ECER KEYNOTE, CADIZ 2012

What is Wrong with the 'What-Went-Right' Approach in Educational Policy?

GITA STEINER-KHAMSI
Teachers College, Columbia University, New York, USA

ABSTRACT This article critically examines how 'what-went-right' analyses are used to subsequently justify the transfer of reform packages or 'best practices' from one country to another. Similar to evidence-based policy planning, the what-went-right approach needs to be criticized for being presumptuous. There are three fallacies of the what-went-right analysis that the article dismantles: rationality, precision and universality. The article focuses on the façade of universality and examines how the claim to universal solutions is methodologically sustained. First, the author shows how standardized or normative comparison has in recent years overshadowed the other two types of comparison: comparison across time (historical analyses) and comparison across contexts ('simple comparison'). Then, she elaborates on why the what-went-right approach requires policy analysts to downplay differences between educational systems in order to establish comparability between cases. The emphasis on comparability and similarity of cases is a prerequisite to importing 'best practices' from vastly different educational systems. But what if transfer occurs regardless of difference? There is a curious phenomenon that the article addresses in greater detail: the retrospective definition of a local problem. Given the worldwide circulation of 'best practices' and traveling reform packages, policy analysts sometimes are under pressure to align their analyses of local problems with already existing global solutions. The article ends with a reflection on policy borrowing and lending research and situates the what-went-right approach in the broader question of why and how policy analysts 'buy' or 'sell' reform packages that worked well in one context for transfer into another.

In this article I examine the reasons why the 'what-went-right' approach has become so popular, the methodological assumptions upon which it depends, as well as its impact upon agenda setting and policy formulation. This approach often leads to the erroneous assumption that what went right in one educational system will inevitably work well in another. Because it is associated with policy borrowing and learning, the approach has drawn the attention of comparative education researchers.

The what-went-right approach is related to two other practices in policy studies: evidence-based policy planning and lesson-drawing. The latter is nowadays more commonly defined as learning from 'best practices.' If we were to place the three terms on a continuum ranging from analytical to prescriptive, evidence-based policy planning would be considered the most analytical, lesson-drawing the most prescriptive, and the what-went-right approach would fall somewhere in between. Evidence-based policy planning is, at least in its intention, an instrument for evaluating and understanding the effectiveness of a reform. At the other end of the continuum are 'best practices' that organizations export or import, lend or borrow or, more generally, transfer from one context to another.

It would make sense if policy analysts first evaluated reforms, and then identified features that account for their effectiveness. If they see themselves as policy brokers they could then put a stamp of approval, or certify them as 'best practices' worth emulating in other contexts. From a rational perspective, differentiating between evaluation, interpretation, and certification would be highly desirable. However, it is rarely done. Arguably, the sequence evidence-based policy planning, followed by what-went-right analysis, and finally identification of 'best practices' only exists on paper. In practice, the three instruments of education policy are intertwined. I will therefore use the terms interchangeably and also inject a fourth term – 'standards' – that has experienced in recent years inflationary usage. As I will explain in this article, there is a thin line between what organizations or businesses claim as 'best practices' and what they later export or sell as international standards.

Three Façades of Evidence-based Policy Planning

Evidence-based policy planning is often viewed, justifiably, as presumptuous. There are three façades, in particular, which deserve closer scrutiny: the façade of rationality, the façade of precision, and the façade of universality. I will outline each of these below.

The *façade of rationality* has been thoroughly dismantled in policy studies, including by critics who shed doubt on whether 'governance by numbers' (Ozga, 2009) is less political or more rational than other modes of regulation. Skeptics scrutinize evaluations of charter schools, vouchers, and other controversial reforms to demonstrate convincingly that such studies are agenda-driven, in that researchers often 'spin' their interpretation to please the architects and financiers of the reforms (Gewirtz et al, 2007; Henig, 2008). The assertion that, despite the claims of its advocates, evidence-based policy planning is deeply political is premised upon the analysis that political manipulation has operated under the guise of scientific rationality.

Relatively less known in the field of educational policy studies are a second group of researchers who demystify statistics, illuminate the *'façade of precision'* (Samoff, 1999), and problematize the uncontested authority attached to numbers. In my studies on educational reform in Mongolia, I noticed a 'statistical eradication' of pressing social issues through these methods. For example, I noticed vast discrepancies in reports on dropout statistics, even among departments within the same ministry (Steiner-Khamsi & Stolpe, 2006, pp. 181ff.). Whereas the Nonformal Education Department reported 40,000 dropouts, the department in charge of educational statistics at the Ministry of Education reported only 11,953 – that is, a figure that is nearly four times smaller.

Given that governments in developing countries are forced to wear the veil of statistical precision in order to present baseline data and subscribe to measurable benchmarks when they seek external financial assistance, the second critique is by no means inconsequential. 'Managing for results' is considered paramount for education planners in developing countries, yet their measures are often deeply flawed.[1] The Global Partnership for Education Secretariat, a believer in helping fund and implement Education Management Information Systems (EMIS) around the world, acknowledges challenges with such systems.[2] For example, there is a 21% gap between reported primary completion rates in Djibouti and Ethiopia, and other significant discrepancies for educational statistics exist in Kenya and Rwanda (see EFA-FTI Secretariat, 2007, p. 9). The primary completion rate for Djibouti ranges from 31% (UNESCO Institute of Statistics) to 53% (Djibouti administrative data), depending on the data source. In the case of Ethiopia, the UNESCO Institute of Statistics reports a higher completion rate than the local national statistical office (55% versus 34%).

Some of these 'errors' are predictable. Ministries of finance and education periodically contradict one another on issues like student enrollment. The problem is particularly acute in educational systems that use per-capita financing, where the head count of students determines the amount of the allocated budget. In developing countries, ministries of education systematically over-report enrollment statistics, whereas finance ministries under-report them as a matter of principle. Similarly, it is not uncommon for national statistical offices to receive a mandate to change the 'calculation method' of poverty, or measures related to other controversial issues, shortly before an election. In Mongolia, for example, household income was replaced with

household expenditures as a poverty measure at a politically critical period during the first years of the new millennium.

Government offices in developing countries are by no means the only ones who construct indicators strategically. In *Far-fetched Facts,* Richard Rottenburg describes how international development experts help manipulate data in sophisticated ways, consolidating them in elaborate knowledge banks to justify the need for urgent and immediate action (Rottenburg, 2009). Similarly, it would be wrong to assume that policy makers in developing countries are alone in relying on unreliable data. The *Charter School Dust-up* (Carnoy et al, 2005) is just one of numerous works in policy studies that demonstrate the poor quality of the data enlisted to make important decisions in school reform.

In contrast to the first two fallacies of evidence-based policy planning (the façade of rationality and the façade of precision), the third – the *façade of universality* – is seriously under-examined. Up till now the critique has been primarily voiced among comparative education researchers who are sensitized to the detrimental effects of universal claims. The false claim of universality is epitomized by the what-went-right approach, and will be explained in greater detail in the next section. For now, it is simply worth pointing out that it provides legitimacy for the uncritical import of policies from elsewhere, or for the export of reform packages from one country to another. Traveling reforms are themselves a compelling phenomenon, and become even more so when actively promoted by funding agencies and international donors, pushing their portfolios of 'best practices' – packaged and framed as 'international standards' – from one country to another.

The Making of Universality: two methodological observations

How is the universal claim – an essential component of the what-went-right approach – (quasi-) scientifically sustained? What methods are used to make such claims appear legitimate? To answer these questions, I will focus on two major methodological tools that proponents of the approach often use in order to elevate a local solution to the realm of universal applicability: (1) standardization of comparison; and (2) retroactive establishment of case similarity. It is necessary to provide background information on comparative methodology in order to situate my methodological observations.

Standardizing Comparison

The method of comparison has undergone a fundamental transformation over the past few decades. A glance at the adjacent comparative social sciences – comparative sociology, comparative political science, comparative economics – reveals the rapid pace with which standardized or normative comparison has permeated social analyses. Similarly in education, global monitoring of national developments, as reflected in Organisation for Economic Cooperation and Development (OECD), UNESCO or World Bank studies [3], has taken on monumental significance as a tool for education planning. Today this tool is routinized, and now more than ever before national education systems are monitored in terms of how they perform on popular benchmarks (e.g. Millennium Development Goals), or broadly defined international standards.

Figure 1 provides an overview of the three types of comparison. They are listed in order of increased de-contextualization. The first type of comparison – *comparison over time* – qualifies as the most contextualized form of comparison. The emphasis is on a particular case. Case-study researchers typically analyze changes over time *in* a particular bounded system. Such single-country case studies operate with a research design that draws on a sample size of one (N=1), and include many variables – that is, they follow a design of *one N and many variables*. This particular type of comparison is more sensitive to culture, context or system in comparison with the other two types of comparison.

This approach, consisting of 'thick' description and dense historical analysis, is typical in historiography and ethnography. Naturally there exists a broad range of methods of inquiry – drawing on both qualitative and quantitative data – that use case-study methodology. For an extensive period in the history of comparative education, implicit comparison leading to '*Education in…*' studies (documentation of national educational systems) was the preferred type of inquiry.

The comparison was implicit because the focus was on particular variables or features of other educational systems that were deemed relevant to learn or borrow from. For example, the well-documented nineteenth-century British interest in the German educational system (see Phillips, 2004) applied such a meliorist approach, and was based on an implicit rather than explicit comparison of the two systems. Nowadays, simple impact evaluations follow the same design: baseline data are collected and then compared with information gathered months or years later. The focus of simple impact evaluations is on changes over time; even though they are by far more quantitative than the early comparative education studies.

Comparison over Time	Implicit comparison
	'Education in …': identifying changes over time in *one* context
	Single case-study methodology
	Simple impact evaluation
Comparison across Contexts	Simple comparison
	Cross-national comparison: comparing system variables
	Multiple case-study methodology
	Impact evaluation with quasi-experimental design
Comparison against Standards	Standardized comparison
	Comparison against agreed norms, standards, or benchmarks
	Comparison of a case or cases against a 'best case scenario'
	Evaluation against specific league leaders

Figure 1. Methods of comparison.

The second type of comparison deals with *comparison across space or contexts*. Any study that compares two and more systems potentially serves as an entry point to understanding culture and structure. This also applies to studies which focus on practices as banal as how schools are cleaned. In fact, 'Pupil Participation in School Cleaning' by Yukata Okihara is an excellent example of the analyses this type of comparison can yield.

In the 1970s, Yutaka Okihara (1978) pursued an ambitious research project to understand the impact of cultural values (Confucian beliefs in particular) on educational practices. He examined how schools are kept clean to illustrate important differences in how children are educated, by identifying three separate cultural belief systems spanning the globe. Type 1 countries he defined as those where students are primarily in charge of cleaning. In type 2 countries professionals are employed to clean the schools, and type 3 countries use both systems. Using his data, Okihara drew a map of the world, illustrated according to how schools are cleaned. He showed, for example, that in South and East Asian countries, as well in pockets of Africa and Latin America, students are put in charge of cleaning the classroom and school grounds. Meanwhile, janitors or professional cleaners are hired to perform the same task in Australia, New Zealand, Europe and North America. The third approach, according to Okihara, was used in the Soviet Union. Okihara's map (as well as the assumptions behind his analysis) may be criticized for many reasons, not the least of which is his category 'the rest', which he assigned to USSR and other former communist countries behind the Iron Curtain.[4] Nevertheless, this example shows how any type of educational practice, compared across space or context, can be deemed relevant to understanding 'underlying culture'.

I cite Okihara's study to demonstrate how the act of comparing can add weight or endow with meaning data we might otherwise ignore. The second type of comparison – comparison across contexts – operates with *many N but few variables*. The reduction of complexity leads to a de-contextualization that increases the chances of false interpretation. As the school-cleaning example shows, one could easily narrow down myriad explanations to just two – wealth and culture – to group countries into three different clusters or 'types'. According to Okihara (1978), some countries cannot afford to hire cleaners, and therefore require students to carry out the task. In countries such as Japan, on the other hand, maintaining a clean learning environment is linked with spiritual values. Students in East and Southeast Asia are thus expected to learn about the importance of taking care of things and of sharing collective responsibility.
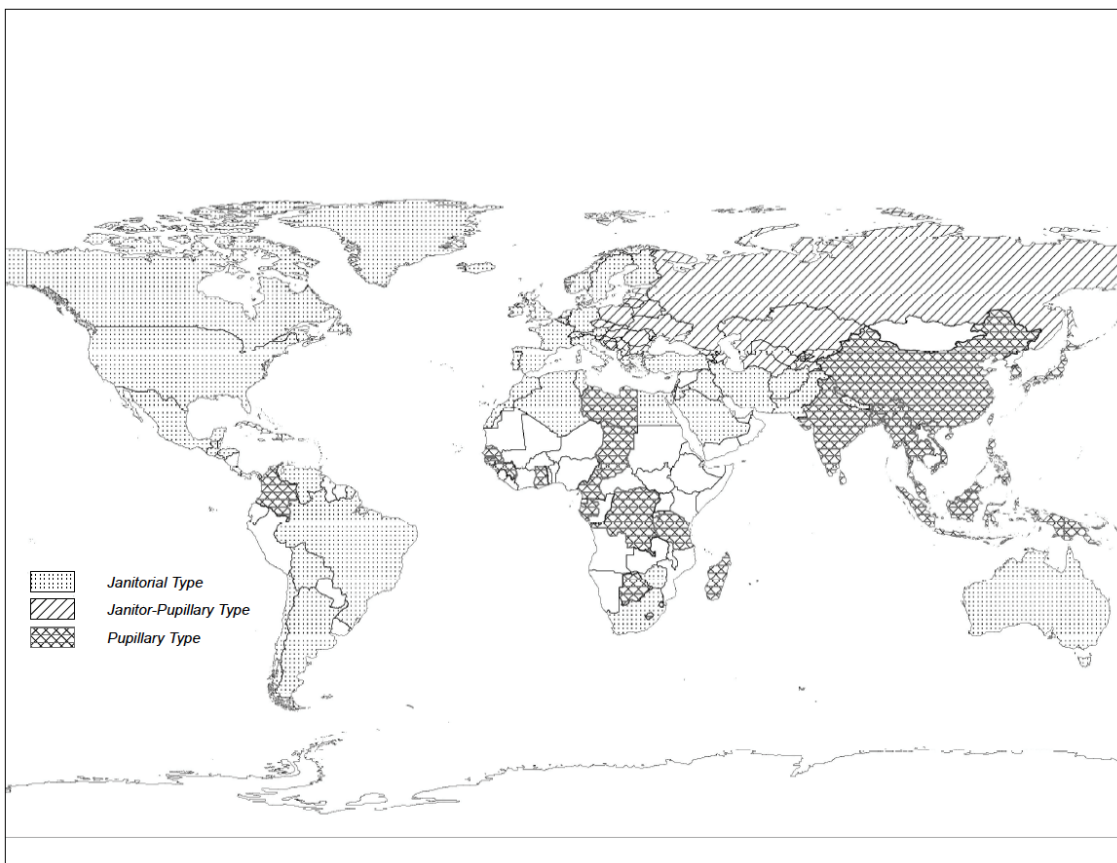
Figure 2. Three types of school cleaning in the world. Source: Okihara (1978).

Okihara's (1978) study is methodologically interesting because it exemplifies the risk of decontextualized interpretation in research designs characterized by simple comparison. The same concern applies to quasi-experimental designs that rest on two cases (control group and treatment group), and draw on a time-series analysis (typically before and after the treatment/intervention/ reform). As with simple impact evaluations, there is a risk that the difference in context may be reduced to a difference in the degree of external intervention, whereby other cultural, contextual or systemic factors are downplayed or neglected. Okihara's research constitutes an important, if problematic, study that applied simple comparison or comparison across contexts. As a result of his research design (many *N*, few variables), he oversimplified the explanation of differences and, by implication, risked producing stereotypes and false interpretations.

It is important to point out that simple comparisons are not always simplistic. In fact, there are many promising methodological approaches that compare countries, cases or systems in a contextualized manner. An example is the method of 'video-cued multi-vocal ethnography' presented by Tobin et al (1989) and Tobin et al (2009). Tobin examined whether preschools in the United States, Japan and China have converged – rhetorically, practically or both – towards a shared international understanding of education for young children. Rather than analyzing video-recorded sequences of preschool practices, he asked preschool educators to interpret their own practices, and compare them with what they saw recorded in preschools in the other two countries. In doing so, he exposed one of the main challenges of country comparison: the tendency to unnecessarily contrast, stereotype and overemphasize differences.

Finally, the third type of comparative research – *standardized comparison* – has experienced unprecedented popularity over the past few years. Of the three types of comparison, standardized comparison represents the method of inquiry that is most prescriptive and normative, and least sensitive to context, culture or system. It measures outcomes in relation to a norm as expressed, for example, in the format of an index (0-1), ratio (0-100%) or average. Examples of standardized

comparison include the gender parity index, the gross enrollment ratio, and empirically established normative averages such as above/below OECD average.

Standardized comparison privileges international over national developments, in that globalization is presented as a pervasive external force overwhelming local influences, which somehow renders the nation-state motionless by paralyzing policy actors. Unsurprisingly, indicator research is at the core of this type of comparison. Much emphasis is placed on how a system scores on a range of socially agreed-upon global indicators in a given supranational setting or 'educational space' (see Nóvoa & Lawn, 2002). Researchers identify, in the case of a what-went-wrong analysis, the reasons why the system failed to perform better or, in the case of a what-went-right analysis, why it has been 'on track' in achieving international benchmarks.

The third type of comparison is a normalization technique that is well analyzed in history and philosophy by Michel Foucault, and in critical curriculum studies by authors such as Thomas Popkewitz. However, standardization or normalization is somewhat less discussed in policy studies. Arguably, it should be seen as a new policy tool or technique that makes use of comparison to either generate or alleviate reform pressure on a system. It is important to note that standardized comparison has by no means replaced other types of comparison. More often than not, the three methods of comparison are used in combination with one another. In fact, many comparative studies include, begin with or end up with an in-depth analysis of what went right in a particular case, context or system.

Generating Comparability

Naturally, the act of lesson-drawing provokes the expectation that the transfer of 'best practices' only occurs between educational systems that are alike. This is due to strongly held beliefs that only educational systems with similar challenges are receptive to the same solutions – that is, open to importing 'best practices' from each other. In reality, however, this is not the case. What if the challenges differ and yet the same solutions – 'best practices' or reforms – are imported anyway? In other words, how is the legitimacy problem resolved? How is a transfer justified or, acknowledging agency in the policy process, how do policy makers explain to their constituents that they imported 'best practices' from a system that is completely different from their own?

One way of solving this dilemma is to deny that policy borrowing actually occurred. An early study of this phenomenon was carried out by Carol-Anne Spreen in her dissertation on the import of outcomes-based education (OBE) from Australia and North America to South Africa (see Spreen, 2004). As Spreen described, once opponents of OBE argued that the educational systems were incompatible, local policy actors who favored the reform claimed it was designed and initiated in South Africa, rather than imported from elsewhere. Even though such retroactive indigenization or reframing techniques are frequently put to work a posteriori to appease critics, the issue at hand is still the legitimacy of policy attraction across dissimilar contexts. How do policy analysts and makers justify their interest in educational systems – whether located in Finland, Singapore or Shanghai – that are so different from their own?

Another way of downplaying difference is to use uniform measurement that makes systems *appear* comparable. As explained above, standardized comparison does indeed generate the appearance of commensurability of educational systems, but this does not mean they are comparable. Even if the same indicators are used to measure certain concepts, the concepts still have – depending on the context, case or system – a different meaning. This applies not only to broad concepts such as 'quality education', but also to more narrow notions such as 'teacher shortage'. For example, educational statistics for the Kyrgyz and other Central Asian republics may show the same level of teacher shortage (under 5%) as any European country, but a closer examination at the school level reveals that as the term began, directors hastily filled vacant positions with substitute teachers, teachers of retirement age, university students, teachers with excessive teaching loads, as well as others who gladly took on additional hours to boost their salaries. In total we found ten such coping strategies at school level or, put differently, ten indicators of latent teacher shortage. This is in stark contrast to the one and only indicator for teacher shortage that usually makes it into national educational statistics: the number of unfilled positions.

25

Our methodological approach to supplementing the standardized indicator of teacher shortage with ten other equally important yet context-specific indicators has been widely discussed in the research community. We labeled our study '10+1 indicators of teacher shortage' to signal the importance of combining context-specific (ten indicators) and standardized (one indicator: unfilled teaching positions) measures of teacher shortage (see Steiner-Khamsi et al, 2011; see also UNICEF CEECIS, 2012). The more subtle indicators of teacher shortage matter a great deal in terms of contextual comparison. In theory it would *seem* that the standard measure of teacher shortage – vacant positions – is universally applicable. In practice, however, this problem manifests itself differently in Kyrgyzstan compared with other countries. For example, in Kyrgyzstan, teachers being brought out of retirement back into the classroom is a good indicator that there is a shortage. Meanwhile, in Swaziland vacancies are usually filled not by retirees, but by teachers who are un- or under-qualified (UNICEF Swaziland, 2010). In any case, it is obvious that measuring teacher shortage simply according to the number of unfilled positions or cancelled classes is a poor approach to understanding how serious the problem might be in a given context.

Another example is the reform package, or 'best practices' associated with 'teacher accountability', that multilateral organizations (e.g. OECD, World Bank) and multinational businesses (e.g. McKinsey) advocate as effective ways to improve the quality of education, despite evidence to the contrary. Universality claims are made despite vast differences in actual problems and needs at the local and national level. For example, in *Making Schools Work* (see Bruns et al, 2011), the World Bank provides a long list of 'best practices' that purportedly bring inferior education systems up to par with those that perform well. These include punitive measures for teachers who underperform, an approach that seems both cruel and redundant if implemented in former socialist countries such as Mongolia, where teachers are tightly monitored, and their salaries or salary supplements can be deducted if they do not show up, come late, grade student notebooks carelessly, or damage school equipment.[5]

The examples of teacher shortage and teacher accountability presented above illustrate the dangers of ignoring local context when measuring problems in a standardized manner or prescribing universal solutions. I used them as examples of how difference is rendered invisible. Returning to our research question, it is important to ask which research design helps substantiate the universal claim attached to the what-went-right approach in education policy. Once again, comparative methodology provides useful insights. Researchers in comparative studies pay close attention to valid case selection, because the correct sample determines the explanatory power attributed to a comparative study.

We should keep in mind how case selection is substantiated in comparative studies of sociology, political science, economy or education. Table I presents the differentiation between systems and outcomes (Berg-Schlosser, 2002, p. 2430; see also Przeworski & Teune, 1970) and organizes them in terms of similarity and dissimilarity or difference. The literature on research methods commonly discusses case selection in qualitative studies under the heading of purposeful sampling. The following paragraphs present the different designs for each cell in Figure 3.

		Outcomes	
		Similar	Different
Systems	Similar	**SS-SO** Similar Systems with Similar Outcomes	**SS-DO** *Similar Systems with* *Different Outcomes*
	Different	**DS-SO** Different Systems with Similar Outcomes	**DS-DO** Different Systems with Different Outcomes

Figure 3. Case selection in comparative policy studies.

Different systems with different outcomes. It comes as no surprise that considerations of sample or case selection are considered key for researchers in globalization studies. Depending on selection technique, researchers may find that systems are becoming increasingly alike. In other publications (e.g. Steiner-Khamsi & Stolpe, 2006), I contended that globalization researchers often resort to the

sampling design DS-SO (different systems with similar outcomes) to substantiate their claim that national education systems are converging – that is, that they are becoming increasingly similar. They purposefully select cases or systems that are very different from each other (DS), in order to demonstrate that they are similar in terms of outcomes (SO). One of the reasons I began studying globalization in Mongolia was the fact that the country was traditionally stereotyped as an uncivilized land of nomadic herders. For this reason, evidence of international convergence is especially compelling if educational structures, beliefs and practices 'even' in Mongolia have become increasingly similar to those elsewhere.[6] No doubt, the DS-SO design makes a strong case for globalization.

Similar systems with similar outcomes. Since the conclusion that similar systems produce similar outcomes seems obvious, the SS-SO design is often dismissed as irrelevant, and is rarely used. Moreover, researchers who draw on three or more similar cases to show that their outcomes are similar risk accusation of over-determination. From an orthodox methodological perspective, it is unnecessary to gather information on cases that are seen as similar because, by definition, a 'case' must represent a wider sample of similar entities. Naturally, similarity of cases only exists from a distance and from the perspective of an outsider. On closer examination and from an insider perspective, vast differences between cases may surface. By the same token, it would be difficult to defend an SS-SO design by claiming that different cases within a larger political system (e.g. former socialist countries), in a similar economic situation (e.g. low-income countries) or with similar religious beliefs (e.g., Muslim countries) constitute a homogeneous group. In its purest form, minor distinctions between systems qualify for this type of design, but if significant differences are uncovered, then it is necessary to replace it with one that acknowledges differences in systems - that is, DS-SO or DS-DO. For these reasons the SS-SO design is, as mentioned above, rarely used.

Different systems with different outcomes. A third cell in Figure 3 presents the DS-DO case selection, also known as the contrastive method. The contrastive method – comparing systems that are most different from one another and that manifest different outcomes – is a popular, or rather, populist, type of comparison. Examples of this approach are the Cold War studies of the 1960s, when researchers from both camps compared the systems of the USA and the USSR; the 1980s, when US researchers became fascinated with the educational system of a country – Japan – they had conquered four decades earlier; as well as the recent studies on the education of women and girls in Muslim countries, in the context of the 'war on terror'. Contrastive analyses, by design, prioritize differences over similarities.

Similar systems with different outcomes. The fourth cell covers the sampling design of the what-went-right approach. This design lends itself to policy learning because it focuses on systems that are supposedly similar (e.g. OECD countries, developing countries, etc.), but that produce different outcomes (SS-DO). An early, well-documented example of the what-went-right approach is the EFA-FTI Indicative Framework, presented as part of the Education for All-Fast Track Initiative (EFA-FTI), by Bruns et al (2003). These World Bank economists examined 155 developing countries with a similar gross domestic product (GDP) (SS), and then identified 69 that had surprisingly high rates of primary school completion (DO). Next, they analyzed what these countries had in common, to extract indicators which could explain why their students performed so well. For example, they found that successful systems pay their teachers 3.5 times the average GDP per capita, exhibit low repetition rates, allocate approximately 60% of the education budget to the primary level, require primary school students to attend for 850-1000 hours of instruction per year, etc. It is worth noting that the list of indicators grew longer over the past decade.

Figure 4 summarizes the different applications of case-study research explored above. The SS-DO design provides the foundation for policy learning, and drives the what-went-right approach used in policy transactions. As explained above and presented in Table II, similarity of systems must be assumed for policy learning to take place. However, the pressure on policy makers to adopt 'best practices' is sometimes so great that they resort to hiring researchers who help construct 'similarities', that may exist only on paper.

27

		Outcomes	
		Similar	Different
Systems	Similar	**SS-SO** Irrelevant	**SS-DO** *What-Went-Right Approach*
	Different	**DS-SO** Globalization Studies	**DS-DO** Contrastive Analyses

Figure 4. Uses/abuses of case-study design in comparative policy studies.

Incentives to Establishing Similarities between Cases

I am not the only one who finds it curious that similar contexts are sometimes retroactively established to justify the import of reform packages. For example, Joel Samoff (1999) forcefully demonstrated the overlap between analysis (problem recognition) and prescription (problem solution) in education sector reviews of African countries. Every education sector review he examined either began or concluded with a reference to 'the crisis' in African education. Talk of 'crisis' can certainly help create pressure for reform and mobilize funds for change; however, we have cause to be suspicious when the same reasons and solutions are put forward to cover a problem diagnosed across such a vast geographic region.

As for the European context, another researcher, Frank-Olaf Radtke (2009, n.14), describes a similar challenge: 'Granted, benchmarks or "best practices" do exist and provide solutions ... but the question is: for which local problems?'

What occurs in practice, both in developing and developed countries, is that the formulation of the (local) problem is aligned with the already existing (global) solution. Problems tend to be redefined in terms of accountability, standards, equity and other common buzzwords. The language changes but the approach – aligning local problem analysis with available global solution – remains constant. For example, just as ten to fifteen years ago project proposals had to address the issue of 'building civil society', today international consultants are hired to formulate education sector reviews that focus on low student achievement in early literacy and numeracy.

There are, arguably, specific reasons why national policy actors '*buy* global education policy' (Verger et al, 2012, p. 19). Politicians who promise to align their systems with airy notions such as 'world class education', 'twenty-first-century skills', 'international standards' or similar concepts elevate themselves into the role of guarantors of economies that are not 'falling behind' in the competitive global market place. This kind of policy talk inevitably attracts broad public support. In developing countries, moreover, the adoption of global education policy is even more complex because dependency on external funding must be taken into account. Policy borrowing in poor countries is to the education sector what structural adjustment, poverty alleviation and good governance are to the public sector at large – namely, a condition for receiving aid. As a requirement for receiving grants or loans, policy borrowing is often coercive and unidirectional. In effect, governments from donor countries lend or give money to low-income governments so that they can afford to import expensive reform packages or 'best practices' from the global North or the global West. The economics of policy borrowing helps explain the international convergence of reform rhetoric and trends in developing countries.

Though the literature on policy borrowing is abundant, there is as yet little research on incentives for organizations and businesses to *sell* global education policies by retroactively establishing a similarity of cases, contexts or systems. As I will outline below, global education policy is apparently good for business and convenient for management. It is obviously cheaper to sell the same reform package worldwide than to individually tailor reforms for a specific context. As always in business, efficiency is key.

Several ingredients are required to make reform packages profitable: training or 'capacity-building' for the 'natives'; stringent quality assurance mechanisms, in the form of licensing and certification, to assure local businesses cannot compete by offering cheaper, watered-down versions; and an international management team to hire and fire local staff. Furthermore, selling a

global education policy is only profitable if it is packaged as a tightly knit, coherent product with interconnected elements. This forces governments to purchase the entire package, rather than selectively pick and bargain over a few elements. Also, implementation has to be sufficiently complex that the 'locals' depend upon international consultancy, and agree to long-term contractual arrangements or service contracts. Finally, standardized student assessments are necessary to prove the effectiveness of the reform, and demonstrate to the client that the import – though more expensive than homespun reforms – was worth it. These elements make up the basic features of global education policy.

This is not to suggest that international reform packages, whether sold by Cambridge Education Services (operating in 160 countries) or by International Baccalaureate (IB; adopted in schools of 143 countries), ignore local needs. As Julia Resnik (2012) pointed out in her study on the global spread of IB schools, local variations remain intact. Naturally, multilateral education companies like IB, Cambridge Education Services or Pearson, who sell products like textbooks, tests, software and laptops to students, soon encounter economies of scale. They have far higher sales margins than companies such as McKinsey or Booz, that sell their consulting services to governments. Standardized comparison in the form of student achievement tests helped to effectively commercialize education. The fixation on outcomes paved the way for education businesses not only to enter the market for education, but to create and augment it.

We are at an interesting crossroads of public-private partnership in education. Needless to say, there are huge commercial interests invested. Education is, after all, one of only two sectors that cater to each individual, the other being healthcare. However, the potential to make money off education is limitless not just because everybody goes to school but, increasingly, because they do so their entire lives.

Across Europe, and from Indonesia to Mongolia, we are witnessing a proliferation of secondary schools emphasizing English language instruction, technology and other curricular aspects that appeal to the upper-middle-class cosmopolitan mindset. In many countries, public funds allocated to expensive pilot schools are justified with reference to competition in the global market. Nevertheless, it must also be acknowledged that these imported curriculum packages resolve one enormous problem that is endemic to curriculum reform in developing countries: the mismatch between standards, curriculum framework, teacher education, textbooks and student assessment. The non-alignment of various elements of one and the same reform domain (in this case curriculum) is a tragedy that reflects the harmful effects of uncoordinated aid. International donors selectively fund aspects of curriculum reform that reflect their own portfolio of 'best practices'. As a result, governments in low-income countries end up with a patchwork curriculum in which the separately financed pieces do not cohere.

In Mongolia, for example, scandals occasionally erupt over students who are assessed in high-stakes secondary school exit exams covering content they were never taught. Unsurprisingly, students are lost, parents are angry, teachers frustrated, and the general public mistrusts the government's ability to steer the educational reform process. It is typically in just such a climate that coherent, high-priced reform packages resonate. From the perspective of policy makers, it is better to import an expensive reform package in which all the elements are aligned than to reinvent everything from scratch, with the support of myriad international donors pulling in different directions, each advocating for their own 'best practices' and funding their own 'international standards'.

Naturally, the high cost of these products is a big concern in developing countries, because they benefit only a few while diverting resources needed to educate the masses. The previous government in Mongolia was heavily criticized for buying into the Cambridge International Education program (referred to as 'Cambridge Standards') and agreeing to fund thirty-two bilingual pilot schools. As with similar reforms carried out in other countries by the multilateral education business complex, the argument was made that the expensive, publicly financed bilingual pilot schools would – by setting higher standards – spill over to regular schools and thereby benefit public schools across the board. The spillover argument is also made elsewhere to justify the import of expensive curriculum packages that only benefit a few. In Indonesia, for example, so-called international standard schools, which provide courses in English, technology and a few other areas, were introduced with the same argument and are similarly under attack for being a drain on the national education budget.

There are also managerial reasons why institutions such as the World Bank, OECD and other international organizations prefer to lend their portfolio of 'best practices' rather than support projects initiated and developed locally. Global reform programs are easier to handle and implement. Although what we could call the 'data-fetishism' of these organizations – that is, amassing statistical information on system variables and collecting data from quasi-experimental impact evaluations - seems expensive, it soon proves cost-effective. Once a reform is successful in one country, it can then be transferred more cheaply to dozens of other countries that are established as being similar in terms of context or system.

There is an organizational logic to global policy lending that a very basic Weberian interpretive framework helps to elucidate: organizations tend to rationalize, routinize, formalize and bureaucratize their operations and programs. This applies to all organizations, including those that are philanthropic and non-governmental. For example, the Open Society Institute developed so-called Network Programs (as opposed to country-specific projects) that are transferred from the headquarters in New York, London or Budapest to global hubs of the Soros Foundation network. These programs are easier to manage, and 'only' require local adaptation, piloting and translation. However, as I have written elsewhere, not enough work has been done on the high transfer cost in terms of both the programs themselves, as well as the loss of human potential, innovation and sustainability.

The education market has, without a doubt, only begun to open up to commercial interests. Soon it will be flooded with entrepreneurs eager to sell their products. When competition is high, product differentiation is essential. Ironically, every business and organization will pitch its own portfolio of 'best practices' by insisting not only that they are different, but that they reflect 'international standards'. That these two selling points should be mutually exclusive will not matter.

Conclusions

This article focused on methodological aspects of the what-went-right approach. The extraction of factors that account for the high performance of an educational system is typically followed by recommendations for what should be borrowed, learned or imported in terms of 'best practices' from league leaders. There is a curious dilemma that this article addressed: what if, despite the fact that educational systems are vastly different, policy borrowing still occurs?

In my previous publications I answered this question from a theoretical perspective, and soon realized that I share a common interest with students of globalization. Many scholars contend that 'traveling reforms' are proof of an international convergence of educational systems, wherein programs are transferred indiscriminately from one context to another. For them, globalization entails – by definition – de-territorialization and a transnational flow of people, finance, goods, services, beliefs, technology, etc. in all sectors, including in education.

Using a sophisticated framework for interpreting change in organizations and society, neo-institutionalist authors are keen to find similar trends cross-nationally. For example, David Suarez (2007) discovered a proliferation of human rights topics in social science textbooks in a wide number of countries. Similarly, Patricia Bromley, John Meyer and Francisco Ramirez (Bromley et al, 2011) have identified a surge in student-centrism in social science textbooks over the past forty years. According to John W. Meyer, his associates and former doctoral students, these patterns have facilitated the development of 'world standards', which are regulated and monitored by a 'global society' (Meyer, 2012).

Neo-institutionalist theory is a coherent framework which manages to uncover isomorphism in organizations and educational systems. However, I have been critical of the neo-institutionalist framework for being oblivious to local policy contexts, and for not accounting for reception or resistance towards global forces in education. For scholars in institutional theory and organizational sociology, the discrepancy between global education policy and local adaptation is attributed to idiosyncratic manifestations of larger trends, and explained in terms of 'loose coupling' between policy and implementation. I criticized neo-institutionalist authors for reverting to loose coupling as a panacea for explaining the differences between a universal standard (e.g.

student-centered teaching, gender awareness, human rights, etc.) and its local manifestation (Steiner-Khamsi, 2012).

Such an approach is also of limited value for understanding cross-national policy attraction or re-contextualization. Ultimately, for neo-institutionalist theory, loose coupling is not the issue that begs explanation – it is, in fact, the explanation itself. The isomorphism argument in neo-institutionalist theory depends upon a particular methodological design. Theorists typically compare systems that are dissimilar, in order to conclude that countries around the world are converging towards a shared understanding of what constitutes good education. In other words, they use the DS-SO (different systems-similar outcomes) case-selection criterion. As Baker and LeTendre (2005) write:

> If current trends persist, what happens in a classroom in Seoul, Paris, Santiago, Cleveland, or
> Tunis will be remarkably similar, most likely even more so than now ... The globalization of
> curricula and its implementation in classrooms will exert a soft but steady pull on nations
> towards a world norm, to the point where little variation in curricula exists across nations. What
> differences remain will be mostly across schools within nations for intentional reasons and some
> idiosyncratic variation introduced by teachers. (Baker & LeTendre, 2005, p. 177)

In this particular article, I briefly mentioned neo-institutionalist theory in terms of its prototypical case-selection design (different systems with similar outcomes), but did not present alternative frameworks for understanding globalization.

In my own work, developed from comparative policy studies, I analyze local policy context to understand why and how global education policies are borrowed, and what impact policy borrowing has on power constellations among policy actors (see Steiner-Khamsi, 2012). I argue, in line with the theory of self-referential systems (Niklas Luhmann) and other comparative education researchers (Jürgen Schriewer), that externalization, or reference to other systems' 'best practices', as well as to international standards, has a salutary effect on protracted policy conflict. It helps to build policy coalitions and enables the mobilization of two or more opposing groups for the support of a third, supposedly more neutral, (de-territorialized) reform initiative.

Despite significant conceptual disagreements, I do take to some extent inspiration from neo-institutionalist explanations for the proliferation of international patterns, standards and organizations. This is not the case, however, with all explanations of global education policy. Perhaps the most naïve interpretation of international convergence is the commonsensical one - that is, that 'best practices' are in fact 'best' – or at least, as evidenced by the findings of education researchers, 'better'.

In contrast to my other work, which often deals with theories and interpretive frameworks, this article represents a reflection on comparative methodology, and elaborates on its uses and abuses for studies on globalization. Standardized comparison has become the most common type, overshadowing comparisons over time or space, or combinations thereof. I make the argument that policy learning not only makes use of standardized comparison but, by necessity, also draws on the case-study design 'similar systems-different outcomes' (SS-DC).

This brings us back to the research question I raised earlier: What if, despite the fact that systems are dissimilar, pressure to borrow, import or 'learn' from one another persists? The methodological maneuvers used to make systems, contexts or cases seem alike - which I discuss in this article - shed light on an absurd dilemma in global education policy: local problems must be redefined in ways that reflect global solutions. This is not easy. However, the difficulty of aligning education sector needs with the prescriptions of universal reform packages is more than made up for by the incentives – for governments, organizations and businesses alike – to do so.

From a distance, it may appear that the what-went-right-approach inevitably leads to a homogenization or isomorphism of systems. Closer examination, however, reveals that the what-the-went-right approach only resonates at particular moments of a policy process: in protracted situations when policy actors generate a quasi-external pressure (globalization) to resolve stalemate. It should come as no surprise that methodological reflections on this theme complement the interpretive framework of policy borrowing and lending research. In an era of evidence-based policy planning, it has become crucial to also examine methodological frameworks of global education policy from a critical perspective.

Notes

[1] The initial international agreement on aid effectiveness, known as the Paris Declaration on Aid Effectiveness (OECD DAC, 2005), includes the following five indicators:
– Ownership: developing countries set their own strategies for poverty reduction, improve their institutions and tackle corruption;
– Alignment: donor countries align behind these objectives and use local systems;
– Harmonization: donor countries coordinate, simplify procedures and share information to avoid duplication;
– Results: developing countries and donors shift focus to development results and results get measured;
– Mutual accountability: donors and partners are accountable for development results.

[2] Until 2011 the Global Partnership for Education was known as the Education for All-Fast Track Initiative (EFA-FTI).

[3] See, for example, studies associated with *PISA* (OECD), *Global Monitoring Report* (UNESCO), or *World Development Indicators* (World Bank).

[4] To be fair, he carried out his study during the Cold War, when information on the Soviet Union was limited and biased.

[5] These countries were examined in the UNICEF CEECIS Study on Teachers (UNICEF CEECIS 2011).

[6] I also wished to show that even though 'global speak' permeates policy discourse in Mongolia, this rhetoric should not be confused with what Mongolians actually believe or implement in practice. It is better to configure the semantics of globalization in terms of multiple audiences: the 'global speak' is addressed to the international community, and intended to secure grants and loans from donors, and signal participation in a larger global 'educational space;. How the Mongolian government communicates with or behaves towards its local constituency ('local speak') is a completely different matter.

References

Baker, D.P. & LeTendre, G.K. (2005) *National Differences, Global Similarities. World Culture and the Future of Schooling*. Stanford, CA: Stanford University Press.

Berg-Schlosser, D. (2002) Comparative Studies: method and design, in *International Encyclopedia of the Social and Behavioural Sciences*, vol. 4, 2427-2433. Amsterdam: Elsevier.

Bromley, P., Meyer, J.W. & Ramirez, F.O. (2011) Student-centeredness in Social Science Textbooks: cross-national analyses, 1970-2005, *Social Forces*, 90(2), 547-570.

Bruns, B., Filmer, D. & Patrinos, A.H. (2011) *Making Schools Work. New Evidence on Accountability Reforms*. Washington, DC: World Bank. http://dx.doi.org/10.1596/978-0-8213-8679-8

Bruns, B., Mingat, A. & Rakotomalala, R. (2003) *Achieving Universal Primary Education by 2015: a chance for every child*. Washington, DC: World Bank. http://dx.doi.org/10.1596/0-8213-5345-4

Carnoy, M., Jacobsen, R., Mishel, L. & Rothstein, R. (2005) *The Charter School Dust-up. Examining the Evidence on Enrollment and Achievement*. New York: Teachers College Press.

EFA-FTI Secretariat (2007) *Quality Education for All Children: meeting the challenge. Annual report 2007*. Washington, DC: Education for All-Fast Track Initiative (EFA-FTI).

Gewirtz, S., Dickson, M. & Power, S. (2007) Unravelling a 'Spun' Policy: a case study of the constitutive role of 'spin' in the education policy process, in Bob Lingard & Jenny Ozga (Eds) *The RoutledgeFalmer Reader in Education Policy and Politics*, pp. 178-197. London: RoutledgeFalmer.

Henig, J.R. (2008) *Spin Cycle: how research is used in policy debates – the case of charter schools*. New York: Russell Sage Foundation.

Meyer, John W. (2012) Effects of the Global Expansion of Education. Keynote Presentation at the Conference of the Comparative and International Education Society (CIES 2012), San Juan, Puerto Rico.

Nóvoa, A. & Lawn, M. (Eds) (2002) *Fabricating Europe: the formation of an education space*. Dordrecht: Kluwer. http://dx.doi.org/10.1007/0-306-47561-8

Okihara, Y. (1978) Pupil Participation in School Cleaning: a comparative survey, *Comparative Education*, 14(1), 33-40. http://dx.doi.org/10.1080/0305006780140104

Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) (2005) *Paris Declaration on Aid Effectiveness*. Paris: OECD DAC.

Ozga, J. (2009) Governing Education through Data in England: from regulation to self-evaluation, *Journal of Education Policy*, 24(2) (Special Issue), 149-163.

Phillips, D. (2004) Toward a Theory of Policy Attraction in Education, in G. Steiner-Khamsi (Ed.) *The Global Politics of Educational Borrowing and Lending*, pp. 54-67. New York: Teachers College Press.

Przeworski, A. & Teune, H. (1970) *The Logic of Comparative Social Inquiry*. New York: Wiley.

Radtke, F.-O. (2009) Evidenzbasierte Steuerung: Der Aufmarsch der Manager im Erziehungssystem, in R. Tippelt (Ed.) *Steuerung durch Indikatoren*. Leverkusen-Opladen: Budrich.

Resnik, J. (2012) The Denationalization of Education and the Expansion of the International Baccalaureate, *Comparative Education Review*, 56(2), 248-269. http://dx.doi.org/10.1086/661770

Rottenburg, R. (2009) *Far-fetched Facts. A Parable of Development Aid*. Cambridge, MA: MIT Press.

Samoff, J. (1999) Education Sector Analysis in Africa: limited national control and even less national ownership, *International Journal of Educational Development*, 19, (4/5), 249-272. http://dx.doi.org/10.1016/S0738-0593(99)00028-0

Spreen, C.-A. (2004) Appropriating Borrowed Policies: outcomes-based education in South Africa, in G. Steiner-Khamsi (Ed.) *The Global Politics of Educational Borrowing and Lending*, pp. 101-113. New York: Teachers College Press.

Steiner-Khamsi, G. (2012) Understanding Policy Borrowing and Lending. Building Comparative Policy Studies, in G. Steiner-Khamsi & F. Waldow (Eds) *Policy Borrowing and Lending. World Yearbook of Education 2012*, pp. 3-17. London: Routledge.

Steiner-Khamsi, G. & Stolpe, I. (2006) *Educational Import: local encounter with global forces in Mongolia*. New York: Palgrave Macmillan.

Steiner-Khamsi, G., Teleshaliyev, N., Sheripkanova-MacLeod, G. & Moldokmatova, A. (2011) Ten-Plus-One Ways of Coping with Teacher Shortage: a school-level analysis of teacher shortage in Kyrgyzstan, in Iveta Silova (Ed.) *Globalization on the Margins. Education and Postsocialist Transformations in Central Asia*, pp. 203-232. Charlotte, NC: Information Age Publishing.

Suarez, D. (2007) Human Rights and Curricular Policy in Latin America and the Caribbean, *Comparative Education Review*, 51 (3), 329-352. http://dx.doi.org/10.1086/518480

Tobin, J., Hsueh, Y. & Karasawa, M. (2009) *Preschool in Three Cultures Revisited: China, Japan, and the United States*. Chicago: University of Chicago Press.

Tobin, J., Wu, D. & Davidson, D. (1989) *Preschool in Three Cultures: Japan, China, and the United States*. New Haven, CT: Yale University Press.

UNICEF CEECIS (2011) *Teachers: a regional study on recruitment, development and salaries of teachers in the CEECIS region*. Geneva: UNICEF CEECIS.

UNICEF Swaziland (2010) *Teacher Recruitment, Development and Retention*. Mbabane: UNICEF Swaziland.

Verger, Antoni, Novelli, Mario & Altinyelken, Hulya Kosar (2012) Global Education Policy and International Development: an introductory framework, in A. Verger, M. Novelli & H.K. Altinyelken (Eds) *Global Education Policy and International Development: new agendas, issues and practices*, pp. 3-31. London: Continuum.

GITA STEINER-KHAMSI is Professor at Teachers College, Columbia University, New York, USA, past president of the Comparative and International Education Society (CIES), and co-editor of the series *World Yearbook of Education* (with Jenny Ozga and Terri Seddon). She completed her studies in Switzerland and worked close to ten years in the Ministry of Education of the Canton of Zurich before her emigration to the United States in 1995. She has published several books with Routledge, Palgrave Macmillan, and Teachers College Press on policy borrowing and lending, comparative policy studies, South–South and East–East cooperation, and globalization in education. *Correspondence*: gs174@columbia.edu