

Revista Interamericana de Educación para la Democracia

RIED  **IJED**

Interamerican Journal of Education for Democracy



**The Pull of PISA:
Uncertainty, Influence, and
Ignorance**

Vol 3, No. 1

April, 2010

Document available in:
www.ried-ijed.org

ISSN: 1941-1799



The Pull of PISA: Uncertainty, Influence, and Ignorance

Sharon Murphy
York University

Abstract:

If participation rates are any indication, the Programme for International Assessment (PISA), sponsored by the Organization for Economic Co-operation and Development (OECD), is increasing in influence, and the OECD plans for further expansion of the scope of PISA. Following a discussion of some validity issues in relation to PISA, several reasons are postulated for the popularity or “pull” of PISA, among which are: borrowing from other jurisdictions in the face of the uncertainties of globalization, the political leverage nation-states may obtain because they can compare themselves to others both in terms of PISA as well as economically, and the common discourse that surrounds PISA, which is the product of the professionalization of educational assessment. It is argued that the influence of PISA may jeopardize the democratization of education policy insofar as it allows elites to pursue their own agendas with little public input.

In 2000, the Organization for Economic Co-operation and Development launched the first round of PISA—the Programme for International Assessment (OECD, n.d.). Unlike some other large scale international assessments such as Trends in Mathematics and Science Study (TIMSS), PISA is distinguished by its connection with the OECD, which offers to this assessment the machinery of the OECD and, as such, guarantees a greater presence for this test on the world stage. As its influence grows, signs indicate that the OECD is contemplating the extension of PISA’s reach to younger populations and possibly to other aspects of education (Governing Board, 2005), making the assessments spawned within the PISA framework unlike any other large scale international assessment.¹ Also, by rapidly increasing the number of participating countries across a relatively short period of time, an implicit message of credibility for the assessment instrument is being built. The consequence of this credibility is that the “comparative advantage’ or ‘comparative disadvantage’ of each system can

be determined and politically and economically utilized” (Steiner-Khamsi, 2004, p. 207).

The then-28 member countries of the OECD, including the United States, Canada, and Mexico, participated in PISA’s first administration. Since that time, participation has more than doubled to include approximately 60 countries. From the Americas, the OECD member nations of United States, Canada and Mexico continue to participate, but participation since 2000 has expanded to include Brazil (Krawchuk & Rust, 2002), Uruguay (OECD, 2005), the Dominican Republic (OECD, 2007), Argentina, Chile, Colombia, Panama, and Peru (OECD, n.d.; OECD, 2009b). In short, in only a few cycles of administration, PISA has made great inroads into the Americas, so great that non-OECD “partner-nations”—nations who need not participate—feel compelled to do so. What is the pull of PISA and what are the consequences of participating in this assessment?



In this paper, I will argue that the motivations for participation in PISA are complex and sometimes contradictory, and the nature and consequences of that participation are underpinned by the undemocratic formation of educational policy. I begin my arguments by first addressing the credibility of PISA as an assessment instrument, since other arguments are premised on PISA's "meticulous... methodology" (Steiner-Khamsi, 2004, p. 207). I will argue that PISA is a very "managed instrument" that ultimately privileges certain types of knowledge and, despite a façade of inviting participation in its design, effectively limits it. Then, using theories of how organizations react in the face of similar sets of environmental conditions, as well as how educational policies are borrowed and loaned, I will examine PISA's attractiveness as a policy solution for nation-states. I will conclude by highlighting the common threads linking the design of PISA and the uses of PISA by nation states in terms of the erosion of basic democratic principles.

A PISA Primer

Although PISA could be considered just one other in a growing list of international assessment measures, such as the Trends in Mathematics and Science Study (TIMSS) or the Progress in International Reading Literacy (PIRLS), PISA is distinguished by its sponsorship by the OECD, an organization with the self-described mission of:

Bring[ing] together the governments of countries committed to democracy and the market economy from around the world to:

- Support sustainable economic growth
- Boost employment
- Raise living standards

- Maintain financial stability
- Assist other countries' economic development
- Contribute to growth in world trade (OECD, 2009a)

The OECD (2009c) frames the purpose of PISA in the following way:

Are students well prepared for future challenges? Can they analyze, reason and communicate effectively? Do they have the capacity to continue learning throughout life? The OECD Programme for International Student Assessment (PISA) answers these questions and more, through its surveys of 15-year-olds in the principal industrialised countries. Every three years, it assesses how far students near the end of compulsory education have acquired some of the knowledge and skills essential for full participation in society. (p. 1)

PISA's Features

The description of PISA is designed to both worry and reassure, and speaks to the anxieties individuals and nation states have about living in contemporary society. The promise of PISA is delivered in the form of paper-and-pencil tests (on mathematical, reading and scientific literacy, as well as, most recently, problem solving) in the form of multiple choice and constructed response (short answer) questions.

A total of 7 hours of test material has been developed for PISA; however, any single student completes only 2 hours of PISA testing. Comparisons can be made of different students who have taken different test questions through a statistical procedure in which the assumption is that a specific trait accounts for responses to test



items (Adams, 2002; McQueen & Mendelovits, 2003). A component of PISA intended to help contextualize the results are surveys of students and school principals. A survey of parents was added in 2006. Surveys focus on student characteristics, perceptions, and backgrounds, school characteristics, and parents' background and perceptions.

Although not specifically a part of PISA, the statistical data collected by the OECD in other capacities since the 1960s is part of the landscape against which PISA data are often discussed. Even though the scope and nature of the statistics has changed in recent times and now emphasizes "human capital theory, which implies a clear link between the competencies and qualifications of individuals at a micro level and economic growth at a macro level" (Ioannidou, 2007, p. 337), it is also the case that, because of its past data gathering efforts, many nation states consider the OECD to have a sound track record in data collection.

The Quality of PISA as an Assessment Tool

If PISA were an extraordinary assessment tool of great quality in the landscape of assessment design and development, being literally "put in place," or rank ordered, by the results of such an assessment might warrant serious consideration. However, close examination of PISA by scholars uninvolved with its development reveals not only that the validity of PISA has come under serious question on several grounds but that some of these validity issues raise basic questions about who is afforded the opportunity to genuinely participate in educational policy formation.

Participation in the making of PISA

From a validity perspective, a number of the literacy test items on PISA are identical to

the International Adult Literacy Survey (IALS), thus raising the question of whether PISA is a test for adults or teens. The fact that IALS developers won the bid for the development of PISA (Rochex, 2006) might explain general similarities between PISA and IALS. Indeed, developers were keen to compare IALS and PISA performances. However, the embedding of "a subset of the literacy tasks from the IALS prose literacy scale" (Yamamoto, 2002, n.p.; see also Grisay, 2002, p. 59) makes one wonder about the suitability of items originally targeted for participants aged 16 to 65 (as in IALS), instead of the 15-year-old PISA participants. Although, much is made in PISA documents about the amount of input nation-states are purported to have in the development of test items, the inclusion of these IALS items receives but a brief mention in PISA Technical Manuals (e.g., Wu, 2000, p. 21) and no evidence is provided of whether nation-states objected to these items. If one takes as a basic democratic principle that those being affected by a decision should have an opportunity to have a say in that decision, then this principle appears to be "managed" and perhaps even thwarted when nation states send their representatives to participate in meetings about PISA's design.

The content of PISA

The issue of what knowledge is valued within PISA raises validity concerns as well as fundamental democratic concerns. Three different aspects of PISA validity are of interest: a) the areas of knowledge 15-year-olds are thought to need for "full participation in society," b) the ways in which language represents that knowledge, and c) the ways in which some aspects of cultures are chosen or not chosen for representation as the basis of PISA test items.



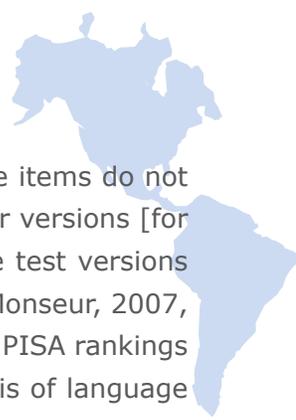
The knowledge needed by 15-year-olds. Several issues arise in terms of ways in which knowledge presumed to be needed by 15-year-olds appears in PISA. First of all, PISA test materials, it is asserted, are intended to be “suitable for and of interest to 15-year-olds regardless of nationality, culture, socio-economic level, or gender,” and are not supposed to “date rapidly” (McQueen & Mendelovits, 2003, p. 214). Support for the “suitability and interest” criteria being at odds with the “should not date rapidly” criterion was reported by Rochex (2006), who observed that students were “puzzled” by the “irrelevant or counterintuitive nature of some of the viewpoints” (p. 189) in a reading task. If one of the motivations of PISA, as declared by developers, is to be “forward-looking” (Adams, 2002, p. 15), then basing a part of a test for 15-year-olds on past items relating to 16 to 65-year-olds seems contradictory. Added to this complexity is that a 15-year-olds’ world view may be at odds not only with the IALS population but with the test developers’ world views. Furthermore, this view seems to suggest that little development occurs throughout the lifespan. Little discussion is afforded this topic in PISA documents. Yet, such issues do appear to make a difference. For example, when one sample of PISA participants was interviewed about individual items on PISA, they had difficulty suspending their own viewpoints in favor of those built into the test (Rochex, 2006).

The knowledge of 15-year-olds is circumscribed in a different way with the criterion of enrollment in compulsory schooling. Yet it is puzzling that test designers repeatedly make comments that they are not interested in “the extent to which these students have mastered a specific school curriculum,” but instead are “concerned with what students can do with what they learn at school” (Adams, 2002, p. 15). Without some discussion of what kinds of engagements

15-year-olds might encounter in the world, and the basis upon which those engagements were selected, it seems reasonable to question what exactly PISA assesses. Indeed, implied in Dohn’s (2007) comment that “PISA assesses, with some degree of reliability, knowledge and skills for PISA. No more, no less” (p. 10) is the suggestion that a “PISA world” is created through these tests.

Some might argue that even if a PISA world is created, performance on PISA tests at least could be viewed as providing comparable indicators of student performance within that artificial world. However, here one must look to the statistical procedure underlying some of the inferences about student achievement. PISA uses a common test development procedure referred to as item response theory (IRT). IRT is based on the assumption that “a single latent trait accounts for the responses to items on a test... [based on] a model of how examinees at different ability levels for the trait should respond to an item” (Crocker & Algina, 1986, p. 339). In this way, the performance of students who have taken different versions of a test, as is the case in PISA, can be compared. However, when Rochex (2006) examined student responses to see if there were any predictable relationships based on similarity of student performance on items, subtests, or even parts of tests, such relationships were at best “difficult to predict” (p. 185). This lack of predictable relationships suggests that more than one trait is at work in the items in PISA and makes questionable the assumption that students taking different versions of the test are taking “the same” test.

Not only are different forms of the test, assumed equivalent using IRT, not equivalent, but substantive questions have been raised as to whether the test assesses what it purports to assess. For example, when Rochex (2006) interviewed students about their performance,



he found that students talked about being able to answer some questions without reading the passages on the test. Such possibilities effectively take the “reading” out of a reading test. Any inferences about passage reading for such items are suspect and the reading test becomes a test of background knowledge and experience.

Language and the representation of knowledge. PISA documentation is largely self-congratulatory about the hurdles surmounted with respect to the enormity of the task of the translation of items. Consider, for example, that in the end, 47 different language versions were used, including 7 variants of English, 4 of French, 4 of German, and 5 of Spanish (Grisay & Monseur, 2007). Idioms and nuance were avoided. Some words or phrasings were replaced by local ones (e.g., currency, place names, people’s names, syntax, spellings), and participating countries were asked to rate the items on a number of features. The tests were then revised according to these ratings. PISA developers make much of the “double translation procedures” in that they are more accurate than alternatives, but they also acknowledge the impact that language factors can have on achievement (Grisay, 2002). For instance, PISA tests are timed and are based on a reading of the English version at a rate of 3500 words per hour (McQueen & Mendelovits, 2003). The technical manual for PISA 2000 notes that, in some instances, lengthier versions of texts resulted from translation and, while the burden on test takers for timely completion did not “seem to be substantial... the hypothesis of some effect on the students’ performance cannot be discarded” (Grisay, 2002, p. 66).

If, as Rochex (2006) argues, translation is key to the ability to make reasonable comparative inferences about PISA, what impact does translation have on PISA results? A comparative study of Indo-European languages and non-Indo-European languages, for example,

reports that “many multiple choice items do not function in equivalent ways in their versions [for Asian countries], compared to the test versions in Western languages” (Grisay & Monseur, 2007, p. 77). Such findings suggest that PISA rankings of countries that differ on the basis of language may well be an exercise in nonsense.

Culture and the representation of knowledge. If language is problematic for PISA test items, so too is the less-discussed representation of culture. Despite a goal of representing cultural diversity rather than cultural neutrality in test items, this goal was difficult to meet. During the test development phase of PISA, representatives of countries involved in the PISA assessment were asked to comment upon cultural diversity with respect to the test. The result was that “not every unit retained in the item pool was favored by all countries, but every concern expressed was carefully explored and a judgment made on the basis of all available information” (McQueen & Mendelovits, 2003, p. 216).

PISA’s oblique type of approach, characterized by the non-informative phrase “a judgment made,” makes it difficult to unmask who made the judgment and what the criteria are underlying such judgments. An example of a unit that was deemed inappropriate was one about the alternative independent singer Ani DiFranco. That this unit was deemed inappropriate, because “feminism, the focus of the song’s lyrics, was an unfamiliar concept to students in certain countries” (McQueen & Mendelovits, 2003, p. 216), makes one wonder how cultural diversity was imagined in general. With only a few examples provided of the items that were dropped from the PISA test, little sense can be made either of individual countries’ concerns with respect to specific items or with the test as a whole. Even so, the small amount of information provided by PISA developers with



respect to this aspect of test design raises more questions than it answers about cultural factors influencing test design. Given these types of issues, PISA test results should be considered “fragile evidence” (Murphy, Shannon, Johnston, & Hansen, 1998), at best, or should have suffered “death by a thousand cuts” in that the many different limitations associated with the test, when aggregated, should have meant that the test as a whole was no longer defensible. However, PISA has somehow managed to hold sway despite its limitations. PISA, with its:

fuzziness of design... treats the links between student, school, and national achievement as self-evident, thus allowing for a black-box approach to schooling itself where the coincidence of results and factors is transformed into correlations and causalities, without establishing how this linearity comes into being. (Hopmann, 2008, p. 444)

In discussions about PISA, technical flaws are often at best uninteresting to many, and, at worst, ignored by most. However, at the heart of these validity issues are fundamental questions about principled-decision making that is in keeping with democratic principles. If PISA tests contain content based on ageist assumptions, contradictory positions on the role of schooling, insufficient similarity among items supposedly measuring the same trait, tasks that may not require of participants the umbrella skill purported to be assessed, variations based on the language in which the test is printed, and highly selective cultural elements, then decisions are being made that privilege some while marginalizing or excluding others. Furthermore, the structure and complexity of the PISA enterprise is such that participants do not appear to have channels for action when faced with a series of decisions which, when taken in the aggregate, will likely

portray the 15-year-olds of their nation states in unwarranted ways. Why then do nation-states participate in PISA even when not obligated to do so? One answer may lie in the worries nation-states have over the future.

PISA Participation as a Response to Uncertainty

Apart from lauding the assessment system itself, PISA’s webpage exploits worries, including worries about the future and the well-preparedness of students, implicit worries about preparedness relative to other countries in the world (because of the international context in which the OECD operates), and worries about economic development through its focus on industrialized countries. As such, the OECD’s positioning of PISA operates as an example of mimetic isomorphism, in which “uncertainty... encourages imitation” (DiMaggio & Powell, 1983, p. 151). The implied forces of uncertainty in PISA’s self-described purposes are globalization and economics.

The term “globalization,” in its simplest rendering, refers to “political and cultural changes...[which] affect in common ways large segments of the world’s peoples” (Spring, 2008, p. 331). Globalizing forces have always existed. However, according to Dale (2000), these forces are worrisome in contemporary contexts for several reasons: a) previously, through colonialism and imperialism, globalization occurred one country at a time, but now globalization is “not reducible to the intentions of any nation” (p. 94); b) no longer are the nation-state’s interests supreme; rather “the survival of the system is more important than the interests of individual nations” (p. 94); and c) “the state can discharge its national responsibilities only by prioritizing its extra-national responsibilities” (p. 95). Nation-states are put in the unenviable



position of losing control while seemingly being in control, a situation which undoubtedly breeds uncertainty. When faced with uncertainty, one of the responses of organizations is modeling or borrowing practices from other jurisdictions (DiMaggio & Powell, 1983, p. 151; Steiner-Khamsi, 2004), which, in turn, could provide a possible explanation for the participation of non-OECD member countries in PISA.

However, by engaging in this borrowing, participating states cede some of their internal sovereignty (Robertson, 2009), creating thereby “a constrained form of democracy where decisions are no longer made by elected representatives” (Kelly, 2009, p. 52). Constraints on decision-making are made starkly visible in descriptions of PISA’s governing board and National Project meetings. Participating nation-states send representatives to these meetings ostensibly to participate in decisions (including general design decisions) about PISA. While published reports do not exist for participants from the Americas, interviews by Grek, Lawn and Ozga (2009) of Scottish and English participants in these meetings candidly reveal how countries are managed by the OECD through PISA. Grek et al. (2009) are quoted at length here because their descriptions are so compelling:

[In] the Board meetings... the members from each country often appear to represent their national ‘stereotypes’ and argue for ‘national’ recognition.... [E]ducation tradition and values are the stereotypes which are experienced repeatedly in the OECD PISA meetings... In a sense, the Board meetings were described as the place where national differences and traditions are ‘ironed’ out, in order to reach a consensus. Nonetheless, it was also very interestingly noted that the ideas put forward are

those that are more likely to lead to a compromise amongst the members. The meetings therefore were described as heavily managed and controlled by the OECD Secretariat and Andreas Schleicher himself.

In terms of technical issues arising in meetings, another interviewee commented that technical issues are almost never discussed—instead, ACER experts often offer technical presentations which Board members never challenge but ‘trust’. The process appears to translate political participation (i.e., through representation of system values and practices) into technical processes through the ritual of enactment of stereotype and the ‘trust’ in technical expertise to reduce divergence and difference to manageable and comparable systems. (p. 8)

In effect, the “consensus building” considered a hallmark of many transnational education schemes (Ioannidou, 2007) is little more than a sham in the context of the meetings described by Grek et al. (2009), and the descriptions raise even deeper questions about who actually is making decisions about PISA. When the pseudo-consensus-building is layered on top of the fact that PISA meetings are held in either English or French, participants from non-English speaking nation-states are further disadvantaged, an observation also offered by Grek et al.’s (2009) Scottish and English participants. Perhaps the failure of the representatives of nation-states to be more forthcoming about the nature of these meetings is analogous to the tale of the emperor’s new clothes in which the populous can see that the emperor is not wearing new clothes but each individual is afraid to speak and unmask his or her



seeming ignorance. Alternatively, other reasons can be offered for the continued participation of nation states in PISA.

PISA Participation as a Source of Influence and Legitimacy

Embedded within worries over globalization are worries over the economy. This relationship is born out of the fact that contemporary globalization is seen by some as having “its origins in economic, and especially financial and production, factors” (Dale, 2000, p. 95). For PISA, formal and informal economic pressures of globalization lead to a kind of coercive isomorphism (DiMaggio & Powell, 1983, p. 150) for non-OECD member countries, who feel pressured to participate in PISA. The nation-state’s loss of autonomy resulting from responding to this pressure can have the effect of eroding the sovereignty of the nation. Grek et al.’s (2009) interviewees provide evidence that the hypothesis that participation in PISA is motivated by the economic pressures of globalization bears substantial merit, since the pairing of economic growth with educational performance was *exactly* why the interviewees saw participation in PISA as more valuable than participation in TIMSS or other assessments. In short, “OECD was able to offer a much greater spread of comparison, both for the more and the less successful education systems and hence economies” (Grek et al., 2009, p. 4).

Even though states may feel coerced into participating, the results of PISA, when combined with economic indicators, do not necessarily hold up to how nation-states may be imagining the consequences of participation. For instance, for the 2006 reading tests of PISA, Canada fell into the cluster of countries scoring above the OECD average, no country in the Americas fell within the OECD average, while Uruguay, Mexico, Brazil, Colombia, and Argentina fell below the

OECD average (OECD, 2007, p. 47). For the 2006 mathematics and science tests, these patterns are almost identical, with the exception that the United States, which is not part of the reading data because of booklet errors, appears towards the top of the list of those countries falling below the OECD average (OECD, 2007, pp. 22 & 53). Similar patterns have persisted across various administrations of PISA (Vegas & Petrow, 2008).

Gross Domestic Product and PISA

Gross Domestic Product (GDP) is sometimes offered as an “explanation” of the patterns of PISA results and, in turn, can become a motivating force for participating in PISA with an unspoken assumption that if PISA scores improve so too will GDP. However, the fact that a large economy such as the United States falls below the average range raises questions about whether GDP is indeed a driver for PISA participation. Those reluctant to give up on the association of GDP and PISA achievement offer explanations for the performance of the United States. For instance, Hopmann (2008) suggests that the United States is very focused on the high stakes assessments associated with the No Child Left Behind legislation, so PISA performance is not a high priority, while Pelham, Crabtree, and Nyiri (2008) indicate that the most common explanation for the performance of the United States is “the relatively inequitable distribution of resources between the haves and the have-nots in American society. American children in affluent areas perform as well as children anywhere in the world, while those in impoverished neighborhoods on average do more poorly, lowering the overall country average” (p. 76). Alternatively, Smithson (2009) suggests as an explanation for performance on the mathematics portion of PISA that there is a mismatch in content between PISA and the curriculum delivered in the United States.



Other explanations for differential performance across the Americas include the fact that the investment in education by Latin American countries is lower as a proportion of GDP than in countries with similar GDPs; further, mean scores on PISA are below those one might predict based on per pupil expenditure levels (Vegas & Petrow, 2008). OECD documents, in particular, provide an explicit motivation for participation in PISA in that they frame test results as mechanisms by which nation states can improve “learning” (e.g., Vegas & Petrow, 2008); however, learning, in turn, is circularly defined as an improvement of future scores on PISA, and GDP is not addressed. But neither GDP nor below-mean per-pupil expenditure scores tell the full story of PISA.

Take, for example, the case of Finland, a country with a reasonably healthy GDP that routinely has performed well across a number of PISA administrations. The success of Finland has been explained by “a web of interrelated factors having to do with comprehensive pedagogy, students’ own interests and leisure activities, the structure of the education system, teacher education, school practices and, in the end, Finnish culture” (Valijarvi, Linnakyla, Kupari, Reinikainen, & Arffman, 2002, p. 8). Yet Finland has found itself under pressure to maintain its high ranking position (Livingston & McCall, 2005) to such an extent that new educational policies with an emphasis on individualism and competition—the opposite of the conditions under which Finland achieved its success (Kivirauma & Ruoho, 2007, p. 298)—are now being advocated. Like some of its counterparts in the Americas, despite achieving high scores, Finland too seems to have been pulled into policy borrowing by PISA; in its responsiveness to the pull of PISA, Finland is at risk of ceding some of its autonomy as a nation-state with respect to education by

borrowing educational practices in line with PISA’s underlying philosophy.

The Crisis of the Welfare State and PISA

A different economic explanation for the pull of PISA can be found in discussions which characterize PISA as a tool for the “management of expectations” in an “age of accountability” (Hopmann, 2008). At the root of this characterization is an economic argument based on the crisis of the welfare state, in which the limits of the provisioning of social institutions became bounded by the nation’s ability to pay for them. Hopmann (2008) suggests that the crisis of the welfare state² shifted how nation states handled supporting their citizenry, from “management of placements” to “management of expectations” (p. 424). In the management of placements, citizens could turn to specific social institutions (education, health care, etc.) with the expectation that their needs would be met; the professionals within these institutions determined how resources were used and the range of issues they dealt with. Confronted with a growing expansion of identified needs, a shrinking purse, and a citizenry worried about the sustainability of their social supports, the modern state now manages the *expectations* of its citizenry. Such management across a variety of sectors means benchmarks and standards of delivery and performance which “allow for more target-oriented management and accountability that, however, comes at the price: whatever does not fit into an expectation regime becomes marginalized” (Hopmann, 2008, p. 424). PISA and education fit well into a pared-down welfare-state model, but sacrificed, then, are larger social purposes of education, such as ethical conduct, social responsibility, and citizenship, as the discourse about education focuses on the narrow indicators exemplified in PISA.



PISA, Hopmann (2008) argues, takes education—something that is relatively large and hard to contain—and transforms it into something that, superficially at least, appears well defined and attainable. The accountability mechanism provided by PISA also minimizes the independent factors (e.g., gender, social class, immigration) that operate within and across the educational systems of nation-states. Instead, “PISA-using economists calculate the transaction costs of schooling and the ways and means by which the principals (parents, the state) might maximize the effectiveness of the chosen agents (i.e., teachers, schools, or school system” (Hopmann, 2008, p. 425). Ease of measurability also constrains what is identified as benchmarkable. Recalling the comments of Grek et al.’s (2009) Scottish and English interviewees, it seems that when the OECD is confronted with the hopes and desires of nation states about their educational systems, hopes and desires that were articulated by the democratic processes within those states, the OECD cannot begin to consider how such disparate elements could be assessed. Instead, in board and managers’ meetings, the combination of management by consensus and the highly statisticalized framing of assessment leaves nation-state representatives listening, but participating in very limited ways, in the development of PISA.

PISA as an Artifact of the Professionalization of Educational Assessment

Complex procedures and statistical protocols permeate discussions of PISA’s qualities as an assessment instrument, and are artifacts of the burgeoning discourse of large-scale assessment used by educational assessment professionals. The professionalization of the act of educational assessment is an example of “normative isomorphism” (DiMaggio & Powell,

1983, p. 150). Because of the professionalization of educational assessment, a shared discourse and shared practices operate across national boundaries and these have a normative influence on the value accorded assessments with the consequence that policy makers and nation-states see participation in such assessments to be of merit. For outcomes-based education on the international scene, Steiner-Khamsi (2004) refers to this shared discourse as a kind of “Maris O’Rourke effect” (p. 211), in which the policies advocated by O’Rourke when she led the outcomes-based movement in New Zealand followed her, and permeated assessment discourse and policy, as she moved to her posting at the World Bank. In countries such as the United States, however, discourse in the area of educational assessment has had a much lengthier time to form.

The Case of Large Scale Testing in the United States

The United States has been described as having a culture of testing (Hanson, 1993), and large scale educational testing has been prevalent there for some time (Clarke, Madaus, Horn, & Ramos, 2000). Although the normalizing forces of professionalization may encounter resistance in some scholarly writing, it is less common for voices from inside the educational assessment community to publicly worry about or protest the uses of large scale tests in the United States.

However, recently, the Board on Testing and Assessment (BOTA) of the National Research Council in the United States wrote a publicly available letter to the U.S. Department of Education’s Race to the Top fund. In this letter, BOTA (2009) cautions the Obama administration about the use of single measures in decision-making and argues for a much more nuanced approach to the use of test results. The letter



comes under the signature of Edward Haertel, a well-recognized name in the field of tests and measurements, and it represents not only his sentiments but those of the Board as well. This letter is important because it comes at a time of frenzied preoccupation with testing in the United States as a consequence of the No Child Left Behind legislation. The letter is also important because it comes not from critics of standardized testing but from tests and measurements professionals themselves.

Efforts by educational assessment professionals to leaven the impact of standards tests within the realm of public policy are not new. For example, nearly 20 years ago another eminent tests and measurements expert, Robert Linn (1991), wrote a paper for the Office of Technology Assessment of the United States Congress entitled, "Test Misuse: Why is it So Prevalent?" And some 40 years ago one of the principal conceptualizers of contemporary test validity theory, Samuel Messick, co-authored a paper (Messick & Anderson, 1967) that raised questions about the consequences of test use practices. In short, numbers of scholars from within the field of tests and measurements have raised their own concerns, within their own discourse communities, about the uses to which test results are put. Yet, if anything, the use of large scale standardized tests across the past 40 years in the United States has risen exponentially (Clarke, et al., 2000). What distinguishes the Haertel letter is the fact that the letter was a very public act, covered by *Education Week*, and not made within the confines of an academic journal or a technical report to a sitting government administration. Given these recent actions by scholars raising cautions about educational assessment, it seems somewhat ironic that the professionalization of the field of educational assessment itself may provide another possible explanation for the pull of PISA.

Professionalization of Assessment as the Promulgation of Ignorance

Recent theorizing by historians of science working in the emerging area of "agnotology" is the source for this further explanation for the pull of PISA. Agnotology, simply described, is the study of the making of ignorance (Proctor, 2008). Smithson (2008) argues that "ignorance, like knowledge, is largely socially constructed" (p. 212). In thinking about ignorance as a feature of our social world, Proctor (2008) focuses on several different types of ignorance: a) ignorance in the sense of "a place where knowledge has not yet penetrated" (p. 4), as happens when a person simply does not know about something; b) ignorance as the result of "selective choice" (p. 6), as in being the "product of inattention" (p. 7), as happens when individuals focus on one aspect of information while ignoring another; c) ignorance as "a strategic ploy or active construct" wherein ignorance is "actively engineered as part of a deliberate plan" (p. 9), as happens when misinformation or doubt is used to lead people to see something a particular way (such as the North American tobacco industry's commentaries suggesting that smoking was not harmful to health despite its knowledge of evidence to the contrary).

For PISA, it seems that the production of the second kind of ignorance, ignorance of selective attention, is at work in several different ways. Consider, for example, two co-existing facts: a) the United States has tended to fall below the average range for successive different administrations of PISA, and b) the United States had the highest gross national income in 2008 of any country in the world (as reported by the Atlas method of the World Bank [2009]). The economic competitiveness argument supporting the pull of PISA is much weakened when the performance of the United States is considered.



This is an example of selective attention at work—nation states are failing to attend to the contradictory elements of PISA.

However, Steiner-Khamsi (2000) goes much further than suggesting that the pull of large scale assessments is due to selective attention. She argues that nation-states are much more deliberate in their use of the results of assessments in their policy borrowing, and she uses the example of the German response to a series of international test results to illustrate this very point. In the name of test achievement, crises may be perpetrated and at other times test results are quietly ignored as governments pursue their own agendas. While governments on the international stage may appear coerced into PISA, on the national stage they are using PISA results towards their own ends, and are taking advantage of the cloak of educational assessment discourse to do so. In this scenario, the opportunity for democratic participation is doubly thwarted; it is thwarted once by the use of a test which, despite the façade of encouraging participation in development by nation-states, seems to be managed and controlled by elites behind the scenes, and it is thwarted again by elected political representatives who take advantage of the professionalization of assessment to use test data to satisfy their own political agendas.

Conclusion

Large-scale assessments like PISA are flawed instruments. They are not robust enough to bear the kinds of comparisons and policy decisions that flow from them. Nation-states may find themselves coerced by uncertainty into participation in these assessments; they may find political or economic leverage or influence in participation; or they may find professionals with strong bureaucratic connections moving

them towards participation. As Steiner-Khamsi (2000) suggests, nation-states may themselves strategically utilize test results to forward their own agendas. Such a move might seem like an appropriate response to PISA results given the managed participation afforded nation states in the development phases of PISA.

However, these moves by nation-states are highly similar to the moves by OECD officials at development and board meetings involving PISA. The nation-states appear to be open but are effectively reducing the transparency of processes. These moves allow for manipulation of the public and foreclose opportunities for deep discussion and debate, qualities necessary in any democracy, by taking advantage of the complex and sometimes confusing discourse created by the professionalization of assessment. In both the development of PISA and, for some states at least, in the uses of PISA, participation is being managed— for test development, by the OECD and test developers, and for test uses, by some nation-states themselves. Both practices effectively silence those who wish to engage in substantive discussion about the underlying values and judgments flowing into and out of PISA.

While further research detailing the uses nation-states make of PISA would be highly beneficial and provide a more nuanced and complete understanding of PISA and its uses, OECD operations continue to march on. For large nations like the United States, which is so dominant that it can turn its attention away from PISA to other matters, the consequences of PISA's pull are not great, but for less affluent democracies within and beyond the Americas, two issues must be addressed. If the diversion of scarce resources towards PISA, and reforms that flow from participation in PISA, are a result of the coercive effects of PISA, then, following

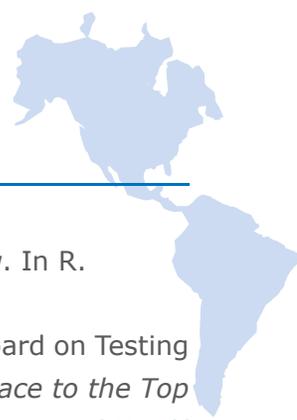


the example of BOTA, the pull of PISA should be a source of determination for educational assessment professionals as well as policy analysts to extend their discourse into more public forums than just the typical academic discourse communities. These professionals need to act in epistemically responsible ways (Code, 1987, 2006) to educate all involved as to the limitations of assessment instruments, as well as the appropriate use of such instruments so that broader and genuine participation in discussion is made possible. In other words, educational assessment professionals must not only be sensitive to creating knowledge about assessment but they must be alert to

the uses to which such knowledge is put. In effect, they have an ethical responsibility with respect to the knowledge domain that they are creating and working within. If governments are engaging in the use of assessment results when it is convenient to do so, such actions also need to be addressed by educational assessment as well as educational policy professionals. If the technological discourses of educational assessment are dismantled by educational professionals, and educational assessment becomes more transparent, then perhaps attention can be returned to educational questions larger and more important in scope than how a country performs on PISA.

Endnotes

1. A recent report (Lederman, 2010) indicates that the Australian Council for Educational Research—the same group associated with PISA development—is one of two groups funded by the OECD to design what essentially amounts to a higher-education variant of PISA called The Assessment of Higher Education Learning Outcomes (AHELO). AHELO will assess general skills and the disciplines of engineering and economics. Richard Yelland, the head of the Education Management and Infrastructure Division of OECD, positions AHELO as “a tool that will help us to help those who are responsible for higher education in the various countries” (Lederman, 2010, n.p.).
2. For a discussion particular to Latin America, see Schugurensky (2003).



References

- Adams, Ray. (2002). The programme for international student assessment: An overview. In R. Adams. & M. Wu (Eds.), *PISA 2000 technical report* (pp. 15-20). Paris, France: OECD. Board on Testing and Assessment. (2009). *Letter report to the U.S. Department of Education on the Race to the Top fund*. Washington: National Academy of Sciences. Retrieved October 28, 2009 from URL <http://www.nap.edu/catalog/12780.html>
- Clarke, Marguerite, Madaus, George, Horn, Catherine , & Ramos, Miguel . (2000). Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum Studies*, 32(2): 159-181.
- Code, Lorraine. (1987). *Epistemic responsibility*. London: University Press of New England.
- Code, Lorraine. (2006). *Ecological thinking: The politics of epistemic location*. New York: Oxford University Press.
- Crocker, Linda, & Algina, James. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Dale, Roger. (2000). Globalization: A new world for comparative education? In J. Schriewer (Ed.), *Discourse formation in comparative education* (pp. 87-109). Frankfurt: Peter Lang.
- DiMaggio, Paul J., & Powell, Walter W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2):147-160.
- Dohn, Nina Bonderup. (2007). Knowledge and skills for PISA--Assessing the assessment. *Journal of Philosophy of Education*, 41(1): 1-16.
- Governing Board. (2005). *Longer term strategy for the development of PISA*. 20th Meeting of the PISA Governing Board, Reykjavik, Iceland. OECD Directorate for Education. Retrieved October 31, 2009, from <http://www.pisa.oecd.org/dataoecd/18/1/35519715.pdf>
- Grek, Sotria, Lawn, Martin & Ozga, Jenny, (2009). Fabrication, circulation and use of a supra-national instrument of regulation based on knowledge—Education Sector: Production of OECD's 'Programme for International Student Assessment (PISA)'. Edinburgh: Centre for Educational Sociology, University of Edinburgh. Retrieved October 31, 2009, from http://www.knowandpol.eu/fileadmin/KaP/content/Scientific_reports/Orientation3/PISA.WP12.Scotland.pdf
- Grisay, Aletta. (2002). Translation and cultural appropriateness of the test and survey material. In R. Adams, Ray & Wu, Margaret. (Eds.), *PISA 2000 technical report* (pp. 57-70). Paris: OECD.
- Grisay, Aletta, & Monseur, Christian. (2007). Measuring the equivalence of item difficulty in the various



versions of an international test. *Studies in Educational Evaluation*, 33(1): 69-86.

Hanson, F. Allan. (1993). *Testing testing: Social consequences of the examined life*. Berkeley: University of California Press.

Hopmann, Stefan Thomas. (2008). No child, no school, no state left behind: schooling in the age of accountability. *Journal of Curriculum Studies*, 40: 417-456.

Ioannidou, Alexandra. (2007). A comparative analysis of new governance instruments in the transnational educational space: A shift to knowledge-based instruments? *European Educational Research Journal*, 6(4): 336-347.

Kelly, Anthony. (2009). Globalisation and education: A review of conflicting perspectives and their effect on policy and professional practice in the UK. *Globalisation, Societies and Education*, 7(1): 51- 68.

Kivirauma, Joel, & Ruoho, Kari. (2007). Excellence through special education? Lessons from the Finnish school reform. *International Review of Education*, 53(3): 283-302.

Krawchuk, Sheila, & Rust, Keith. (2002). Sample design. In Ray Adams & Margaret Wu (Eds.), *PISA 2000 technical report* (pp. 39-56). Paris: OECD.

Lederman, Doug. (2010, January 28). Measuring student learning, globally. *Inside Higher Education*. Retrieved February 9, 2010 from <http://www.insidehighered.com/news/2010/01/28/oecd>

Linn, Robert L. (1991). *Test misuse: Why is it so prevalent?* Washington, DC: Office of Technology Assessment.

Livingston, Kay, & McCall, Jim. (2005). Evaluation: Judgmental or developmental? *European Journal of Teacher Education*, 28(2): 165-178.

McQueen, Joy, & Mendelovits, Juliette. (2003). PISA reading: Cultural equivalence in a cross-cultural study. *Language Testing*, 20(2): 208-224.

Messick, Samuel & Anderson, Scarvia. (1967). Educational testing, individual development, and social responsibility. *The Counseling Psychologist*, 2(2): 80-87.

Murphy, Sharon., with Shannon, Patrick., Johnston, Peter., & Hansen, Jane. (1998). *Fragile evidence: A critique of reading assessment*. Mahwah, NJ: Lawrence Erlbaum.

OECD (n.d.). *PISA: The OECD programme for international student assessment*. Paris: OECD. Retrieved October 21, 2009, from <http://www.pisa.oecd.org/dataoecd/51/27/37474503.pdf>

OECD (2005). *PISA 2003 technical report*. Paris: OECD.



- OECD (2007). *Executive summary—PISA 2006: Science competencies for tomorrow's world*. Retrieved October 21, 2009 from <http://www.pisa.oecd.org/dataoecd/15/13/39725224.pdf>
- OECD (2009a). *About OECD*. Retrieved October 20, 2009, from http://www.oecd.org/pages/0,3417,en_36734052_36734103_1_1_1_1_1,00.html
- OECD (2009b). *What's new?* Retrieved October 20, 2009, from http://www.oecd.org/home/0,2987,en_2649_201185_1_1_1_1_1,00.html
- OECD. (2009c). *OECD: Programme for international student assessment (PISA)*. Retrieved October 22, 2009, from http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html
- Pelham, Brett, Crabtree, Steve, & Nyiri, Zsolt. (2008). Gallop presents... technology and education: The power of the personal computer. *Harvard International Review*, Summer: 74-76. Retrieved October 23, 2009, from www.gallup.com/poll/File/122171/TechnologyAndEducation.pdf
- Proctor, Robert N. (2008). Agnotology: A missing term to describe the cultural production of ignorance (and its study). In Robert N. Proctor & Londa Schiebinger (Eds.), *Agnotology: The making and unmaking of ignorance* (pp. 1-33). Stanford, CA: Stanford University Press.
- Robertson, Susan L. (2009). Editorial: The shifting politics of governance, sovereignty, and citizenship. *Globalization, Societies and Education*, 7(1): 1-3.
- Rochex, Jean-Yves. (2006). Social, methodological, theoretical issues regarding assessment-lessons from a secondary analysis of PISA 2000 literacy tests. *Review of Research in Education*, 30(1): 163-212.
- Schugurensky, David. (2003). Two decades of neoliberalism in Latin America: Implications for adult education. In Stephen J. Ball, Gustavo Fischman, & Silvina Gvirtz (Eds.), *Crisis and hope: The educational hopscotch of Latin America* (pp. 45-65). London: Routledge.
- Smithson, John. (2009). *Describing the academic content of PISA mathematics and science item pools*. Madison, WI: Wisconsin Center for Education Research, University of Wisconsin-Madison. Retrieved October 31, 2009, from https://edsurveys.rti.org/PISA/documents/SmithsonNCES_PISA_Conference_Paper_FINAL.pdf
- Smithson, Michael J. (2008). Social theories of ignorance. In R.N. Proctor & L. Schiebinger (Eds.), *Agnotology: The making and unmaking of ignorance* (pp.209-229). Stanford, CA: Stanford University Press.
- Spring, Joel. (2008). Research on globalization and education. *Review of Educational Research*, 78 (2): 330-363.
- Steiner-Khamsi, Gita. (2000). Transferring education: Displacing reforms. In J. Schriewer (Ed.), *Discourse formation in comparative education* (pp. 155-187). Frankfurt: Peter Lang.



- Steiner-Khamsi, Gita. 2004). Blazing a trail for policy theory and practice. In Gita Steiner-Khamsi (Ed.), *The global politics of educational borrowing and lending* (pp.201-220). New York: Teachers College.
- Valijarvi, Jouni, Linnakyla, Pirjo, Kupari, Pekka, Reinikainen, Pasi, & Arffman, Inga. (2002). *The Finnish success in PISA--and some reasons behind it: PISA 2000*. Jyvaskyla, Finland: Institute for Educational Research, University of Jyvaskyla, Finland.
- Vegas, Emiliana, & Petrow, Jenny. (2008). *Raising student learning in Latin America: The challenge for the 21st century*. Washington, DC: World Bank.
- World Bank. (October, 2009). *Data and statistics: Quick Reference Tables—GNI 2008 Atlas method*. Retrieved October 29, 2009, from <http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,,contentMDK:20399244~menuPK:1504474~pagePK:64133150~piPK:64133175~theSitePK:239419,00.html#ranking>
- Yamamoto, Kentaro (2002). *Estimating PISA students on the International Adult Literacy Survey (IALS) prose literacy scale*. Retrieved November 2008, from http://www.pisa.oecd.org/document/61/0,3343,en_32252351_32236159_33680573_1_1_1_1,00.html
- Wu, Margaret. (2000). Test design and test development. In R. Adams & M. Wu (Eds.), *PISA 2000 technical report*. Paris: OECD.