator’s decision, and that such decisions have to be made in order to run a profitable business. The utilitarian could also concede though that the administrator should have preferred Form B to Form A as it had utility. The contractarian could argue that the administrator did not act morally as she did not uphold the rights of all test takers to a fair assessment by holding defective assessments. These arguments could lead us to some important questions: What should the administrator have done? Which of these perspectives appeals to us? What is the right thing to do?

Vignette 4: Compensation for misclassification

Continuing with Vignette 3 and expanding it . . . Imagine further that the administrator was convinced of her error and agreed to pay compensation to the test takers who were misclassified as failed. She offered a free retake of the assessment at a later date (as per the contract issued by the agency) but the results would be available only after the completion of the university

admissions cycle. The test takers were not satisfied with the remedy offered: Some test takers wanted more compensation, while others planned to file a law suit in court against the assessment agency.

This action raises additional questions regarding the right thing to do. The utilitarian could claim that the consequences of the assessment were mostly successful as most test takers were assessed appropriately, thus satisfying the principle of maximizing utility and the maxim of “greatest good for the greatest number of people.” Further, as the assessment did not provide sufficient utility for these test takers who were erroneously misclassified, they were offered compensation as per the contract. The contractarian could argue that the administrator did not carry out her duty and therefore should be tried for dereliction or breach of duty, as she did not uphold the rights of all test takers to a fair assessment. The argument could then be made that the administrator and her agency should face a tort, product liability, or a similar lawsuit that would be available in a just institution. These are simplified arguments from different perspectives, but they nevertheless indicate how difficult it is to do the right thing. Both the vignettes above pose problems for the principle of fairness and in particular subprinciples 3 and 4, and the principle of justice and in particular subprinciple 2.

Vignette 5: Selecting an assessment

Now imagine a school teacher—or a group of teachers—were authorized by the school principal to choose an assessment for a grade 8 reading class in English. The teachers had to choose from three assessments that were commercially available: Test A was developed by a well-established company known for its quality products, the test was traditional and was broadly suitable in terms of content, it was normed for the national population, and it cost \$500 for a class of 40. Test B was developed by a small local company, it was highly suitable in terms of content, it was normed for the local population and checked for fairness, it was proven to offer accurate results and useful diagnostic information, and it cost \$800 for a class set of 40. Test C was an innovative test developed by teachers from another local school, it had not been analyzed yet but was available for free for the class.

The main question is: On what grounds should the teachers choose from these assessments? Should the decision be based on cost? If the decision were to be made on this ground, the choice would be Test C even though the assessment had not been analyzed; the argument could be that the assessment was not high stakes in grade 8. Another consideration would be consequences and fairness—whether the assessment would have consequences that were beneficial to the students and the community. If the decision were to be made on this ground, the choice would be Test B. This would appeal to the consequentialist doctrine. Yet another consideration would be the quality of the organization developing the assessment. If this was the ground, then it would be Test A. You will see that the choice is not straightforward and the teachers have to weigh several factors such as quality, cost, and so on in order to make their choice.

Vignette 6: What was the quality of the assessment?

Imagine a high stakes high school exit examination that is conducted by the ministry of education in a country. The examination had been in use for many years and students worked hard during the months prior to the examination. After the examination, some students got together and exchanged thoughts on the examination. They concluded that some of the questions were tasks and topics that were new to them. When the results were announced, it turned out these students had received low grades. Apart from feeling upset, the students could do nothing else (there was no review or appeal process in place) but their parents went to the ministry and complained that something was wrong with the examination. The ministry officials said that there could not be anything wrong as their examinations were written by expert senior teachers who had been doing this for decades. When pressed to show that the examination was an appropriate assessment procedure, the ministry officials defended their examination by saying there were no prior complaints and therefore no analysis of the examination was conducted as it was not necessary.

The main question here is whether the ministry had the motivation and expertise to provide the best possible examination. There were numerous problems with the examination from an equal rights perspective (a Rawlsian concern): Did some of the students not have the opportunity to learn all the material? Did their school or teacher perhaps not cover all the topics? In which case, did the assessment have utility (a utilitarian concern)? Did the ministry's regulations not have any provision for review or appeals? Further, were there no analyses of the examination tasks conducted although they were written by experts? 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 ministry owe the student community public justification (Rawlsian requisites)? Finally, was the ministry acting responsibly? In general, Rawlsian theorists would in particular be up in arms with the ministry's lack of provision for basic rights to the students and its dereliction of duty to the students and community.

Vignette 7: Public reasoning of the assessment

Continuing with Vignette 7 and expanding it . . . Imagine that the parents of the students who received low grades protested against 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.

Once again, the 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. They could also argue that public reasoning (in public forums) would be the only way to ensure that the assessment is fair and the institution is just. Both these vignettes pose problems in terms of the principle of fairness and all its subprinciples. Vignette 7 in particular poses a problem for the principle of justice and its subprinciple 2.

Vignette 8: The role of differential pricing

Imagine you received a questionnaire from a well-known assessment developer and publisher who is planning on introducing differential pricing to test takers for different services. Which of these would you find acceptable? The proposal is higher test-taker fees for new services. Thus, there would be two levels of pricing: regular pricing for regular assessment and premium pricing for additional services. Here are the additional services for premium pricing: Better assessments with higher reliability and thorough validation and fairness studies, individual diagnostic feedback instead of generic feedback, better raters who are experienced and not severe in their ratings, faster turn-around time for results, front row seating for the listening section (where audio speakers are in front of the test room), fast-track line for speaking tests/interviews, better test room facilities (air-conditioning, heating, plus seats), relaxed time conditions (more and frequent breaks), assistance from spell and grammar checks for the writing test, upgraded technology (computers, monitors, keyboard and mouse, color photos, and video), accommodations for test takers with disabilities, no experimental section included in the assessment, repeat assessment within a few days, re-evaluation of assessment performance by two human raters, and return of responses to tasks to test taker (selected and constructed responses on items/tasks).

The main questions are: For which of these services would we consider differential pricing appropriate? On what grounds would we accept or not accept differential pricing? Is the market forcing us to change our ethical behavior? Is there an obligation on the part of the assessment developer to offer some of these services without differential pricing? Would the differential pricing for some of these services violate the principles of fairness and justice?

Advancing Fairness and Justice

The earlier section considered how Rawls's and Sen's ideas can be applied to language assessments in order to design and establish just institutions by examining hypothetical vignettes. But it is also important to simultaneously explore how institutions can advance the principles of fairness and justice and remove existing unfairness and injustice in current society. Any example of unjust practice should motivate skeptics about the need for action should such practices be identified. For example, recall the discriminatory practice behind the dictation test given to immigrants in Australia in the 1900s during the country's White Australia policy (McNamara & Ryan, 2011). If such a policy were to be in place now, we could ask on what moral principles such an assessment could be defended. One real-world example from language assessment for immigration and citizenship that presents a serious challenge to fairness and justice is important to consider.

Assessing Immigrants in the Netherlands

A new real-world example is unfolding in the Netherlands (see Kunnan, 2012, for the context). Immigrants to the Netherlands now have to pass three stages of

testing in order to become citizens: admission to the country, civic integration after arrival, and naturalization to citizenship. The Law on Integration Abroad passed in 2006 described what applicants for admission to the Netherlands need to do—they have 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, education, health care, history, and living. This type of requirement is the first in the modern world as it clearly presents barriers to family unification (particularly for women in Morocco and Turkey) and has been criticized on grounds of human rights. Extra and Spotti (2009) cite a Human Rights Watch (2008) report that considered this testing regime as discriminatory "because it explicitly applies to particular 'non-Western' countries and because it violates the qualified human right to marry and start a family" (p. 133). A legal challenge in a Dutch court in 2008 of this regulation resulted in a Dutch court providing relief to a Moroccan woman plaintiff who challenged the admission tests in Morocco because she failed the test and thus was not admitted to the country. After an appeal by the state, a higher state court ruled in February 2010 that applicants for admission to the Netherlands will no longer face a Dutch language test and questions about knowledge of Dutch society. The court's decision now will allow immigrants from Turkey and Morocco to apply for temporary residence on the basis of family reunification, and Dutch citizens would be able to bring their spouses from outside the country without having to pass the test of Dutch language and knowledge of Dutch society. Such testing was already considered to be in violation of the equality principle because citizens from European countries and others (such as Australia, Canada, Japan, New Zealand, the USA, etc.) do not need to take these tests.

This example highlights the right of a sovereign state to demand certain abilities (including language and knowledge of history and civics or culture),¹² on the one hand, and the right of an individual to join his or her family based on marriage (irrespective of any abilities that may be needed excluding any moral grounds). The decision of the Dutch government to deny this basic freedom demonstrates that they are denying respect for human dignity. Further, the law only applies these requirements to some groups of immigrants, but allows others to immigrate without the required Dutch language ability.

Leaving aside legal issues, what are the ethical issues here? Is this institution's requirement that a spouse demonstrate a certain level of language ability prior to traveling 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? If this is the case, what could language assessment professionals do about this?

Conclusion

This chapter has provided a principled foundational basis for fairness and justice in language assessment by drawing on work from Rawls and Sen. Applying arguments from Rawls and Sen to language assessment has provided the background that led to the principles of fairness and justice as instruments being used

to evaluate assessments and assessment institutions. The methodology of how principles of fairness and justice may be derived by assessment agencies remains a concern. Rawls and Sen offer ways in which this can be done through the original position/veil of ignorance and the impartial spectator. These or other methods should help put in place mechanisms that enable the development of fair assessments and just institutions.

As Sen argues, it is not sufficient to use principles to design and establish just institutions: efforts should be made to remove manifest injustice that exists in the world today. He argues that a nonparochial global justice view would be best suited to establish and review unjust institutions. This is critical with globalized assessment institutions. These institutions need to be evaluated by enforcing categorical imperatives with obligations and *ought-to* principles, particularly general principles. The chapter also put forth the idea that there should be public reasoning of assessments. This would mean that whether an assessment is fair or not and whether an institution is just or not should be a matter of public discourse for which public reasoning is necessary. This is a critical part of justifying fair assessments and just institutions.

Finally, the main point of this chapter is not to debate whether the putative principles of fairness and justice (proposed in this chapter) are appropriate or workable for all contexts but to find principles that can guide us to right action when we encounter examples of unfair assessments and unjust institutions. We hope therefore the answers to the preliminary questions that were raised at the beginning of the chapter can now be answered by focusing on the right thing to do in setting up fair assessments and just institutions and the right thing to do to remove any unfairness and injustice.

SEE ALSO: Chapter 65, Evaluation of Language Tests Through Validation Research; Chapter 76, Differential Item and Testlet Functioning Analysis; Chapter 85, Philosophy and Language Testing; Chapter 94, Ongoing Challenges in Language Assessment

Notes

- 1 ETS, Princeton, is one of a few agencies that have publicly declared they have a sensitivity review phase in which all items are subject to careful review.
- 2 Other perspectives include virtue ethics, cosmopolitanism, communitarianism, and postmodernism.
- 3 This does not include traditional and theological ethics such as the divine command theory or the natural law theory.
- 4 Rawls's reflective equilibrium is remarkably similar to Habermas's idea of discourse ethics in which principles are not mirrors of truth but something that emerges from fair argumentation by members in a society.
- 5 Rawls is a secular theorist and his theory is in this tradition. There are traditional virtues from religion-based ethics that might overlap with the major secular theories. These could include virtues of consciousness, benevolence, and self-restraint (from Buddhist ethics), humanity and goodness, rightness and duty, consideration and

- reciprocity, loyalty and commitment (Chinese ethics), neighborly love, natural morality (Judeo-Christian ethics), social and individual duties (classic Hindu ethics), and charity, kindness, and prayer (Islamic ethics).
- 6 This idea is similar to the thinking of Schopenhauer, although Sen does not mention him.
 - 7 Although Sen does not put it this way, we can assume he advocates, in Cohen and Sabel's words, "equal concern, equal respect, and equal opportunity regardless of any background conditions" (2006, p. 148). Sen cites many examples of this parochial nature: the common practice of the murder of newborn infants in ancient Greece despite the presence of Aristotle and Plato in their midst; more recently, stoning of adulterous women in Taliban Afghanistan, selective abortion of female fetuses in China, Korea, and parts of India, female genital mutilation in parts of Africa, capital punishment in China and the USA and restrictions of personal and media freedoms in North Korea and China.
 - 8 Nagel (2005) argues against global justice: that outside the state, there is no justice, and therefore as there is no global state, there can be no global justice. But does this mean that a state does not have any humanitarian obligations to its citizens even if the obligations do not lead to egalitarian justice?
 - 9 Sen (1999) illustrates this through a variety of examples that show the use of democratic ways from the past in Asian states. Examples include democratic Buddhist councils in the first and second century AD; Japanese Prince Shotoku's "Constitution of 17 Articles" in the seventh century AD; Emperor Ashoka of India in the third century BC; and the Moghul Emperor Akbar in the 16th century. It is also not true that all Western or European states are liberal democracies with media freedoms and all Asian states are authoritarian. There are counterexamples too: the early to mid-20th century saw authoritarian regimes in Germany, Italy, Spain, and apartheid South Africa. In contrast, in the mid- to late 20th century, there have been Asian democracies with full media freedoms in India, Japan, South Korea, and Taiwan.
 - 10 But Sen, like Rawls, for different reasons shies away from a full cosmopolitan approach to justice which holds moral universalism as paramount to moral value.
 - 11 Previous versions of the principles were presented in Kunnan (2000, 2004). This revision is more extensive.
 - 12 Whether these abilities as assessed currently around the world can contribute to social integration of immigrants is an important but different question.

References

- APA (American Psychological Association), AERA (American Educational Research Association), & NCME (National Council for Measurement in Education). (1999). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2, 1, 1–30.
- Cohen, J., & Sabel, C. (2006). Extra rempublicam nulla justitia? *Philosophy & Public Affairs*, 30, 147–75.
- Davies, A. (2010). Test fairness: A response. *Language Testing*, 27, 171–6.
- Extra, G., & Spotti, M. (2009). Testing regimes for newcomers in the Netherlands. In G. Extra, M. Spotti, & P. Van Avermaet (Eds.), *Language testing, migration and citizenship* (pp. 125–47). London, England: Continuum.
- Ferne, T., & Rupp, A. (2007). A synthesis of 15 years of research on DIF in language testing: Methodological advances, challenges, and recommendations. *Language Assessment Quarterly*, 4, 113–48.

- Hartigan, J., & Wigdor, A. (1989). *Fairness in employment testing*. Washington, DC: National Academy Press.
- Kane, M. (2010). Validity and fairness. *Language Testing*, 27, 177–82.
- Kunnan, A. J. (1997). Connecting validation and fairness in language testing. In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S. Luoma (Eds.), *Current developments and alternatives in language assessment* (pp. 85–105). Jyväskylä, Finland: University of Jyväskylä.
- Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 1–14). Cambridge, England: Cambridge University Press.
- Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context* (pp. 27–48). Cambridge, England: Cambridge University Press.
- Kunnan, A. J. (2008). Towards a model of test evaluation: Using the test fairness and wider context frameworks. In L. Taylor & C. Weir (Eds.), *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity: Papers from the ALTE Conference in Berlin, Germany* (pp. 229–51). Cambridge, England: Cambridge University Press.
- Kunnan, A. J. (2010). Fairness matters and Toulmin's argument structures. *Language Testing*, 27, 183–9.
- Kunnan, A. J. (2012). Language assessment for immigration and citizenship. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 162–77). New York, NY: Routledge.
- Kunnan, A. J., & Davidson, F. (2004). Situated ethics in language assessment. In D. Douglas (Ed.), *English language tests and testing practice* (pp. 115–32). Washington, DC: NAFSA.
- McNamara, T., & Ryan, K. (2011). Fairness versus justice in language testing: The place of English literacy in the Australian Citizenship Test. *Language Assessment Quarterly*, 8, 161–78.
- Nagel, T. (2005). The problem of global justice. *Philosophy & Public Affairs*, 33, 113–47.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Rawls, J. (2001). *Justice as fairness: A restatement* (E. Kelly, Ed.). Cambridge, MA: Harvard University Press.
- Sen, A. (1999). *Development as freedom*. New York, NY: Random House.
- Sen, A. (2009). *The idea of justice*. London, England: Penguin Books.
- Walters, F. S. (2012). Fairness. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 469–94). New York, NY: Routledge.
- Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27, 147–70.

Suggested Readings

- Freeman, S. (2007). *Rawls*. New York, NY: Routledge.
- Rawls, J. (1999). *The law of peoples*. Cambridge, MA: Harvard University Press.
- Sandel, M. (2009). *Justice: What's the right thing to do?* London, England: Penguin.
- Smith, A. (1759/2009). *The theory of moral sentiments*. London, England: Penguin Books.