Cross-national indices with gender-differentiated data: What do they measure? How valid are they?

Asia-Pacific Human Development Report Background Papers Series 2010/01

**Cross-national indices with gender-differentiated data:
What do they measure? How valid are they?**

Angela Hawken • Gerardo L. Munck

Angela Hawken, born in South Africa, received her Ph.D. in policy analysis at the RAND Graduate School and is associate professor of economics and policy analysis at the School of Public Policy at Pepperdine University. Her research interests are primarily in drugs, crime, and corruption. She has published in the *Journal of Drug Policy Analysis,* the *Journal of Addiction Medicine*, and *American Prospect*, and she is co-author of *Drugs and Drug Policy: What Everyone Needs to Know* (Oxford, 2011). E-mail: Angela.Hawken@pepperdine.edu

Gerardo Munck, Argentinian by birth, received his Ph.D. in political science from the University of California, San Diego (UCSD) and is professor in the School of International Relations at the University of Southern California (USC). His research focuses on democracy, corruption, methodology, and Latin America. His books include *Measuring Democracy: A Bridge Between Scholarship and Politics* (Johns Hopkins University, 2009); *Regimes and Democracy in Latin America* (Oxford, 2007); *Passion, Craft, and Method in Comparative Politics* (with Richard Snyder; Johns Hopkins, 2007); and *Authoritarianism and Democratization. Soldiers and Workers in Argentina, 1976-83* (Penn State, 1998). E-mail: munck@usc.edu

Comments or clarifications can be addressed by email to the authors.

Abstract

The two cross-national indices with gender-differentiated data introduced by the UNDP in 1995, as well as several other such indices developed subsequently, are an important resource for researchers and policy makers interested in gender disparities. Yet questions linger about how these indices should be interpreted and how valid they are. Relying on a framework that synthesizes key guidelines concerning the methodology of measurement, this paper offers an assessment of indices currently used to study gender disparities on a global scale and sheds light on these unresolved questions. That is, we answer the question, what do these indices with gender-differentiated data actually measure? and, how valid are these indices? In addition, we spell out the stakes of research on measures disaggregated by gender, and an agenda for work geared to producing good cross-national gender-differentiated indices.

Key words: Gender disparity, gender-differentiated data, indices, methodology

Cross-national indices with gender-differentiated data: What do they measure? How valid are they?

Driven largely by the desire to provide follow-up to, and monitor compliance with, regional and world agreements that focus on gender equality and other gender-related issues, and by the Beijing Platform for Action of 1995 in particular, a multitude of researchers and organizations have carried out serious work to develop measures disaggregated by gender. Several projects have considered a broad range of indicators and raised the question—sometimes implicitly, sometimes explicitly—of what set of indicators are needed to cover the main aspects of social life (UNECLAC 1999; UNECE 2001; UNIFEM 2002; UNESCAP 2003; UNRISD 2005; UNDP 2006). Other works have focused on how to develop measures focused on specific issues, such as health, including reproductive health (Abdullah 2000; Yinger et al. 2002), and violence against women (Walby 2007; Kelly et al. 2008). Moreover, much effort has gone into collecting and disseminating gender-differentiated data, that is, data on some issues that includes separate measures for both sexes.[1] As a result of this collective endeavor, the amount of data and the quality of data that incorporates a gender perspective has increased considerably.

A distinctive facet of this broad challenge of measurement has been the development of indices, that is, compound measures that aggregate multiple indicators. The first indices that explicitly included gender-differentiated data were the United Nations Development Programme's (UNDP) Gender-related Development Index and Gender Empowerment Measure, both launched in 1995 (UNDP 1995). These pioneering UNDP indices gained some public visibility, in part because they were reported on an annual basis in the UNDP's Human Development Reports. But they also triggered a more academic debate about how to construct a valid and reliable index with gender-differentiated data. Some contributions to this debate have concentrated on assessing the two pioneering UNDP indices, highlighting some of their weaknesses and suggesting how the UNDP might improve their indices. But much effort by other organizations and by independent researchers has also gone into the development of new indices. Thus, today, various indices with gender-differentiated data are broadly available and used for academic and public policy purposes.[2]

---

[1] For an overview of resources on gender-differentiated data, including the production of data, the actors involved in the dissemination of data, and statistical databases, see UNECLAC (2002).

[2] The number of proposed gender indices is by now quite long. Tin addition to the two UNDP indices, the list of proposed indices includes the Gender Equality Index (White 1997), the Gender Inequality Index (Forsythe, Korzeniewick and Durrant 1998), the Relative Status of Women index (Dijkstra and Hanmer 2000), the

Indices are likely to play a growing role in discussions of gender. All indices serve to summarize a complex picture by aggregating the information of multiple indicators. And this role is all the more important in the study of gender because the solution to the lack of information on gender—the growing availability of gender indicators—has created a new problem—the sheer clutter produced by the growing availability of gender indicators. Indeed, a comprehensive list of data sources reveals a total of 303 indicators (UNECLAC 2002: Annex 6), making it difficult to study a particular country, let alone to compare several countries, by considering all the available information. However, by itself an index is not a solution to the problem brought about by the plethora of indicators. For an index to serve as a synthetic measure, it is indispensable that the index be a good measure of the issue one is interested in addressing.

This paper seeks to contribute to the collective effort to develop cross-national indices that aggregate gender-differentiated data, as well as to offer guidance to users of these indices, by offering an assessment of the most visible indices that provide up-to-date data on most countries of the world.[3] We focus on five indices: the UNDP's Gender-related Development Index (GDI), the UNDP's Gender Empowerment Measure (GEM), Social Watch's Gender Equity Index (GEI), the World Economic Forum's Global Gender Gap Index (GGGI), and the OECD's Social Institutions and Gender Index (SIGI).[4] As is standard in evaluations of measures, we seek to answer the question, *how valid are these indices?* And, as is conventional in evaluations of validity, we address the methodology used to generate the index, seeking to ascertain whether the measure captures variations in the phenomena being studied rather than being an artifact of the methodology used.

Standardized Index of Gender Equality (Dijkstra 2002), the European Union Gender Equality Index (Plantenga, Figueiredo, Remery and Smith 2003). the African Gender and Development Index (UNECA 2004), the Gender Equity Index (Social Watch 2004), the Social Institutions and Gender Index (Jütting, Morrisson, Dayton-Johnson, and Drechsler 2006), the Gender Equality in Education Index (Unterhalter 2006), the Global Gender Gap Index (Hausmann, Tyson, and Zahidi. 2007b), and the Multidimensional Gender Equality Index (Permanyer 2008).

[3] For other assessments of indices, see Bardhan and Klasen (1999), Dijkstra (2002, 2006), Charmes and Wieringa (2003), Chant (2006), Klasen (2006a, 2006b, 2006c), Shuller (2006), Permanyer (2007), and Jütting, et al. (2008).

[4] Though many other indices have been proposed, as pointed out, not all are equally relevant to the current assessment. Some of these indices are presented as purely theoretical exercises. Others constitute merely a recalculation of the data used in computing the UNDP's two pioneering indices, the Gender-related Development Index and Gender Empowerment Measure. And yet others present data that have become dated or cover only a small number of countries. All these exercises are interesting and useful, and they form part of the broad debate about gender measures. But a core challenge in the development of suitable gender indices concerns the practical challenge of developing data about sometimes hard to measure concepts across a large number of countries in a sustainable manner.

But we also address a more basic, logically prior, question. These five indices are all measures that use gender as a basis for differentiation. Yet the issue of how these indices should be interpreted is hardly a straightforward matter. Indeed, confusion about the interpretation of these indices has been common (Shuller 2006). And some authors go as far as to suggest that this problem may have hindered the visibility and impact of these indices (Klasen 2006a). Thus, rather than proceeding as is typical in validity assessments, assuming that a measure is, say, a measure of the concept x and that the key problem is to ascertain how validly x has been measured, we also address the conceptual question, *what do these indices actually measure?*

The problem of interpretability of these indices is not altogether surprising. After all, gender-differentiation can be linked to all sorts of matters of interest, such as the functioning of the brain, roles of men and women within the family or the military, attitudes toward violence, job advancement, participation in politics, and so forth. Gender is, to use a familiar phrase, a cross-cutting issue. And since a concern about gender is hence linked with many other concepts, interpreting a measure that is presented as a "gender index" is not necessarily easy. But the problem of interpretability of gender indices is compounded by methodological issues. The generation of indices involves complex methodological choices, and these choices, even if technical in nature, determine the meaning of the resulting measure. Thus, whether the goal is to evaluate the validity of an index, or to understand how to interpret an index, one must consider the methodology used to generate the index.

To study the methodology used in constructing the five indices under review, so as to answer our two questions—*what do these indices measure?* and, *how valid are they?*—we rely on a simple framework. Measures are produced using a measuring instrument. And, at its core, the construction of a measuring instrument involves three tasks: the identification of dimensions of the concept that is being measured or what might be called the overarching concept; the selection of indicators to measure each conceptual dimension; and the design of scales for each indicator. Finally, inasmuch as the measures under consideration are indices, that is, compound measures that aggregate the values of multiple indicators, a fourth task that must be considered concerns the rule used to aggregate the indicators (see Table 1).

Table 1. A Methodological Framewoirk for the Production and Evaluation of Data

Measuring Instrument Element	Key Question	Assessment Criteriia
1. Conceptual Dimensions	Are different dimensions of the overarching concept to be measured identified? If so, what are these conceptual dimensions and how are they formulated?	• Identification of a mutually exclusive and jointly exhaustive set of conceptual dimensions, in light of theory and the feasibility of collecting data • Avoidance of extraneous conceptual dimensions, that is, conceptual dimension of a different overarching concept
2. Indicators	What indicators are selected to measure each conceptual dimension? How do the indicators connect to each conceptual dimension?	• Selection of mutually exclusive and jointly exhaustive indicators, linked to each conceptual dimension • Avoidance of extraneous indicators, that is, indicators of a different conceptual dimension or a different concept
	How are values assigned to each indicator?	• Selection of indicators for which reliable data are available or can be generated through a replicable value assignment process
3. Indicator Scales	How are the scales designed so as to distinguish the possible values of each indicator?	• Indicator scales that are consistent with the concept being measured • Indicator scales that offer as much nuance (as many distinctions) as is justified
4. Rule to Aggregate Indicators	How are the values of the indicators combined to form an index?	• A theoretically-justified aggregation procedure • Robustness of the aggregate data and replicability of aggregation process

Each task involves distinct considerations and thus can be addressed one at a time. Indeed, there are many advantages to breaking down the overall task of generating an index into the core tasks connected to each element of a measuring instrument. But, as suggested by the very criteria for assessing how each task is tackled (see column 3 in Table 1), these tasks are interrelated. For example, the selection of indicators must be made both in light of the conceptual dimensions that have been identified and the choice of aggregation rule must refer back to the theory that drives the identification of conceptual dimensions. Moreover, though each task involves distinct considerations and can be assessed according to distinct criteria, the interpretation of an index is influenced by decisions made regarding all four elements of a measuring instrument, and the validity of an index is jointly determined by the extent to which the criteria regarding all tasks is adequately met.[5]

---

[5] Another way to state this point is that any discussion of measurement must focus broadly on measurement error, as highlighted in standard discussions of measurement methodology, but in a broader manner than is customary in this literature. That is, our comprehensive framework seeks to show that there are many sources of error, that go beyond the standard ones discussed in the literature.

This methodological framework, which draws on and integrates different bodies of literature on measurement methodology,[6] serves as the organizing principle for the paper. Section 1 focuses on how the five indices under review tackle the initial three tasks in the construction of a measuring instrument: i) the identification of conceptual dimensions, ii) the selection of indicators, and iii) the design of indicator scales. Section 2 addresses the particularly complicated aggregation of indicators, introducing—as a way to make these indices more comprehensible—a distinction between two steps in the aggregation of gender-differentiated data: first, the aggregation of the male and female values for each indicator and, second, the aggregation of the values of all indicators. In both these sections, we elaborate as necessary the barebones guidelines presented in Table 1, describe and compare the choices made by the index creators, and evaluate compliance with methodological criteria. We consider, in particular, whether each choice contributes to the generation of a measure that is clear and consistent in meaning, whether an appropriate justification is offered for each choice, and whether any empirical tests were used to validate the choices.[7]

Section 3 provides an overall assessment of the five indices under review. We offer our response to the question, *what do these indices measure?* suggesting how each index should be interpreted and what ambiguities cloud a clear interpretation. We also answer the question, *how valid are these indices?* summarizing the discussion of methodological strengths and weaknesses of each index presented in sections 1 and 2.

Finally, in the conclusion, we articulate the stakes of research that aims at the development of measures that differentiate by gender and spell out some implications of our assessment for this research agenda. As we emphasize, the stakes of generating better measures, especially of a global scope, are considerable. Thus, it is critical to recognize the progress that has been made in generating measures, but also to acknowledge that many existing measures are still affected by serious shortcomings, not to get wedded to existing measures out of inertia, and to see the production of measures that incorporate a gender perspective as an evolving research agenda that calls for considerable efforts and innovation.

---

[6] On measurement methodology, see Blalock (1982), Campbell (2001), and Kempf-Leonard (2005).

[7] In this assessment, we rely on texts about the methodology followed to produce each index as well as the data on each index and the data on the indicators used in each index On the methodology used to constructing each index, we consulted Anand and Sen (2003a, 2003b), Hausmann, Tyson and Zahidi (2007b), Jütting et al (2006, 2008), Porter and Schwab (2008), Social Watch (2005, 2008a: 68-79, 2008b), and UNDP (1995, 2007: 225-27, 358-60). The data on each index are generally available in reports and/or the websites of each organization.

1. From Overarching Concept to Data on Indicators

This section focuses on how the five indices under review tackle the initial three tasks in the construction of a measuring instrument: i) the identification of conceptual dimensions, ii) the selection of indicators, and iii) the design of indicator scales.

1.1. Conceptual Dimensions

The first task in developing a measuring instrument—the identification of the dimensions of the overarching concept to be measured—is the most purely theoretical and logical one. The identification of conceptual dimensions hinges on a decision that should be justified in light of a theory about the overarching concept, for it is in terms of such a theory that measures are interpreted. Moreover, this decision should follow a basic rule of logic— that good definitions consist of a mutually exclusive and jointly exhaustive set of elements that do not smuggle in any extraneous element. However, even at this stage a practical matter, the feasibility of collecting data that measures the full meaning of the overarching concept cannot be fully forgotten.

This first task is tackled formally in a fairly similar way by the constructors of the five indices under review. In each case, the index's overarching concept is disaggregated into three or four conceptual dimensions (see column 2 in Table 2). Moreover, in each case, this process of disaggregation relies on broad distinctions between spheres of social life that are well rooted in social science analysis and that identify mutually exclusive conceptual dimensions. In this way, a firm basis for measuring is established. That is, in all five indices, the disaggregation of the index's overarching concept into component parts breaks down and makes more concrete an otherwise extremely broad and abstract concept and hence helps to orient the subsequent task of indicator selection.

Table 2. Indices with Gender-differentiated Data I: Key Methodological Features

Index Producer and Name	Conceptual Dimensions	Indicators *	Indicator Value Assignment	Indicator Scales	**Rule to Aggregate Indicators**	
					Step 1: Aggregation of Male and Female Values for Each Indicator	Step 2: Aggregation of the Values of All Indicators
UNDP, Gender-related Development Index (GDI)	A long and healthy life	Life expectancy (at birth) (Male and female)	National registration systems, population censuses and surveys	Scales: predetermined by data collectors; rescaling: normalization of scales (as in HDI)	Standard of parity: women live 5 years more than men (adjusted for population share); deviations from standard favoring men and women weighted equally; combines relational and absolute measure using the harmonic mean	Weighted average, at the level of indicators (for the knowledge component), and then average, at the level of conceptual dimensions
	Knowledge	Literacy rate (for adults, age 15 and over) (Male and female)	National censuses or surveys	"	Standard of parity: 50/50 ratio (adjusted for population share); deviations from standard favoring men and women weighted equally; combines relational and absolute measure using the harmonic mean	
		Primary level enrolment (net) (Male and female)	Administrative records and population censuses	"	"	
		Secondary level enrolment (net) (Male and female)	"	"	"	
		Tertiary level enrolment (gross) (Male and female)	"	"	"	
	A decent standard of living	Earned income (PPP US$) (Male and female) **	National surveys, accounts and population censuses; International Comparison Program surveys	Scales: predetermined by data collectors; rescaling: normalization of scales (as in HDI) and log trasnformation	"	
UNDP, Gender Empowerment Measure (GEM)	Power over economic resources	Earned income (PPP US$) (Male and female) **	"	Scales: predetermined by data collectors; rescaling: normalization of scales (as in HDI)	Standard of parity: 50/50 ratio (adjusted for population share); deviations from standard favoring men and women weighted equally; combines relational and absolute measure using the harmonic mean	Average, at the level of indicators, and then at the level of conceptual dimensions
	Economic participation and decision-making	Professional and technical positions (Male and female)	National surveys	Scales: predetermined by data collectors	Standard of parity: 50/50 ratio (adjusted for population share); deviations from standard favoring men and women weighted equally	
		Positions as legislators, senior officials and managers (Male and female)	"	"	"	
	Political participation and decision-making	Parliamentary seats (Male and female)	Administrative records	"	"	
Social Watch, Gender Equity Index (GEI)	Education	Literacy rate (for youths, age 15 to 24) (Male and female)	National censuses or surveys	Scales: predetermined by data collectors	Standard of parity: 50/50 ratio (adjusted for population share); deviations from standard favoring women have no weight	Average, at the level of indicators, and then at the level of conceptual dimensions
		Primary level enrolment (net) (Male and female)	Administrative records and population censuses	"	"	
		Secondary level enrolment (net) (Male and female)	"	"	"	
		Tertiary level enrolment (gross) (Male and female)	"	"	"	
	Economic activity	Labor force participation (Male and female)	National surveys	"	"	
		Earned income (PPP US$) (Male and female) **	National surveys, accounts and population censuses; International Comparison Program surveys	"	"	
	Empowerment	Professional and technical positions (Male and female)	National surveys	"	"	
		Positions as legislators, senior officials and managers (Male and female)	"	"	"	
		Parliamentary seats (Male and female)	Administrative records	"	"	
		Ministerial level positions (Male and female)	"	"	"	

Index Producer and Name	Conceptual Dimensions	Indicators *	Indicator Value Assignment	Indicator Scales	Rule to Aggregate Indicators	
					Step 1: Aggregating Male and Female Values for Each Indicator	Step 2: Aggregating the Values of All Indicators
World Economic Forum, Global Gender Gap Index (GGGI)	Health and survival	Number of births (male and female) at birth	National registration systems	Scales: predetermined by data collectors	Standard of parity: women live 5 years more than men; deviations from standard favoring women have no weight	Variation-weighted indicators, and then average at the level of conceptual dimensions
		Healthy life expectancy (Male and female)	National registration systems, population censuses and surveys	"	Standard of parity:1.06 males for every female born; deviations from standard favoring women have no weight	
	Educational attainment	Literacy rate (for adults, age 15 and over) (Male and female)	National censuses or surveys	"	Standard of parity: 50/50 ratio; deviations from standard favoring women have no weight	
		Primary level enrolment (net) (Male and female)	Administrative records and population censuses	"	"	
		Secondary level enrolment (net) (Male and female)	"	"	"	
		Tertiary level enrolment (gross) (Male and female)	"	"	"	
	Economic participation and opportunity	Labor force participation (Male and female)	National surveys	"	"	
		Earned income (PPP US$) (Male and female) **	National surveys, accounts and population censuses; International Comparison Program surveys	"	"	
		Wage equality for similar work	Expert survey	"	"	
		Professional and technical positions (Male and female)	National surveys	"	"	
		Positions as legislators, senior officials and managers (Male and female)	"	"	"	
	Political empowerment	Parliamentary seats (Male and female)	Administrative records	"	"	
		Ministerial level positions (Male and female)	"	"	"	
		Heads of state (number of years over last 50 years) (Male and female)	Historical records	Count of years	"	
OECD, Social Institutions and Gender Index (SIGI)	Family law	Early marriage	National registration systems	Rescaling: Ratio one-sided scale (restriction on rights of girls)	Not applicable	Average, at the level of indicators, and then at the level of conceptual dimensions
		Polygamy	Expert coding	Ordinal scale with end points, with no gender distinction	"	
		Parental authority	"	Ordinal one-sided scale (iinequality disadvantaging women), with end points	"	
		Inheritance	"	"	"	
	Physical integrity	Female genital mutilation	National surveys	Rescaling: Ratio one-sided scale (restriction on rights of girls)	"	
		Violence against women	Expert coding	5-point ordinal one-sided scale (restriction on rights of women)	"	
		Missing women	National registration systems	Rescaling: Ratio one-sided scale, with normalization based on goalpost idea (restriction on rights of females)	"	
	Civil liberties	Freedom of movement	Expert coding	3-point ordinal one-sided scale (restriction on rights of women), with end points	"	
		Freedom of dress	"	Ordinal one-sided scale (restriction on rights of women), with end points	"	
	Ownership rights	Access to land	"	"	"	
		Access to bank loans	"	"	"	
		Access to property	"	"	"	

Notes: (*) Indicators are understood here as the concepts on which data for males and/or females are gathered. In some cases, the index creators present their list of indicators in a way that already combines the values of men and women, say, in the form of a ratio. Nonetheless, to arrive at these combined values, data on men and women has to be collected originally. (**) The Earned income (PPP US$) indicator is itself calculated on the basis of multiple indicators: on nonagricultural wages, the economically active population, the total population, and GDP per capita (PPP US$). This indicator is calculated by the UNDP based on multiple data sources.

Sources: On the UNDP indices, see UNDP (1995, 2007: 225-27, 358-60) and Anand and Sen (2003a, 2003b). On the Social Watch index, see Social Watch (2008a: 68-79, 2008b). On the World Economic Forum index, see Hausmann, Tyson and Zahidi (2007b) and Porter and Schwab (2008). On the OECD index, see Jütting et al (2008).

These similarities notwithstanding, as Table 2 shows, there are some notable differences in terms of the extent to which the index's overarching concept is theorized and the content of the conceptual dimensions of each index. The UNDP's GDI has a key virtue, that it focuses on a well theorized overarching concept, development and, moreover, that it is conceptualized and measured in exactly the same terms as the UNDP's Human Development Index (HDI), except for the way it differentiates between genders and uses gender inequalities to adjust the HDI scores downwards (UNDP 2007). The clarity of purpose evident in the design of the GDI makes it relatively easy to interpret. But this choice also makes the GDI vulnerable to the main conceptual critique that was voiced against the HDI: the failure to include several key conceptual dimensions of human development (Fukuda Parr 2003).[8] After all, the concept of human development is very broad, including factors essential to the "formation of human capabilities", that were explicitly included in the HDI and the GDI, but also a series of "political, economic and social freedoms" (UNDP 1990; Sen 1999), that were not included in the HDI and the GDI due to, among other reasons, the lack of data on the relevant indicators at the time the HDI was constructed (Fukuda Parr 2003). In a nutshell, the GDI's conceptual strengths but also its limitations are quite obvious.

The other indices are not as clearly rooted in a well-theorized overarching concept. But, interestingly, they all appear to take the GDI as a point of reference and partly explicitly, partly implicitly, they seek to overcome the narrowness of the GDI. The first index to tap into conceptual dimensions excluded from the GDI is the UNDP's own Gender Empowerment Measure (GEM), launched in 1995 along with the GDI. This index goes beyond the GDI's focus on socio-economic aspects and explicitly addresses issues of rights and of access to power. But the concept of empowerment used in this index is not developed and, as a result, the relationship between the GDI and the GEM is hard to pin down. Subsequently, the Social Watch's Gender Equity Index (GEI), first released in 2004, and the World Economic Forum's Global Gender Gap Index (GGGI), introduced in 2006, offered, as their key conceptual innovation, the inclusion of conceptual dimensions presented separately in the two UNDP indices.[9] However, the creators of both these indices do not explicitly state what overarching concept these indices measure, and hence do not offer a theoretical rationale for choosing the conceptual dimensions they include. Indeed, besides the GDI, only the OECD's Social Institutions and Gender Index (SIGI), made public in 2008, is based on an explicit

---

[8] The one problem of the HDI the GDI gets around, in a partial way, through its focus on gender, is the critique that the HDI does not address distributional issues.

framework built around the overarching concept of social institutions (Morrisson and Jütting 2005). This concept serves as the point of reference in the identification of the SIGI's conceptual dimensions. Moreover, it helps to highlight the purpose and novelty of this index. The SIGI does not seek to subsume previous efforts, and thus drops all the conceptual dimensions introduced by the UNDP indices, the GEI and the GGGI. Rather, its contribution lies in offering a supplementary measure that encompasses a range of issues pertaining to social institutions and norms other indices ignore.

In sum, the history of indices with gender-differentiated data can be summarized, in large part,[10] as a series of successive efforts to address ever-expanding spheres of social life in which gender differences could be a factor that had not been included in the GDI. These other indices do not always rest on as clear an overarching concept as the GDI, and thus it is hard to state with certainty whether the entire meaning of each index's overarching concept has been grasped. In this regard, it is probably safe to state that, in addition to identifying conceptual dimensions that are mutually exclusive, all five indices avoid the introduction of extraneous elements, which are not part of the concept's meaning. But it is also tempting to posit that all five indices could be understood as measures of the concept of development, as conceived by the UNDP and that, even though none of these indices offer a jointly exhaustive measure, a combination of the World Economic Forum's GGGI, which subsumes the two UNDP indices and the Social Watch's GEI, and the OECD's SIGI would be an important step in that direction.

1.2. Indicators

The second task in developing a measuring instrument concerns the selection of indicators to measure each of the conceptual dimensions identified during the prior step (see Table 1). This task is an extension of the prior more purely conceptual task, in that indicators are but concrete, empirically grounded concepts. Thus, a fundamental concern in the selection of indicators is whether they appropriately build a bridge between the abstract concepts that are part and parcel of theorizing and the observables that are essential to measurement. Moreover, much as with the identification of conceptual dimensions, indicators should be selected so as to capture the full meaning of the conceptual dimensions they purportedly measure, using mutually exclusive indicators, that is, that do not duplicate each

---

[9] The main conceptual difference between these two indices being that the Social Watch index dropped the dimension focused on health while the World Economic Forum index retained it.

[10] The other main conceptual differences concern how these indices conceive of gender equality and whether they focus on absolute and/or relative levels of attainment of goals, points addressed in Section 2.2.

other, and avoid extraneous indicators, that is, indicators of a different conceptual dimension or a different concept.[11] In other words, the selection of indicators should be made, and justified, in light of the link between conceptual dimensions and indicators, a link that directly affects the validity of eventual measures.

In addition, practical considerations, particularly as they relate to the process whereby values are assigned to each indicator, play a heavy role in the selection of indicators. The process of assigning values to indicators affects whether the data generated on each indicator are replicable, that is, whether it is feasible for independent researchers to reproduce the data generating process, and reliable, that is, whether an independent attempt to assign values to the indicators produces the same data. Hence, the method whereby values are assigned to indicators should be a factor in itself in considering whether certain indicators should be included or not. In short, the selection of indicators must be made looking backwards, to the concepts that are being measured and the theoretical content of these concepts, and forwards, to the measurability of these concepts and the quality of the relevant information that is available or can be collected.

With regard to this task, some interesting points emerge from an inspection of the indices under review (see column 3 in Table 2). An obvious conclusion to be drawn is that, differences concerning the various indices' overarching concept and conceptual dimensions aside, most of them (the exception is the OCED's SIGI) rely largely on the same indicators, such as Literacy Rate; Primary (net), Secondary (net), and Tertiary (gross) Level Enrolment; Earned income (PPP US$); Labor Force Participation; Professional and Technical Positions; and Parliamentary Seats (see Table 3).[12] These indicators have some advantages, in that data on these indicators are readily available for most of the world and the process whereby the data are generated is, within reasonable limits of feasibility, a reliable one (see column 4 in

---

[11] One way to think about the selection of indicators is that it constitutes the operational definition of the concept being measured and hence that, in making choices about indicators, it is imperative to avoid a mismatch between the theoretical and operational definition of a concept. The use of multiple indicators to measure the same concept or conceptual dimension, that is, that are not mutually exclusive, can be justified as a way to produce a more robust measure. But if the goal is to produce an index that measures many conceptual dimensions, and an assumption is made that each indicator measures a different part of a conceptual dimension, the use of multiple indicators leads to the problem of redundancy and double counting in the aggregation process. The concern about extraneous indicators is driven by the need to avoid introducing indicators that do not belong and that would therefore thwart the meaning of the measure and cloud its interpretation.

[12] Capitalization is used to distinguish references to the conceptual dimensions and indicators of the indices under review from other uses of the same words.

Table 2).[13] But the weaknesses associated with these choices out of convenience should be recognized.

Table 3. Indices with Gender-differentiated Data II: A Comparison of Indicators

	Indices			
Indicators	UNDP, Gender-related Development Index (GDI)	UNDP, Gender Empowerment Measure (GEM)	Social Watch, Gender Equity Index (GEI)	World Economic Forum, Global Gender Gap Index (GGGI)
Life expectancy (at birth) or healthy life expectancy	x			x
Number of births				x
Literacy rate (for adults or youths)	x		x	x
Primary level enrolment (net)	x		x	x
Secondary level enrolment (net)	x		x	x
Tertiary level enrolment (gross)	x		x	x
Earned income (PPP US$)	x	x	x	x
Professional and technical positions		x	x	x
Positions as legislators, senior officials and managers		x	x	x
Labor force participation			x	x
Wage equality for similar work				x
Parliamentary seats		x	x	x
Ministerial level positions			x	x
Heads of state				x

Note: An "x" means that this indicator in included in an index. Each of these indices calls for data on males and females for each of these indicators.
Source: See Table 2

One problem is that the selected indicators fail to capture the full meaning of the conceptual components they are supposed to measure. Thus, to take the most extreme example, it is hard to avoid the sense that Parliamentary Seats, as an indicator by itself, is too narrow a measure of the UNDP's GEM conceptual component Political Participation and Decision-making. But other problems emerge as index creators attempt to make do with readily available information. One problem is the use of what could be labeled extraneous indicators, those that fully or partly, are measures of conceptual components other than those

---

[13] This assessment does not seek to dismiss the serious concerns expressed by Srinivasan (1994) about the reliability and comparability of many of the data used in these indices. Moreover, it does not take into account some blatant errors introduced by those responsible for generating gender indices. For example, with regard to the indicator Positions as Legislators, Senior Officials and Managers (female/male ratio), the World Economic Forum data point to sudden changes, from 0.27 to 0.48 in Italy between the 2007 and 2008 report, from 0.09 to 0.30 for Bangladesh between the 2006 to 2007 reports, and from 0.08 to 0.58 for France between the 2006 to 2007 reports! (Hausmann, Tyson, and Zahidi 2006: 37, 64; 2007b: 42, 72, 87; 2008: 94). These changes are, on the face of things, incredible and indeed the data for the early year differ from those reported by the World Economic Forum's own data source, the International Labor Organization's LABORISTA datasite. In other words, concerns about data reliability cannot be put to rest.

they are claimed to measure. This is the case with the decision to use the indicator Legislators, Senior Officials and Managers as a measure of the UNDP GEM's conceptual component Economic Participation and Decision-making, and of the World Economic Forum GGGI's conceptual component Economic Participation and Opportunity. In both instances, strictly political information about legislators, is mistakenly presented as a measure of an economic dimension.[14] And this is the case with the World Economic Forum GGGI's indicator Heads of State (number of years over last 50 years), a slightly misnamed indicator, in that it focuses on presidents and prime ministers, that uses information about the past as an assessment of the present, counter to what is done for all the other GGGI indicators.

The OCED's SIGI differs considerably from the other indices with regard to the selection of indicators. This index breaks new ground by not letting the scarcity of readily available data constrain the selection of indicators and by calling for indicators that address key spheres of social life in which gender differences could be a factor even though they are not readily available. However, the generation of data on most of the SIGI's indicators through expert coding is a matter of concern. The use of expert coding is a well-established method of generating data. Moreover, the sources of information used to carry out the coding are made public and the information itself is public in nature (Jütting et al. 2008). But the standard practice of reporting the results of an intercoder reliability test is not followed by the SIGI's creators. Especially because some of the indicator scales are vaguely formulated, we are left with serious doubts about the reliability of the data on the SIGI's indicators.[15]

In sum, the five indices under review rely on indicators that exemplify some important strengths, but also have shortcomings. Probably most significantly, while the selection of indicators for which data are easily accessible is associated with a failure to tap into the full meaning of each conceptual dimension, a problem most evident in the GDI, we find that attempts to overcome this problem by selecting indicators for which data are not readily available has led to a different problem—the use of indicators of dubious reliability.

1.3. Indicator Scales

The third task in developing a measuring instrument concerns the design of scales for each indicator (see Table 1). As discussions about levels of measurement highlight, decisions

---

[14] The simultaneous use of the Legislators, Senior Officials and Managers and the Parliamentary Seats indicators by the GEM and GGGI, as well as the Social Watch's GEI, also leads to a problem of partial redundancy.

[15] A similar critique applies to the GGGI's indicator Wage Equality for Similar Work. Data on this indicator is collected through a question in an expert survey (Porter and Schwab 2008: question 9.13) about which little is

regarding scales should be guided by the criteria that indicator scales should offer as much nuance, that is, as many distinctions, as is justified. But it is also critical to emphasize that indicator scales should be consistent with the concept being measured, that is, that the choices about scales—much like choices about the indicators themselves—should be made in light of the concept that is being measured.

The issue of indicator scales is a moot question for most indices. This is so because most indices, as pointed out earlier, rely on data that are readily available, as opposed to data that have been collected specifically for the purpose of generating an index, and simply operate with the scales designed by the original data collectors. Thus, in these cases, the indicator scales could be considered to be predetermined and that the choice about scales was made by index creators when the indicators were selected. However, index creators did make distinct scaling choices regarding all indicators in two indices, the UNDP's GDI and the OCED's SIGI,[16] and interestingly approached this decision in quite different ways.

The UNDP's GDI relies on data collected by sources other than the UNDP and hence on indicators measured according to predetermined scales. But because the GDI is conceived as an index to be compared to the Human Development Index (HDI) and hence retains a focus on absolute levels of attainment of certain goals, the GDI's indicators are rescaled using the same procedures as in the HDI. Specifically, rescaling is carried out to normalize scales that have different metrics, using the notion of goalposts with minimum and maximum levels, and to adjust the Earned Income indicator, in line with the idea of diminishing returns, through the log transformation of this indicator. That is, the rescaling of the GDI's indicators is well justified.

The OCED's SIGI is similar to the GDI in that it rescales three of its 12 indicators, in a way that is quite reasonable. But the SIGI differs from the GDI and the other indices in that it does not rely on data from external sources with predetermined scales for most of the indicators and is based instead on scales designed from scratch (see column 5 in Table 2). These scales are all ordinal scales, which reflect a sensitivity about the distinctions that can justifiably be made in light of available information (Jütting et al. 2008). Moreover, the work involved in developing such scales should be duly recognized. But their weaknesses should

---

known. For example, the degree of agreement among the respondents is not reported and, since the data are not publicly available, independent researchers are not able to conduct test to assess the reliability of the data.

[16] Beyond these two indices, only two scaling decisions were made, The UNDP's GEM Earned Income indicator is rescaled by using the same notion of goalposts as the GDI. And he World Economic Forum GGGI includes a scaling choice with regard to the Heads of State indicator, which is measured with a simple count, of the number of years over the past 50 years during which a country has had a male and a female president or prime ministers.

also be noted. The main problem is that most of the scales are one-sided, addressing inequalities that disadvantage girls or women. Thus, potential inequalities that work against boys and men are not even envisioned and the possibility of empirically studying gender equality as it is broadly understood is closed.[17] In addition, some confusion is introduced because the one-sided scales for some indicators, such as Parental Authority and Inheritance, posit a relationship between the situation of men and women; while the one-sided scales for other indications, such as those for the Female Genital Mutilation and Violence Against Women indicators, focus on restrictions to the rights of girls and women while making no contrast to the situation of boys and men.[18] As a result, the SIGI's indicator scales do not prepare the way for systematically collecting gender-disaggregated data and, furthermore, combines, in a largely implicit manner, truncated measures of gender inequality and measures of distinctive women's rights.

1.4. An Interim Reflection

The amount of attention index creators put into the three elements of a measuring instrument discussed thus far, relative to what they dedicate to the fourth element, the choice of the rule to aggregate indicators, varies considerably. In the case of the OCED's SIGI, the generation of data on indicators is clearly the core concern. The SIGI's real value added is the production of new data on indicators of key spheres of social life; in contrast, the aggregation of these indicators is hardly discussed and appears as little more than an afterthought.[19] For the other four indices (GDI, GEM, GEI and GGGI), the reverse is true. The generation of data on indicators is not a goal for the creators of these indices—indeed, the work done by data collectors and standardizers is a precondition for these indices—and hence the value added is given by the methodology of aggregation. But even in these cases, it is important to recognize the impact of decisions, explicit or implicit, regarding conceptual dimensions, indicator selection, and indicator scales (we return to this point in Section 3). As the methodological framework presented in Table 1 seeks to show, these three elements are

---

[17] This problem also affects the GGGI's indicator Wage Equality for Similar Work, which is based on a question about women's earnings relative to men, and thus excludes a priori the possibility of collecting data that might show that women earn more than men (Porter and Schwab 2008).

[18] Finally, the Polygamy indicator scale is simply puzzling, because no reference to men or women is made and hence it is unclear if it refers only to polygyny, only to polyandry, or to both.

[19] The rule to aggregate indicators into sub-indices at the level of conceptual dimensions is simply mentioned in a footnote in the article on the SIGI's methodology (Jutting et al. 2008) and the rule to aggregate these sub-indices into the SIGI index is not mentioned in any publication. Rather strangely, an unpublished working paper states that the "social institutions indicator" is calculates by adding the four sub-indices (Jutting et al. 2006: 20) when the available data on SIGI is calculated as the average of the four sub-indices.

integral parts of a measuring instrument designed to produce an index. And they affect both the interpretation and validity of these indices.

2. From Data on Indicators to an Index

Data on indicators and data on an index differ in that the generation of the latter involves an additional task, not relevant to the production of data on indicators: the aggregation of the values of multiple indicators. Aggregation presupposes that the three tasks discussed earlier have been completed and, as mentioned, aggregate data is affected by the choices made in dealing with these tasks. But the task of aggregation involves a distinct choice, concerning the rule to aggregate indicators, and the aggregation rule is the most complex and technical element of a measuring instrument used to generate indices with gender-differentiated data.

Because most of the indices aggregate data on indicators originally disaggregated by gender, aggregation involves two distinct steps. On the one hand, as is standard for any index, aggregation involves the combination of the values of *all indicators and all conceptual dimensions.* An example, to take the UNDP's GEM, is the combination of the values of the Professional and Technical Positions indicator and the Positions as Legislators, Senior Officials and Managers indicator; and the combination of the values of the index's three conceptual dimensions: Power over Economic Resources, Economic Participation and Decision-making, and Political Participation and Decision-making (see column 2 and 3 of Table 2).[20] On the other hand, aggregation of data that has been differentiated by gender also involves the combination of the male and female values for *each indicator.* An example, again referred to the UNDP's GEM, is the combination of the male and female values of the Professional and Technical Positions indicator.

The choice of aggregation rule revolves around two general questions: what is the *relationship* among indicators? and, what *weight* should be assigned to the indicators? And selected aggregation rule must abide by a general criterion: it should be theoretically justified and, inasmuch as possible, empirically tested. But these general questions are formulated in different terms in the specific context of the two steps in the aggregation process. Likewise, compliance with this general criterion can only be assessed within the context of the different issues relevant to each step. Thus, in what follows, we address the delicate choice of the aggregation rule made by the creators of the five indices under review one step at a tine,

---

[20] When a conceptual dimension is measured with only one indicators, as is the case of the GEM's Political Participation and Decision-making dimension, no aggregation of indicators is necessary and the value of the relevant indicator—in this case the Parliamentary Seats indicator—is the value of the conceptual dimension.

starting with the distinctive challenge of choosing a rule to aggregate data disaggregated by gender.

2.1. The Rule to Aggregate Male and Female Values for Each Indicator

All indices with gender-differentiated data have a common goal: to overcome the weakness of depictions of reality that rely on simple averages, that is, an additive aggregation rule, that provides an absolute measure that necessarily overlooks possible differences in the situation of women relative to men.[21] However, in seeking to move beyond simple averages, index creators face several choices. First, to generate a measure that addresses the relationship between women and men, index creators must determine the standard to be used in comparing the data on women and men for each indicator. Second, to generate a relational measure, index creators must decide what weight to assign to deviations from this standard. Third, index creators must decide whether to combine the relational measure generated by the prior two choices with an absolute measure, and if so, what weight to assign to the relational and the absolute measure.

Interestingly, even though there is a large degree of overlap across the UNDP's GDI and GEM, the Social Watch's GEI and the World Economic Forum's GGGI with regard to the selection of indicators and data sources, index creators largely part ways regarding these choices. Regarding the choice of standard, these indices use a 50/50 ratio, understood as the point of parity, practically as a default option. The only exceptions concern the Life Expectancy indicator in the GDI, the Healthy Life Expectancy indicator in the GGGI, and the Number of Births indicator in the GGGI (see column 6 in Table 2).[22] But concerning the weight assigned to deviations from the chosen standard, two options can be distinguished. Both UNDP indices give equal weight to deviations from the standard of parity that favor men *and* women (this option is sometimes refereed to as a "double-sided scale"). That is, they measure inequality as it is generally understood. In contrast, the GEI and GGGI assign no weight at all to deviations from the standard of parity favoring women and only count inequalities that disadvantage women (this option is sometimes referred to as a "single-sided scale"). That is, they measure inequality only from the perspective of women.[23] Finally, with regard to the third choice, three options can be distinguished. The UNDP's GDI seeks to

---

[21] A simple average also gives equal weight to the situation of women and men.

[22] The two UNDP indices, and the GEI, also make an adjustment for the share of women and men in the population.

[23] In the 2005 version of the GGGI, deviations from the standard of parity favoring women were given the same weight as deviations favoring men, but interpreted not as an indicator of inequality but rather as a measure of women's empowerment or of "women's supremacy over men" (Hausmann, Tyson and Zahidi 2007a: 22).

retain the information about the absolute level of attainment of goals conveyed by simple averages and thus aggregates a relational and absolute measure using the harmonic mean of these two measures, an aggregation rule that assigns a "moderate" weight to their relational measure vis-à-vis the absolute measure (UNDP 1995: 73; 2007: 358-60). The UNDP's GEM aggregates a relational and absolute measure of the Earned Income indicator, again using the harmonic mean, but uses only a relational measure for the other indicators. And the GEI and GGGI rely, more simply on relational measures only, and thus do not take the absolute level of the attainment of goals into consideration.

The different options selected by index constructors are relatively well justified and the aggregation process is largely replicable,[24] but basic questions can be raised about some choices. Concerning the two UNDP indices, the standard of parity chosen for the GDI's Life Expectancy indicator is convincingly defended (Anand and Sen 2003a: 142, 2003b: 211, 213-14). But the reliance on a 50/50 ratio for the GEM's Parliamentary Seats indicators could be disputed. For example, if women fall short of gaining 50 per cent of the seats in parliament, this may be due to the lack of a level playing field but also to a whole host of other factors. Thus, this standard is more consistent with a measure focused on equality of results than a measure of "opportunities", which the UNDP suggests that the GEM is (UNDP 2007).[25] The decision to give equal weight to deviations from the standard of parity that favor men and women is justified in light of the stated purpose of the UNDP to address gender equality in general as opposed to, say, the disadvantages of women relative to men. Finally, the rationale for use of the harmonic mean to aggregate relational and absolute measures in the GDI and GEM is spelled out in a paper by Anand and Sen (2003b). As these authors explain, the core idea is that this aggregation rule is meant to capture the degree of inequality aversion in a society and impose a penalty for deviations from equality, such that the measure that is produced is, to use the UNDP's vocabulary, moderately lower than a measure solely based on a simple average. But no justification is offered for assigning a moderate weight to inequalities as opposed to any other weight and, even though this weighting choice is shown to have a large potential impact (Anand and Sen 2003b), no results of an empirical test that

---

[24] The OECD's SIGI is similar, with regard to this step in the aggregation process, to the GEI and the GGGI, but the choice of aggregation rule is embedded in the design of their scales, a decision that hinders the SIGI's replicability.

[25] Another problem with this indicator is that not all countries have a parliament. In those instances, the UNDP simply drops the country from the GEM, a rather drastic decision probably driven by the fact that Parliamentary Seats is the only indicator of the Political Participation and Decision-making conceptual component.

would gauge the impact of picking different weighting are reported. In short, it is hard to dispel the sense that this weighting choice is quite arbitrary.[26]

Concerning the Social Watch's GEI, though its use of a 50/50 ratio as a standard of parity is a reasonable choice for most indicators, the same question raised regarding the GEM's Parliamentary Seats indicators applies to it. In turn, though the GEI's assignment of no weight to deviations from the standard of parity favoring women is justified by Social Watch's stated interest to draw "conclusions about critical deficiencies in what women are able or allowed to do" (Social Watch 2005), it is important to note that such a decision makes it less general than the UNDP indices and hence limits the index's use. Additionally, though the decision to rely solely on a relational measure, a notable difference between the GEI and GDI, is founded in Social Watch's desire to empirically assess the relationship between gender equity (or, to be more precise, the deficiencies of women's status relative to men's status) and absolute levels of development (Social Watch 2005), such a decision is not problem-free. This decision gets around the problem the UNDP faces in assigning weights to relative and absolute measures, but it creates a new problem, most evident with regard to the Parliamentary Seats indicator. The problem is that a country may or may not have a parliament and, to get around an awkward conclusion,[27] Social Watch opts for a costly solution: countries lacking a parliament are categorized as missing data on the Parliamentary Seats indicator, which leads either to dropping that country from the index—a choice that diminishes the GEI's substantive utility—or (affecting the step in the aggregation process discussed next) dictates an ad hoc reweighting of the indicators of the Empowerment conceptual dimensions for that particular country—a choice that reduces the GEI's comparability.

Finally, concerning the World Economic Forum's GGGI, though the use of standards of parity that diverge from a 50/50 ratio for the Healthy Life Expectancy and Number of Births indicators reflects considerable research (Hausmann, Tyson and Zahidi 2007), the use of a 50/50 ratio for the Parliamentary Seats and Heads of State indicators is, for the reason given above, far from unimpeachable. The option to assign no weight to deviations from the standard of parity favoring women entails, as with the GEI, a loss of generality but is justified

---

[26] For a critique of this weighting scheme, that also questions the use of the same weight for all of the GDI's indicators, see Bardhan and Klasen (1999).

[27] Logically speaking, given Social Watch's commitment to developing a relative measure, they would in effect have to say that a country in which men and women hold the same number of seats in parliament is indistinguishable from, to give a real example, a country in which the military has taken power and decided to rule without a parliament.

cryptically, with nothing more than the enunciation that the index creators made this choice because they find it "more appropriate for [their] purposes" (Hausmann, Tyson and Zahidi 2007). In contrast, the decision to rely solely on a relational measure is better justified, in terms similar to those put forward by Social Watch (Hausmann, Tyson and Zahidi 2007). But, as with the GEI, the sole reliance on relational measures creates its own problem. Specifically, when a country has no parliament, the GGGI drops this indicator and reweights the remaining indicators of the Political Empowerment conceptual dimension in what amounts to an ad hoc and unacknowledged procedure that reduces the comparability of the data. All indices, in sum, entail some questionable choices regarding how they aggregate male and female values for each indicator.

2.2. The Rule to Aggregate the Values of All Indicators and Conceptual Dimensions

The aggregation of the male and female values for each indicator results in combined male/female values for each indicator, which themselves must be aggregated in a second step to generate the index value. The relationship among the indicators (and among conceptual dimensions), and the weight of each indicator (and each conceptual dimension) is, again, the key issue. However, in dealing with part of the aggregation process, index creators are not concerned with the relationship between, and the weight of, male and female values of each indicator. Instead, the focus turns to the aggregation of groups of indicators into sub-indices at the level of conceptual dimensions, and the aggregation of the sub-indices into a single index. Thus, attention focuses sequentially on: i) the relationship among, and the weight of, indicators used to measure each conceptual dimension (if only one indicator is used to measure a conceptual dimensions, this step is skipped), and ii) the relationship among, and the weight of, each conceptual dimension.

With regard to the procedure used at this stage in the aggregation process, the five gender indices under review select, with only a few exceptions, the same option: an additive aggregation rule that assigns the same weight to all the indicators that measure the same conceptual dimension, and the same rule to aggregate all the conceptual dimensions (see column 7 in Table 2). The only exceptions concern the decision to give equal weight to the set of indicators that measure a conceptual dimension. A minor exception in this regard is provided by the UNDP's GDI, which assigns different weights to the indicators of the conceptual dimension Knowledge.[28] A more significant departure from the norm is offered by the World Economic Forum's GGGI, which assigns variable weights to each indicator

---

[28] The Literacy Rate component is assigned double the weight of the combined three enrollment indicators.

according to its standard deviations relative to that of other indicators of the same conceptual dimension, giving more weight to indicators with a lower standard deviation and using the weights calculated for the 2006 index in subsequent versions of the index (Hausmann, Tyson and Zahidi 2007).[29]

The clarity regarding the aggregation procedure adopted by each index and the relatively high replicability of the aggregation process notwithstanding,[30] the extent of theoretical justification and empirical testing is quite low. The UNDP does offer a cogent justification for the aggregation rule used in the HDI, and hence by implication to the GDI, based on an assessment of their importance, the distinction between stock and flow indicators, and the possibility of substitutability across indicators (Haq 2003; Jahan 2003). But the use of an additive aggregation rule that assigns equal value to each set of indicators and to each conceptual dimension is usually treated more or less as a default option, needing little justification. And where explicit justifications have been offered, they have not been convincing. For example, the World Economic Forum GGGI's weighting scheme amounts to saying that greater weight should be assigned to indicators for which the situation around the world is fairly similar (Hausmann, Tyson and Zahidi 2007). But it is not clear, for example, why a ban forbidding women to hold seats in parliament in a certain country should be considered less of a problem simply if it is widely practiced in other countries. Furthermore, the GGGI's use of weights calculated for the 2006 index in subsequent versions of the index—the rationale being the desire to ensure the comparability over time of the index—has the effect of allowing the situation in an arbitrarily picked year to determine whether certain problems will be given greater or lesser weight in the future. The GGGI anchors weighting decisions in global trends, which might be positive or negative, offering thus a relative as opposed to a universal standard, and then obfuscates matters more by contradictorily disregarding future trends—a questionable procedure.[31]

Another problem is that index creators appear to be unaware of how choices regarding the way indicators are grouped under different conceptual dimensions and the number of indicators affects the aggregation process. The distinction among conceptual dimensions and

---

[29] As a result of this procedure, the maximum difference in weight between indicators is 2.4. For the weights assigned to each indicator, see Hausmann, Tyson and Zahidi (2007).

[30] The replicability of the aggregation process is ensured because both the aggregation procedure is articulated clearly and the underlying data, that is, the data on each indicator, is made public by the index producers or is otherwise publicly available. However, when comparing the data used by the index producers to the data from the sources cited by the index producers, we have found numerous discrepancies.

[31] To be consistent, given the justification offered by the creators of the GGGI, the empirically derived weights would have to recalculated each year, using the data on all the years that were being compared.

the groupings of indicators is important in that it provides a framework within which choices about aggregation can be addressed in a more nuanced way than if all the indicators were treated as direct measures of an overarching concept. But the impact on the process of aggregation of the selection of extraneous indicators (on these indicators, see Section 1.2) is not acknowledged.[32] And, more broadly, the impact on the weights of indicators on the way indicators are grouped into conceptual dimensions, and the number of indicators used to measure each conceptual dimension, is likewise never addressed. These two factors indirectly but strongly affect the weighting of indicators. Indeed, as a result of these factors, in addition to the formal aggregation rule, the actual weight of each indicator varies considerably both within each index and across the indices: within these indices some indicators have eight times the weight of other indicators, and across these indices some indicators have 11 times the weight of other indicators (see Table 4).[33] Yet this important aspect of the weighting process goes practically unnoticed.

---

[32] The use of the indicator Legislators, Senior Officials and Managers as a measure of an economic-oriented conceptual dimension in the UNDP's GEM and the World Economic Forum's GGGI undermines the theoretical justification for the grouping of indicators into distinct conceptual dimensions and the internal balance among indicators within each conceptual dimension. Moreover, the inclusion in these two indices of another indicator that measures the presence of men and women within the legislature—the Parliamentary Seats indicator—leads to a slight problem of redundancy that affects, surreptitiously, the weighting of indicators by double counting the same information. The Social Watch's GEI uses the indicator Legislators, Senior Officials and Managers but it is not an extraneous indicator in that case because it is used as an indicator of the broader but vaguer conceptual component Empowerment. Nonetheless, because this indicator is included alongside the Parliamentary Seats indicator, the GEI also suffers from a slight problem of redundancy.

[33] Another carryover factor from other elements of the measuring instrument that is relevant to this discussion concerns the indicator scales. Specifically, in the case of the OCED's SIGI, no consideration is given to the appropriateness of aggregating indicators measures with different scales, including various ordinal scales.

Table 4. Indices with Gender-differentiated Data III: A Comparison of Indicator Weights (%)

Indicators	Indices			
	UNDP, Gender-related Development Index (GDI)	UNDP, Gender Empowerment Measure (GEM)	Social Watch, Gender Equity Index (GEI)	World Economic Forum, Global Gender Gap Index (GGGI)
Life expectancy (at birth) or healthy life expectancy	33			8
Number of births				17
Literacy rate (for adults or youths)	22		8	5
Primary level enrolment (net)	4		8	11
Secondary level enrolment (net)	4		8	6
Tertiary level enrolment (gross)	4		8	3
Earned income (PPP US$)	33	33	17	6
Professional and technical positions		16	8	3
Positions as legislators, senior officials and managers		16	8	4
Labor force participation			17	5
Wage equality for similar work				8
Parliamentary seats		33	8	8
Ministerial level positions			8	6
Heads of state				11

Note: The percentages have been rounded off and do not necessarily add up to 100%. The weight of the OCED indicators range from 6% for the indicators of family law, to 8% for the indicators of physical integrity and of ownership rights, to 12.5 for the indicators of civil liberties.

Source: Authors' calculations based on Hausmann, Tyson and Zahidi (2007b: 5-6), Jütting et al (2006: 20, 2008: 79, Social Watch (2008b), and UNDP (2007: 358-60).

Finally, and most generally, the constructors of the five indices under review do little to present ideas about alternative ways in which they might have constructed the index and, more pointedly, no tests are reported that compare the impact of different aggregation rules and estimate the robustness of the chosen aggregation rule.[34] Such tests could focus on different parts of the measuring instrument that affect the process of aggregation. And such tests could play an important role in the validation of an index. For example, the weighting of the World Economic Forum GGGI's Heads of State indicator (see Table 4) raises a red flag, which reinforces doubts posed earlier (see Section 1.2) about the suitability of this indicator. Does it make sense to posit, as this index does, that considering the year 2009 the difference between having 0 or 29 percent female enrollment in universities is equivalent to the difference between never having had a female head of state and having had a female head of state during four of the past 50 years, maybe all in the 1980s? Thus, one legitimately wonders how the World Economic Forum's GGGI would look if this indicator were weighted differently or if it were altogether excluded from the index, and expects index

producers to address these concerns. Indeed, it is the responsibility of index creators to, at the very least, consider alternative but feasible ways of constructing an index and to present the results of tests that assess the impact of these differences. Yet the producers of the reviewed indices have all failed to shoulder the burden of performing such sensitivity analyses.

3. An Overall Assessment

The assessment of the five indices under review can be summarized in the form of answers to the two questions posed in the introduction: What are the five reviewed indices a measure of? And, how well do the five reviewed indices stand up to the criteria we highlight in our methodological framework?

3.1. What do These Indices Measure?

To answer the first of these questions, it is fundamental to note that the two key factors shaping the meaning of these indices are the overarching concept the index seeks to measure, and the choices made in aggregating the male and female values for each indicator. Indeed, these indices differ in their meaning—a matter of keen interest to the users of these indices—but their meaning is not evident in all cases from a reading of the index's name or even of the description offered by the index creators. The GDI is the best explained index and probably the easiest one to understand, no doubt because of its clear focus on the concept of development and the wide dissemination of the HDI. As the UNDP states in the report that introduced the GDI, this index "measures the same basic capabilities as the HDI [human development index] does, but takes note of inequalities in achievement between women and men…The GDI is simply the HDI, discounted, or adjusted downwards, for gender inequality" (UNDP 1995). But, to correctly interpret the GDI, it is also important to recall that, inasmuch as it is a gender measure, it is a general measure of gender inequality, which considers inequalities favoring men and women.[35]

The GEM has not been as well explained as the GDI by the UNDP. The concept of empowerment used in the index's label was never developed as fully as the concept of development. Moreover, the explanation that the GEM focuses "on women's opportunities rather than their capabilities" (UNDP 2007) is not very helpful. The GEM's indicators, as

---

[34] The only partial exception concerns the GDI, in that the UNDP states that tests were conducting to assess the impact of assigning different weights to the HDI's indicators (Haq 2003: 128). However, no results were reported and similar tests appear not to have been done with the GDI indicators proper.

[35] The GDI's general conception of gender inequality could be seen as making it hard to understand, because it is not directly apparent whether inequalities favor men or women. But this problem of interpretation is resolved through an analysis of disaggregate level data.

pointed out above, include some indicators that address results more than opportunities. In addition, the GEM's inclusion of one of the same indicators used in the GDI—Earned Income—makes this contrast between opportunities and capabilities somewhat puzzling, explanations notwithstanding (UNDP 1995). Finally, since the GEM relies on the same general concept of gender equality as the GDI, this emphasis on women is misleading. To make sense of the GEM it helps to think of it as a measure that taps into some dimensions of the concept of human development not included in the GDI. But the distinction between the GEM and the GDI is not as neat as that between the GDI and the HDI. There is some overlap due to the inclusion of one common indicator. And, to make matters more complicated, there is also some overlap between the GDI and GEM in that the GEM follows the GDI in combining relational and absolute measures for one indicator but not for the others. Thus, it is not possible to interpret the GEM as a precise complement to the GDI.

The GEI and GGGI are rather similar indices. Their creators do not explicitly state what concept is being measured. That is, though the indices advertize that they are measures of gender equality, they do not start out, as the GDI does, with a clearly formulated and well theorized concept and hence leave implicit what the index's overarching concept is. Thus, the best one can do is consider the conceptual dimensions of these indices and impute that they are measures of human development, as defined by the UNDP and interpret these indices as measures of development that conceptually subsume (especially in the case of the GGGI) both the GDI and the GEM. But there are two other key differences between the GEI and GGGI, on the one hand, and the GDI, on the other hand. One difference is that the GEI and GGGI, as their creators clearly state (Social Watch 2005; Hausmann, Tyson and Zahidi 2007), are measures of the situation of women relative to men that do not take into consideration the aggregate absolute situation of women and men. The other difference, somewhat ambiguously addressed by the index creators, is that the GEI and GGGI consider inequalities favoring men but exclude from their index any inequalities that favor women. Indeed, though Social Watch (2005) accurately argues that the GEI can be used to draw "conclusions about critical deficiencies in what women are able or allowed to do", the name of their index—the Gender Equity Index—suggests that it is a measure of gender equity plain and simple. Likewise, the name of the World Economic Forum's index, the Global Gender Gap Index, is only somewhat accurate—gender gaps can go in either direction—and the creators of this index confusingly suggest that they are measuring gender equality broadly speaking when they prominently state that the GGGI is "a tool for…tracking global gender-based inequalities" (Hausmann, Tyson and Zahidi 2007). Thus, in interpreting the GEI and

27

GGGI, it bears remembering that unlike the UNDP indices, these two indices measure the narrower concept of female disadvantages relative to men.

Finally, the interpretability of the OECD's SIGI is hampered by other problems. The point of departure for the measurement of the SIGI is a critique of the narrowness of the indicators used by the other four indices and the intent to provide a measure that goes beyond the other indices is unambiguously stated. Indeed, the overarching concept of social institutions is elaborated sufficiently to make it clear how SIGI is an index that supplements the other indices (Jutting et al. 2008). Nonetheless, the interpretation of the SIGI is not straightforward. As the GEI and GGGI, the SIGI only considers inequalities favoring men. But it includes relational measures, based on a comparison of the rights of women relative to men, as well as absolute measures, of the rights of women that have no obvious male counterpart. Moreover, the lack of symmetry in the type of indicators and indicator scales makes it quite cumbersome to get a clear idea of what precisely is being measured. Thus, the SIGI can be seen as a loose measure of female disadvantages relative to men concerning some basic rights and the fulfillment of rights distinctive to women.

3.2. How Well are Methodological Criteria Met?

To answer the second question, *how well do the indices stand up to the criteria we highlight in our methodological framework?* it is important to highlight some strengths, including the reliance of well established distinctions to differentiate conceptual dimensions, the inclusion of indicators that incrementally tap into more aspects of gender, the use of indicators for which reliable data on most countries of the world are available, and the replicability of the aggregation process (see Table 5). In short, the valuable contributions made by the creators of indices with gender-differentiated data should be duly recognized. Succinctly put, these indices constitute a resource for the study of development and a range of rights, which has allowed analysts to portray gender disparities in a more comprehensive and integrated manner than was possible prior to 1995.

However, a full answer to the second question must acknowledge several methodological weaknesses in all five indices. Problems were detected regarding the theorization of the index's overarching concept, the choice of indicators, the reliability of indicators, the aggregation procedure, and the lack of sensitivity analyses that are essential elements in the validation of measures. Indeed, some standard and basic methodological criteria are simply ignored. And these shortcomings are costly. In some cases, they make it difficult to interpret the indices clearly and see how these measures might be considered as partial but supplementary measures. And in all cases, they cast doubts about the quality of the

data that are produced with these measuring instruments. Undoubtedly, having an index, especially one of global empirical scope, is better than having none, even if the index suffers from methodological weaknesses. But users of these indices should be cautious when interpreting these indices—they measure different things and are anything but interchangeable—and when conveying the sense of confidence they have in the conclusions drawn from an analysis of these indices.[36]

Table 4. Indices with Gender-differentiated Data IV: Methodological Strengths and Weaknesses

Index	Assessment	Conceptual Dimensions	Indicators	Indicator Scales	Rule to Aggregate Indicators
UNDP, Gender-related Development Index (GDI)	Strength	• A well theorized concept • The introduction of well-established conceptual dimensions	• Use of (relatively) reliable indicators	• Well justified normalization of scales	• Replicability of the aggregation process • Overall cogent justification for aggregation procedure
	Weakness	• Dropping of conceptual dimensions due to a lack of available data	• Narrow set of indicators		• No information on the robustness of the aggregate data
UNDP, Gender Empowerment Measure (GEM)	Strength	• The introduction of well-established conceptual dimensions	• Use of (relatively) reliable indicators	• Well justified normalization of scales	• Replicability of the aggregation process
	Weakness	• The undertheorization of the overarching concept of empowerment	• Selection of a (partly) extraneous indicator • Selection of a (partly) redundant indicator		• Partial justification for aggregation procedure • No information on the robustness of the aggregate
Social Watch, Gender Equity Index (GEI)	Strength	• The introduction of well-established conceptual dimensions • Relatively broad broad set of indicators	• Use of (relatively) reliable indicators		• Replicability of the aggregation process
	Weakness	• The lack of specification of the index's overarching concept	• Selection of a (partly) redundant indicator		• Partial justification for aggregation procedure • Ad hoc reweightng compromises comparability of the index • No information on the robustness of the aggregate
World Economic Forum, Global Gender Gap Index (GGGI)	Strength	• The introduction of well-established conceptual dimensions • Relatively broad broad set of indicators	• Use of (relatively) reliable indicators		• Replicability of the aggregation process
	Weakness	• The lack of specification of the index's overarching concept	• Selection of (partly) extraneous indicators • Selection of a (partly) redundant indicator • No information on the reliability of data on one indicator	• Poorly designed indicator scale: gender-disaggregated data is not systematically collected	• Disputable justification of the aggregation procedure • Ad hoc reweightng compromises comparability of the index • No information on the robustness of the aggregate
OECD, Social Institutions and Gender Index (SIGI)	Strength	• A moderately theorized concept • The introduction of well-established conceptual dimensions	• Inclusion of not-readily-available indicators on key spheres of social life		
	Weakness		• No information on the reliability of data on most indicators	• Poorly designed indicator scales: gender-disaggregated data are not systematically collected • Some vaguely formulated scales	• No justification for aggregation procedure • Only partial replicability of the aggregation process • No information on the robustness of the aggregate

---

[36] Going beyond the methodological issues we addressed in this paper, we have serious reservations about using data on these indices as time series. Yearly data on all indicators are simply not available and in some cases the data on indicators used to calculate the index for a certain year can span up to 10 years. Nonetheless, the producers of these indices sometimes use their data to discuss address "evolution and trends," to "track progress over time" and to identify "countries progressing and regressing" (Hausmann, Tyson, and Zahidi 2007a; Social Watch 2005; 2008a). We would suggest that the data we currently have does not allow us to talk with confidence about trends over time.

4. Conclusion

Cross-national data on the multiple aspects of gender has a growing relevance to a range of policy areas. In addition to their standard uses in academic research on the causes and consequences of gender inequality and other gender-related matters, these data are used as tools to set agendas, especially by raising awareness of problems, and to keep governments and a variety of organizations accountable, particularly by monitoring compliance with certain standards and goals. Moreover, beyond the fairly visible global initiatives mentioned in the introduction, there is evidence that in some world regions and countries, data on gender equality are becoming part of the policy process.[37] These are undoubtedly positive indirect effects of the production of data, that confirm that there is something to the old saying, "What gets measured gets done". But the use of data on gender also draws attention to the quality of the data and the need to both conduct evaluations of commonly used measures that differentiate by gender and to engage in efforts to generate better measures. Indeed, if the data that are used in the policy arena are questioned, the very legitimacy of using data in the pursuit of collective goals suffers. Thus, issues of measurement, though involving intricate methodological questions, are anything but merely academic curiosities. Rather, there are real stakes to getting measurement right.

This paper has sought to contribute to this challenge through an assessment of five commonly used indices of nearly global scope, identifying three broad implications for further work in this area. First, we have emphasized the importance of taking measurement methodology seriously and, in particular, drawn attention to the need for an explicit methodological framework in work on measurement. Though the value of such a framework has been recognized in discussions of measures of development (Booysen 2002), current discussions of measures with gender-differentiated data do not rely on such a framework. Yet, as we show, the meaning and validity of a measure can only be ascertained if consideration is given to the range of methodological choices highlighted in our framework as well as to the relationship among the different elements of a measuring instrument. Hence, future research on data that incorporates a gender perspective should build more explicitly on a broad and integrated methodological framework.

Second, the assessment of indices provided in this paper has implications for users of these indices. The current availability of multiple indices is a great resource. But it also

---

[37] Within the European Union, a gender equality index in currently under development (Plantenga, Figueiredo, Remery and Smith 2003). On the use of data on gender equality in the UK and Ireland, see Breitenbach and Galligan (2006).

creates a problem for users. Indeed, the availability of multiple indices—which could support different conclusions—places a burden on users to offer a rationale for opting for one index over the others. That is, before putting a measure to substantive use, it is necessary to understand what each index actually measures, a more complicated matter than may seem at first sight, and to assess how valid and reliable these measures are. Thus, our assessment, briefly summarized in section 3, serves the immediate purpose of clarifying how the five reviewed indices should be interpreted and how much confidence we should have in them—two critical matters which have not always been well-addressed by the producers of indices that rely on gender-differentiated data.

Finally, as this paper illustrates, the development of measures is an incremental process in which a lot of the learning comes from doing and necessarily making mistakes. Thus, it is critical that researchers acknowledge the uncertainty they have regarding their data and the need to make changes to improve the quality of the available data. Admittedly, it is not always easy to make such an admission when the data are used in a policy environment, given that policy makers tend to equate quantitative data with science and science with certainty. And unwilling to make changes—making a change is akin to admitting error—is a common occurrence in data initiatives that have a large institutional backing. But this conservatism is unwarranted. To be sure, data producers who are interested in having an impact on policy must offer sufficient justification to warrant the judgment that the quality of their data that is good enough to be used at the present in the public sphere. But such a commitment to act responsibly is not weakened by a public recognition that the data, not despite being knowledge-produced through a scientific method but precisely because it is a product of a scientific method, is open to debate and subject to revision in the future.

Bibliography

Abdullah, Rashidah. 2000. *A Framework on Indicators for Action on Women's Health Needs and Rights After Beijing.* Kuala Lumpur, Malaysia: Asian- Pacific Resource and Research Centre for Women.

Anand, Sudhir, and Amartya Sen. 2003a. "Human Development Index: Methodology and Measurement". In S. Fukuda-Parr and S. AK Kumar, eds. *Readings in Human Development: Concepts, Measures, and Policies for a Development.* Oxford: Oxford University Press. [Originally published as Human Development Report Office *Occasional Paper* Nª 12. New York: UNDP, 1994.]

-------.2003b. "Gender Inequality in Human Development: Theories and Measurement". In S. Fukuda-Parr and S. AK Kumar, eds. *Readings in Human Development: Concepts, Measures, and Policies for a Development.* Oxford: Oxford University Press. [Originally published as Human Development Report Office *Occasional Paper* Nª 19. New York: UNDP, 1995.]

Bardhan, Kalpana, and Stephan Klasen. 1999. "UNDP's Gender-related Indices: A Critical Review." *World Development* 27 (6): 985–1010.

Blalock, Hubert M. 1982. *Conceptualization and Measurement in the Social Sciences.* Beverly Hills, Cal.: Sage Publications.

Booysen, Frederik. 2002. "An Overview and Evaluation of Composite Indices of Development." *Social Indicators Research* 59 (2): 115–51.

Breitenbach, Esther, and Yvonne Galligan. 2006. "Measuring Gender Equality: Reflecting on Experiences and Challenges in the UK and Ireland." *Policy and Politics* 34 (4): 597-614.

Campbell, Donald Thomas. 2001. *Social Measurement.* Thousand Oaks, Cal.: Sage Publications.

Chant, Sylvia. 2006. "Re-thinking the ''Feminization of Poverty'' in Relation to Aggregate Gender Indices." *Journal of Human Development* 7 (2): 201-20.

Charmes, Jacques, and Saskia Wieringa. 2003. "Measuring Women's Empowerment: An Assessment of the Gender-related Development Index and the Gender Empowerment Measure." *Journal of Human Development* 4 (3): 419–35.

Dijkstra, A. Geske. 2002. "Revisiting UNDP's GDI and GEM: Towards an Alternative." *Social Indicators Research* 57 (3): 301–38.

-------.2006. "Towards a Fresh Start in Measuring Gender Equality: A Contribution to the Debate." *Journal of Human Development* 7 (2): 275-83.

Dijkstra, A. Geske, and Lucia Hanmer. .2000. "Measuring Socio-economic Inequality: Towards an Alternative to the UNDP Gender-related Development Index." *Feminist Economics* 6 (2): 41–75.

Forsythe, Nancy, Roberto Patricio Korzeniewick, and Durrant, Valerie. 1998. "Gender Inequalities, Economic Growth and Structural Adjustment: A Longitudinal Evaluation." Paper presented to XXI Conference of the Latin American Studies Association, Washington, September 24-26, 1998.

Fukuda-Parr, Sakiko. 2003. "Rescuing the Human Development Concept from the HDI: Reflections on a New Agenda". In S. Fukuda-Parr and S. AK Kumar, eds. *Readings in Human Development: Concepts, Measures, and Policies for a Development.* Oxford: Oxford University Press.

Haq, Mahbub ul. 2003. "The Birth of the Human Development Index". In S. Fukuda-Parr and S. AK Kumar, eds. *Readings in Human Development: Concepts, Measures, and Policies for a Development.* Oxford: Oxford University Press.

Hausmann, Ricardo, Laura D. Tyson, and Saadia Zahidi. 2006. *The Global Gender Gap Report 2006.* Switzerland: World Economic Forum.

-------. 2007a. "The Global Gender Gap Index 2007". In R. Hausmann, L. D. Tyson, and S. Zahidi, *The Global Gender Gap Report 2007*. Switzerland: World Economic Forum.

-------. 2007b. *The Global Gender Gap Report 2007*. Switzerland: World Economic Forum.

-------. 2008. *The Global Gender Gap Report 2008*. Switzerland: World Economic Forum.

IPU (Inter-Parliamentary Union). 1997. "Men and Women in Politics: Democracy Still in the Making: A World Comparative Study." Geneva: Inter-Parliamentary Union, IPU Study Nº 28.

Jahan, Selim. 2003. "Evolution of the Human Development Index". In S. Fukuda-Parr and S. AK Kumar, eds. *Readings in Human Development: Concepts, Measures, and Policies for a Development.* Oxford: Oxford University Press.

Jütting, Johannes P., Christian Morrisson, Jeff Dayton-Johnson, and Denis Drechsler, 2006. *Measuring Gender (In)equality: Introducing the Gender, Institutions and Development Data Base (GID).* Paris: OECD Development Centre Working Papers 247, OECD Development Centre.

-------. 2008. "Measuring Gender (In)Equality: The OECD Gender, Institutions and Development Data Base." *Journal of Human Development* 9 (1): 65-86.

Kelly, Liz, with Lorna Kennedy, and Miranda Horvath. 2008. *The Next Step: Developing Transnational Indicators on Violence Against Women*, addendum to the Report of the Special Rapporteur on Violence Against Women, its Causes and Consequences, Yakin Ertürk, UN General Assembly, A/HRC/7/6/Add.5, February 25, 2008. New York: UN.

Kempf-Leonard, Kimberly, ed. 2005. *Encyclopedia of Social Measurement* 3 Vols. San Diego, Cal.: Academic Press.

Klasen, Stephan. 2006a. "UNDP's Gender-related Measures: Some Conceptual Problems and Possible Solutions." *Journal of Human Development* 7 (2): 243-74.

-------. 2006b. "Gender-related Indicators of Well-being". In M. McGillivray, and A. Shorrocks, eds. *Measuring Human Well-Being*. London: Palgrave-Macmillan.

Klasen, Stephan, ed. 2006c. "Special issue: Revisiting the Gender-related Development Index (GDI) and Gender Empowerment Measure (GEM)." *Journal of Human Development* 7 (2).

Morrisson, Christian, and Johannes P. Jütting. 2005, "Women's Discrimination in Developing Countries: A New Data Set for Better Policies." *World Development* 33 (7): 1065-81.

Permanyer, Iñaki. 2007. "On the Measurement of Multidimensional Gender Inequality: Continuing the Debate." Unpublished paper, Universitat Autònoma de Barcelona.

-------. 2008. "On the Measurement of Gender Equality and Gender-related Development Levels." *Journal of Human Development* 9 (1): 87–108.

Plantenga, Janneke, Hugo Figueiredo, Chantal Remery, and Mark Smith. 2003. "Towards an EU Gender Equality Index." A feasibility study commissioned by and presented to the European Commission, Utrecht School of Economics, Utrecht. December 2003.

Porter, Michael, and Klaus Schwab. 2008. "Executive Opinion Survey 2008." Available at: http://www.kadin-indonesia.or.id/enm/images/dokumen/KADIN-117-2573-29022008.pdf

Sen, Amartya. 1999. *Development as Freedom*. Oxford: Oxford University Press.

Shuller, Dana. 2006. "The Uses and Misuses of the Gender-related Development Index and Gender Empowerment Measure: A Review of the Literature." *Journal of Human Development* 7 (2): 161-81.

Social Watch. 2004. *Social Watch Annual Report 2004.* Montevideo: Instituto del Tercer Mundo.

-------. 2005. "No Country Treats its Women the Same as its Men. The Gender Equity Index: A New Perspective". In Social Watch, *Social Watch Annual Report 2005*. Montevideo: Instituto del Tercer Mundo.

-------. 2008a. *Social Watch Annual Report 2008.* Montevideo: Instituto del Tercer Mundo.

-------. 2008b. "Technical Notes: The Construction of the GEI." Available at: http://www.socialwatch.org/en/avancesyRetrocesos/IEG_2008/tablas/technicalNotes.html

Srinivasan, T.N. 1994. "Data Bases for Development Analysis: An Overview." *Journal of Development Economics* 44 (1): 3-27.

UNDP. 1990. "Defining and Measuring Human Development". In UNDP, *Human Development Report 1990.* New York: Oxford University Press.

-------. 1995. "Measuring Gender Inequality". In UNDP, *Human Development Report 1995. Gender and Human Development.* New York: UNDP and Oxford University Press.

-------. 2006. *Measuring Democratic Governance: A Framework for Selecting Pro-poor and Gender Sensitive Indicators.* Oslo: UNDP Oslo Governance Centre.

-------. 2007. "Calculating the Human Development Indices". In UNDP, *Human Development Report 2007/2008. Fighting Climate Change: Human Solidarity in a Divided World.* New York: UNDP.

UNECA (United Nations Economic Commission for Africa). 2004. *The African Gender and Development Index.* Addis Ababa, Ethiopia: Economic Commission for Africa.

UNECE (United Nations Economic Commission for Europe). 2001. "Final Report. ECE/UNDP Task Force Meeting on a Regional Gender Web-Site." Geneva: UNECE, Statistical Division.

UNECLAC (Economic Commission for Latin America and the Caribbean). 1999. *Indicadores de Género para el Seguimiento y la Evaluación del Programa de Acción Regional para las Mujeres de América Latina y el Caribe, 1995–2001, y la Plataforma de Acción de Beijing.* Santiago, Chile: LC/L.1186.

-------. 2002. *Tools and Indicators for Gender Impact Analysis, Monitoring and Evaluation.* New York, Inter-Agency Network on Women and Gender Equality, IANWGE/2002/6, February 2002.

UNESCAP (Economic and Social Commission for Asia and the Pacific). 2003. *Gender Indicators for Monitoring the Implementation of the Beijing Platform for Action in the Asia-Pacific Region.* New York: United Nations.

UNIFEM (United Nations Development Fund for Women) 2002. "Innovations in Measuring and Monitoring," pp. 55-61, in UNIFEM, *Progress of the World's Women 2002* Volume 2: *Gender Equality and the Millennium Development Goals.* New York: UNIFEM.

UNRISD (United Nations Research Institute for Social Development). 2005. *Gender Equality: Striving for Justice in an Unequal World.* Geneva: UNRISD.

Unterhalter, Elaine. 2006. *Measuring Gender Inequality in Education in South Asia.* Kathmandu: The United Nations Children's Fund [UNICEF], Regional Office for South Asia; and United Nations Girls' Education Initiative [UNGEI].

Walby, Sylvia. 2007. *Indicators to Measure Violence Against Women.* Working Paper Nº 1, prepared for the Expert Group Meeting on indicators to measure violence against women, organized by the United Nations Division for the Advancement of Women, UNECE and UNSD, in collaboration with ECLAC, ESCWA, ESCAP, and ECA, Geneva, Switzerland, October 8-10, 2007.

White, Howard. 1997. "Patterns of Gender Discrimination: An Examination of the UNDP's Gender Development Index." Mimeo. The Hague: Institute of Social Studies.

Yinger, Nancy, with Anne Peterson, Michal Avni, Jill Gay, Rebecca Firestone, Karen Hardee, Elaine Murphy, Britt Herstad, and Charlotte Johnson-Welch. 2002. *A Framework To Identify Gender Indicators For Reproductive Health and Nutrition Programming.* Washington, DC: USAID, Interagency Gender Working Group, Subcommittee on Research and Indicators.