Annual Review of Applied Linguistics (2009) **29**, 3–20. Printed in the USA.
Copyright © 2009 Cambridge University Press 0267-1905/09 $16.00
doi:10.1017/S0267190509090023

1. TEST USE AND POLITICAL PHILOSOPHY

Glenn Fulcher

Since Messick (1989) included test use in his validity matrix, there has been extensive debate about professional responsibility for test use. To theorize test use, some researchers have relied upon Foucault's social criticism, thereby stressing the negative role of tests in the surveillance of the marginalized. From a wider perspective, Shohamy (2001a) sees negative test impact as stemming from centralizing agencies, which still leaves open the possibility of positive test use. In this article I argue that how tests are used is a reflection of the wider political philosophy of a society. Political philosophy can generally be characterized as placing more emphasis on either the state or the citizen, leading to collectivist or individualist solutions to problems, be they real or perceived. In collectivist societies, tests, like history, are used to achieve conformity, control, and identity. In individualistic societies, they are used to promote individual progress. The role of tests within each broad approach will be described and illustrated. Finally, I briefly describe effect-driven test architecture as a method for testers to proscribe unintended uses of their tests.

Introduction

In recent years language testers have taken a critical interest in the *use* of tests. Motivated by the notion of consequential validity (traceable to Messick, 1989), there has been an active debate about how far language testers should become involved in questions of test use and ethics (Davies, 1997a, 1997b). The judgment appears to be that these are issues language testers cannot ignore (Davies, 2008). For example, there is concern about the surreptitious use of language tests by states to achieve political goals (Fulcher, 2004, 2008; Shohamy, 2001a), or as tools in immigration policy to achieve ends that would otherwise appear illiberal (McNamara, 2005, 2008). The "meaning" of a test is being conceptualized either in terms of its role in policy (McNamara & Roever, 2006), where the test is usually co-opted into service for which it was not designed (Fulcher & Davidson, 2009), or in terms of its intended effects on stakeholders (Davidson and Fulcher, 2006; Fulcher & Davidson, 2007).

One problem with the current discussion of test misuse is that examples are cited in isolation from a theory of *why* tests are misused, even if discussions of these

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

examples draw upon ethical theory, or the less developed and controversial notions of "test fairness" (Xi, forthcoming). Where social theory comes into play, the appeal is to Foucault's view of a test as the normalizing judgment of authority (Foucault, 1975). Although providing useful insights, this is too narrow a theoretical base, for Foucault represents a recent self-destructive existentialist tradition in European philosophy that accepts the utter inability of the marginalized to take action against the unknowable and unchallengeable powers of the state, society, and institutions. The best we can do is "problematize" our condition, but not free ourselves from it, as any such attempt would itself be conditioned by the social power relations embedded in our culture (Foucault, 1990). So while Foucault's philosophy is a call for us to "uncover" the hidden uses of tests, it cannot provide a guide to positive, ethical test use (Fulcher & Davidson, 2008).

Shohamy (2001a, pp. 28–35) takes us further by considering the use of tests in centralized and decentralized systems. She argued that centralized systems use tests to control large-scale educational systems by defining what kind of knowledge is prestigious. In these systems "the primary goal is to make teachers teach and students study specific topics" (p. 34). The tests are externally imposed, standardized, one-shot, high-stakes assessments. This analysis begins to provide a theory of test use in society, but it does not provide an analysis that relates test use to the larger goals of states, of which centralization or decentralization are but symptoms; what we require is an understanding of the political philosophies that lead to these tendencies and associated ways of using tests as policy tools.

The starting point for any political philosophy is our view of the relationship between the state and the individual. The first discussion we have is Plato's *Republic*, in which it is argued that individualism puts the state in danger of disintegration. Who we are is defined by the state and our role within it. First, as Plato's Socrates reminds us, "a state is larger than an individual" (Plato, trans. 1987, p. 117), and only if individuals "devote their full energy to the one particular job for which they are naturally suited" can "the integrity and unity of both the individual and the state . . . be preserved" (p. 190). To achieve this, a state education requires a strong central curriculum. Any innovation that would encourage corrosive individualism is proscribed, so that each person is educated to fulfill the role allotted to them (pp. 191–193). Plato also makes explicit the role testing plays in such a society, for in choosing the leaders, "we must see how they stand up to hard work and pain and competitive trials. . . . And any Guardian who survives these continuous trials in childhood, youth, and manhood unscathed, shall be given authority in the state. . . . Anyone who fails them we must reject" (p. 180).

The opposite view is most clearly articulated by Pericles in the funeral oration recorded by Thucydides in Book II (1954, pp. 37–46) of the *History of the Peloponnesian War*. In the Athenian democracy the starting point was the plurality of individuals, in which "everyone is equal before the law" (p. 145), and in which everyone may participate. Individualism is seen to strengthen democracy because options and risks can be considered by many minds before decisions are taken. Each person can contribute according to his own abilities without restriction: "I declare that

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

in my opinion each single one of our citizens, in all the manifold aspects of life, is able to show himself the rightful lord and owner of his own person, and do this, moreover, with exceptional grace and versatility" (pp. 147–148). In a free democracy an uncensored education is available for all, so that society is strengthened by drawing on the diverse talents of the many.

All political philosophy since the 4th century BC has been concerned with the fundamental balance between state and individual. Depending on where a political philosophy stands on the cline between the two, we can identify the kind of government likely to be favored, and the kind of society valued. It is my contention that it also explains (and predicts) the uses of tests that we are likely to find.

Collectivism and Testing

Collectivist societies are those in which the identity, life, and value of the individual is determined by membership of the state and its institutions. Decisions are made to benefit the collective and its survival rather than its individual members. Collectivism has been the predominant political philosophy throughout human history, allowing individualism to take front stage only in the coming together of Enlightenment ideas, late protestant individualism in Europe, and the beginnings of the industrial revolution. In the late 18th century individuals in Europe and the United States were no longer willing to play their part in collectives that ignored their rights and brutalized their lives. Locke (1690/2002, p. 22) had already set the tone: "Lay then the axe to the root, and teach governments humanity." The Enlightenment was subversive. Divine rights were no longer tenable. The security of an elite, unrepresentative, and unelected political class was undermined. But long before the Enlightenment project could be finished, neocollectivist movements appeared in the form of extreme political ideologies in the 19th and early 20th centuries. Collectivisms prior to the Enlightenment were untroubled by notions of freedom, democracy, reason, and individual liberty. Neocollectivisms had to survive in a world that had experienced the American and French revolutions, as well as the popular antigovernment uprisings across Europe in the mid-19th century (Sperber, 1994). Tillich argued that these new collectivisms, which we still face today, had to adapt in order to control populations in more subtle ways:

> First, neocollectivism is preceded by the liberation of autonomous reason and the creation of a technical civilization. It uses the scientific and technical achievements of this development for its purposes. Second, neocollectivism has arisen in a situation where it meets many competing tendencies, even within the neocollectivist movement. Therefore it is less stable and safe than the older forms of collectivism. This leads to the third and most conspicuous difference: the totalitarian methods of present collectivism in terms of a national state or a supra-national empire. The reason for this is the necessity for a centralized technical organization and even more for the suppression of tendencies which

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

> could dissolve the collectivist system by alternatives and individual decisions. (Tillich, 1952, p. 97)

The most obvious example of extreme neocollectivism is that of Germany from 1933 to 1945, which also used education to achieve control over its people. "The regime endeavored to assert its control over the education system through reorganization and centralization" (Noakes & Pridham, 1983, p. 433) through the nationalization of tests and the introduction of a strict standards-based system, in which curriculum and content was set by the Ministry of Education. Teachers were not allowed to innovate, and the state controlled both teacher education and certification (Cecil, 1971). Personal intellectual development was abandoned as a goal of learning, and "it was preferred that people should not have a will of their own and should totally subordinate themselves" (Cecil, p. 428). The authorities created a new myth of the state based on a biological reinterpretation of history (Cassirer, 1946) that had begun in the 1850s (Gobineau, 1999), and in which:

> The people and the fatherland place their hands on the shoulders of the youth and determine what educational and cultural values and goals are necessary for this youth to meet the needs of people and fatherland. This requires a truly national and social education for the German youth, and all involved in education have to serve these educational and cultural goals with their full energy. (German Propaganda Archive, 1937)

Education and testing, along with a mythological reading of history, are organized to create subservience to the collective. Individuality is ignored, and educational policies are designed to protect and intensify the identity and future of the collective in terms of its past. Popper (2002) called this "historicism," and argued that equating "true happiness" and "justice" with the well-being of the collective links Plato's view of education and society to that of National Socialism and other forms of neocollectivist government (Popper, p. 186). Thus, centralization of educational systems and testing is but a symptom of a deeper political philosophy that can be traced directly to Plato.

Nor are established democracies immune from neocollectivism, for collectivist tendencies thrive wherever there is a perceived external threat. In the United Kingdom, for example, at the turn of the 20th century there was a fear of the loss of empire after the disastrous Boer War (1899–1902) and the loss of economic supremacy to the emerging American and German superpowers. Policymakers in the United Kingdom drew upon American eugenics (Black, 2003) when they began to see the degeneration in national intelligence as the cause of state decline. For four decades, the solution to this perceived problem was testing for allocation to suitable education and employment (Evans & Waites, 1981, pp. 48–49). Spearman (1904, p. 271) argued that language tests (particularly Greek) were more saturated with "g" (general intelligence) than other tests, and therefore especially valuable. Educational practitioners such as Cattell (1937), working in the United Kingdom during the 1930s, drew on evidence from mass testing in the First World War (Yerkes, 1921) to

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

argue that structuring an efficient society through testing was essential to national survival. Along with many others he saw an opportunity to regiment society, even to the extent of following Terman's (1919, p. 17) advice on "determining vocational fitness" and placing the least able in the lowest ranks of the military—a policy for which he praised Mussolini (Cattell, 1937, p. 51). Cattell's work shows that by the 1930s educational and psychological testing had become part of the technology of a neocollectivist democracy, with a tendency to become less and less democratic:

> Many people argue today that because the management of a modern society in any true sense demands considerable appreciation of economic and biological technical matters, the day of democracy is at an end. They point to oligarchic Germany where the community boldly acts upon the wisdom of the biologist and the medical man. . . . I am convinced that the issues are not by any means either beyond the interest or the capacity of a democracy such as ours. (Catell, pp. 88–89)

Although such sentiments could not be expressed after the Second World War, the selective and elitist education advocated by Cattell and his colleagues (such as Burt) were embedded into the British Education system until the advent of Comprehensive Education in the 1970s.

Weber (1947) understood the educational policies of Germany and the United Kingdom, different as they were in scale, in terms of the drive for modern technical collectives to maximize production and efficiency. This involves the application of science to both industry and war, the specialization of labor, and bureaucratization. Testing is the technology that allows a collective to select the most appropriate person for a task: "Aptitude, regardless of whether it is the product of hereditary or environmental and educational influences, can only be determined by testing" (Weber, p. 261).

Since the war, collectivist governments have been benign, for the most part. But they embody many features that make Foucault's (1975) social critique of their use of tests highly relevant. Indeed, testing has become ubiquitous as all skills, including language learning, are interpreted in terms of their economic value (Fulcher, 2008). Even universities are subject to "the new regimes of accountability, and strict adherence to the economic imperative" (Harris, 2008, p. 347). Examples of centrally controlled standards-based education systems, with a high level of control over teacher training and school learning, are not hard to find (Brindley, 2008). The clearest example is that of the United Kingdom, which has systematically introduced standards-based testing in an accountability framework that ensures total state control over the national curriculum and national tests, as well as teacher training; even educational staff are rewarded or disciplined based on national league tables (Mansell, 2007). Collectivist authorities operating within modern democracies have not learned the lessons from totalitarian collectives, that "whatever the exact character of the built-in safeguards, the best Ministry of Education is that which interferes least in the operation of the system" (Cecil, 1971, p. 4). The reason why such collectivist

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

authorities continue to pursue hyperaccountability policies is because there is once again a sense that the state is failing to perform well in the global market place, and the educational system is reengineered to deliver the kinds of people who will serve the perceived needs of the economy (Fulcher, 2008).

At the supranational level one example of the use of a system that is increasingly being used to harmonize and control language learning, to deal with perceived threats such as a weakened position in global markets, is the Common European Framework of Reference (CEFR) (Council of Europe, 2001; Fulcher, 2004). Its primary use is emerging as a tool for designing curricula, reporting both standards and outcomes on its scales, and for the recognition of language qualifications through linking test scores to levels on the CEFR scales. We now see stronger evidence for more intrusive collectivist policy emerging in calls for claims of linkage to the CEFR to be approved by a central body (Alderson, 2007), and the removal of the principle of subsidiarity from language education in Europe (Bonnet, 2007). If realized, these changes would lead to unaccountable centralized control of education and qualification recognition across the continent. However, many language educators have been convinced that without implementing European language policy in full and without question, language education will not serve the European economy well (Kelly, 2004). This is closely tied to the implementation of the Bologna declaration (1999) to harmonize and standardize higher education, including the structure and content of programs. With this package come "new systems of quality assurance" and a "restrictive view of education" (Harris, 2008, p. 349) that can only be understood as the action of a neocollective mobilizing its human resources to combat perceived economic threats through standardization. The language of Bologna and other documentation surrounding standardization fabricates a set of specifically "European values" that a European citizen will hold; simultaneously "such ways of speaking also create 'the other,' the one who does not have this European essence" (Fejes, 2008, p. 525). References to "toleration" and "appreciation of diversity" create illusory symbols to act as "a technique which fosters nations to desire to participate in standardizing their higher education system at the same time as it fosters students to desire to become a specific European citizen" (Fejes, p. 526).[1] The contradictory rhetoric that diversity can only be achieved through standardization to a common model is all too often left unchallenged.[2]

Using tests to impose standardization and achieve political goals in this way is not new. Charlemagne is also known to have imposed a centrally controlled curriculum upon the monastic schools of Europe as he developed an educational system to control his empire: "In textbooks, as in fundamental practice, the Carolingians pressed for uniformity, and as such an examination was one form of pressure" (Jones, 1963, p. 23). The earliest example of this is a report of an examination board to an unknown central authority in 809 CE. The test was a group oral examination for "computists" who taught principles of calculating the time relevant to the celebration of moveable festivals. There are no objectively "correct" answers to the questions set, but as they are all based on Bede's *On the Reckoning of Time*, the calculations of which were being imposed on the Western liturgical calendar, "the examiners had as one of the several aims the enforcing of a strict party

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

line" (Jones, p. 23). This example should also make us wary of the current tendency to appeal to a so-called golden age of European unity in the construction of a new myth for the foundation of a superstate.[3]

In this section, I have explored the nature of collectivist societies and their use of tests. Modern collectives use testing as a "scientific and technical tool" to control the educational system with the intention of (1) creating or reinforcing the identity of the state, (2) increasing the sense of belonging to the state, (3) selecting and allocating individuals to roles or tasks that benefit the collective, and (4) introducing hyperaccountability to ensure uniformity and standardization. This may not imply centralization where the state does not have the power to centralize, but in such cases it may create social and economic structures that would lead to anyone who does not comply being at a significant disadvantage. This may be the case with the financial and other penalties available under the No Child Left Behind legislation in the United States (Chalhoub-Deville & Deville, 2008, p. 512). It is certainly the case with the CEFR, where recognition of curricula or tests increasingly requires linkage to the CEFR, and "the main goal of this linking exercise is to demonstrate compliance with a mandate" (Kaftandjieva (2007, p. 35). Validity is conceptualized as test recognition by institutions, and validity evidence becomes the extent to which linkage is demonstrated by institutionally approved procedures (North, Figueras, Takala, Van Avermaet, & Verhelst, 2003). Whether centralization has been achieved or not, test use to achieve collectivist goals usually occurs where there is a sense of external threat, which can co-occur with an "insider–outsider" mentality.

Individualism and Testing

Modern individualism has its origins in the writings of Locke (1690/2002), Paine (1791/1999), the Declaration of Independence (Maier, 1998), and the liberalism of the late 18th and 19th centuries. The starting point is radically different: "Men being, as has been said, by nature all free, equal, and independent, no one can be put out of this estate, and subjected to the political power of another, without his own consent" (Locke, 1690/2002, p. 44) and that should he consent, there are limits upon the authority of the state, such that laws apply to all equally, that they protect the rights of individuals, that taxation should only be raised by consent, and that the legislative "neither must nor can transfer the power of making laws to anybody else, or place it anywhere but where the people have" (Locke, p. 66). Individualists also use history in support of their position, and it is also sometimes romanticized, but the sources are different: Magna Carta 1215, Henry III and the first parliaments from 1237 (Carpenter, 2004), and the Levellers in Cromwell's army who demanded equality and universal suffrage (Benn, 1976; Sharp, 1998). The central pillar of Enlightenment individualism is the right of each person to be free from control or oppression from a state that acquires too much power and begins to control the lives of citizens. Mill (1859) argued that individuals must even be protected from the tendency of liberal democracies to control them, for "whatever crushes individuality is despotism, by whatever name it may be called" (Mill, p. 71). This is not a right-wing position, as attested to by the fact that Tony Benn[4] is found among its most outspoken advocates. Rather, it is a nonparty political struggle to escape oppression,

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

currently benign or otherwise. Attempts to summarily dismiss individualistic critiques of test use as right-wing reactionism by labeling them "Eurosceptic" (Alderson, 2007, p. 660) therefore fail to engage with the social consequences of test use and misuse.

The purpose of education and language learning within a liberal individualistic tradition is the growth of the individual. The classic position was first put by von Humboldt:

> It seems to follow, even from these few and general reflections, that national education—or that which is organized or enforced by the State—is at least in many respects very questionable. The grand, leading principle, towards which every argument hitherto unfolded in these pages directly converges, is the absolute and essential importance of human development in its richest diversity. (von Humboldt, 1854, p. 48)

Mill (1859, p. 81) took two principles from von Humboldt as necessary for human development, and the advancement of society: freedom and variety. For von Humboldt and Mill it is difference and divergence that create the conditions for a healthy, questioning society. For where the individual is limited in thought or choice of how to live by "rule of conduct, there is wanting one of the principal ingredients of human happiness, and quite the chief ingredient of individual and social progress" (Mill, 1859, p. 63). Mill (1873, p. 257) believed that difference and divergence could be achieved through "a systematic National Education" that he believed would provide a stepping stone to universal suffrage. The link between individualism, education, and democracy with severe limits to the power of the state, particularly with regard to noninterference in education and testing, is not incidental. Dewey (1888, p. 191) argued that the purpose of education is to develop individual personality, that "personality is the one thing of permanent and abiding worth, and that in every human individual there lies personality"; and for Dewey "democracy is not just a form of social life among other workable forms of social life, it is the precondition for the full application of intelligence to the solution of social problems" (Putnam, 1990, p. 185).

In this tradition, it is argued that a move away from individualism creates societies that do not experiment, lack tolerance, and tend to stagnate. Individuals grow through experiment, and experimentation is the most efficient way of changing and developing society. Similarly, a variety of school types with the freedom to experiment with curriculum and methodology is necessary "for the purpose of example and stimulus, to keep others up to the standard of excellence" (Mill, 1859, p. 118). This happens through freedom, not control. Thus, states should have little say in what is taught, or how. Within this paradigm tests should only be used in ways that promote personal growth or provide individuals with new learning opportunities. They should only be used in gate-keeping mode when they are essential for the protection of individuals. Mill argued that public tests should be given to children at set ages to ensure the acquisition of key abilities that are essential for individuals to play an active role in society, but that "the knowledge for passing an examination

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

(beyond the merely instrumental parts of knowledge, such as languages and their use) should, even in the higher classes of examinations, be confined to facts and positive science exclusively," for the examination of anything upon which there could be disagreement requires a political decision about what views an individual should hold, and "all attempts by the state to bias the conclusions of its citizens on disputed subjects, are evil" (Mill, 1859, p. 119). When it comes to gate keeping, the state should once again play no role in test use; rather "public certificates of scientific or professional acquirements, should be given to all who present themselves for examination, and stand the test; but that such certificates should confer no advantage over competitors, other than the weight which may be attached to their testimony by public opinion" (Mill, 1859, p. 119). Tests can therefore play an important positive role in society, but not as Foucault's "normalizing gaze" (Fulcher & Davidson, 2008).

Examples of tests being used in keeping with individualism include the original Binet tests, designed for the sole purpose of identifying children in need of additional help. Work on diagnostic and classroom testing, loosely defined as "low-stakes formative assessment" would also fall into this category. Language assessment in the classroom is an underresearched area, but what has been done shows that its purpose is to act as a way of providing individual learners with feedback that helps them to improve in an ongoing cycle of teaching and learning (Rea-Dickens, 2001). In such a context Dewey's notion of personal growth as a validity criterion is echoed by current researchers, such as Moss (2003). The recent interest in "dynamic assessment" in language testing is not different in principle from this position (Lantolf & Poehner, 2008a, 2008b). However, it draws upon Vygotsky's zone of proximal development and defines validity as the extent to which testing improves learner performance; it therefore differs from other approaches in its conceptualization of the educational intervention: "In DA [dynamic assessment], assessment and instruction are a single activity that seeks to simultaneously diagnose and promote learner development by offering learners mediation, a qualitatively different form of support from feedback" (Lantolf & Poehner, 2008a, p. 273).

When tests become high stakes and their primary purpose is still to provide access to education or employment for nonelite individuals, even though what is "valued knowledge" is still decided upon by society, they still serve the individual. This is arguably the case with the Indian civil service examinations, which some see as having introduced liberty and equality to selection (Spolsky, 1995, p. 17–19). That this was termed "the Chinese principle" (Spolsky, p. 18) indicates that the Chinese Civil Service Examinations fulfilled a similar purpose, even though they are further away from the individualistic end of the cline. Miyazaki (1981, p. 111) stated that "the purpose of instituting the examinations . . . was to strike a blow against government by the hereditary aristocracy . . . and to establish in its place an imperial autocracy." This was no democratic system, but the tests, as Miyazaki pointed out (pp. 118–121) were open to any candidate, there was no entry fee, the tests were conducted under standardized conditions, and scripts were blind marked.

Many modern language tests for academic and other specialist purposes may be categorized in this way when they are used for their original explicitly stated

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

purposes (Fulcher & Davidson, 2009), especially those built on careful analyses of how individuals need to use language successfully in a specific context (e.g., Chapelle, Enright, & Jamieson, 2008), and avoid bias against individuals or subgroups of the test taking population (Elder, 1997).

In this section I have explored the nature of individualism and the use of tests that is likely to be made within this paradigm. Specifically, tests are used to (1) help individuals to develop their own potential without reference to others, (2) ensure that individuals acquire the key knowledge and skills they will need to innovate in their own lives and participate in democratic societies, and (3) provide access to employment through the assessment of critical skills where practicing without those skills would be detrimental to others. This may or may not imply centralization, but if it does, the state should not specify the content of the tests; this should be left to professional educators and stakeholders. Validity is assessed in terms of the success in helping individuals to achieve their goals and develop necessary skills, and validity evidence will be collected in relation to the claims of intended test effect. External systems are never imposed upon teachers. Teachers are involved in the process of defining the knowledge and skills to be taught and assessed, or design their own assessments as part of the learning process. One of the criteria for success is therefore the empowerment of professional educators to make their own judgments and decisions in their own contexts of work (Fulcher & Davidson, 2007, p. 34–35).

A Way Forward

Although a Foucaultian critique provides insights into testing in collectivist states, it offers few solutions; the marginalized cannot do anything to ameliorate their condition. One way forward is to rediscover the unfinished business of the Enlightenment.

Such an agenda would place the needs of the individual first, but recognize that the individual is only able to grow within a community. This is the essential insight of Dewey: Individual growth is about personal experiment and the implementation of intelligent learning through inference; society is strengthened when individuals within communities are allowed to inference, and communities can "try out" innovative solutions to social problems. Educational systems that value standardization of outcomes produced according to a mould created by "juggernaut-like centralizing institutions" (Davies, 2008, p. 438) have tended to stifle the variety and divergence that are prerequisites for innovation, experimentation, and growth.

Shohamy's (2001b) view of "democratic assessment" highlights a way forward, calling as it does for participatory democracy, the responsibility of language testers for the use of powerful tests, the consideration of the views of stakeholders, and the protection of individuals from powerful institutions. Building upon Shohamy (2001a, 2001b), Fulcher and Davidson (2007) introduced the notion of effect-driven testing, which holds that

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

> Tests have outcomes and impacts on the world, and, ultimately, it should be these test effects that drive the final design decisions about crafting particular items and tasks. We believe in "effect-driven testing": test creation in which the ultimate test design decisions are driven by the impacts that the test will have on stakeholders. Fulcher and Davidson (2007, pp. 50–51)

This position explicitly links the consequences of test use to test design, and it requires the statement of intended effects as drivers of design decisions. Decisions with regard to intended effects and how these are implemented in design are arrived at through open, democratic, and inclusive discussion (Fulcher & Davidson, 2007, p. 147). The practical implications are spelled out in Fulcher and Davidson (2009). Test design is not dissimilar from architecture, in which the design of a building is directly related to its intended effect on a defined user group. The specification of the user group and the intended effect dictates the shape, size, and layout of the building or test. The explicit intended effect will produce a very different architecture in each case. This is not only an important principle of test design (see Mislevy & Riconscente, 2005, p. 3), but also imposes restrictions on the post-hoc use of tests for purposes other than those for which they were originally designed.

Aligning explicit statements of intended test-effect to test design decisions through explicit validity arguments and test architectures will have a number of corollaries. First, any collectivist use of a test will either be made explicit, or will be ruled out, at the design stage. This introduces the potential for accountability and denunciation of illegitimate test use. Second, as the use of a test for an originally unintended purpose would require an architectural retrofit with a new validity argument, the misuse of existing tests for collectivist purposes would be reduced.

Mill (1873, pp. 253–254) predicted that his work on "the liberty" would be rediscovered whenever "this noxious power" of collectivism is exercised. As Foucault (1975) correctly observed, the operations of neocollectivist authorities in the postwar world are more subtle (even subliminal) and manipulative than their predecessors. Prophetically, Mill urged us to be sensitive to the impact of collectivist policies and practices on individual liberties, ways of thinking and behaving. All growth, including economic development, comes from variety and innovation, not from uniformity and standardization. Effect-driven testing is a way of revitalizing the Enlightenment project in the understanding of test use by making explicit the intended effects of test use, thus making it possible to more easily identify the misuse of these powerful tools for collectivist ends.

Notes

1. The Bologna process is specifically designed to remove diversity in content, qualifications, and assessment, because diversity is perceived to result in lack of transparency, or as Rauhvargers (2004, p. 339) wrote, there may be "huge differences between degrees that bear the same name in terms of admission requirements, content, learning objectives and function, as well as in the rights they confer." Rauhvargers

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

approves the Bologna solution of standardizing program structure, content, and the introduction of a single quality assurance system for the whole of Europe.

2. The Council of Europe Recommendation on the use of the CEFR (Council of Europe, 2008) states that the CEFR is "purely descriptive—not prescriptive, nor normative," and says that its use should be selective, to promote "social inclusion, intercultural dialogue, active democratic citizenship, language diversity, plurilingualism, learner autonomy and lifelong learning." It is difficult to disagree with the idealism behind such language. However, the recommendation itself is for governments to adopt the CEFR as the key tool in language policy. Specifically, governments are requested to promote the use of the CEFR and to "ensure that the document is used in a valid way" as defined by the Council of Europe, that it becomes the basis for language teacher training, the organization of course books and curricula, the comparison of educational institutions, and the recognition of language qualifications. It suggests that governments may wish "to establish, if required, the legislative and administrative framework for implementing the Recommendation." Such a contradiction between the rhetoric and the collectivist political intention to harmonize and control is highly suited to a Foucaultian inspired critical discourse analysis.

3. The last appeal to the golden age of Charlemagne and European Union in the 9th century CE was by the Vichy regime in France to create an ideology in support of German occupation (Black, 2005, pp. 106–108). Today, the Charlemagne prize is given each year to the politician judged to have done the most for "European integration."

4. Antony Wedgewood Benn was a socialist member of the UK parliament from 1950 to 2001. His opposition to European institutions and integration is explicitly on *democratic*, rather than *nationalist*, grounds. He argues with Locke that the transferral of legislative powers from those whom the people can elect and remove from government to a nonrepresentative body is illegitimate, as is removing the right of a subsequently elected government to repeal laws passed by a previous government. This is a classic libertarian position that has little to do with modern party politics, and frequently politicians of many political persuasions find themselves on the same side for philosophical reasons. To hear Benn put this argument for himself, see his videos on YouTube: http://www.youtube.com/watch?v=GTjg-vt0Ao4 and http://www.youtube.com/watch?v=J-hyxK9JB4Q&feature=related

ANNOTATED REFERENCES

Fulcher, G., & Davidson, F. (2008). Tests in life and learning: A deathly dialogue. *Educational Theory and Philosophy, 40*(3), 407–417.

In the recent social critique of language testing and the use of tests, most researchers turn to Foucault's (1975) analysis of tests as the "normalizing gaze" of powerful institutions. This article attempts to uncover

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

the assumptions that come as part of the baggage of Foucault's social theory by placing him in an imaginary Socratic dialogue with John Stuart Mill. The two meet in a corner of heaven and discuss their views on testing, with, as one would expect, a knowledge of what has happened (and been published) since their own deaths. The dialogue explores issues of pessimism and power, optimism and freedom, and whether tests can play a useful role in society.

Fulcher, G., & Davidson, F. (2009). Test architecture. Test retrofit. *Language Testing, 26*, 1 123–144.

The design and construction of language tests is compared, by analogy, to architecture. Architects design buildings for a clearly defined set of users with very specific needs. Similarly, language testers should create architectural drawings for tests that articulate the test users, the test purpose, and the uses to which it will be put. Developing the idea of "retrofit" for upgrading a building, or changing the use of a building, the article suggests that changes in the use of tests should be accompanied by architectural changes to the design layers of the test, including a reworking of a validity argument. In this way test design is intimately linked with test use in a much closer way than in traditional accounts of the primacy of test purpose.

Mansell, W. (2007). *Education by numbers: The tyranny of testing*. London: Methuen.

Unlike the other authors in the annotated bibliography, Mansell is an investigative journalist. Concerned at the growth in England's "testing regime," he went into the field to discover just what the impact of national testing was having upon schools and classrooms around the country. The findings relate to teacher demotivation, cheating on coursework, teaching to the test, focusing on borderline students (at the expense of gifted or less able students) to raise the school's position in league tables, and pushing learners to take tests that carry higher league table weightings. The findings are interpreted with reference to a standards-based education system that aims at "hyperaccountability," and which Mansell claims is failing to deliver a quality education for England.

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. London: Blackwell.

This volume takes a very wide view of the "social dimension" of language testing, from situating *language testing* as a social activity, as distinct from language testing as a psychological discipline, to the testing of *language* as a social phenomenon. Within a larger discussion of notions of fairness and codes of practice, the role of language testing is considered as a tool to investigate identity in immigration and asylum cases, and the role of tests in educational policy.

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

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

Shohamy's groundbreaking book is still the single most important work in this area. It contains a masterful sketch of the ways in which tests exercise power over the lives of test takers and other stakeholders; it is extensively illustrated with examples of powerful test use, and the associated intended and unintended consequences. Although Shohamy uses Foucaultian analysis, she does not fall into pessimism, or assume that the impact of test use must always be negative. Given the purpose of the book there are, of course, many negative examples; but in her discussion of the responsibilities of testers, ethics, the right of test takers, and the requirement for collaboration and fairness, there is optimism that despite all the problems, testing can be practiced for the benefit of the test takers.

Shohamy, E. (2001b). Democratic assessment as an alternative. *Language Testing, 18*(4), 373–392.

This article succinctly sets out the goals of critical language testing (CLT). Starting from the assumption that tests are tools of control used by powerful institutions, Shohamy offers CLT as a method of uncovering the hidden agendas of the powerful within a democratic and inclusive framework. The article provides a list of questions that stakeholders need to ask to reveal concealed purposes, and suggests ways in which consequences can be investigated. It is argued that within a democratic approach language testers have responsibilities, and test takers have rights, many of which are articulated in the Code of Ethics of the International Language Testing Association (ILTA).

OTHER REFERENCES

Alderson, J. C. (2007). The CEFR and the need for more research. *Modern Language Journal, 91*(4), 659–663.

German Propaganda Archive. (1937). The educational principles of the new Germany. *Frauen-Warte, 22*, 692–693. Retrieved July 15, 2008, from http://www.calvin.edu/academic/cas/gpa/frau01.htm

Benn, T. (1976). *The legacy of the levellers*. [Speech delivered on May 15, 1976, at Burford, Oxfordshire] Retrieved July 12, 2008, from http://www.channel4.com/history/microsites/H/history/a-b/pdf/benn.pdf

Black, E. (2003). *War against the weak: Eugenics and America's campaign to create a master race*. New York: Four Walls Eight Windows.

Black, J. (2005). *Using history*. London: Hodder Arnold.

Bonnet, G. (2007) The CEFR and education policies in Europe. *Modern Language Journal, 91*(4), 669–672.

Brindley, G. (2008). Educational reform and language testing. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7.*

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

Language testing and assessment (2nd ed., pp. 365–378). New York: Springer.
Carpenter, D. (2004). *The struggle for mastery: Britain 1066–1284*. London: Penguin.
Cassirer, E. (1946). *The myth of the state*. New Haven, CT, and London: Yale University Press.
Cattell, R. B. (1937). *The fight for our national intelligence*. London: P. S. King and Son.
Cecil, R. (1971). Education and elitism in Nazi Germany. *ICR Monograph Series No. 5*. London: Institute for Cultural Research.
Chalhoub-Deville, M., & Deville, C. (2008). Nationally mandated testing for accountability: English language learners in the U.S. In B. Spolsky & F. Hult (Eds.), *The handbook of educational linguistics* (pp. 510–522). London: Blackwell.
Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008). *Building a validity argument for the test of English as a foreign language*. London and New York: Routledge.
Council of Europe. (2001). *Common European framework of reference for languages: Learning, Teaching and assessment*. Cambridge, England: Cambridge University Press. Retrieved July 12, 2008, from http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf
Council of Europe (2008). *Recommendation CM/Rec(2008)7 of the Committee of Ministers to member states on the use of the Council of Europe's common European framework of reference for languages (CEFR) and the promotion of plurilingualism*. Strasbourg, France: Council of Europe.
Davidson, F., & Fulcher, G. (2006). The Common European Framework of Reference (CEFR) and the design of language tests: A matter of effect. *Language Teaching, 40*, 231–241.
Davies, A. (1997a). Introduction: The limits of ethics in language testing. *Language Testing, 14*(3), 235–241.
Davies, A. (1997b). Demands of being professional in language testing. *Language Testing*, *14*(3), 328–339.
Davies, A. (2008). Ethics and professionalism. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (2nd ed., pp. 429–443). New York: Springer.
Dewey, J. (1888). The ethics of democracy. In L. Menand (Ed.), *Pragmatism: A reader* (pp. 182–204). New York: Vintage Books.
Elder, C. (1997). What does test bias have to do with fairness? *Language Testing, 14*(3), 261–277.
Evans, B., & Waites, B. (1981). *IQ and mental testing: An unnatural science and its social history*. London: Macmillan.
Fejes, A. (2008). European citizens under construction: The Bologna process analysed from a governmentality perspective. *Educational Philosophy and Theory, 40*(4), 515–530.
Foucault, M. (1975). *Discipline and punish: The birth of the prison*. New York: Vintage.
Foucault, M. (1990). *The history of sexuality: An introduction*. New York: Vintage.

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly, 1*(4), 253–266.
Fulcher, G. (2008, July 20–24). Testing times ahead? *Liaison Magazine, 1*. Retrieved July 12, 2008, from www.llas.ac.uk/resourcedownloads.aspx?resourceid=179&filename=liaison_july08.pdf
Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London and New York: Routledge.
Fulcher, G., & Davidson, F. (2008). Tests in life and learning: A deathly dialogue. *Educational Theory and Philosophy, 40*(3), 407–417.
Fulcher, G., & Davidson, F. (2009). Test architecture. Test retrofit. *Language Testing, 26*, (1) 123–144 .
Gobineau, de, A. (1999). *The inequality of human races*. New York: Howard Fertig.
Harris, S. (2008). Internationalising the university. *Educational Philosophy and Theory, 40*(2), 346–357.
Jones, C. W. (1963). An early medieval licensing examination. *History of Education Quarterly, 3*(1), 19–29.
Kaftandjieva, F. (2007). Quantifying the quality of linkage between language examinations and the CEF. In C. Carlsen & E. Moe (Eds.), *A human touch to language testing* (pp. 33–43). Oslo: Novus Press.
Kelly, M. (2004) Engaging with Europe. *Liaison Magazine, 5*(1), 1. Retrieved August 11, 2007, from www.llas.ac.uk/resourcedownloads.aspx?resourceid=179&filename=liaison_feb04.odf
Lantolf, J. P., & Poehner, E. (2008a). Dynamic assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (2nd ed., pp. 273–284). New York: Springer.
Lantolf, J. P., & Poehner, E. (2008b). *Sociocultural theory and the teaching of second languages*. London: Multilingual Matters.
Locke, J. (2002). *The second treatise of government*. In T. Crawford (Ed.), *The second treatise of government and a letter concerning toleration*. New York: Dover. (Original work published 1690)
Maier, P. (Ed.). (1998). *The Declaration of Independence and the Constitution of the United States*. New York: Bantam Books.
Mansell, W. (2007). *Education by numbers: The tyranny of testing*. London: Methuen.
McNamara, T. (2005). 21st century Shibboleth: Language tests, identity and intergroup conflict. *Language Policy, 4*(4), 351–370.
McNamara, T. (2008). The socio-political and power dimensions of tests. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (2nd ed., pp. 415–427). New York: Springer.
McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. London: Blackwell.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). New York: Macmillan/American Council on Education.
Mill, J. S. (1859). *On liberty*. In J. Gray (Ed.), *John Stuart Mill on liberty and other essays*. Oxford, England: Oxford University Press.

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

Mill, J. S. (1873). *Autobiography*. London: Longmans, Green, Reader and Dyer.
Mislevy, R. J., & Riconscente, M. M. (2005). *Evidence-centered assessment design: layers, structures, and terminology*. Menlo Park, CA: SRI International.
Miyazaki, I. (1981). *China's examination hell* (2nd ed.) New Haven, CT, and London: Yale University Press.
Moss, P. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice, 22*(4), 13–25.
Noakes, J., & Pridham, G. (1983). *Nazism: A history in documents and eyewitness accounts, 1919–1945*. Exeter, England: Exeter University Press.
North, B., Figueras, N., Takala, S., Van Avermaet, P., & Verhelst, N. (2003). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEF): Manual preliminary pilot version*. Strasbourg, France: Language Policy Division of the Council of Europe.
Paine, T. (1999). *Rights of man*. New York: Dover. (Original work published 1791)
Plato. (1987). *The republic* (2nd rev. ed.; Desmond Lee, Trans.). London: Penguin Classics. Retrieved July 12, 2008, from http://classics.mit.edu/Plato/republic.html (Original work published 360 BC)
Popper, K. (2002). *The open society and its enemies*. London: Routledge.
Putnam, H. (1995). A reconsideration of Deweyan democracy. R. B. Goodman (Ed.), *Pragmatism: A contemporary reader* (pp. 183–204). London: Routledge. (Original work published 1990: *Southern Californian Law Review, 63*, 1671–1697)
Rauhvargers, A. (2004). Improving the recognition of qualifications in the framework of the Bologna process. *European Journal of Education, 39*(3), 331–347.
Rea-Dickens, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429–462.
Sharp, A. (1998). *The English levellers*. Cambridge, England: Cambridge University Press.
Shohamy, E. (2001a). *The power of tests: A critical perspective on the uses of language tests*. London: Longman.
Shohamy, E. (2001b). Democratic assessment as an alternative. *Language Testing, 18*(4), 373–392.
Spearman, C. (1904). General intelligence: Objectively determined and measured. *American Journal of Psychology, 15*, 201–292.
Sperber, J. (1994). *The European revolutions, 1848–1851*. Cambridge, England: Cambridge University Press.
Spolsky, B. (1995). *Measured words*. Oxford, England: Oxford University Press.
Terman, L. (1919). *The measurement of intelligence*. London: Harrap.
Thucydides. (1954). *History of the Peloponnesian war* (Rex Warner, Trans.). London: Penguin Classics. [Cira. 404 BC]
Tillich, P. (1952). *The courage to be* (2nd ed.). New Haven, CT, and London: Yale University Press.
von Humboldt, W. (1854). *The limits of state action*. Indianapolis, IN: Liberty Press.

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press

Weber, M. (1947). *The theory of social and economic organization*. (A. M. Henderson & T. Parsons, Trans.). Oxford, England: Oxford University Press.

Xi, X. (forthcoming). How do we go about investigating test fairness? *Language Testing*.

Yerkes, R. M. (1921). *Psychological examining in the United States Army. Memoirs of the National Academy of Sciences, Vol. XV*. Washington, DC: Government Printing Office.

https://doi.org/10.1017/S0267190509090023 Published online by Cambridge University Press