



Education Policy Analysis
Archives/Archivos Analíticos de Políticas
Educativas

ISSN: 1068-2341

epaa@alperin.ca

Arizona State University
Estados Unidos

Smith, William C.
The Global Transformation Toward Testing for Accountability
Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas, vol. 22,
2014, pp. 1-34
Arizona State University
Arizona, Estados Unidos

Available in: <http://www.redalyc.org/articulo.oa?id=275031898141>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal
Non-profit academic project, developed under the open access initiative

SPECIAL ISSUE
**The Comparative and International History
of School Accountability and Testing**

education policy analysis
archives

A peer-reviewed, independent,
open access, multilingual journal



Arizona State University

Volume 22 Number 116

December 8th, 2014

ISSN 1068-2341

The Global Transformation Toward Testing for Accountability

William C. Smith

The Pennsylvania State University
United States

&

National Research University – Higher School of Economics
Russia

Citation: Smith, W. C. (2014). The global transformation toward testing for accountability. *Education Policy Analysis Archives*, 22(116). <http://dx.doi.org/10.14507/epaa.v22.1571>. This article is part of EPAA/AAPE's Special Issue on *The Comparative and International History of School Accountability and Testing*, Guest Co-Edited by Dr. Sherman Dorn and Dr. Christian Ydesen.

Abstract: To ensure equal access to high quality education, the global expansion of universal basic education has included accountability measures in the form of academic tests. Presently the majority of countries participate in national testing; however, the past two decades have seen a substantial shift in test characteristics and aims. This article investigates the global transformation toward testing for accountability, where intentional or unintentional positive or negative consequences are applied to educators (teachers and administrators) based on their student's test scores, in light of the emerging global culture, identified by World Culture theorists. Elements of the world culture – including the expansion of western education models, an emphasis on academic intelligence, faith in science as a rational path to truth, and the decentralization of authority to the local level – justify the establishment of testing for accountability systems. Descriptive evidence from regional and international datasets, such as PISA, PIRLS, and TIMSS, illustrate the speed at which this transformation occurs. The

convergence of countries toward testing for accountability and its position as an increasingly normative policy lever is illustrated in brief vignettes from the diverse systems of Hungary, Mexico, and South Korea. As testing for accountability becomes embedded in the world culture as a legitimate tool for education reform it is less prone to critical reflection. If the potential benefits and concerns of testing for accountability, outlined in this article, are not thoughtfully evaluated this global transformation will lead to a testing culture that is internalized as normative and adopted as individual values.

Keywords: accountability; testing; education policy; New Right; globalization; World Culture

La transformación global hacia los exámenes de rendición de cuentas.

Resumen: Para garantizar la igualdad de acceso a una educación de alta calidad, la expansión global de la educación básica universal ha incluido medidas de rendición de cuentas en forma de exámenes académicos. En la actualidad la mayoría de los países participan en las exámenes nacionales; Sin embargo, las últimas dos décadas se ha producido un cambio sustancial en las características y los objetivos de los exámenes. En este artículo se investiga la transformación global hacia exámenes de rendición de cuentas, donde se aplican las consecuencias positivas o negativas intencionales o no-intencionales a educadores (maestros y administradores) sobre la base de los resultados de las pruebas de sus estudiantes, a la luz de una cultura global emergente, identificada por los teóricos de la cultura mundial. Elementos de la cultura del mundo - incluyendo la expansión de los modelos de educación occidentales, un énfasis en la inteligencia académica, la fe en la ciencia como un camino racional a la verdad, y la descentralización de la autoridad a nivel local – justifican el establecimiento de exámenes de sistemas de rendición de cuentas. Evidencia descriptiva de conjuntos de datos regionales e internacionales, como PISA, PIRLS, y TIMSS ilustran la velocidad a la que se produce esta transformación. La convergencia de los países hacia exámenes de rendición de cuentas y su posición como una palanca política cada vez más normativa se ilustra en breves viñetas de los diversos sistemas de Hungría, México y Corea del Sur. Como los exámenes de rendición de cuentas se incrustan en la cultura mundial como una herramienta legítima para la reforma de la educación esos exámenes son menos propensos a reflexiones críticas. Si los beneficios y las preocupaciones de los exámenes de rendición de cuentas potenciales, descritos en este artículo, no se evalúan cuidadosamente esta transformación global conducirá a una cultura de exámenes que se internaliza como normativa y adoptada como un valor individual.

Palabras clave: la rendición de cuentas; las pruebas; política educativa; Nueva Derecha; la globalización; Cultura Mundial.

A mudança global para exames de responsabilização.

Resumo: Para garantir a igualdade de acesso à educação de qualidade, a expansão global da educação básica universal incluiu medidas de responsabilização como exames acadêmicos. Atualmente, a maioria dos países que participam nos exames nacionais; No entanto, nas últimas duas décadas, houve uma mudança substancial nas características e objetivos dos testes. Este artigo analisa a transformação global para a prestação de contas, onde as consequências não intencionais positivos ou negativos ou intencionais não-educadores (professores e administradores) com base nos resultados dos ensaios de seus alunos aplicam, investiga a luz de uma cultura global emergente, identificado por teóricos da cultura mundial. Elementos da cultura em todo o mundo - incluindo a expansão dos modelos ocidentais de educação, a ênfase na inteligência acadêmica, a fé na ciência como uma forma racional para a verdade, ea descentralização da autoridade para o nível local - justificam o estabelecimento exame dos sistemas de prestação de contas. Prova descritiva dos dados regionais e internacionais comuns, como o PISA, PIRLS e TIMSS ilustrar a velocidade com que essa

transformação ocorre. A convergência de países para os testes de prestação de contas e sua posição como uma alavanca política cada vez mais regras ilustradas em breves vinhetas dos vários sistemas de Hungria, México e Coréia do Sul. Como os testes de prestação de contas incorporado na cultura mundial como uma ferramenta legítima para a reforma da educação esses testes são menos propensas a reflexões críticas. Se os benefícios e as preocupações dos exames de prestação de contas de contas potenciais descritas neste artigo não for cuidadosamente avaliar essa transformação global vai levar a uma cultura de testes que é internalizado como normativo e adotada como um único valor.

Palavras-chave: prestação de contas; exames; política de educação; Nova Direita; globalização; Cultura Mundial.

Introduction

With the importance of education increasing globally, international debate has turned from providing educational access to ensuring efficiency and equity in educational outcomes. To certify the available education is of high quality, the global expansion of universal basic education has included accountability measures in the form of academic tests. Understanding the transformation from traditional high-stakes exams that place responsibility of test scores on the student to testing for accountability, which places intentional or unintentional positive or negative consequences on educators (teachers and administrators) for their student's performance, is important because different approaches to testing are likely to lead to variant student outcomes (Harris & Herrington, 2006). This article investigates this global transformation through an exploration of the components of the emerging global culture, as identified by World Culture theorists. World Culture theory is often criticized by both proponents and opponents as merely a descriptive theory that fails to consider the potential outcomes of the emerging world culture (Carney, Rappleye, & Silova, 2012; Schofer et al., 2012), this article explains how the self-proclaimed components of world culture support and perpetuate the movement toward testing for accountability.

The article starts by using data from international assessments to illustrate the rapid expansion of testing and contrast the current turn toward testing for accountability with more traditional understandings of high-stakes examinations. This is followed by a historical look at the early adopters of testing for accountability, the United States and the United Kingdom, and the role of New Right ideology in their testing reform. Sections four and five situates the testing for accountability trend within the contemporary "global educational reform movement" (see Sahlberg, 2010, p. 47), introduces World Culture theory, and explains how multiple elements of the emerging world culture legitimate testing for accountability systems. To understand the shift toward greater accountability, national testing policy categories are outlined in section six and illustrative national examples from Hungary, Mexico, and South Korea are provided in section seven. Finally, the concluding section asks whether this global transformation is moving the world toward a normative testing culture that has the potential to influence multiple facets of society.

Expansion of Testing

Testing has long been used to assess student understanding, inform instruction, and identify students for academic advancement. However, the latter half of the 20th century signified a shift in the type of test administered, illustrated by a sharp rise in the use of large-scale standardized tests. In investigating the educational systems of 21 industrialized countries between 1974 and 1999, Phelps (2000) found that 18 increased the number of annually administered large scale tests, leading him to

conclude that there is a “clear trend towards adding, not dropping testing programs” (p. 19). Since 1980 nearly all European countries have adopted national testing policies (Eurydice, 2009a). Additionally, this trend is not limited to the industrialized north as educational reformers around the globe insist that “improving national (or state) testing systems is an important, perhaps the key, strategy for improving educational quality” (Chapman & Snyder, 2000, p. 457). The acceleration in national test policy adoption is perhaps best illustrated in the work of Benavot and Tanner (2007). They found that between the years 1995 and 2006 the number of countries worldwide that participate in an annual national testing program more than doubled from 28 to 67. As of 2006, 81% of developed countries and 51% of developing countries have conducted at least one national test.

Concurrent with the rise in national testing programs is increasing participation in cross-national assessments. The first cross-national assessments were initiated in the 1960s and were originally regionally focused with participation solely from industrialized countries. For example, twelve countries participated in the First International Mathematics Study in 1964 with only Japan and the United States located outside of Europe. However, since the 1990s international assessments have included a diverse array of countries outside the industrialized world as well as provincial economies, such as Shanghai, China and Dubai, United Arab Emirates. As illustrated in Figure 1, this has resulted in a steady increase in the number of participants in the three largest international studies: Trends in International Mathematics and Science Study (TIMSS), Program for International Student Assessment (PISA), and Progress in International Reading Literacy Study (PIRLS). All studies show a roughly 50% increase in participation between 1995 and 2012.

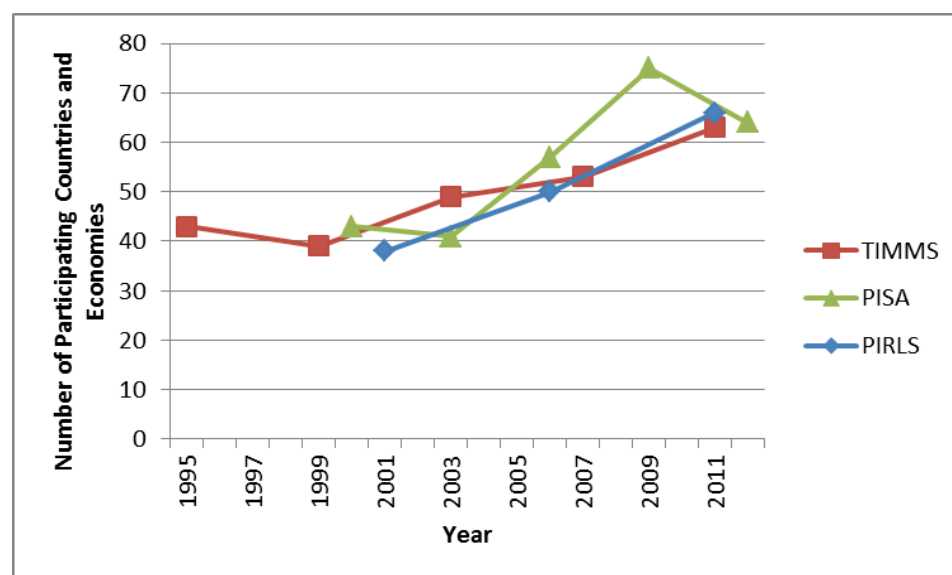


Figure 1. National participation in select international student assessments

With the support of international agencies the late 1990s also saw the creation of regional assessments in developing regions. For example the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) completed their first round of data collection in seven countries in 1999. Since the initial assessment of reading literacy among sixth graders, the number of participant countries has more than doubled, with 16 countries currently partaking in SACMEQ IV, scheduled for completion in 2014. In Latin America, participation in the Latin America Laboratory for Assessment of the Quality of Education (LLECE) has also increased, although at a slower rate, from 13 countries in 1997 to 17 countries in 2006.

At the school or classroom level a shift is also evident in the types of tests students complete, moving away from strictly teacher-administered tests to large scale standardized tests. Using data from PISA questionnaires, figure 2 illustrates that the percentage of schools that participate in more than two standardized tests a year have increased from 13.7% to 29.9% over just a ten year span. The number of standardized tests students take varies greatly across countries with many students exposed to a substantial amount of standardized tests over their compulsory school career. For example in Denmark students take 36 national tests during their time in primary and lower secondary school, while in China students take up to nine standardized subject tests each year (Schmidt, Houang, & Shakrani, 2009).

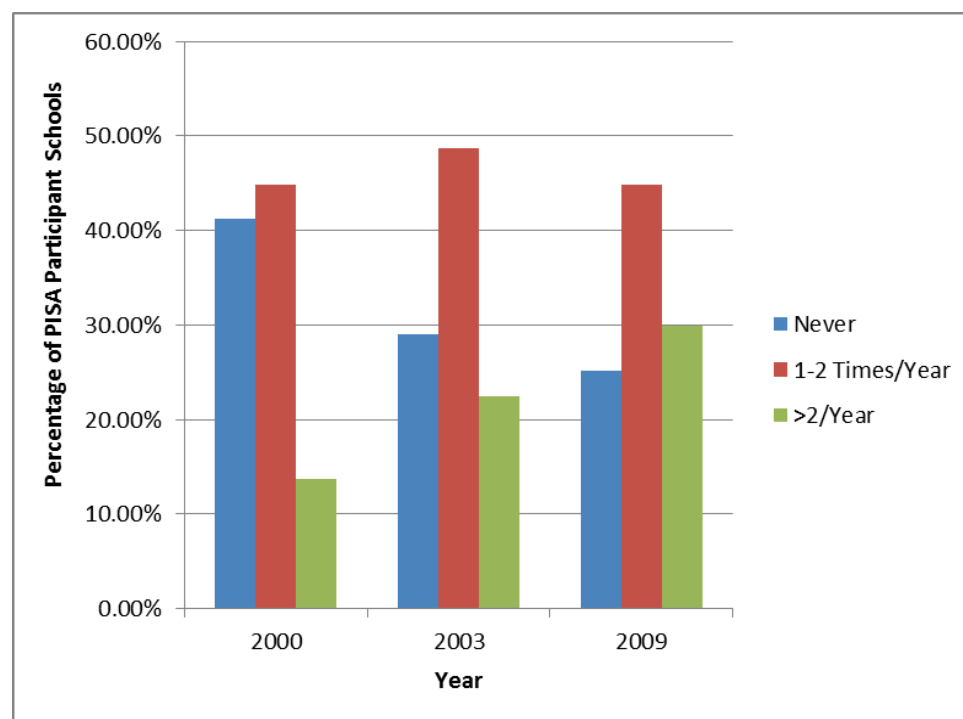


Figure 2. School participation in standardized external tests

Turn Toward Testing for Accountability

This global transformation is not limited to the number of tests but instead encapsulates a qualitative shift in testing characteristics and aims. “High-stakes exams” have a long history in many countries. Traditional high stakes exams focus on student knowledge with test outcomes determining student’s academic and career trajectory (Eckstein & Noah, 1993). Perhaps the best historical example of a high-stakes test comes from the Chinese Civil Service Exam. Originating in the third century B.C. and formally instituted during the T’ang dynasty (618-907AD), the Chinese Civil Service Exam was based on the Confucian belief that selection into the ruling class should be based on individual merit (Eckstein & Noah, 1993). The test was originally composed of six distinct examinations but was later narrowed down to one exam, the chin-shih examination, which remained in place until 1904. By the Ming period (1368-1644) the exam was seen as the only legitimate route to government positions as it was “more objective and less open to particularistic influence than the recommendatory system” (Ho, 1964, p. 15). Additionally high-stakes testing for students has been in place in countries throughout Europe, including Iceland, Portugal, and the United Kingdom (U.K.) since at least the mid-1940s (Eurydice, 2009a).

While much of the early literature muddles high stakes testing by combining effects on students and teachers into one general category (see Chapman & Snyder, 2000; Pari & McEvoy, 2000), the current transformation toward testing for accountability shifts the high stakes from students to educators. Teasing out tests that place high stakes on students from those that place high stakes on educators is important because they provide a different set of motivations and are likely to lead to diverse student outcomes (Harris & Herrington, 2006). Identification of the dominant features of a test, including where the test is administered and to whom the test results focused on, can illuminate these differences. Table 1 identifies tests that are administered within the K12 structure that make educators responsible for their students test scores as *testing for accountability*. Testing for accountability includes the application of formal or informal, positive or negative consequences on educators dependent on their students' performance measures (Figlio & Loeb, 2011). Educational systems that apply testing for accountability are interested in ensuring those that are delegated authority to educate children are "answerable to another level of authority for their prescribed responsibility" (Smeed & Victory, 2010, p. 28), explicitly answering the essential questions linked to accountability: accountability "to whom" and accountability "for what" (William, 2010). The contentious nature of testing for accountability indicates that answers to these questions remain challenged (Dorn, 1998; Kornhaber, 2004a).

Table 1

Test Foci and Administration

	Focus on Students	Focus on Teachers/Schools
Administration Within K12 School Structure	Graduation Exam	Testing for Accountability
Administration Outside K12 School Structure	College Entrance Exam	Teacher Certification Exam

Tests typically associated with high-stakes testing are illustrated in the first column of table 1. Within the K-12 school structure these tests are compulsory for all students (Eurydice, 2009a). High stakes tests that determine the academic trajectory (via tracking) of a student or provide them with access to a subsequent education level can be identified as *testing for advancement*. This differs from *testing for assessment*, low stakes exams which focus on students but are designed to assess their academic progress and direct instruction. *Testing for accreditation* is present when the aim of a test is to provide a credential, identifying an individual as a member of a distinct social group. Teacher certification exams and legal Bar exams are both examples of testing for accreditation. The categorization of testing aims is not mutually exclusive; for example, a high school graduation exam performs the function of both testing for advancement and testing for accreditation.

Holding different actors responsible for student achievement scores has shifted the blame from low performing students to low performing schools (Apple, 1999). When test scores remain aggregated at the individual level, parents often feel a personal interest in improving the quality of education in the classroom, leading to a more collaborative relationship with the teacher as both the parent and teacher work together to improve student learning. In contrast, the aggregation of results to a classroom or school level allow parents to place blame squarely on the teacher/school, leading to a more conflictual relationship in which the community questions why the school isn't maximizing the students learning while taking little responsibility on itself. This hostile relationship may become increasingly common as tests for advancement and tests for assessments are increasingly transformed into tests for accountability through school level aggregation. In some

countries this evolution is so great that testing has become synonymous with accountability (Froese-German, 2001), suggesting testing no longer has other aims.

The ultimate outcome of testing for accountability reforms is largely shaped by educators who play a dual role as policy implementer and student influencer. Teachers, through their position in the classroom hierarchy and student's perception of teacher as legitimate classroom authority, are in a position to significantly influence their student's behavior and cultural understanding (Smith, 2012). The educator position encompasses both autonomy and obligation. From this position educators have to prioritize often-competing demands while continuing practices that are in the best interest of their students (McLaughlin, 1991; Shulman, 1983). The perceived best interests of the student may include adapting classroom routines, structures, or instructional practices to individual student's academic needs. Therefore, unlike student focused testing which affects one student at a time, testing for accountability affects a classroom full of students simultaneously by leveraging behavior change in educators.

Educators as local policy enactors have been identified as "street level bureaucrats"—individuals that have the ultimate responsibility of policy implementation (Shulman, 1983). With autonomy, educators interpret or make sense of the policy, shaping how it is implemented as well as the resulting consequences (Rosen, 2009). Differences in the beliefs and experiences of educators can lead to policy enactment that is substantially different than the originally articulated policy. Although testing for accountability imposes on and restricts educators' professional autonomy (Luna & Turner, 2001), external accountability measures have been successful in altering the in class and administrative decisions of educators (Booher-Jennings, 2005). The greatest concern for teachers is their personal and professional survival (Gilles, Cramer, & Hwang, 2001) and when faced with testing for accountability this survival instinct is heightened (Nicols & Berliner, 2007), leading "educators who feel oppressed by an ineffective and potentially harmful evaluation system [to] feel justified" (Paris & McEvoy, 2000, p. 150) in altering their behavior to maintain their livelihood. Given the autonomy and authority of educators, the linkage of student success with educator survival has the potential to create large scale changes in the academic quality and success of multiple students simultaneously; making testing for accountability policies substantially different from student focused testing.

Early Adopters of Testing for Accountability

The shift towards testing for accountability has been a relatively recent phenomenon, seen first in the United States and the United Kingdom. The movement towards testing for accountability in the U.S. is rooted in a national history that wants "both a system that rewards merit and a system that generates equality" (Dorn, 2007, p. xiv). Linking tests to student accountability in the U.S. started in the 1960s and by the end of the 1970s many states established a link between test scores and school accountability (Dorn, 2007). The 1970s was characterized by the entrenchment of human capital arguments for education as an investment (Becker, 1962; Schultz, 1961). When combined with the shift in school funding from primarily local sources to a mix of local and state control, this investment perspective created an atmosphere where "legislators wanted some quid pro quo for spending more on education" (Dorn, 2007, p. 6). The 1970s was dominated by minimum competency standardized tests which mirrored the decline in intelligence testing, the later due to civil rights challenges against using IQ for student placements and an increasing desire to dispel human potential as fixed (Kornhaber, 2004a).

The 1980s saw the rise of the New Right in the U.S. and the U.K.. Beginning with the governments of President Reagan in the U.S. and Prime Minister Thatcher in the U.K. (Figlio &

Loeb, 2011), New Right calls for improved schooling were articulated “by national policymakers into an umbrella of neoliberal and neoconservative...reforms” (Carl, 1994, p. 315). This ideological shift was largely accepted by the public because of an increasing discontent and distrust of public administrators (Dorn, 2007), due in part to: (1) the economic recession of the 1970s which led to a decline in social services as the “crisis of the welfare state” led to calls for increased fiscal accountability (Hopmann, 2008), (2) an increasing anxiety about the state of U.S. schools and the ability of U.K. schools to equip students with the skills needed to excel in the economy (Fitz, 2003), (3) the inability of desegregation attempts to open up mobility paths for minority students while concurrently threatening the privileged position of the middle class and white parents, and (4) the alienation of working class and minority families from the traditional education experience (Carl, 1994).

The neoliberal push in education from the New Right was dependent on their belief that private schools were more dynamic and innovative than the rigid and bureaucratic public schools, largely because they were situated within a market (Carl, 1994). New Right supporters used the work of Friedman (1955, 1962) and Chubb and Moe (1990) to justify their position that markets are the solution to a failing school system. Friedman (1962) believed the privatization of education would create a higher quality, more efficient product while Chubb and Moe (1990) suggested private schools were more efficient due to their organizational structure that provides principals with more autonomy and power. Forcing public schools to compete with private schools would, therefore, result in either an improvement in the product or a closure of poor performing schools. The neoliberal emphasis results in the promotion of individual responsibility amongst self-interested actors, effectively removing any potential societal blame (Hursh, 2007).

The neoconservative call for uniformity and standardization complimented the neoliberal push for market-based accountability. Concerned with the relative permissiveness of the 1960s and 1970s, neoconservatives identified the school system as one in crisis (Carl, 1994). To remedy the crisis neoconservatives believed the education system must create uniformity in classroom curriculum and increase enforcement mechanisms (Hill, 2006). In the 1980s, instead of reflecting nostalgically on the past, neoconservatives pressed for “the development of coherent prescriptions for change – usually by hitching the neoconservative cart to the neoliberal horse” (Carl, 1994, p. 300).

New Right Policy Shifts in the U.S.

The racial achievement gap in the U.S. was part of a publicly perceived crisis, increasing concerns about the state of schooling in America during the 1970s and 1980s (Tyack & Cuban, 1995). Measuring the achievement gap required a shift in funding and reporting from inputs of education to outputs – typically measured in test scores (Hanushek & Raymond, 2004; Supovitz, 2009). The failing of the American school system and its inability to reduce the achievement gap were captured in the 1983 report, *A Nation at Risk* (Hopmann, 2008). The report and supporting rhetoric of the New Right helped shape the problem of U.S. education by “implicitly or explicitly attributing responsibility to particular individuals, institutions, or conditions” (Rosen, 2009, p. 276). The association of a problem with a solution is more easily accepted by the public when it is put in simplistic terms and aligns with already established cultural beliefs (Rosen, 2009). During the 1980s, the simple problem was poor performing schools, shifting the target of accountability from the individual to the schools (Lee, 2008). This “distinct change in direction and philosophy” (Harris & Herrington, 2006, p. 227) resulted in an enormous increase in the number of states with an accountability system from four in 1993 to forty in 2000 (Hanushek & Raymond, 2005). Testing for accountability was also linked to the modern school choice movement, a New Right push to put

increased pressure on schools through market based consumer choice. Comparative school level results, produced in testing for accountability systems, would help inform parental decision-making. The result was a significant increase in choice options in the U.S. in the 1980s and early 1990s (Eurydice, 2009a; Smith & Rowland, 2014).

Neoliberal ideas have dominated education reform in the U.S. since *A Nation at Risk* (Hursh, 2007). In the 1980s, Texas became the first state to implement testing for accountability (Yarema, 2010). This movement gained prominence nationally when President Bush and state governors met in 1989 to endorse the tying of student test scores to school performance. At approximately the same time the movement to standardize curriculum and instruction was gaining momentum (Hopmann, 2008). Standards were seen as essential for equity, ensuring that everyone was held to the same high expectations (Stotsky 2000). Testing for accountability was encouraged because aligning tests with higher standards would make the tests worth teaching to (Cohen & Ball, 1999; Spalding, 2000; Viadero, 1994), especially if the high standards measure important content (Koretz, 2008). The idea of national standards was embodied by President Clinton's "Goals 2000" which was supported by the Educate America Act of 1994. Goals 2000 pushed for national standards and implemented voluntary national testing in grades 4, 8, and 12 (Carl, 1994).

In 2001, the reauthorization of the Elementary and Secondary Education Act (ESEA), known as No Child Left Behind (NCLB), became the first national framework linking standards, assessment, and accountability (Datnow & Park, 2009). NCLB linked school performance with student scores on standardized examines and can be understood as "an evolution of previous attempts to use high-stakes tests to improve educational outcomes" (William, 2010, p. 110). Schools were judged on their ability to make adequate yearly progress (AYP) towards 100% student proficiency on achievement tests by 2014. Schools that failed to reach AYP for three consecutive years were subject to corrective action, including potential school closure (Springer, 2008). The emphasis on standardized tests was a boon for the testing industry who recorded a massive increase in test sales in the U.S. from \$260 million annually in 1997 to \$700 million annually in 2008, a lower bound estimate that does not take into account test support materials and services (Frontline, 2008).

New Right Policy Shifts in the U.K.

In a 1976 speech at Ruskin College Prime Minister Callaghan signaled a change in education policy toward the use of market mechanisms in the U.K.. Central to his speech was the notion that "the education system was not providing industry and the economy with what it required in terms of a skilled and well-educated workforce" (Furlong & Phillips, 2001, p. 6). The result of his speech was a shift in blame from larger societal issues, such as poverty and inequality, to ineffective schools (Hursh, 2005b). The subsequent Conservative government, led by Prime Minister Thatcher, made it clear that the individual had the responsibility to combat ineffective schools by making informed consumer choices that would pressure poor performing schools to change their practice.

Influenced by New Right ideology, the 1988 Education Reform Act and the 1992 Schools Act established national testing based on a national curriculum for ages 7, 11 and 14, and required local education authorities to produce school level comparable examination results, known as league tables (Edwards & Whitty, 1992; Teelken, 1999; West & Pennell, 2000). The 1988 Education Act was a simplified version of policy suggestions laid out by the Task Group of Assessment and Testing (TGAT), an expert group of practitioners and policy researchers that proposed testing for instructional and diagnostic purposes as well as systems evaluation and accountability. However, concerned that the TGAT brief was a subversion of left-wing educators, the TGAT brief was dismantled leaving policies that focus primarily on evaluating the system, schools, and educators (James, 2011). The 1988 act had at least four substantial effects on education policy in the U.K..

First, national curriculum was created and designed to occupy 70% of schools instructional time (Hursh, 2005b). Second, standardized tests were established at four Key Stages with the publication of results through league tables providing parents with the information necessary to make an informed consumer choice. National curriculum and standardized tests were necessary in order for decentralize fiscal decision making to the school level (Edwards, 2001). Third, open enrolment was established where students could enroll of the school of their choice, given space was available. However, space is rarely available in high performing schools. Additionally, open enrolment was linked per pupil funding, requiring schools to compete to ensure they have adequate school enrollment (Edwards & Whitty, 1992; Fitz, 2003). Finally, the power of Local Education Authorities (LEAs) was limited as curricular and pedagogic control was taken by the national government and fiscal decision-making was devolved to the school level. This resulted in weaker LEAs that essentially became the deliverer of national level policy (Fitz, 2003).

The entrenchment of testing for accountability continued in the 1990s. In 1991, the Parent Charter enhanced the choice environment by emphasizing the rights of parents as active choosers in their child's education, providing means for parents to evaluate the schools based on league tables and relocate their children if necessary to higher performing schools and establishment of Office for Standards in Education (Ofsted) in 1992 increased presence of accountability in education (James, 2011). The Labour Party further strengthened the national curriculum by specifying teaching methods in math and literacy (Hursh, 2005b). Somewhat surprisingly, it "not only been keenly committed to using the available levers created by the Conservatives, but ... added a raft of its own, to maintain pressure on schools to improve levels of attainment" (Fitz, 2003, p. 234). The Blair government of the late 1990s and early 2000s supported policies that prompted between school competition (DiGaetano, 2014). In an attempt to improve education by weeding out schools that unable to compete, the 1998 School Standards and Framework Act identified special measures that would be taken if a school failed inspection. Those schools that did not show improvement would be shut down (DiGaetano, 2014).

Global Transformation toward Testing for Accountability

Although there are some signs that early adopters (i.e. the U.K. and U.S.) are taking marginal steps away from holding schools accountable this has not slowed the global transformation toward testing for accountability. In the U.K., regional autonomy has led Scotland to scrap its testing program (Volante, 2007) and England to eliminate tests for 14 year olds in 2008 (Eurydice, 2009a), however league tables and between school competition still dominate U.K. education policy. In the U.S., congress has failed to reauthorize ESEA and provide an alternative to NCLB. With congress deadlocked, President Obama pushed through his seminal policy, Race to the Top, and provided waivers to states to circumvent some of the requirements put forth by NCLB, namely the arbitrary 2014 deadline for 100% proficiency. Both policies, however, continue the NCLB emphasis of basing school evaluations on comparable school level data which is available to the public and tying test scores to teacher livelihood through the implementation of "pay for performance" schemes (Dillon, 2011; McNeil & Klein, 2011; Smith & Rowland, 2014).

Regardless of perceived steps away from testing for accountability by the U.S. and U.K., the global expansion continues full speed as countries follow the early examples of the U.S. and U.K. and engage in the "ubiquitous adoption of accountability policies" (Hanushek & Raymond, 2004, p. 407). Testing for accountability is viewed as a common solution to education problems around the world, and an important part in the global education compact (Mundy, 2006) or "global education reform movement" (Sahlberg, 2010, p. 47), as illustrated in Butland (2008), Lemke et al. (2004), and

Figlio and Loeb (2011). Furthermore, accountability mechanisms that leave control to regional or national authorities are substantially similar across countries (Macnab, 2004). The adoption to testing for accountability by diverse countries around the world (see section 7 below for examples) suggests that “the development and implementation of accountability systems has been one of the most powerful, perhaps the most powerful, trend in educational policy in the last 20 years” (Volante, 2007, p. 4).

In many countries donor agencies play a significant role in shaping national policy. Increasingly multilateral organizations and international finance institutions reinforce testing for accountability by linking loan conditions to assessment infrastructure and policy (Kamens & McNeely, 2010). Of note is the World Bank’s movement toward supporting a testing culture. In their content analysis of the World Bank’s Education Sector Strategy 2020, released in 2011, Joshi & Smith (2012) find a nearly 100% increase in terms associated with the testing culture from the prior 1999 Sector Strategy paper. This included an astonishing increase in mentions of “accountability” from twice in the 1999 strategy to 32 times in the latest release. This led the authors to question whether a World Bank focus on “test-based education may be crowding out emphasis on well-rounded skills, personality development, and critical thinking that might come from a more problem-solving or dialogic approach to education” (p. 192). The recent establishment of the World Bank’s (SABER) tool provides additional evidence of a global trend towards increased accountability. SABER is a voluntary tool for nations to evaluate the effectiveness of their education system. It is strongly encouraged by the World Bank and includes provisions for national and international assessments. Countries that do not implement a national testing policy that ensures accountability or fail to participate in international assessments, such as PISA, receive lower grades (Bruns, Filmer, & Patrinos, 2011). Although voluntary in nature, the normative pressure placed on national leaders by SABER pushes countries toward adopting testing for accountability policies.

Explaining the Turn Toward Testing for Accountability

Similar to the convergence of democratic and republican policy in the U.S. or conservative and labour policy in the U.K., the global education compact that emerged at the beginning of the 21st century was a compromise between traditionally neo-liberal institutions, such as the World Bank and International Monetary Fund, and the equity focused United Nations (Daun & Mundy, 2011; Mundy, 2006). The “ideal” governance laid out in the compact largely mirrors the priorities of the New Right; including a focus on decentralization, incorporating the private sector into education, and the use of standardized tests (Mundy, 2006; Rose, 2005). Solidified in the Dakar Forum on Education for All and the Millennium Development Project (Kitaev, 2004; Mundy, 2006), the practices and expectations outlined in the global education compact are part of a larger world culture.

World Culture theory suggests that the global acceptance of testing for accountability can be understood as part of a larger cultural and collective process based around shared global values and ideas of legitimacy (Meyer, 1977). World Culture theory is one strand of a larger theoretical framework known as Neo-institutionalism, Sociological Institutionalism, or World Society theory (Schofer et al., 2012; Wiseman, Astiz & Baker, 2013). Neo-institutionalism sees “social action as deriving from culture, knowledge, and authority rooted in global institutions and structures” (Schofer et al., 2012, p. 57). Institutions, such as the family, religion, and education, help construct culture by expanding social roles and legitimating action and knowledge (Baker et al., 2006; Meyer, 1977). For example, “modern educational systems formally reconstruct, reorganize, and expand the socially defined categories of personnel and of knowledge in society” (Meyer, 1977, p. 72). The

cultural products of institutions are therefore shaped by the institution, strengthening and reinforcing its authority (Baker et al., 2006). Institutions influence actors at the local, regional, national, and international level through molding the culture they are embedded in (Schofer et al., 2012).

Individual actors recognize what is appropriate behavior within a given culture by the normative scripts or cultural models associated with each social role (Schofer et al., 2012). Scripts guide actors, telling them how to feel or act in the world (Baker, 2014; Baker et al., 2006). As within any culture, deviation from the script is subject to public scrutiny. The socialization process through which institutions create socially accepted scripts and culturally enforced role compliance often result in the internalization of acceptable and logical ways to engage in the surrounding social environment (Baker et al., 2006). The result is behavior that is taken for granted or understood as common sense and, therefore, beyond question. Essentially, culture has shaped behavior by identifying some actors and actions as legitimate while dismissing others (Jepperson, 2002).

Cultural institutions spur a process of cultural alignment across populations known as isomorphism (Wiseman, Astiz, & Baker, 2013). Education, as an institution, is an interesting example of isomorphism. From a neo-institutional perspective, the presence of similar education models globally are the result of shared meanings and values that identify appropriate rules and routines (Baker, 2014; Wiseman, Pilton, & Lowe, 2010). Schools and education systems then “become isomorphic with the institutional environment in order to achieve legitimacy and ensure their survival” (Booher-Jennings, 2005, p. 234).

Components of the World Culture

As an increasingly normative policy lever that provides legitimacy to countries that practice it and reconstructs the notion of education, teachers, and students, testing for accountability is one of the components of the dominant world culture. Embedded within the testing for accountability movement is the taken for granted assumption in neoliberalism and the power of competition to produce quality. Kamens & McNeely (2010) suggest that the growth in assessment internationally represents a move towards a world educational ideology that consists of unfailing faith in science as the path for legitimate knowledge and a belief that organizations can be managed to produce a desired outcome. As illustrated by policy practices in the World Bank, there is a growing international consensus that participation in international testing and the use of a national assessment system are essential in a legitimate education system. Additionally the number of countries involved in testing for accountability is likely to increase as more policymakers are “doing what is expected of them by their individual and institutional peers” (Wiseman, 2010, p. 2).

Multiple elements of the emerging world culture provide justification for testing for accountability systems, including: the expansion of western models, an emphasis on academic intelligence, faith in science as the rational path to truth, and the decentralization of authority to the local level.

Western Model. The world culture is western in origin, shaped through the western universities charter to produce legitimate knowledge and spread through the rapid expansion of educational attainment known as the education revolution (Baker, 2014; Ramirez, 2003). As an advanced schooled society, many American ideas of education can foreshadow global outcomes (Baker, 2014). Noticeably the western model preaches education for human development and education as a human right (Baker, 2014; Kamens & McNeely, 2010). Situated within this model is the increasing importance of education in later life outcomes. The position of schooling as both a private and public good encourages countries to implement mandatory policies requiring all children attend (Baker, 2014). Since all students have the ability to achieve, stratified schooling is deemed

unjust and decisions regarding equal rights to an education no longer center on access but quality. When combined with a western view of education as “a ‘technical’ science that can be studied, rationalized, and quantified” (Wiseman, 2010, p. 18), it is not surprising that the right to a high quality education leads to a push toward measurable indicators that can be used for accountability purposes (Kamens & McNeely, 2010).

Academic Intelligence. The education revolution has reinforced the cognitive and scientific dimensions of legitimate knowledge (Baker, 2014). Understood collectively as academic knowledge, this legitimate knowledge emphasizes meta-cognitive skills and the value of empirical evidence (Wiseman, 2010). Subjects that encourage this type of knowledge, namely mathematics and science, are considered valuable as demonstrated by an increasing use of mathematics achievement scores in public comparisons of schools and countries (Baker, 2014). Science and mathematics, as central to academic intelligence, is reflected in the work of Kamens, Meyer, and Benavot (1996) who found these subjects were no longer restricted to specialist knowledge but were now available to and intended for all students. Additionally, less academic subjects, such as the visual arts, have been dismissed for subjects that produce academic intelligence (Baker, 2014).

Faith in Science. Science production measured by the number of scientists, scientific publications, scientific training, and the number of countries with a national science program continues to increase globally (Baker, 2014). The swelling of science production is often called for by policymakers and practitioners who believe science to be an objective arbiter of truth (Rosen, 2009). Critics of the unquestionable faith in science recognize that “as long as the public maintains this irrefutable objectivity of statistics, a graph here and a chart there can leverage support for provincial reforms that could never survive nuanced deliberation” (Robertson, 1999, p. 715). Sciences taken for granted position increases the value placed on education that uses test scores to objectively and accurately measure student knowledge (Paris & McEvoy, 2000). When international test scores are reported their “seemingly authoritative measure of students’ skills and abilities” (Cohen & Rosenberg, 1977, p. 128) prompt nations to respond through the development of appropriate educational reform (Drori et al., 2003).

Decentralization. Decentralization has become a widespread institutional model as “a significant set of nations have responded to the legitimizing global forces within a multinational economy and world institutional system by adopting decentralization” (Astiz, Wiseman, Baker, 2002, p. 86). In Europe the increasing devolution of responsibility to the local level has been met with increased curricular and evaluative control at the regional or national level (Eurydice, 2009a). Testing for accountability is likely in decentralized systems because external exams are required to ensure education quality across diverse communities, where local control often leads to information asymmetry (Woessman, 2004, 2007). From this perspective, “statistical accountability systems” are seen as “one way to resolve the dilemma between granting autonomy and authority to educators and keeping them under some political control” (Dorn, 2007, p. 13).

Neoliberalism. Neoliberalism, as an economic system, has spread to nearly every country in the world (Friedman, 1999). Neoliberalism promotes private property, open markets, and free trade on the basis of three core assumptions: (1) consumers have access to accurate market information, (2) consumers act as self-interested profit maximizers, and (3) private provision and competition will be more efficient than public control (Harvey, 2005; Jolly, 2003). The diffusion of neoliberalism as a legitimate economic approach occurred once it emphasized the general benefits of competition, embraced the role of actor as central to global institutions, and was viewed as a legitimate system by international institutions (Schofer et al., 2012).

As a dominant approach to education policymaking, neoliberal practices are often adopted by countries seeking legitimacy (Wiseman, 2010). The implementation of neoliberal policy reinforces

cultural faith in its promise of increased performance with improved efficiency, solidifying the position of neoliberalism as a “common sense” approach that cannot be questioned (Apple, 1999; Hursh, 2007; Rosen, 2009). Treating education as a market invites the invisible hand of the market to improve the quality of schools through between school competition for students (Apple, 1999; Levin, 1992). For neoliberal policymakers academic gains must be balanced with financial costs (Wiseman, 2010). Public investment in education must be justified (Smeed & Victory, 2010): “the public has a right to expect that its resources are being used responsibly and that the public institutions are accountable for caretaking the public trust” (Supovitz, 2009, p. 215). Test results, therefore, provide an easily measured indicator of quality to ensure the public investment in education is being used efficiently and effectively.

Test Results as Information. The publication of standardized test results provides parents, acting in the role of consumers, with easily interpretable market information allowing them to put more pressure on schools to respond to consumer demand (Ball, 1993; Woessman, 2007). The ability of parents to act rationally as self-interested consumers is a prerequisite to an effective market (Apple, 1999). The use of test results as information helps overcome the principal-agent problem, recognized by many neoliberal economists (Figlio & Loeb, 2011). In situations where a principal (i.e. parents) hire an agent (i.e. educators) to perform a service, if the interests of the principal and the agent do not align, the service may not be performed efficiently. Student test scores provide a common metric of measurement for evaluation, informing the principal of the agents real performance and providing evidence to ensure that the needs of the principal are being met (Woessman, 2004). Test results can also be conceptualized as quality indicators, which can help direct resources (Joshi & Smith, 2012) and aid the government in targeting inefficient schools to be shut down (Lincove, 2009). The use of test scores in this manner is acceptable as they are often considered the only legitimate measure of quality in education. With this authority, the implementation of testing is increasingly seen as an end in education, instead of a means to improve student understanding (Booher-Jennings, 2005).

Test scores are used to spur parent involvement in the education process (Smith & Rowland, 2014). The U.K. used their league tables to employ parents to make market choices, with the government telling parents that “you should get all the information you need to keep track of your child’s progress, to find out how the school is being run, and to compare all schools” (Department of Education, 1994, p. 3). While there is some evidence to support the use of this information in the parent’s decision-making regarding their child’s school attendance (Teelken, 1999), others find parents pay little attention to the publication of data (de Wolf & Janssens, 2007). Janssens and Visscher (2004) suggest that parents pay little attention to this information because they do not have access to the information, they lack the capacity to understand the information, they have limited real choice between schools, and the information does not reflect the factors that parents base their decision on.

National Testing Policy Categories

As the expansion of world culture legitimates and institutes testing for accountability as a taken for granted education policy it is important to remember that variation in national policy remains. Examining national policy is important because “national ministries of education typically act as agents imposing this activity [testing] on schools and education systems” (Kamens & McNeely, 2010, p. 6). Differences in national testing policy can best be seen on a rough continuum based on the presence and intensity of testing for accountability (see Table 2). This categorization

strategy is similar to the scheme of Eurydice (2009a) and the stated purposes of education policymakers often transcend policy categories (Eurydice, 2009a).

Table 2

National Testing Policy Categories

Testing for Assessment	Testing for Assessment, Advancement, or Accreditation	Testing for Accountability	
Formative	Summative	Evaluative	Punitive
The use of national or regional examinations as diagnostic tools that are used internally by schools to inform instruction.	The use of national or regional examinations as a tool that summarizes student learning and is shared with parents; when disseminated is done so at the national or regional level.	The use of national or regional examinations as a tool that summarizes student learning and is disseminated at the school level to allow for between school comparisons.	The use of national or regional examinations as a tool that summarizes student learning, is disseminated at the school level, with school/class level results used to apply rewards or sanctions.

At the far left of the continuum are *formative testing policies* which use tests for assessing the progress of students. In this system, tests are ongoing, informing teacher instruction through direct feedback (Eurydice, 2009a). Formative policies are often professed by policymakers, however, more often than not this gesture is simply policy rhetoric (Irons & Harris, 2006). *Summative testing policies* use tests to assess, provide accreditation, or direct student's educational advancement. They summarize how well an individual is doing at a given point in time (Eurydice, 2009a; Nitko & Brookhart, 2006) and when scores are disseminated they are done so at the national or regional level. Traditional high-stakes student tests that do not aggregate results at the school level and tests that disaggregate scores by ethnic, economic, or regional subgroup, but not school, fall into this policy category.

Evaluative and punitive testing policies both apply testing for accountability. In *Evaluative testing policies* test scores aggregated at the school level are used by the public to compare schools and evaluate school quality. In this economically based model, "responsibility is devolved to the individual consumer and the aggregate of consumer choices provides the discipline, of accountability and demand, that the producer cannot escape from" (Ball, 1995, p. 69). Informal consequences in this system result from a consumer choice market mechanism and public stigmatization through "naming and shaming" (de Wolf & Janssens, 2007) or the "scarlet letter" effect (Harris & Herrington, 2006).

In *Punitive testing policies* school level aggregate data is used to apply consequences through the application of formal rewards or sanctions. The primary difference between evaluative and punitive testing policies can be found in the consequences. Implicit consequences through public pressure

characterize evaluative systems while explicit consequences through formal channels characterize punitive systems. Punitive systems function through a behaviorist model, which suggests individual action can be molded through incentives (Hanushek & Raymond, 2004). Punitive testing policy responds to reformers that believe serious consequences are necessary to transform the education system and has both compliance and avoidance costs, making it expensive to implement (McDonnell & Elmore, 1987).

Examples of National Turns toward Testing for Accountability

To illustrate the turn toward testing for accountability the following section explores three heterogeneous examples of countries that have transitioned toward a punitive testing policy. The three examples of Hungary, Mexico, and South Korea indicate that the transition toward testing for accountability is not limited by a country's economic status or regional position. As illustrated in figure 3 the movement toward a punitive policy is not dependent on the relative achievement of a nation's education system. Similar to table 2, figure 3 uses yellow to indicate an evaluative testing policy and red to designate a punitive testing policy. Looking across mathematics test scores across four rounds of PISA (2000, 2003, 2006, and 2009), no evidence of increasing achievement is observed as countries turn toward more intense measures of accountability.

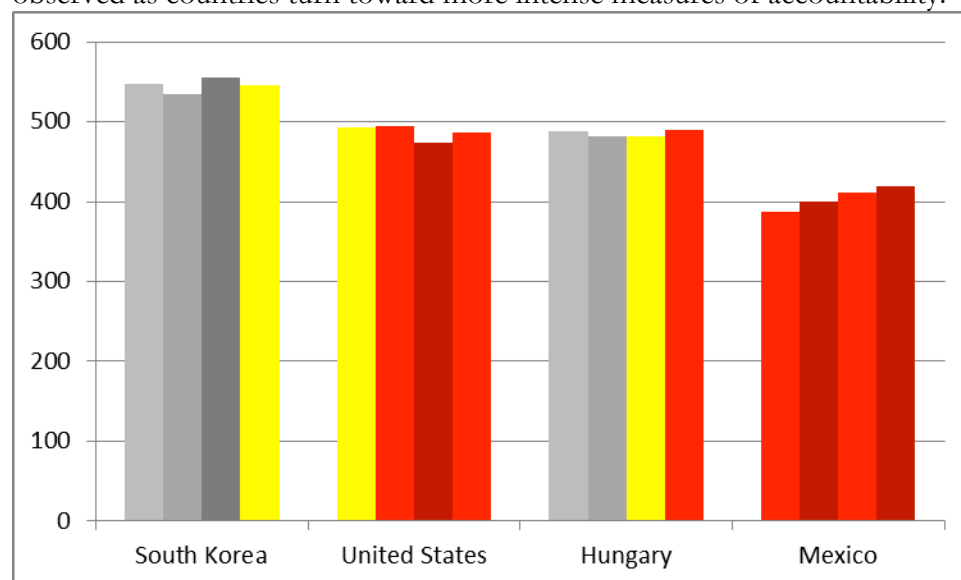


Figure 3. Country mean mathematics scores for South Korea, the United States, Hungary, and Mexico as measured by the 2000, 2003, 2006, and 2009

Source: PISA

Hungary

Hungary is a country that has traditionally been without a national testing system but has seen dramatic changes over the past two decades (Eurydice, 2009b). After spending 33 years as part of the Soviet bloc, education policies in the 1980s and 1990s provided a lot of freedom, if not a lot of guidance. The 1985 Education Act abolished the previous inspection system but failed to replace it with a viable way to measure education quality. The rapid democratization and decentralization of Hungary following the collapse of the Soviet Union led it to have one of the most decentralized education systems in Europe (Eurydice, 2009a, 2009b). During the 1990s, in a shift that was partially fueled by national performance on international assessments, it became clear that without a national test and with dissension in defining education quality assessment reform was needed. Starting in

2001, 'monitoring surveys' were implemented yearly and the National Assessment of Basic Competencies (NABC) was established (Eurydice, 2009a, 2009b). Originally testing 5th and 9th graders, the NABC quickly expanded to include 4th, 6th, and 8th grade. The early goal of the test was to develop a within school evaluation culture and allow schools to compare their results to nationally aggregated sub-groups. However, since 2006 results have been disseminated to the general public, providing the school's clients (parents) with information on school effectiveness (Eurydice, 2009a). This evaluative policy shifted quickly to a punitive policy when in 2008 schools were mandated to incorporate test scores into internal quality reports. Low achieving schools were then required to use this report to prepare and implement an action plan for remediation (Eurydice, 2009b). As illustrated in Figure 3, during this period of rapid transition toward a punitive testing system Hungary saw a less than 0.5% improvement in PISA mathematics scores.

Mexico

Since the passage of the National Agreement for the Modernization of Basic Education in 1992 Mexico has applied a punitive national testing policy by tying student test scores to explicit rewards for teachers. Interestingly, this policy, usually opposed by teacher unions was adamantly support by the Mexican teachers union. Hecock (2014) suggests that this support was partially due to the union's ability to use their position as an entrenched powerful institution, at a time when democratization in the country was in its infancy, to co-opt the policy. Specifically the union was able to ensure that once a performance raise was given it could not be rescinded and that administrators would be subject to a separate merit pay system.

The pay for performance policy, known as Carrera Magisterial (CM), links the results of the Instrument for Testing New Secondary School Pupils (IDANIS) to teacher bonuses (Ferrer, 2006). Teachers volunteer to participate in CM, which provides rewards of up to 300% of their base wage (Hecock, 2014). Teachers are evaluated on a one hundred-point scale, covering a myriad of outcomes. Over time student test scores have played a larger role in teacher evaluation; points associated with student test scores increased from 20 between 1999 and 2011 to 50 today (Hecock, 2014). Student test scores are also publicly available. After initial differences due to capacity at the regional level, by 2000 nearly all of Mexico participated in publishing school level data (Ferrer, 2006; Hagerstrom, 2006). Additional regional variation is present in the concentration of high-performing teachers, with the segregation of high-performing teachers in more advantaged communities exasperating between school differences (Luschei, 2012).

In 2001 the quality schools program (PEC) was established as a voluntary program in which schools could apply for a competitive grant by submitting a five-year improvement plan. Increasing in popularity, there are concerns about the financial feasibility of the program (Hagerstrom, 2006). National mean scores in mathematics over the past decade have increased by roughly 8% (see figure 3), however, considering the early establishment of a punitive system more information is needed to see whether the improvement can be partially attributed to the systems presence, duration, or other unrelated factors.

South Korea

In 1991 South Korea decentralized their education system. Since that time the Korean Institute for Curriculum and Evaluation (KICE) has been responsible for the administration of national assessments. The National Assessment of Educational Achievement (NAEA) is a criterion reference test administered in the 6th, 9th, and 10th grade focusing on Korean, mathematics, science, social studies, and English. Initially the NAEA was a sample-based test with results aggregated and disseminated at the national level (Chung, 2014). In 2007, plans were unveiled to move to a census test and publish results at various levels, including the school. President Lee Myung-bak declared

that moving to an evaluative policy was essential for school choice to work. Teachers and teacher unions opposed the move on the grounds that Korea was already a high scoring country and national policy should be focusing on “creative” skills, and as a result school level publication was delayed until 2011 (Schmidt, Houang, & Shakrani, 2009). In 2011, the government began to target low performing schools known as ‘creative management schools that pursue academic ability enhancement’ (Kim et al., 2010) and “similar to the provisions in the No Child Left Behind Act in the U.S., Korea’s core plan is to provide additional support to schools with lots of children who are underperforming, but only for a specific period of time” (Schmidt, Houang, & Shakrani, 2009, p. 56). The transition into a punitive policy, however, slowed in 2013 as new president, Park Geun-hye, felt pressure from the public and teachers and eliminated tests at the elementary school level and cut back middle school assessments (Chea, 2014). During the period Korean leadership pushed toward more intense forms of testing for accountability national mathematics scores on PISA saw essentially no change. However, in the years directly before and after the President Myung-bak’s announced change in school test score reporting, the national mean score dropped by approximately 2% (see figure 3).

Toward a Normative Testing Culture?

With more countries transitioning toward testing for accountability systems it is important to move beyond describing testing as part of the world culture to investigate the substantive outcomes of this global trend (Schofer et al., 2012). World Culture theory is often criticized for its lack of attention to power and its tendency to describe the components and breadth of the culture without explicitly addressing the potential consequences. For example, Carney, Rapple, and Silova (2012) state that World Culture theory undertheorizes aspects of agency and power and suggest that through recognizing practices such as shadow education and decentralization as part of the larger world culture, World Culture theorists are implicitly endorsing such practices as efficient and effective. While one can argue that origins and outcomes are not central in a theory focused on describing commonalities across heterogeneous environments, an important next step, once legitimate practices embedded in the world culture are identified, is the examination of potential benefactors of this world culture. As testing for accountability expands globally it is important for researchers and policymakers to take this next step and explore potential positive and negative consequences of tying students test scores to educator livelihood.

Past meta-analyses identify a positive effect of accountability on student test scores ranging from a marginal effect size of 0.10 (Belfield & Levin, 2002) to a medium effect size of 0.55 (Phelps, 2012)¹. However, these studies were dominated by U.S. examples and did not distinguish between high stakes testing for advancement and educator focused testing for accountability. Additionally, the type of testing for accountability applied can lead to divergent results. Studies suggest that evaluative policies, in which schools are compared through the publication of results, has a positive effect on student test scores, although the “practical significance of this gain is negligible” (Springer, 2008, p. 5). Additionally, in punitive systems, where explicit consequences are present, student achievement is higher (Dee & Jacob, 2009). When evaluative and punitive systems are compared, the relative advantage of punitive policy outweighs the market pressure of evaluative systems (Bishop et al., 2001; Hanushek & Raymond, 2005). However, a recent study in which participants from the 2009 PISA were categorized into the four national testing policy categories described above found no difference between student math achievement in Summative, Evaluative, and Punitive systems

¹ The effect size in Lee (2008) lies between these two studies at 0.27

once schools that use student achievement as a criterion in admission was accounted for (Smith, 2014).

Concurrent with the potential achievement gains from increased accountability are concerns that educators game the system to ensure their survival and that testing for accountability systems and their corresponding school practices increase inequity. Past studies suggest that educators in testing for accountability systems are more likely to shape the testing pool by increasing student repetition (Hursh, 2005a; Kornhaber, 2004a, 2004b), reclassifying students into special education (Cullen & Reback, 2006; Haney, 2000; Jacob, 2005), or suspending low achieving students during testing periods (Figlio, 2006). Schools in testing for accountability systems are also more likely to narrow the curriculum by shifting their resources toward testing subjects and away from non-tested subjects (Rentner et al., 2006; Supovitz, 2009), participate in increased test practice (Au, 2007; Cuban, 2007; Nicols & Berliner, 2007; Yarema, 2010), including explicit direct instruction (Certo, 2006), and focus instructional energy on those students closest to the desired proficiency level (Booher-Jennings, 2005; Gillborn & Youdell, 2000; Lipman, 2004; Reback, 2008). Finally, upsetting trends in equity are present as the effect of accountability on student achievement varies significantly by student ethnicity (Kornhaber, 2004a; McNeil, 2000; McNeil & Valenzuela, 2001; Wheelock, Bebell & Haney, 2000), native language (Monk, Sipple, & Killeen, 2001), and prior achievement (Jacob, 2001; West & Pennell, 2000).

The global transformation toward testing for accountability outlined in this article should lead policymakers to question whether the potential benefits of implementing such a program outweigh the concerns, especially among the most marginalized student groups. Unfortunately rich discussions of this nature are less likely among national decision makers in the future, as testing for accountability is increasingly legitimated as a neoliberal script, which lays out appropriate action for nation-states within a world culture that emphasizes faith in science, academic knowledge, and western style education. Testing as the way to acquire valued knowledge is now taken for granted. As testing for accountability becomes a normative practice, engrained in the educational landscape of more and more countries, it has the potential to reconstruct education and how it is perceived by its actors and the general public.

Testing for accountability, as a neoliberal policy, reinforces the reconceptualization of students, parents, and teachers, as products, consumers, and producers (Carl, 1994). With educator survival on the line, students are judged by their proclivity to pass a test. Remedial students, with long odds of passing the test, are considered a liability. The social construction of this student group as a “liability” happens early in schooling and may follow the path of similarly constructed social categories that are now engrained in the rhetoric of education, “dropouts” and “at-risk” students (Baker, 2014; Fine, 1991; Swadener & Lubeck, 1995). In a schooled society, students that struggle in school will be identified as deviant and with education increasingly used as the only legitimate form of stratification, ‘liability’ students will be marred by that status for the rest of their life (Baker, 2014).

As a centerpiece of testing for accountability systems, between school competition shifts “schools *modi operandi* from those based on moral purpose towards those that emphasize productivity and efficiency” (Sahlberg, 2010, p. 48). Once productivity and efficiency are established as the essential aim of schooling, testing for accountability policies may be criticized but will rarely be abolished (Tyack & Cuban, 1995). Competition between educators sharply contrasts with more cooperative models that are important for healthy school climates and student and teacher motivation. In systems where scores are aggregated at the class level teachers may be concerned about peer judgment and being stigmatized as ‘bad’ teachers (Booher-Jennings, 2005). This tumultuous situation leads to educators blaming others, especially those in earlier grades, for not

adequately preparing students (Wiggins & Tymms, 2000). Internal feelings of anxiety and shame among teachers are exasperated by a belief that an emphasis on testing for accountability has a negative effect on public education, often stymieing their motivation (Certo, 2006; Jones & Egley, 2006; Smeed & Victory, 2010). Additionally, the focus on test scores suggests that everyone is welcome in teaching as long as they can produce the test gains needed in an accountability system (Hopmann, 2008). This assumption of importance threatens the professional position of teachers, delegitimizing it as a profession that requires long-term pedagogical training. The importance of test scores as information assumes that parents engage with the data and use it to drive their children's school placement. Parents that are uninformed, do not use evidence based decision-making, or do not participate in school choice will be shunned by society (Ball, 1993). Evidence of subjective evaluations can already be found in the literature. For example, Woessman (2004) suggests that parents ability and willingness to make use of available information is a measure of "how strongly parents care for their children's progress" (p. 4).

The long-term societal effect of treating education as a market, concerned primarily with private returns, is a reduction in public spending on social services, such as education and health (Hopmann, 2008). These reductions may be partially ameliorated by increased private support, accelerating the movement toward privatization. Evaluating quality in education is a challenging endeavor due to education's multifaceted short term and long-term outcomes. Test results provide a simple, relatively easy to understand measure that can be used with investors who want to ensure their money is going towards a high quality product. Recent literature suggests this is the case, as the publication of results increase the amount of voluntary contributions a school receives (Figlio & Kenny, 2009). The publication of results can also shape society through residential segregation. Families use school level test scores to judge the quality of the school system leading to the establishment of more "highly desirable" neighborhoods and effecting real estate prices (Figlio & Lucas, 2004).

Testing for accountability is so engrained in many countries that it is partially self-perpetuating (Dorn, 2007). The use of assessment data reinforces the testing culture (Baker & Wiseman, 2005) and the public view of testing for accountability as synonymous with high expectations makes it challenging for policymakers to alter established practices, whether or not they want to, in fear of constituents labeling them soft on education (Paris & McEvoy, 2000). If the potential benefits and concerns of testing for accountability are not critically evaluated this global transformation will lead to a testing culture that is internalized as normative and adopted as individual values. This shift in constitutional mind-sets has the ability to affect the whole of society as "deeply engrained ways of understanding the relationship between the public and its institutions" (Hopmann, 2008, p. 425) are altered. Recognizing the cultural diffusion of testing for accountability and evaluating it before its cultural entrenchment is essential for the world to avoid the challenges faced by early adopters.

References

- Apple, M. (1999). Rhetorical reforms: Markets, standards, and inequality. *Current Issues in Comparative Education*, 1(2), 6-17.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267. <http://dx.doi.org/10.3102/0013189X07306523>
- Baker, D. (2014). *The schooled society: The educational transformation of global culture*. Stanford, CA: Stanford University Press.
- Baker, D., & LeTendre, G. (2005). *National differences, global similarities: World culture and*

- the future of schooling*. Stanford, CA: Stanford University Press.
- Baker, D., Thorne, S., Blair, C., & Gamson, D. (2006). Cognition, culture, and institutions: Affinities within social construction of reality. Unpublished paper. University Park, PA: The Pennsylvania State University.
- Baker, D., & Wiseman, A. W. (Eds.). (2005). *Global trends in educational policy* (Vol. 6, International Perspectives on Education and Society Series). Oxford, U.K.: Elsevier Science.
- Ball, S. J. (1995). Parents, schools and markets: The repositioning of youth in United Kingdom education. *Young*, 3(3), 68-79. <http://dx.doi.org/10.1177/110330889500300305>
- Ball, S. J. (1993). Education markets, choice and social class: The markets as a class strategy in the U.K. and the USA. *British Journal of Sociology of Education*, 14(1), 3-18. <http://dx.doi.org/10.1080/0142569930140101>
- Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *The Journal of Political Economy*, 70(5), 9-49. <http://dx.doi.org/10.1086/258724>
- Belfield, C. R., & Levin, H. M. (2002). The effects of competition between schools on educational outcomes: A review for the United States. *Review of Educational Research*, 72(2), 279-341. <http://dx.doi.org/10.3102/00346543072002279>
- Benavot, A., & Tanner, E. (2007). The growth of national learning assessments in the world, 1995-2006. Background paper for the Education for All Global Monitoring Report 2008: Education for All by 2015: Will we make it.
- Ben Jaafar, S., & Anderson, S. (2007). Policy trends and tensions in accountability for educational management and services in Canada. *Alberta Journal of Educational Research*, 53(2), 205-225.
- Bishop, J., Mane, F., Bishop, M., & Moriarity, J. (2001). The role of end-of-course exams and minimal competency exams in standards-based reforms. *Brookings Papers in Educational Policy*. Washington, DC: Brookings Institution.
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Education Research Journal*, 42(2), 231-268. <http://dx.doi.org/10.3102/00028312042002231>
- Bruns, B., Filmer, D., & Patrinos, H. (2011). *Making schools work: New evidence on accountability reforms*. Washington, DC: The World Bank. <http://dx.doi.org/10.1596/978-0-8213-8679-8>
- Butland, D. (2008). *Testing times: Global trends in marketisation of public education through accountability testing*. Marrickville NSW: NSW Teachers Federation.
- Carl, J. (1994). Parental choice as national policy in England and the United States. *Comparative Education Review*, 38(3), 294-322. <http://dx.doi.org/10.1086/447254>
- Carney, S., Rappleye, J., & Silova, I. (2012). Between faith and science: World Culture theory and comparative education. *Comparative Education Review*, 56(3), 366-393. <http://dx.doi.org/10.1086/665708>
- Certo, J. (2006). Beginning teacher concerns in an accountability-based testing environment. *Journal of Research in Childhood Education*, 20(4), 331-349. <http://dx.doi.org/10.1080/02568540609594571>
- Chapman, D., & Snyder, C. (2000). Can high stakes national testing improve instruction: Re-examining conventional wisdom. *International Journal of Educational Development*, 20, 457-474. [http://dx.doi.org/10.1016/S0738-0593\(00\)00020-1](http://dx.doi.org/10.1016/S0738-0593(00)00020-1)
- Chea, H. (2014). National achievement tests in Korea and the U.S.A.: How does it impact policy, schools, and students? Presented at the Comparative and International Education Society Conference, Toronto, Canada.

- Chubb, J., & Moe, T. (1990). *Politics, markets, and America's schools*. Washington, D.C: Brookings Institution Press.
- Chung, P. J. (2014). National achievement test and supplemental educational service in USA and Korea. Presented at the Comparative and International Education Society Conference, Toronto, Canada.
- Clotfelter, C., Ladd, H., Vigdor, J., & Diaz, R. (2004). Do school accountability systems make it more difficult for low performing schools to attract and retain high quality teachers? *Journal of Policy Analysis and Management*, 23(2), 251–271. <http://dx.doi.org/10.1002/pam.20003>
- Cohen, D. K., & Ball, D. L. (1999). Instruction, capacity, and improvement. In *Consortium for policy research in education (CPRE) report series RR-43* (pp. 1-41). Philadelphia, PA: University of Pennsylvania Graduate School of Education.
- Cohen, D. K., & Rosenberg, B. H. (1977). Functions and fantasies: Understanding schools in capitalist America. *History of Education Quarterly*, 17(2), 113-137. <http://dx.doi.org/10.2307/368122>
- Cuban, L. (2007). Hugging the middle. Teaching in an era of testing and accountability, 1980–2005. *Education Policy Analysis Archive*, 15(1). <http://epaa.asu.edu/ojs/article/view/49/175>
- Cullen, J., & Reback, R. (2006). Tinkering towards accolades: School gaming under a performance accountability system. In T. Gronberg & D. Jansen (Eds.) *Advances in applied microeconomics* (pp. 1-34). London: Emerald Group Publishing Limited.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64(1), 5-30.
- Datnow, A., & Park, V. (2009). Conceptualizing policy implementation: Large-scale reform in an era of complexity. In G. Sykes, B. Schneider & D. Plank (Eds.) *Handbook on Education Policy Research* (pp. 348-361). New York: Routledge Publishing.
- Daun, H., & Mundy, K. (2011). *Education governance and participation with focus on developing countries*. Stockholm: Stockholm University, Institute of International Education.
- Dee, T., & Jacob, B. (2009). The impact of No Child Left Behind on student achievement. *National Bureau of Economic Research Working Paper No. 15531*.
- Department of Education. (1994). *Our children's education: The updated parent's charter*. London: DFE.
- De Wolf, I., & Janssens, F. (2007). Effects and side effects of inspection and accountability in education: An overview of empirical studies. *Oxford Review of Education*, 33(3), 379-396. <http://dx.doi.org/10.1080/03054980701366207>
- DiGaetano, A. (2014). Accountability school reform in comparative perspective. *Urban Affairs Review*. Published online 25 June 2014. <http://dx.doi.org/10.1177/1078087414537606>
- Dillon, S. (2011). Obama to waive parts of No Child Left Behind. *New York Times* (September, 22, 2011).
- Dorn, S. (2007). *Accountability Frankenstein: Understanding and taming the monster*. Charlotte, NC: Information Age Publishing.
- Dorn, S. (1998). The political legacy of school accountability systems. *Education Policy Analysis Archives*, 6(1). Retrieved from <http://epaa.asu.edu/ojs/article/view/568/691>
- Drori, G., Meyer, J.M., Ramirez, F.O., & Schofer, E. (Eds.) (2003). *Science in the world polity: Institutionalization and globalization*. Stanford, CA: Stanford University Press.
- Earl, L., & Torrance, N. (2000). Embedding accountability and improvement into large-scale assessment: What differences does it make? *Peabody Journal of Education*, 75(4), 114-141. http://dx.doi.org/10.1207/S15327930PJE7504_6
- Eckstein, M., & Noah, H. (1993). *Secondary school examinations*. New Haven, CT: Yale

- University Press.
- Edwards, T. (2001) Educational performance, markets and the state: Present and future prospects. In R. Phillips & J. Furlong (Eds.) *Education reform and the State: Twenty-five years of politics, policy, and practice* (pp. 239-254). New York: Routledge.
- Edwards, T., & Whitty, G. (1992). Parental choice and educational reform in Britain and the United States. *British Journal of Educational Studies*, 40(2), 101-117.
<http://dx.doi.org/10.1080/00071005.1992.9973916>
- Eurydice. (2009a). *National testing of pupils in Europe: Objectives, organisation and use of results*. Education, Audiovisual and Culture Executive Agency, European Commission.
- Eurydice. (2009b). *National testing of pupils in Europe: Objectives, organisation and use of results – Country level reports: Hungary*. Education, Audiovisual and Culture Executive Agency, European Commission.
- Falk, B. (1996). Issues in designing a learner-centered assessment system in New York state: Balancing reliability and flexibility, authenticity, and consequential validity. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.
- Feng, L., Figlio, D., & Sass, T. (2009). School accountability and teacher mobility. *University of Wisconsin Working Paper*.
- Ferrer, G. (2006). *Educational assessment systems in Latin America: Current practice and future challenges*. Washington, D.C.: Partnership for Educational Revitalization in the Americas.
- Figlio, D. (2006). Testing, crime, and punishment. *Journal of Public Economics*, 90, 837-851.
<http://dx.doi.org/10.1016/j.jpubeco.2005.01.003>
- Figlio, D., & Kenny, L. (2009). Public sector performance measurement and stakeholder support. *Journal of Public Economics*, 93(9-10), 1069-1077.
<http://dx.doi.org/10.1016/j.jpubeco.2009.07.003>
- Figlio, D. & Loeb, S. (2011). School accountability. *Handbook of the Economics of Education*, 3, 383-421. <http://dx.doi.org/10.1016/B978-0-444-53429-3.00008-9>
- Figlio, D., & Lucas, M. (2004). What's in a grade? School report cards and the housing market. *American Economic Review*, 94(3), 591-604. <http://dx.doi.org/10.1257/0002828041464489>
- Fine, M. (1991). *Framing dropouts: Notes on the politics of an urban public high school*. New York, NY: SUNY Press.
- Fitz, J. (2003). The politics of accountability: A perspective from England and Wales. *Peabody Journal of Education*, 78(4), 230-241. http://dx.doi.org/10.1207/S15327930PJE7804_12
- Freidman, M. (1962). *Capitalism and freedom*. Chicago, IL: University of Chicago Press.
- Freidman, M. (1955). The role of government in education. In Solo, R. (ed.) *Economics and the public interest*. New Brunswick, NJ: Trustees of Rutgers College in New Jersey.
- Friedman, T. (1999). *The Lexus and the olive tree*. New York: Random House.
- Froese-Germaine, B. (2001). Standardized testing + high-stakes decisions = educational inequity. *Interchange*, 32(2), 111-130. <http://dx.doi.org/10.1023/A:1011985405392>
- Frontline. (2008). *The testing industries Big Four*. From
<http://www.pbs.org/wgbh/pages/frontline/shows/schools/testing/companies.html>
- Furlong, J. & Phillips, R. (2001) Introduction and rationale. In R. Phillips & J. Furlong (Eds.) *Education reform and the State: Twenty-five years of politics, policy, and practice* (pp. 3-11). New York: Routledge.
- Gillborn, D. (1996). *Exclusions from School*. London: University of London.
- Gillborn, D., & Youdell, D. (2000). *Rationing education: Policy, practice, reform and equity*. Buckingham, U.K.: Open University Press.

- Gilles, C., Cramer, M., & Hwang, S. (2001). Beginning teacher perception of concerns: A longitudinal look at teacher development. *Action in Teacher Education*, 23(3), 92-98. <http://dx.doi.org/10.1080/01626620.2001.10463079>
- Gipps, C. (2003). Educational accountability in England: The role of assessment. Paper presented at the Annual Meeting of the American Education Research Association, New York, NY.
- Gipps, C., & Murphy, P. (1994). *A Fair Test?* Philadelphia: Open University Press.
- Hagerstrom, M. (2006). Basic education. In *Decentralized service delivery for the poor* (pp. 1-47). Washington, DC: The World Bank.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41).
- Hanushek, E., & Raymond, M. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297-327. <http://dx.doi.org/10.1002/pam.20091>
- Hanushek, E., & Raymond, M. (2004). The effect of school accountability systems on the level and distribution of student achievement. *Journal of the European Economic Association*, 2(2-3), 406-415. <http://dx.doi.org/10.1162/154247604323068096>
- Harris, D., & Herrington, C. (2006). Accountability, standards, and the growing achievement gap: Lessons from the past half-century. *American Journal of Education*, 112(2), 209-238. <http://dx.doi.org/10.1086/498995>
- Harvey, D. (2005). *A brief history of neoliberalism*. New York, NY: Oxford University Press.
- Hecock, R. D. (2014). Democratization, education reform, and the Mexican teachers' union. *Latin American Research Review*, 49(1), 62-82. <http://dx.doi.org/10.1353/lar.2014.0008>
- Hill, D. (2006). Class, capital and education in this neoliberal and neoconservative period. *Information for Social Change*, 23, 11-34.
- Ho, P. (1964). *The ladder to success in imperial China: Aspects of social mobility*. New York: Science Editions.
- Hopmann, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. *Journal of Curriculum Studies*, 40(4), 417-456. <http://dx.doi.org/10.1080/00220270801989818>
- Hursh, D. (2007). Assessing No Child Left Behind and the rise of neoliberal education policies. *American Educational Research Journal*, 44(3), 493-518. <http://dx.doi.org/10.3102/0002831207306764>
- Hursh, D. (2005a). The growth of high-stakes testing in the USA: Accountability, markets and the decline of educational equality. *British Educational Research Journal*, 31(5), 605-622. <http://dx.doi.org/10.1080/01411920500240767>
- Hursh, D. (2005b). Neo-liberalism, markets and accountability: Transforming education and undermining democracy in the United States and England. *Policy Futures in Education*, 3(1), 3-15. <http://dx.doi.org/10.2304/pfie.2005.3.1.6>
- Irons, E. J., & Harris, S. (2006). *The challenges of No Child Left Behind: Understanding the issues of excellence, accountability, and choice*. Lanham, MD: Rowman and Littlefield Education.
- Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics*, 89(5-6), 761-796. <http://dx.doi.org/10.1016/j.jpubeco.2004.08.004>
- Jacob, B. A. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, 23(2), 99-121. <http://dx.doi.org/10.3102/01623737023002099>
- Jacob, B.A., & Levitt, S.D. (2003). Rotten apples: An investigation of the prevalence and

- predictors of teacher cheating. *Quarterly Journal of Economics*, 118(3), 843–878.
<http://dx.doi.org/10.1162/00335530360698441>
- James, M. (2011). Assessment for learning: Research and policy in the (Dis)United Kingdom. In R. Berry & B. Adamson (Eds.) *Assessment Reform in Education* (pp. 15-32). New York: Springer.
- Janssens, F., & Visscher, A. (2004). Towards a school report card for primary education. *Educational Studies*, 81, 371-383.
- Jepperson, R. L. (2002). The development and application of sociological neoinstitutionalism. *New Directions in Contemporary Sociological Theory*, 229-266.
- Jolly, R. (2003). Human development and neoliberalism: Paradigms compared. In S. FU.K.uda Parr & A. Shiva Kumar (Eds.) *Readings in Human Development* (pp. 82-92). New York, NY: Oxford University Press.
- Jones, B. & Egle, R. (2006). Looking through different lenses: Teachers' and administrators' views of accountability. *Phi Delta Kappan*, 87(10), 767-771.
<http://dx.doi.org/10.1177/003172170608701012>
- Joshi, D. & Smith, W. (2012). Education and inequality: Implications of the World Bank's Education Strategy 2020. In Wiseman, A. & Collins, C. (Eds.), *Education strategy in the developing world: Revising the World Bank's education policy* (pp. 173-202). United Kingdom: Emerald Publishing.
[http://dx.doi.org/10.1108/S1479-3679\(2012\)0000016013](http://dx.doi.org/10.1108/S1479-3679(2012)0000016013)
- Kamens, D. H., & McNeely, C. L. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review*, 54(1), 5-25.
<http://dx.doi.org/10.1086/648471>
- Kamens, D. H., Meyer, J. W., & Benavot, A. (1996). Worldwide patterns in academic secondary education curricula. *Comparative Education Review*, 40(2), 116-138.
<http://dx.doi.org/10.1086/447368>
- Kim, K., Kim, G., Kim, S., Kim, J. et al. (2010). *OECD review of assessment and evaluation frameworks for improving school outcomes: Country background report for Korea*. Paris: OECD.
- Kitaev, I. (2004). EFA and private education: Some regional experiences and findings. *IEA Economic Affairs*, December, 27-30.
- Kleinman, M., West, A., & Sparkes, J. (1998). *Investing in employability: Theories of business and government in transition to work*. Commissioned by BT. London: LSE.
- Koretz, D. (2008). Test-based educational accountability. Research evidence and implications. *Zeitschrift für Pädagogik*, 54(6), 777-790.
- Kornhaber, M. L. (2004a). Assessment, standards, and equity. *Handbook of research on multicultural education*, 2, 91-109.
- Kornhaber, M. L. (2004b). Appropriate and inappropriate forms of testing, assessment, and accountability. *Educational Policy*, 18(1), 45-70. <http://dx.doi.org/10.1177/0895904803260024>
- Lashway, L. (2001). *The new standards and accountability: Will rewards and sanctions motivate America's schools to peak performance?* Washington, DC: ERIC Clearinghouse of Educational Management.
- Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies. *Review of Educational Research*, 78(3), 608-644. <http://dx.doi.org/10.3102/0034654308324427>
- Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., Kastberg, D., & Jocelyn, L. (2004). *International outcomes of learning in mathematics, literacy and problem solving: PISA 2003 Results from the U.S. Perspective*. National Center for Education Statistics. Washington, DC: US Government Printing Press.

- Levin, H. (1992). Market approaches to education: Vouchers and school choice. *Economics of Education Review*, 11(4), 279-285.
- Lewin, T., & Medina, J. (2003, July 31). To cut failure schools shed students. *The New York Times*, p. A1.
- Lincove, J. (2009). Are markets good for girls? The World Bank and neoliberal education reforms in developing countries. *The Whitehead Journal of Diplomacy and International Relations*, x(1), 19-35.
- Lipman, P. (2004). *High-stakes education: Inequality, globalization, and urban school reform*. New York: Routledge. <http://dx.doi.org/10.4324/9780203465509>
- Luna, C., & Turner, C.L. (2001). The impact of the MCAS: Teachers talk about high-stakes testing. *Engl. J.*, 91(1), 79-87. <http://dx.doi.org/10.2307/821659>
- Luschei, T. F. (2012). In search of good teachers: Patterns of teacher quality in two Mexican states. *Comparative Education Review*, 56(1), 69-97. <http://dx.doi.org/10.1086/661508>
- Macnab, D. (2004). Hearts, minds, and external supervision of schools: Direction and development. *Educational Review*, 56(1), 53-64. <http://dx.doi.org/10.1080/0013191032000152273>
- McDonnell, L., & Elmore, R. (1987). Getting the job done: Alternative policy instruments. *Educational Evaluation and Policy Analysis*, 9(2), 157-184. <http://dx.doi.org/10.3102/01623737009002133>
- McLaughlin, M. (1991). Test-based accountability as a reform strategy. *Phi Delta Kappan*, November, 248-251.
- McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- McNeil, L., & Valenzuela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In G. Orfield & M.L. Kornhaber (Eds.) *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 127-150). New York: Century Foundation Press.
- McNeil, M., & Klein, A. (2011). Obama outlines NCLB flexibility: Plan waves cornerstone provisions of law. *Education Week*, 31(5), 1 & 20-22.
- Meyer, J. (2005) *Weltkultur: wie die westlichen Prinzipien die Welt durchdringen*, ed. G. Krücken (Frankfurt, Germany: Suhrkamp).
- Meyer, J. (1977). The effects of education as an institution. *American Journal of Sociology*, 83(1), 55-77. <http://dx.doi.org/10.1086/226506>
- Meyer, J., Ramirez, F, Frank, D., & Schofer, E. (2006). Higher education as an institution. Working Paper. Center on Democracy, Development, and the Rule of Law: Stanford University.
- Monk, D., Sipple, J., & Killeen, K. (2001, September 10). Adoption and adaptation, New York state school districts' responses to state imposed high school graduation requirements. An eight year retrospective. Working Paper: University of Albany.
- Mundy, K. (2006). Education for All and the new development compact. *Review of Education*, 52, 23-48.
- Nichols, S., & Berliner, D. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.
- Nitko, A. J., & Brookhart, S. M. (2006). *Educational assessment of students* (5th ed.). New York: Prentice Hall.
- Paris, S., & McEvoy, A. (2000). Harmful and enduring effects of high-stakes testing. *Issues in Education*, 6(1), 145-159.

- Phelps, R. P. (2000). Trends in large-scale testing outside the United States. *Educational Measurement: Issues and Practice*, 19(1), 11–21. <http://dx.doi.org/10.1111/j.1745-3992.2000.tb00018.x>
- Ramirez, F. P. (2003). The global model and national legacies. In K. Anderson-Levitt (Ed.), *Local meanings, global schooling. Anthropology and world culture theory* (pp. 239–254). New York: Palgrave Macmillan.
- Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92, 1394–1415. <http://dx.doi.org/10.1016/j.jpubeco.2007.05.003>
- Rentner, D. S., Scott, C., Kober, N., Chudowsky, N., Chudowsky, V., Jofus, S., et al. (2006). *From the capital to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Center on Education Policy.
- Robertson, H. (1999). In Canada – bogus points. *Phi Delta Kappan*, 80(9), 715–716.
- Robertson, H. (1998). *No more teachers no more books: The commercialization of Canada's schools*. Toronto, ON: McClelland & Stewart.
- Rose, P. (2005). Is there a ‘fast track’ to achieving education for all? *International Journal of Educational Development*, 25(4), 381–394. <http://dx.doi.org/10.1016/j.ijedudev.2005.04.002>
- Rosen, L. (2009). Rhetoric and symbolic action in the policy process. In G. Sykes, B. Schneider & D. Plank (Eds.) *Handbook on education policy research* (pp. 267–285). New York: Routledge Publishing.
- Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. N. (2007). *Success under pressure? How low-performing schools respond to voucher and accountability pressure*. Mimeo. Princeton University and University of Florida.
- Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, 11, 45–61. <http://dx.doi.org/10.1007/s10833-008-9098-2>
- Schmidt, W., Houang, R., & Shakrani, S. (2009). *International lessons about national standards*. Washington, DC: Thomas B. Fordham Institute.
- Schofer, E., Hironaka, A., Frank, D. J., & Longhofer, W. (2012). Sociological institutionalism and world society. *The Wiley-Blackwell companion to political sociology*, 33, 57–69. <http://dx.doi.org/10.1002/9781444355093.ch6>
- Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, 51(1), 1–17.
- Shulman, L. (1983). Autonomy and obligation: The remote control of teaching. In L. Shulman & G. Sykes (Eds.) *Handbook of teaching and policy* (pp. 484–505). New York: Longman.
- Smeed, J., & Victory, M. (2010). Testing for accountability. *TLN Journal*, 17(3), 28–29.
- Smith, W. (2014, Forthcoming). National testing policies, school practices, and student outcomes: An analysis using data from PISA 2009. *Education Working Papers #xx*. Paris: OECD Publishing.
- Smith, W. (2012). Conceptualizing soft power through education. *International and Comparative Education Magazine* 3. Published online at <http://www.icemag.org/2/post/2012/02/conceptualizing-soft-power-through-education.html>
- Smith, W., & Rowland, J. (2014). Parent trigger laws and the promise of parental voice. *Journal of School Choice*, 8(1), 94–112.
- Springer, M. G. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review*. doi: 10.1016/j.econedrev.2007.06.004
- Stotsky, S. (2000). *What's at stake in the K12 standards war?* New York: Peter Lang

Publishing.

- Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, 10, 211-227. <http://dx.doi.org/10.1007/s10833-009-9105-2>
- Swadener, B. B., & Lubeck, S. (1995). The social construction of children and families “at risk”: An introduction. In *Children and families “At Promise”: Deconstructing the discourse of risk* (pp. 1-16). New York, NY: SUNY Press.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- Volante, L. (2007). Evaluating test-based accountability perspectives: An international perspective. Paper presented at the Association for Educational Assessment – Europe, Stockholm, Sweden.
- Von Zastrow, C. & Janc, H. (2004). *Academic atrophy: The condition of the liberal arts in America’s public schools*. Washington DC: Council for Basic Education.
- Waterreus, I. (2003). *Lessons in Teacher Pay: Studies on Incentives and the Labour Market for Teachers*. Doctoral Thesis: University of Amsterdam.
- West, A. & Pennell, H. (2000). Publishing school examination results in England: Incentives and consequences. *Educational Studies*, 26(4), 423-436. <http://dx.doi.org/10.1080/03055690020003629>
- West, A., Pennell, H., & Noden, P. (1998). School admissions: Increasing equity, accountability and transparency. *British Journal of Educational Studies*, 46(2), 188-200. <http://dx.doi.org/10.1111/1467-8527.00078>
- Wheelock, A., Haney, W., & Bebell, D. (2000). What can student drawings tell us about high-stakes testing in Massachusetts?. *The Teachers College Record*.
- Wiggins, A. & Tymms, P. (2000). Dysfunctional effects of public performance indicator systems: A comparison between English and Scottish primary schools. Paper presented at the European Conference on Educational Research, Edinburgh.
- William, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, 45(2), 107-122. <http://dx.doi.org/10.1080/00461521003703060>
- Wiseman, A. (2010). The uses of evidence for educational policymaking: Global contexts and international trends. *Review of Research in Education*, 34, 1-24. <http://dx.doi.org/10.3102/0091732X09350472>
- Wiseman, A. W., Astiz, M. F., & Baker, D. P. (2013). Comparative education research framed by neo-institutional theory: a review of diverse approaches and conflicting assumptions. *Compare: A Journal of Comparative and International Education*, (ahead-of-print), 1-22.
- Wiseman, A. W., Pilton, J., & Lowe, J. C. (2010). International educational governance models and national policy convergence. *International Perspectives on Education and Society*, 12, 3-18. [http://dx.doi.org/10.1108/S1479-3679\(2010\)0000012004](http://dx.doi.org/10.1108/S1479-3679(2010)0000012004)
- Woessman, L. (2007). International evidence on school competition, autonomy, and accountability: A review. *Peabody Journal of Education*, 82(2-3), 473-497. <http://dx.doi.org/10.1080/01619560701313176>
- Woessman, L. (2004). The effect heterogeneity of central exams: Evidence from TIMSS, TIMSS-Repeat and PISA. *CESifo Working Paper No. 1330*.
- Yarema, C. H. (2010). Mathematics teachers’ views of accountability testing revealed through lesson study. *Mathematics Teacher Education and Development*, 12(1), 3-18.
- Ysseldyke, J., Dennison, A., & Nelson, R. (2004). *Large-scale assessments and accountability*

systems: Positive consequences for students with disabilities. Minneapolis, MN: National Center on Educational Outcomes.

About the Author

William C. Smith
The Pennsylvania State University
wcs152@psu.edu

William C. Smith is an assistant professor in education policy at the National Research University – Higher School of Economics in Moscow, Russia. He recently completed a dual-title Ph.D. in Education Theory and Policy and Comparative International Education at the Pennsylvania State University. His research is situated around measuring the influence of education on social and economic development with particular emphasis on the influence of school accountability and choice policies on school practices and student outcomes and the effect of education on public health.

Acknowledgements

Financial support for this research was provided by the OECD through a Thomas J. Alexander Fellowship. The author would like to thank Tracey Burns, David Baker, and Bryan Mann for their thoughtful comments on earlier drafts.

About the Guest Co-Editors

Sherman Dorn
Arizona State University
sherman.dorn@asu.edu

Sherman Dorn is the author of *Accountability Frankenstein* (2007) as well as a number of articles on the history of accountability and related policies in the United States. He is currently a professor of education and director of the Division of Educational Leadership and Innovation at Arizona State University's Mary Lou Fulton Teachers College.

Christian Ydesen
Aalborg University
cy@learning.aau.dk

Christian Ydesen is the author of *The Rise of High-Stakes Testing in Denmark, 1920-1970* (2011) as well as a number of articles on the history of educational testing and accountability in Denmark and wider Scandinavia. He is currently an assistant professor of evaluation and testing at the Department of Learning and Philosophy at Aalborg University, Denmark.

education policy analysis archives

Volume 22 Number 116 December 8th, 2014

ISSN 1068-2341



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-sa/3.0/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A2 (Brazil), SCImago Journal Rank; SCOPUS, SOCOLAR (China).

Please contribute commentaries at <http://epaa.info/wordpress/> and send errata notes to Gustavo E. Fischman fischman@asu.edu

Join **EPAA's Facebook community** at <https://www.facebook.com/EPAAAPE> and **Twitter feed** @epaa_aape.

education policy analysis archives
editorial board

Editor **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **Audrey Amrein-Beardsley** (Arizona State University) **Rick Mintrop**, (University of California, Berkeley) **Jeanne M. Powers** (Arizona State University)

Jessica Allen University of Colorado, Boulder

Gary Anderson New York University

Michael W. Apple University of Wisconsin, Madison

Angela Arzubiaga Arizona State University

David C. Berliner Arizona State University

Robert Bickel Marshall University

Henry Braun Boston College

Eric Camburn University of Wisconsin, Madison

Wendy C. Chi University of Colorado, Boulder

Casey Cobb University of Connecticut

Arnold Danzig Arizona State University

Antonia Darder University of Illinois, Urbana-Champaign

Linda Darling-Hammond Stanford University

Chad d'Entremont Strategies for Children

John Diamond Harvard University

Tara Donahue Learning Point Associates

Sherman Dorn University of South Florida

Christopher Joseph Frey Bowling Green State University

Melissa Lynn Freeman Adams State College

Amy Garrett Dikkers University of Minnesota

Gene V Glass Arizona State University

Ronald Glass University of California, Santa Cruz

Harvey Goldstein Bristol University

Jacob P. K. Gross Indiana University

Eric M. Haas WestEd

Kimberly Joy Howard* University of Southern California

Aimee Howley Ohio University

Craig Howley Ohio University

Steve Klees University of Maryland

Jackyung Lee SUNY Buffalo

Christopher Lubienski University of Illinois, Urbana-Champaign

Sarah Lubienski University of Illinois, Urbana-Champaign

Samuel R. Lucas University of California, Berkeley

Maria Martinez-Coslo University of Texas, Arlington

William Mathis University of Colorado, Boulder

Tristan McCowan Institute of Education, London

Heinrich Mintrop University of California, Berkeley

Michele S. Moses University of Colorado, Boulder

Julianne Moss University of Melbourne

Sharon Nichols University of Texas, San Antonio

Noga O'Connor University of Iowa

João Paraskveva University of Massachusetts, Dartmouth

Laurence Parker University of Illinois, Urbana-Champaign

Susan L. Robertson Bristol University

John Rogers University of California, Los Angeles

A. G. Rud Purdue University

Felicia C. Sanders The Pennsylvania State University

Janelle Scott University of California, Berkeley

Kimberly Scott Arizona State University

Dorothy Shipps Baruch College/CUNY

Maria Teresa Tatto Michigan State University

Larisa Warhol University of Connecticut

Cally Waite Social Science Research Council

John Weathers University of Colorado, Colorado Springs

Kevin Welner University of Colorado, Boulder

Ed Wiley University of Colorado, Boulder

Terrence G. Wiley Arizona State University

John Willinsky Stanford University

Kyo Yamashiro University of California, Los Angeles

archivos analíticos de políticas educativas
consejo editorial

Editor: **Gustavo E. Fischman** (Arizona State University)

Editores. Asociados **Jason Beech** (Universidad de San Andrés) **Alejandro Canales** (UNAM) y **Jesús Romero Morante** (Universidad de Cantabria)

Armando Alcántara Santuario Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

Claudio Almonacid Universidad Metropolitana de Ciencias de la Educación, Chile

Pilar Arnaiz Sánchez Universidad de Murcia, España

Xavier Besalú Costa Universitat de Girona, España

Jose Joaquín Brunner Universidad Diego Portales, Chile

Damián Canales Sánchez Instituto Nacional para la Evaluación de la Educación, México

María Caridad García Universidad Católica del Norte, Chile

Raimundo Cuesta Fernández IES Fray Luis de León, España

Marco Antonio Delgado Fuentes Universidad Iberoamericana, México

Inés Dussel DIE, Mexico

Rafael Feito Alonso Universidad Complutense de Madrid, España

Pedro Flores Crespo Universidad Iberoamericana, México

Verónica García Martínez Universidad Juárez Autónoma de Tabasco, México

Francisco F. García Pérez Universidad de Sevilla, España

Edna Luna Serrano Universidad Autónoma de Baja California, México

Alma Maldonado Departamento de Investigaciones Educativas, Centro de Investigación y de Estudios Avanzados, México

Alejandro Márquez Jiménez Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Felipe Martínez Fernández University of California Los Angeles, USA

Fanni Muñoz Pontificia Universidad Católica de Perú

Imanol Ordorika Instituto de Investigaciones Economicas – UNAM, México

Maria Cristina Parra Sandoval Universidad de Zulia, Venezuela

Miguel A. Pereyra Universidad de Granada, España

Monica Pini Universidad Nacional de San Martín, Argentina

Paula Razquin Universidad de San Andrés

Ignacio Rivas Flores Universidad de Málaga, España

Daniel Schugurensky Arizona State University

Orlando Pulido Chaves Universidad Pedagógica Nacional, Colombia

José Gregorio Rodríguez Universidad Nacional de Colombia

Miriam Rodríguez Vargas Universidad Autónoma de Tamaulipas, México

Mario Rueda Beltrán Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Luis San Fabián Maroto Universidad de Oviedo, España

Yengny Marisol Silva Laya Universidad Iberoamericana, México

Aida Terrón Bañuelos Universidad de Oviedo, España

Jurjo Torres Santomé Universidad de la Coruña, España

Antoni Verger Planells University of Amsterdam, Holanda

Mario Yapu Universidad Para la Investigación Estratégica, Bolivia

arquivos analíticos de políticas educativas
conselho editorial

Editor: **Gustavo E. Fischman** (Arizona State University)
Editores Associados: **Rosa Maria Bueno Fisher** e **Luis A. Gandin**
(Universidade Federal do Rio Grande do Sul)

Dalila Andrade de Oliveira Universidade Federal de Minas Gerais, Brasil
Paulo Carrano Universidade Federal Fluminense, Brasil

Alicia Maria Catalano de Bonamino Pontifícia Universidade Católica-Rio, Brasil
Fabiana de Amorim Marcello Universidade Luterana do Brasil, Canoas, Brasil
Alexandre Fernandez Vaz Universidade Federal de Santa Catarina, Brasil
Gaudêncio Frigotto Universidade do Estado do Rio de Janeiro, Brasil
Alfredo M Gomes Universidade Federal de Pernambuco, Brasil
Petronilha Beatriz Gonçalves e Silva Universidade Federal de São Carlos, Brasil
Nadja Herman Pontifícia Universidade Católica –Rio Grande do Sul, Brasil
José Machado Pais Instituto de Ciências Sociais da Universidade de Lisboa, Portugal
Wenceslao Machado de Oliveira Jr. Universidade Estadual de Campinas, Brasil

Jefferson Mainardes Universidade Estadual de Ponta Grossa, Brasil
Luciano Mendes de Faria Filho Universidade Federal de Minas Gerais, Brasil
Lia Raquel Moreira Oliveira Universidade do Minho, Portugal
Belmira Oliveira Bueno Universidade de São Paulo, Brasil
António Teodoro Universidade Lusófona, Portugal

Pia L. Wong California State University Sacramento, U.S.A
Sandra Regina Sales Universidade Federal Rural do Rio de Janeiro, Brasil
Elba Siqueira Sá Barreto [Fundação Carlos Chagas](#), Brasil
Manuela Terrasêca Universidade do Porto, Portugal

Robert Verhine Universidade Federal da Bahia, Brasil

Antônio A. S. Zuin Universidade Federal de São Carlos, Brasil