

# Cross-National Indices with Gender-Differentiated Data: What Do They Measure? How Valid Are They?

Angela Hawken · Gerardo L. Munck

Accepted: 2 April 2012 / Published online: 24 April 2012  
© Springer Science+Business Media B.V. 2012

**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 remain regarding 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 article offers an assessment of indices currently used to study gender disparities on a global scale and sheds light on these unresolved questions. We answer two questions—what do these indices with gender-differentiated data actually measure? and, how valid are these indices?—and discuss the implications of our assessment for users and producers of gender indices.

**Keywords** Gender indices · Measurement validity · Cross-national data · Global data sets

Cross-national data on multiple aspects of gender disparities are used in academic research on the causes and consequences of gender inequality and other gender-related matters, but is also of growing relevance to policy discussions. These data are used for agenda setting by raising awareness of gender gaps, and help to keep governments and a variety of organizations accountable for compliance with certain standards and goals. In addition to fairly visible global initiatives, such as the Beijing Platform for Action of 1995, there is evidence that in many countries and regions, data on gender equality are becoming part of the policy process.<sup>1</sup> These are undoubtedly positive indirect effects of the production of

---

<sup>1</sup> Within the European Union, a gender-equality index is under development (Plantenga et al. 2003). On the use of data on gender equality in the UK and Ireland, see Breitenbach and Galligan (2006).

A. Hawken  
School of Public Policy (SPP), Pepperdine University, 24255 Pacific Coast Highway,  
Malibu, CA 90263-7490, USA

G. L. Munck (✉)  
School of International Relations, University of Southern California, 3518 Trousdale Parkway,  
VKC 330, Los Angeles, CA 90089-0043, USA  
e-mail: munck@usc.edu

data, for, as the adage goes, “What gets measured, gets done.” But the use of data on gender also draws attention to the issue of data quality and the need both to evaluate measures of gender disparities and to engage in efforts to generate improved measures. If the data that are used in the policy arena are of questionable quality, the legitimacy of using data in the pursuit of collective goals suffers. Thus, issues of measurement, though involving intricate methodological questions, are more than merely academic curiosities. It is of policy importance to get measurement right.

Many researchers and organizations have been involved with developing measures disaggregated by gender.<sup>2</sup> Several projects have focused on measuring specific issues, such as reproductive health (Abdullah 2000; Yinger et al. 2002) and violence against women (Walby 2007; Kelly et al. 2008). Others have taken advantage of the growing availability of a broad range of indicators, and sought to identify which indicators can be used to measure diverse aspects of social life and to identify areas where measures are lacking (UNECLAC 1999; UNECE 2001; UNIFEM 2002; UNESCAP 2003; UNRISD 2005; UNDP 2006). As a result of a broad collective endeavor, the amount of data that incorporates a gender perspective has increased considerably.

A distinctive facet of this 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 a yearly 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. Contributions to this debate have been focused on assessing the two UNDP indices, highlighting some of their weaknesses and suggesting how the UNDP might improve their indices. And much effort by organizations and independent researchers has gone into developing new indices. Thus, today, a number of indices with gender-differentiated data are broadly available.

Indices are likely to play a growing role in discussions of gender, both in academic and public policy circles. All indices serve to summarize a complex phenomenon by aggregating multiple indicators. This function is all the more important in the study of gender disparities because the successful response to the lack of information on gender in the 1980s has created a new problem: the intractable number of gender indicators. Indeed, the current availability of gender indicators—a comprehensive list of data sources reveals a total of 303 indicators (UNECLAC 2002: Annex 6)—makes 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 overwhelming number of indicators. For an index to serve as a synthetic measure of gender disparities, the construction of the index must take as its point of departure a clearly formulated concept and comply with basic methodological criteria. And, as the ongoing discussion about how best to craft a gender index underscores,<sup>3</sup> the construction of a sound index with gender-differentiated data is not an easy task.

<sup>2</sup> 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).

<sup>3</sup> This discussion has led to the proposal of multiple gender index, in addition to the two UNDP indices: the Gender Equality Index (White 1997), the Gender Inequality Index (Forsythe et al. 1998), the Relative Status of Women index (Dijkstra and Hanmer 2000), the Standardized Index of Gender Equality (Dijkstra 2002), the European Union Gender Equality Index (Plantenga et al. 2003), the African Gender and Development Index (UNECA 2004), the Gender Equity Index (Social Watch 2004), the Social Institutions and Gender

This article contributes to this discussion by offering a systematic assessment of the most visible gender indices that provide data on most countries of the world. 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).<sup>4</sup> The purpose of our comparison of these five indices is to provide useful guidance to users of gender indices and to contribute to the collective effort to further develop cross-national indices that aggregate gender-differentiated data.

Comparing these gender indices is not straightforward. The description of the methodologies underlying these indices do not share a common terminology, do not adequately describe methodological choices and often remain silent on potential sources of measurement error. Comparing the five gender indices is further complicated as not all index developers are explicit about the overarching concept they seek to measure. A careful inspection of the methodological choices underlying the construction of these indices often shows that what the index actually measures is different from what it purportedly measures. As such, our assessment addresses two related questions: *what do these indices actually measure?* and, *how valid are these indices?*

To answer these questions, in Sect. 1 we present a framework for evaluating gender indices that integrates a large literature on the methodology of measurement and identifies the key methodological choices underlying gender indices. This framework serves as the organizing principle for the article. Sections 2, 3 and 4 focus on the distinct methodological tasks involved in the construction of the five indices. In these three sections, we describe and compare the choices made by the index creators, and we assess whether each choice contributes to a measure that is clear and consistent in meaning, whether an appropriate justification is offered for each choice, whether the index creators offer empirical tests that validate their choices, and whether tests we conduct suggest that these indices are well constructed.<sup>5</sup>

By way of a conclusion, 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. Finally, we outline the implications of our assessment for users and producers of gender indices.

---

Footnote 3 continued

Index (Jütting et al. 2006, 2008; Branisa et al. 2009), the Gender Equality in Education Index (Unterhalter 2006), the Global Gender Gap Index (Hausmann et al. 2007b), and the Multidimensional Gender Equality Index (Permanyer 2008). This discussion also includes various assessments of gender indices (Bardhan and Klasen 1999; Dijkstra 2002, 2006; Charmes and Wieringa 2003; Chant 2006; Klasen 2006a, b, c; Shuller 2006; Jütting et al. 2008; Permanyer 2010).

<sup>4</sup> Though many other indices have been proposed, 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.

<sup>5</sup> In this assessment, we rely on texts about the methodology underlying each index as well as the data from which indicators and, thence, indices are constructed. On the methodology used to construct each index, we consulted Anand and Sen (2003a, b), Branisa et al. (2009), Hausmann et al. (2007b), Jütting et al. (2006, 2008), Porter and Schwab (2008), Social Watch (2005, 2008a: 68–79, b) and UNDP (1995, 2007: 225–227, 358–360). The data on each index are generally available in reports or the websites of each organization.

### 1 A Framework for the Evaluation of Gender Indices

A framework to evaluate gender indices need not be overly elaborate. In the broadest terms, producing a gender index involves two stages—securing raw data on indicators, and then combining the values of these indicators into an index—tackling a series of distinct tasks, and meeting certain criteria (see Table 1). Yet, assessing the validity of

**Table 1** A methodological framework for the evaluation of indices

| Index construction stage | Task                                    | Key question  | Assessment criteria  |
|--------------------------|---|---|--|
| Overarching concept<br>↓ | Identification of conceptual dimensions | Are different dimensions of the <i>overarching concept</i> to be measured identified? If so, what are these conceptual dimensions and how are they formulated?  | Identification, in light of theory, of a mutually exclusive and jointly exhaustive set of conceptual dimensions<br><br>Avoidance of extraneous conceptual dimensions, that is, conceptual dimension of a different overarching concept                   |
|                          | Raw data on indicators<br>↓             | Selection of indicators   | What indicators are selected to measure each conceptual dimension? How do the indicators connect to each conceptual dimension?<br><br>Avoidance of extraneous indicators, that is, indicators of a different conceptual dimension or a different concept |
| Data on indices<br>↓     | Design of 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<br>Indicator scales that offer as much nuance (as many distinctions) as is justified  |
|                          | Assignment of values to indicators      | How are values assigned to each indicator?  | Use of a method for assigning values to indicators that is replicable and that generates reliable and valid measures   |
|                          | Choice of rescaling rule                | If the original indicator scales (that is, those used to collect the raw data) are rescaled, what rescaling rule is used? What rule is used to normalize the indicator scales, that is, ensure that all indicators are measured with the same unit? Is any other transformation of the original indicator scales carried out? | A theoretically-justified rule<br>Robustness of the data to changes in rule, and replicability of the procedures   |
|                          | Choice of weighting rule                | What weight is assigned to each indicator?  | A theoretically-justified rule<br>Robustness of the data to changes in rule, and replicability of the procedures   |
|                          | Choice of aggregation rule              | What is the relationship among the indicators being aggregated?   | A theoretically-justified rule<br>Robustness of the data to changes in rule, and replicability of the procedures   |

gender indices is tricky. The validity of an index is influenced by many methodological choices and these choices are interrelated. Thus, a brief discussion of the methodological choices involved in index construction sets the stage for the analysis of gender indices that follows.

### 1.1 From Overarching Concept to Raw Data on Indicators

All measures, inasmuch as they are theoretically interpretable, explicitly or implicitly take as a point of reference an overarching concept. Hence the first task in developing a measuring instrument—identifying the dimensions of the overarching concept—is the most purely theoretical one. The first desideratum of a measuring instrument is a clear, theoretically justified definition, consisting of a mutually exclusive and jointly exhaustive set of conceptual dimensions that avoid contamination by extraneous concepts.

The second task is the selection of indicators to measure the conceptual dimensions identified in the prior step. Since indicators are simply concrete, empirically grounded concepts, this task is an extension of the prior more purely conceptual task. Thus, the selected indicators should appropriately bridge the abstract concepts that are based on theorizing and the observables that are essential to measurement. The selected indicators should also capture the full meaning of the conceptual dimensions they purportedly measure, avoiding duplication and extraneous indicators.<sup>6</sup>

The third task concerns the design of scales for each indicator. These, too, should be consistent with the concept being measured, and should offer as much nuance, that is, as many distinctions, as is justified. Finally, the fourth task focuses on the process whereby values are assigned to each indicator and should aim to maximize reliability, that is, that an independent attempt to assign values to the indicators would produce similar data; and validity, that is, that the values assigned to each indicator are true measures of the concept being measured. It is also important to ensure the replicability of this process, that is, that the data generating process can be reproduced, so that the reliability and validity of the data on the indicators can be verified by independent researchers.

### 1.2 From Raw Data on Indicators to Data on an Index

The choices involved in securing raw data on indicators set the initial parameters of an index. But the transformation of raw data on indicators into data on an index involves several additional choices. First, an index developer faces some choice regarding rescaling. Specifically, rescaling might be necessary if the indicators have been measured with different scales and might be called for if the form of the function between the values being aggregated and the aggregate value is non-linear. Second, the index developer must consider what weight should be assigned to each indicator. And, third, the index developer must address the relationship among indicators and determine the appropriate aggregation rule. Moreover, when producing indices with gender-differentiated data, each of these

---

<sup>6</sup> In making choices about indicators, it is important to avoid a mismatch between the theoretical and operational definition of a concept. The use of multiple indicators to measure the same conceptual dimension, that is, indicators 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 distinct 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 that they distort the meaning of the measure and cloud its interpretation.

choices must be confronted in the context of two distinct steps: (i) the aggregation of the male and female values for *each* indicator (for example, the percentage of men and women holding seats in parliament), and (ii) the aggregation of the values of *all indicators*.

Here again, theory and testing is important. For example, the choice of aggregation rule should be theoretically justified. Moreover, sensitivity tests that assess how robust the aggregated values are to changes in the aggregation rule should be part of the validation of an index. Indeed, if the theory guiding the choice of aggregation rule is weak, and if alternative, but plausible, aggregation rules result in considerable changes in the data that are generated, the validity of the index should be questioned.

This framework is not exhaustive, but it presents the key issues involved in the production of data, and draws attention to choices—such as those involved in aggregation—that are rarely treated explicitly, and highlights issues—such as the many links between theory and data—that are frequently overlooked. Moreover, this framework is suitable to the job at hand: a comparison of gender indices and an assessment of the strengths and weaknesses of these indices. Thus, we turn next to an analysis of gender indices, starting with the choices involved in securing data on the indicators used, addressing subsequently the two-step aggregation process that completes the production of these indices with gender-differentiated data.

## 2 From Overarching Concept to Raw Data on Indicators

### 2.1 Identifying Conceptual Dimensions

The creators of the UNDP's GDI and GEM, Social Watch's GEI, the World Economic Forum's GGGI, and the OECD's SIGI identified the dimensions of their indices' overarching concepts in a similar way. Each overarching concept is disaggregated into three or four mutually exclusive conceptual dimensions (see column 2 in Table 2). This helps to orient the subsequent task of indicator selection. Nonetheless, there are notable differences in terms of the level of theorization of the overarching concept, and the content of the conceptual dimensions, of each index. And these differences are consequential.

The GDI has a key virtue: its clarity of purpose. The index's overarching concept, development, is well theorized and is disaggregated and measured in much the same terms as the UNDP's Human Development Index (HDI).<sup>7</sup> As a result, the GDI is relatively easy to interpret. But the explicit grounding of the GDI in a theorized concept also draws attention to a key limitation. The concept of human development is very broad, encompassing factors essential to the "formation of human capabilities" that are included in the HDI and the GDI, but also a series of "political, economic and social freedoms" (UNDP 1990: 10, Sen 1999: 3–53) that, due to the lack of data on the relevant indicators at the time the HDI was constructed (Fukuda-Parr 2003: 119), are excluded from the HDI as well as the GDI. Thus, the conceptual critique of the HDI, its failure to include several key conceptual dimensions of human development (Fukuda-Parr 2003), applies equally to the GDI.<sup>8</sup>

<sup>7</sup> The difference between the GDI and the HDI is that the GDI differentiates between genders and uses gender inequalities to adjust the HDI scores downwards (UNDP 2007).

<sup>8</sup> The one problem of the HDI that the GDI partially skirts, through its focus on gender, is the critique that the HDI does not address distributional issues.

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

| Index producer and name                      | From overarching concept to raw data on indicators |   |  |  | From raw data on indicators to data on an index   |  |
|--|--|---|--|--|---|--|
|  | Conceptual dimensions                              | Indicators*   | Indicator scales                       | Indicator value assignment                                     | 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)                  | Scale predetermined by data collectors | National registration systems, population censuses and surveys | Normalization of scale (as in HDI); equal weighting of deviations from standard of parity (women live 5 years more than men) favoring men and women, adjusted for population share; harmonic mean of relational and absolute measures | Additive and unweighted (except for the dimension) aggregation rule at the level of indicators; and additive and unweighted aggregation rule at the level of conceptual dimensions |
|  | Knowledge  | Literacy rate (for adults, age 15 and over) (Male and female) | Scale predetermined by data collectors | National censuses or surveys                                   | Equal weighting of deviations from standard of parity (50/50 ratio) favoring men and women, adjusted for population share; harmonic mean of relational and absolute measures  |  |
|  |  | Primary level enrolment (net) (Male and female)               | Scale predetermined by data collectors | Administrative records and population censuses                 | Equal weighting of deviations from standard of parity (50/50 ratio) favoring men and women, adjusted for population share; harmonic mean of relational and absolute measures  |  |
|  |  | Secondary level enrolment (net) (Male and female)             | Scale predetermined by data collectors | Administrative records and population censuses                 | Equal weighting of deviations from standard of parity (50/50 ratio) favoring men and women, adjusted for population share; harmonic mean of relational and absolute measures  |  |
|  |  | Tertiary level enrolment (gross) (Male and female)            | Scale predetermined by data collectors | Administrative records and population censuses                 | Equal weighting of deviations from standard of parity (50/50 ratio) favoring men and women, adjusted for population share; harmonic mean of relational and absolute measures  |  |

**Table 2** continued

| Index producer and name                | From overarching concept to raw data on indicators |   |  | From raw data on indicators to data on an index  |   |   |
|--|--|---|--|--|---|---|
|  | Conceptual dimensions                              | Indicators*   | Indicator scales                       | Indicator value assignment   | Step 1: Aggregation of male and female values for each indicator  | Step 2: Aggregation of the values of all indicators   |
|  | A decent standard of living                        | Earned income (PPP US\$) (Male and female)**                              | Scale predetermined by data collectors | National surveys, accounts and population censuses; International Comparison Program surveys | Normalization of scale (as in HDI) and log transformation; equal weighting of deviations from standard of parity (50/50 ratio) favoring men and women, adjusted for population share; harmonic mean of relational and absolute measures |   |
| UNDP, Gender Empowerment Measure (GEM) | Power over economic resources                      | Earned income (PPP US\$) (Male and female)**                              | Scale predetermined by data collectors | National surveys, accounts and population censuses; International Comparison Program surveys | Normalization of scale (as in HDI); equal weighting of deviations from standard of parity (50/50 ratio) favoring men and women, adjusted for population share; harmonic mean of relational and absolute measures                        | Additive and unweighted aggregation rule, at both the level of indicators and conceptual dimensions |
|  | Economic participation and decision-making         | Professional and technical positions (Male and female)                    | Scale predetermined by data collectors | National surveys   | Equal weighting of deviations from standard of parity (50/50 ratio) favoring men and women, adjusted for population share; harmonic mean of relational and absolute measures  |   |
|  | Political participation and decision-making        | Positions as legislators, senior officials and managers (Male and female) | Scale predetermined by data collectors | National surveys   | Equal weighting of deviations from standard of parity (50/50 ratio) favoring men and women, adjusted for population share; harmonic mean of relational and absolute measures  |   |
|  |  | Parliamentary seats (Male and female)                                     | Scale predetermined by data collectors | Administrative records   | Equal weighting of deviations from standard of parity (50/50 ratio) favoring men and women, adjusted for population share; harmonic mean of relational and absolute measures  |   |



**Table 2** continued

| Index producer and name                 | From overarching concept to raw data on indicators |   | From raw data on indicators to data on an index |  |
|---|--|---|---|--|
|   | Conceptual dimensions                              | Indicators*   | Indicator scales                                | Indicator value assignment   |
| Social Watch, Gender Equity Index (GEI) | Education  | Literacy rate (for youths, age 15–24) (Male and female) | Scale predetermined by data collectors          | National censuses or surveys   |
|   |  | Primary level enrolment (net) (Male and female)         | Scale predetermined by data collectors          | Administrative records and population censuses   |
|   |  | Secondary level enrolment (net) (Male and female)       | Scale predetermined by data collectors          | Administrative records and population censuses   |
| Economic activity                       |  | Tertiary level enrolment (gross) (Male and female)      | Scale predetermined by data collectors          | Administrative records and population censuses   |
|   |  | Labor force participation (Male and female)             | Scale predetermined by data collectors          | National surveys   |
|   |  | Earned income (PPP US\$) (Male and female)**            | Scale predetermined by data collectors          | National surveys, accounts and population censuses; International Comparison Program surveys |

Step 1: Aggregation of male and female values for each indicator

Step 2: Aggregation of the values of all indicators

Deviations from standard of parity (50/50 ratio) favoring women have no weight, adjusted for population share

Deviations from standard of parity (50/50 ratio) favoring women have no weight, adjusted for population share

Deviations from standard of parity (50/50 ratio) favoring women have no weight, adjusted for population share

Deviations from standard of parity (50/50 ratio) favoring women have no weight, adjusted for population share

Deviations from standard of parity (50/50 ratio) favoring women have no weight, adjusted for population share

Deviations from standard of parity (50/50 ratio) favoring women have no weight, adjusted for population share

Deviations from standard of parity (50/50 ratio) favoring women have no weight, adjusted for population share

Table 2 continued

| Index producer and name                              | From overarching concept to raw data on indicators   |  |   | From raw data on indicators to data on an index   |  |  |                  |   |
|--|--|--|---|---|--|--|------------------|---|
|  | Conceptual dimensions  | Indicators*                            | Indicator scales  | Indicator value assignment  | Step 1: Aggregation of male and female values for each indicator<br>Step 2: Aggregation of the values of all indicators  |  |                  |   |
| Empowerment  | Professional and technical positions (Male and female)<br>Positions as legislators, senior officials and managers (Male and female)<br>Parliamentary seats (Male and female) | Scale predetermined by data collectors | National surveys  | Deviations from standard of parity (50/50 ratio) favoring women have no weight, adjusted for population share | Step 1: Aggregation of male and female values for each indicator<br>Step 2: Aggregation of the values of all indicators  |  |                  |   |
|  |  |  |   |   |  | Scale predetermined by data collectors | National surveys | Deviations from standard of parity (50/50 ratio) favoring women have no weight, adjusted for population share |
|  |  |  |   |   |  |  |                  |   |
| Ministerial level positions (Male and female)        | Scale predetermined by data collectors   | Administrative records                 | Deviations from standard of parity (50/50 ratio) favoring women have no weight, adjusted for population share |   |  |  |                  |   |
| World Economic Forum, Global Gender Gap Index (GGGI) | Health and survival  | Number of births (male and female)     | Scale predetermined by data collectors  | National registration systems   | Additive and weighted aggregation rule at the level of indicators (higher weights are assigned to indicators with lower standard deviations, the weighting scheme of the 2006 index is used in subsequent years); and additive and unweighted aggregation rule at the level of conceptual dimensions |  |                  |   |

**Table 2** continued

| Index producer and name                | From overarching concept to raw data on indicators            |                          |  | From raw data on indicators to data on an index   |   |
|--|---|--------------------------|--|---|---|
|  | Indicators*   | Indicator scales         | Indicator value assignment                                     | Step 1: Aggregation of male and female values for each indicator                                    | Step 2: Aggregation of the values of all indicators |
|  | Healthy life expectancy (Male and female)                     | Scale by data collectors | National registration systems, population censuses and surveys | Deviations from standard of parity (1.06 males for every female born) favoring women have no weight |   |
| Educational attainment                 | Literacy rate (for adults, age 15 and over) (Male and female) | Scale by data collectors | National censuses or surveys                                   | Deviations from standard of parity (50/50 ratio) favoring women have no weight                      |   |
|  | Primary level enrolment (net) (Male and female)               | Scale by data collectors | Administrative records and population censuses                 | Deviations from standard of parity (50/50 ratio) favoring women have no weight                      |   |
|  | Secondary level enrolment (net) (Male and female)             | Scale by data collectors | Administrative records and population censuses                 | Deviations from standard of parity (50/50 ratio) favoring women have no weight                      |   |
|  | Tertiary level enrolment (gross) (Male and female)            | Scale by data collectors | Administrative records and population censuses                 | Deviations from standard of parity (50/50 ratio) favoring women have no weight                      |   |
| Economic participation and opportunity | Labor force participation (Male and female)                   | Scale by data collectors | National surveys   | Deviations from standard of parity (50/50 ratio) favoring women have no weight                      |   |

Table 2 continued

| Index producer and name | From overarching concept to raw data on indicators                        |  | From raw data on indicators to data on an index  |   |
|-------------------------|---|--|--|---|
|                         | Conceptual dimensions   | Indicators*                            | Indicator scales   | Indicator value assignment  |
|                         | Earned income (PPP US\$) (Male and female)**                              | Scale predetermined by data collectors | National surveys, accounts and population censuses; International Comparison Program surveys | Step 1: Aggregation of male and female values for each indicator<br>Step 2: Aggregation of the values of all indicators |
|                         | Wage equality for similar work  | Scale predetermined by data collectors | Expert survey  | Deviations from standard of parity (50/50 ratio) favoring women have no weight  |
|                         | Professional and technical positions (Male and female)                    | Scale predetermined by data collectors | National surveys   | Deviations from standard of parity (50/50 ratio) favoring women have no weight  |
|                         | Positions as legislators, senior officials and managers (Male and female) | Scale predetermined by data collectors | National surveys   | Deviations from standard of parity (50/50 ratio) favoring women have no weight  |
| Political empowerment   | Parliamentary seats (Male and female)                                     | Scale predetermined by data collectors | Administrative records   | Deviations from standard of parity (50/50 ratio) favoring women have no weight  |

**Table 2** continued

| Index producer and name                           | From overarching concept to raw data on indicators                    |                               | From raw data on indicators to data on an index              |  |
|---|---|-------------------------------|--|--|
|   | Conceptual dimensions   | Indicators*                   | Indicator scales   | Indicator value assignment   |
| OECD, Social Institutions and Gender Index (SIGI) | Ministerial level positions (Male and female)                         | Administrative records        | Scale predetermined by data collectors                       | Deviations from standard of parity (50/50 ratio) favoring women have no weight |
|   | Heads of state (number of years over last 50 years) (Male and female) | Historical records            | Count of years   | Deviations from standard of parity (50/50 ratio) favoring women have no weight |
|   | Family law  | National registration systems | Scale predetermined by data collectors                       | Rescaling: Ratio one-sided scale (restriction on rights of girls)              |
|   | Polygamy  | Expert coding                 | Ordinal scale with end points, with no gender distinction*** | Not applicable   |

Step 1: Aggregation of male and female values for each indicator

Step 2: Aggregation of the values of all indicators

Additive and weighted aggregation rule at the level of indicators (higher weights are assigned to indicators that correlate more), after a dimensionality test on the indicators within each conceptual dimension leads to the introduction of a new conceptual dimension †; non-linear (square function) transformation of subindex scales ††, then additive and unweighted aggregation rule at the level of conceptual dimensions

Table 2 continued

| Index producer and name | From overarching concept to raw data on indicators                         |   | From raw data on indicators to data on an index |  |
|-------------------------|--|---|---|--|
|                         | Conceptual dimensions  | Indicators*   | Indicator value assignment                      | Step 1: Aggregation of male and female values for each indicator   |
| Inheritance             | Parental authority   | Ordinal one-sided scale (inequality disadvantaging women), with end points        | Expert coding                                   | Not applicable   |
|                         | Ordinal one-sided scale (inequality disadvantaging women), with end points | Expert coding   | Not applicable                                  |  |
| Physical integrity      | Female genital mutilation  | Scale predetermined by data collectors  | National surveys                                | Rescaling: Ratio one-sided scale (restriction on rights of girls)  |
|                         | Violence against women   | 5-point ordinal one-sided scale (restriction on rights of women)****              | Expert coding                                   | Not applicable   |
| Civil liberties         | Missing women  | Scale predetermined by data collectors  | National registration systems                   | Rescaling: Ratio one-sided scale, with normalization based on goalpost idea (restriction on rights of females)**** |
|                         | Freedom of movement  | 3-point ordinal one-sided scale (restriction on rights of women), with end points | Expert coding                                   | Not applicable   |
|                         | Freedom of dress   | Ordinal one-sided scale (restriction on rights of women), with end points         | Expert coding                                   | Not applicable   |

**Table 2** continued

| Index producer and name | From overarching concept to raw data on indicators |   | From raw data on indicators to data on an index |                            |
|-------------------------|--|---|---|----------------------------|
|                         | Conceptual dimensions                              | Indicators*   | Indicator scales                                | Indicator value assignment |
| Ownership rights        | Access to land                                     | Ordinal one-sided scale (restriction on rights of women), with end points | Expert coding                                   | Not applicable             |
|                         | Access to bank loans                               | Ordinal one-sided scale (restriction on rights of women), with end points | Expert coding                                   | Not applicable             |
|                         | Access to property                                 | Ordinal one-sided scale (restriction on rights of women), with end points | Expert coding                                   | Not applicable             |

(\* 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. (\*\*\*) It is unclear what scale is used for this indicators, given that two different descriptions are offered (compare Jütting et al. 2008: 68–70, 83 to Branisa et al. 2009) (†) Prior to aggregation, a test of the dimensionality of the indicators within each conceptual dimension was conducted and, given that the indicator Missing Women did not correlate with the other indicators of Physical Integrity, this indicator was included under a new conceptual dimension labeled Son Preference. (††) The non-linear function reduces the degree to which high levels of inequality on one conceptual dimension can be compensated with low levels of inequality on other conceptual dimensions

Sources: On the UNDP indices, see UNDP (1995, 2007: 225–227, 358–360) and Anand and Sen (2003a, b). On the Social Watch index, see Social Watch (2008a: 68–79; 2008b). On the World Economic Forum index, see Hausmann et al. (2007b) and Porter and Schwab (2008). On the OECD index, see Jütting et al. (2008) and Branisa et al. (2009)

The other four indices appear to take the GDI as a point of reference and, partly explicitly, partly implicitly, seek to overcome the narrowness of the GDI. But not all these indices are as clearly rooted in a well-theorized overarching concept as the GDI. The first index to incorporate conceptual dimensions excluded from the GDI was the UNDP's own Gender Empowerment Measure (GEM), launched along with the GDI in 1995. This index explicitly addresses factors excluded from the GDI, such as rights and access to power. But the concept of empowerment used in the GEM was not elaborated and the relationship between the GDI and the GEM is unclear. 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.<sup>9</sup> However, as in the case of the GEM, the overarching concepts of the GEI and the GGGI are not explicitly discussed and thus their developers do not clarify what these indices aim to measure and do not offer a theoretical rationale for the conceptual dimensions included in these indices.

Indeed, besides the GDI, only the OECD's Social Institutions and Gender Index (SIGI), made public in 2009, is based on an explicit framework, built around the overarching concept of social institutions (Morrisson and Jütting 2005: 1066–70). This 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 two 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 that, even if not entirely comprehensive (for example, discrimination in the labor market is not addressed), are ignored by other indices. Thus, the concept of social institutions served as a useful anchor for the SIGI and helps to clarify the meaning of the index.

In sum, the history of gender indices involved a series of efforts to include broader dimensions of social life that had not been included in the GDI. In this regard, it is tempting to posit that all five indices could be understood as measures of the concept of human development, as elaborated by Sen (1999) and Nussbaum (2000), 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. But, because the GEM, GEI and GGGI in particular do not rest on as clear an overarching concept as the GDI, it is difficult to compare these indices to each other conceptually.<sup>10</sup> Moreover, there are a number of other important differences in the choices that go into developing these indices and, as the analysis that follows will show, the meaning of each index is also affected by these choices.

## 2.2 Selecting Indicators

When assessing the selection of indicators to measure each of the conceptual dimensions identified in the prior step, some interesting points emerge (see column 3 in Table 2). The

<sup>9</sup> 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.

<sup>10</sup> Indeed, though it is probably safe to say that all five indices avoid the introduction of extraneous elements that are not part of the concept's meaning, it is not possible to assess with certainty whether the conceptual dimensions the GEM, GEI and GGGI do include are jointly exhaustive of the meaning of their overarching concepts.



OECD's SIGI breaks new ground by calling for indicators that address key spheres of social life in which gender differences could be a factor even if they are not readily available, i.e., the scarcity of readily available data does not constrain the selection of indicators. Indeed, the SIGI shows that data availability does not need to be seen as an overwhelming constraint on how indices are constructed and that new data can be generated to measure certain indicators that are considered central to an index's overarching concept.

In contrast, the other four indices 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).<sup>11</sup> The choices are undoubtedly driven by the desire to include indicators that cover a large number of countries, are publicly available, and are regularly updated. But there are weaknesses associated with these choices of convenience.

One problem is that the selected indicators fail to capture the full meaning of the conceptual components they are supposed to measure. As an example, Parliamentary Seats, by itself, seems too narrow a measure of the GEM's conceptual component Political Participation and Decision-making. Another problem is the use of extraneous indicators, which are measures of conceptual components other than those they are claimed to measure. Indeed, it is clear that the indicator Legislators, Senior Officials and Managers—used in the GEM to measure the conceptual component Economic Participation and Decision-making, and in the GGGI to measure the conceptual component Economic Participation and Opportunity—is problematic in that political information about legislators is presented as a measure of an economic dimension.<sup>12</sup> In sum, the selection of indicators for which data are easily accessible is associated with the common failure to tap the full meaning of each conceptual dimension, a problem most evident in the GDI, and other problems, such as the inclusion of extraneous indicators.<sup>13</sup>

### 2.3 Designing Indicator Scales

The next choice in the production of an index, the design of the scales used in collected raw data for each indicator, is a moot question for most indices. Indeed, because most indices largely rely on data gathered by others, as pointed out above, they simply operate with scales designed by the original data collectors that consistently assess the situation of men and women. In contrast, the developers of the SIGI do make distinct scaling choices. Indeed, the SIGI differs from the other indices in that, because it seeks to respond to the laudable goal of generating new data on some hard to measure indicators, its elaboration includes the specific design of scales for new indicators (see column 5 in Table 2).<sup>14</sup> Nonetheless, from the perspective of the ultimate goal of constructing an index, the SIGI's scales are quite problematic. On the one hand, though the designed scales reflect sensitivity about the number of distinctions that can justifiably be made

---

<sup>11</sup> Capitalization is used to distinguish references to the conceptual dimensions and indicators of the indices under review from other uses of the same words.

<sup>12</sup> The simultaneous use of the Legislators, Senior Officials and Managers and the Parliamentary Seats indicators in the GEM and GGGI, and the GEI, also leads to a problem of partial redundancy.

<sup>13</sup> This problem of "indicator creep" is not unique to indicators that are selected in part because of the ready availability of data. For example, the GGGI's indicator Heads of State uses information about the past 50 years to assess the present, counter to the approach taken for all the other GGGI indicators, and hence introduces an extraneous element into the index.

<sup>14</sup> In addition, the World Economic Forum's 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 in the past 50 during which a country has had a male or female president or prime minister.

**Table 3** Indices with gender-differentiated data II: a comparison of indicators

| 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   | ×  |  |   | ×  |
| Number of births  |  |  |   | ×  |
| Literacy rate (for adults or youths)                    | ×  |  | ×                                       | ×  |
| Primary level enrolment (net)                           | ×  |  | ×                                       | ×  |
| Secondary level enrolment (net)                         | ×  |  | ×                                       | ×  |
| Tertiary level enrolment (gross)                        | ×  |  | ×                                       | ×  |
| Earned income (PPP US\$)                                | ×  | ×                                      | ×                                       | ×  |
| Professional and technical positions                    |  | ×                                      | ×                                       | ×  |
| Positions as legislators, senior officials and managers |  | ×                                      | ×                                       | ×  |
| Labor force participation                               |  |  | ×                                       | ×  |
| Wage equality for similar work                          |  |  |   | ×  |
| Parliamentary seats                                     |  | ×                                      | ×                                       | ×  |
| Ministerial level positions                             |  |  | ×                                       | ×  |
| Heads of state  |  |  |   | ×  |

An “ × ” means that this indicator is included in an index. Each of these indices calls for data on males and females for each of these indicators

Source: See Table 2

in light of available information (Jütting et al. 2008: 68–70, 83–84; OECD Development Centre 2009b), they are all ordinal scales. Thus, their use in an index is highly questionable. On the other hand, most of the SIGI's scales are one-sided, addressing inequalities that disadvantage girls or women. Thus, potential inequalities that disadvantage boys and men are thus ignored and the possibility of empirically studying gender equality broadly understood is closed.<sup>15</sup> Furthermore, the interpretation of the index is muddled due to the combination of different types of indicators: some, such as Parental Authority and Inheritance, that posit a relationship between the situations of men and women, and others, such as Female Genital Mutilation and

<sup>15</sup> 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 (Porter and Schwab 2008: question 9.13), and thus excludes a priori the possibility of collecting data that might show that women earn more than men.

Violence Against Women, that focus on restrictions of the rights of girls and women while making no contrast to the situations of boys and men. As a result, the SIGI's indicator scales do not offer a basis for systematically collecting gender-disaggregated data and fuse, in a largely implicit manner, measures of gender inequality and measures of distinctive women's rights.

## 2.4 Assigning Values to Indicators

Since most of the gender indices rely on readily available data sets (the clearest exception being the OECD's SIGI), the choice concerning the way values are assigned to indicators, much as the choice concerning indicator scales, largely follows the choice regarding which indicators to include. The concerns about the reliability and comparability of many of the data sets used in these indices expressed by Srinivasan (1994) deserve attention.<sup>16</sup> Nonetheless, when seen from a broad perspective, the data on these indicators are relatively reliable (see column 4 in Table 2). Indeed, the most serious doubts regarding the process of data generation pertain less to available data sets than to the efforts by the GGGI and the SIGI developers to generate their own data on indicators.

Data on the GGGI's indicator Wage Equality for Similar Work is collected through a question in an expert survey (Porter and Schwab 2008: question 9.13), of unknown reliability. The degree of agreement among the respondents is not reported and, since the data are not publicly available, independent researchers are not able to assess the reliability of the data. And the data on the SIGI's indicators, generated through expert coding, are even more troubling. Expert coding is a well-established method of generating data. Moreover, the sources of information used to code the SIGI's indicators are made public, and an accompanying narrative provides valuable background information used in the coding of indicators (Jütting et al. 2008: 68–70, 83–84; OECD Development Centre 2009a). But, as some of the indicator scales are vaguely formulated (Jütting et al. 2008: 83–84; OECD Development Centre 2009b), and as the SIGI's developers do not report the results of an intercoder reliability test, we are left with serious doubts about the validity and reliability of the data on the SIGI's indicators. Given that this method of expert coding is used for nine of the SIGI's twelve indicators, this is a major weakness.

## 2.5 An Interim Assessment

Summing up the discussion thus far, the following points deserve emphasis. The UNDP's GDI and the OECD's SIGI take, as a point of departure, a theoretically developed overarching concept and what these indices seek to measure is relatively clear. But the theoretical ideas guiding the design of these two indices are poorly executed; in the case of the GDI, because its indicators fail to capture a large part of the meaning of the concept of development and, in the case of SIGI, because the new data it relies on are of questionable quality. In contrast, the UNDP's GEM, the Social Watch's GEI and the World Economic

---

<sup>16</sup> Some concerns about the reliability and comparability of many of the data sets used in these indices have been expressed by Srinivasan (1994). Moreover, some blatant errors have been 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 from 2006 to 2007, and from 0.08 to 0.58 for France from 2006 to 2007 reports! (Hausmann et al. 2006: 37, 64; 2007b: 42, 72, 87; 2008: 94). These changes are, on the face of things, incredible and indeed the data for the earlier year differ from those reported by the World Economic Forum's own data source, the International Labor Organization's LABORISTA data site. In other words, concerns about data reliability cannot be put to rest.

Forum's GGGI lack clear theoretical foundations. Hence, the meaning of these three indices has to be uncovered, inductively, by considering the content of the selected indicators. Moreover, any theoretical grounds for validating the choice of indicators included in these indices are surrendered. In short, though the creation of these indices reveals a considerable degree of self-awareness about the choices regarding conceptual dimensions, indicators, indicator scales, and the assignment of values to indicators, the discussion thus far also reveals weaknesses in all five indices.

### 3 From Raw Data on Indicators to Data on an Index I: The Rule to Aggregate Male and Female Values for Each Indicator

Indices with gender-differentiated data have a common goal: to move beyond simple averages that overlook possible differences in the situation of women relative to men. This goal simplifies the challenge of forming an index because, in most cases (the two UNDP indices and the OECD's SIGI are partial exceptions), it more or less automatically resolves the need to ensure that all indicators are measured with the same units. Thus, the key decisions tackled in the first step in the transforming the raw data on indicators into data on an index revolve around the issue of how to define a standard of parity, how to weight deviations from the posited standard, and whether and how to aggregate relational and absolute measures of attainment.

With regard to these choices, the gender indices under review—with the exception of the SIGI—share some features. The indices move toward a relational measure by positing a standard to be used in comparing the data on women and men for each indicator—opting, practically as a default option, for a 50/50 ratio as the standard of parity<sup>17</sup>—and calculating deviations from this standard (see column 6 in Table 2). But the gender indices differ with regard to two key choices: whether to assign equal weights to deviations from the posited standard favoring men and women; and whether to combine a relational measure with an absolute measure and, if so, how to combine these measures.

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 *or* women. That is, they measure inequality broadly understood. In contrast, the GEI and GGGI assign no weight to deviations from the standard of parity that favor women and only count inequalities that disadvantage women. That is, they measure inequality only from the perspective of women.<sup>18</sup> As to the combination of relational and absolute measures, three options can be distinguished. The GDI and GEM seek to retain the information about the absolute level of attainment of goals,<sup>19</sup> and aggregate a relational

<sup>17</sup> 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. The two UNDP indices and the GEI also make an adjustment for the share of women and men in the population.

<sup>18</sup> 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 et al. 2007a: 22).

<sup>19</sup> Because the UNDP gender indices are concerned with absolute levels, and not all the indicators use the same metrics, the two UNDP gender indices rely on the rescaling of some indicators. Specifically, the GDI's Life Expectancy and Earned Income indicators and the GEM's Earned Income indicator are normalized using the same reasonable procedures as in the HDI, that is, the notion of goalposts with minimum and maximum levels. In addition, so as to capture the idea of diminishing returns, the GDI's Earned Income indicator is log transformed.

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).<sup>20</sup> In contrast, the GEI and GGGI rely solely on relational measures, and do not take the absolute level of attainment of goals into consideration.

The options selected by the index developers are relatively well justified and the aggregation process is largely replicable, but basic questions can be raised about some choices. With the two UNDP indices, the decision to give equal weight to deviations from the standard of parity that favor men or women in both the GDI and the GEM 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. But the reliance on a strict 50/50 ratio in, for example, the GEM’s Parliamentary Seats indicator could be questioned. For example, if women fall short of gaining 50 percent of the seats in parliament, this may be due to the lack of a level playing field but could also be attributed to a host of unrelated factors. This standard is more consistent with a measure focused on equality of results than a measure of “opportunities,” which the UNDP suggests the GEM to be (UNDP 2007: 360).<sup>21</sup> Finally, the use of the harmonic mean to aggregate relational and absolute measures in the GDI and GEM is open to questioning. The rationale for this choice is spelled out in a paper by Anand and Sen (2003b: 211–13), who explain 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 some other weight; indeed, we are left with the impression that this weighting choice is arbitrary.<sup>22</sup>

Regarding the GEI, though its use of a 50/50 ratio as a standard of parity is a reasonable choice for most indicator, the same question raised about the GEM’s Parliamentary Seats indicator 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: 77), 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 and absolute levels of development (Social Watch 2005: 76–77), 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,<sup>23</sup> 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—

<sup>20</sup> In the case of the GEM, however, only the Earned Income indicator actually contains information about absolute levels of attainment.

<sup>21</sup> In contrast, the standard of parity chosen for the GDI’s Life Expectancy indicator is convincingly defended (Anand and Sen 2003a: 142, b: 211, 213–14).

<sup>22</sup> For a critique of this weighting scheme, see Bardhan and Klasen (1999: 987–90, 994).

<sup>23</sup> 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.

a choice that diminishes the GEI's substantive utility—or to an ad hoc reweighting of the indicators of the Empowerment conceptual dimension for that particular country—a choice that reduces the GEI's comparability across cases.

Finally, as to the GGGI, the use of a 50/50 ratio for the Parliamentary Seats and Heads of State indicators raises similar concerns to those mentioned above.<sup>24</sup> The option to assign no weight to deviations from the standard of parity favoring women entails, as with the GEI, a loss of generality that is somewhat cryptically played down by the index developers with little more than the claim that they find such a choice “more appropriate for [their] purposes” (Hausmann et al. 2007a: 5). 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 et al. 2007a: 3). But, as with the GEI, the sole reliance on relational measures creates its own problems. 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 procedure that reduces the comparability across cases of the data.

The OECD's SIGI is an exception with regard to aggregating male and female values for each indicator. The SIGI's approach is close to the one adopted by the GEI and GGGI, in that it disregards potential inequalities favoring women. But, unlike the other indices, the SIGI does not use one criterion for aggregating male and female values consistently, given that some of the scales focus solely on women and make no comparison with the situation of men. Finally, as mentioned above, data on men and women are not collected separately for the SIGI's indicators; rather, the comparisons are embedded in the scales of each indicator. Thus, unlike the other four indices, independent researchers are unable to test the robustness of SIGI's choice of aggregation rule.<sup>25</sup>

The possibility of conducting sensitivity analyses in the case of four of the five gender indices notwithstanding, it is striking that none of the index developers provides tests to validate their choice of rule to aggregate the male and female values of each indicator. Such tests are particularly relevant to the choice of aggregation used in the two UNDP indices. Indeed, given that the weight of the penalty assigned to deviations from equality in the GDI and GEM is not based on any theoretical grounds, it is important to verify what impact this choice has on the indices.<sup>26</sup> And when we performed such a test, the results were not encouraging (see Table 4). Though the purpose of building the GDI was to address inequalities, the GDI, which differs from the HDI only in its choice of the harmonic mean over the arithmetic mean, is not all that different from the HDI (a measure that does not address inequalities). Assigning the harshest possible penalty for inequalities, as captured by the minimum aggregation rule, may have been more suitable to the purpose. In contrast, when we calculate the GEM using alternative aggregation rules—whether the slightly less-punishing geometric mean or the vastly more-punishing minimum—we show that the GEM is not robust to changes in its aggregation rule. In other words, these tests raise questions about the choice of the use of the harmonic mean as an aggregation rule in both the GDI and GEM.

<sup>24</sup> In contrast, the use of standards of parity that diverge from a 50/50 ratio for the Healthy Life Expectancy and Number of Births indicators draws on considerable research (Hausmann et al. 2007a: 5).

<sup>25</sup> Another problem, as indicated above, is that the SIGI's indicator scales are mainly ordinal scales; yet the normalization and aggregation of these indicators is treated as entirely trouble-free. Some of the problems with the SIGI may be due to its creation in two stages, by two different teams. Specifically, Jütting et al. (2006, 2008) initially created a data set on the indicators and later Branisa et al. (2009) worked on an aggregation scheme.

<sup>26</sup> Even though this weighting choice is shown to have a large potential impact (Anand and Sen 2003b: 216), no results of an empirical test gauging the impact of picking different weighting are reported.

**Table 4** The impact of alternative aggregation rules I: the aggregation of male and female values

| Aggregation rule | UNDP, Gender-related development index (GDI) | UNDP, Gender empowerment measure (GEM) |
|------------------|--|--|
| Arithmetic mean  | 0.671  | ...                                    |
| Geometric mean   | 0.667  | 0.662                                  |
| Harmonic mean    | <b>0.665</b>                                 | <b>0.582</b>                           |
| Minimum          | 0.628  | 0.443                                  |

A comparison of means. The scales are 0.000–1.000 scales with higher values being indicative of a more positive situation. The actual aggregation rule is highlighted in bold. The arithmetic mean for the GDI is the same as the HDI; for the GEM this rule does not make sense since the average of three of the four indicators would always be 50, that is, a perfect balance. In these calculations we follow the exact procedure used by the UNDP except for the population weighting element. All differences of means are significant at a 0.0000 level. The N for the GDI data is 120 and for the GEM 95

Sources: Authors' calculations based on data on indicators, c. 2005; for more, see the data "Appendix"

#### 4 From Raw Data on Indicators to Data on an Index II: The Rule to Aggregate the Values of All Indicators

Turning to the last key choice in the production of gender indices—combining the values of all indicators—the five gender indices share some features. First, they all rely on a distinction between the level of conceptual dimensions and the level of indicators (see columns 2 and 3 in Table 2), and aggregate the group of indicators corresponding to conceptual dimensions into subindices before aggregating these subindices into a single index.<sup>27</sup> Second, they all rely on additive aggregation rules. Beyond this commonality, there are some significant differences (see column 7 in Table 2). The GEM and the GEI use simple averages, that is, an additive and unweighted aggregation rule, at both the level of indicators and the level of conceptual dimensions. The GDI follows suit, except that—following the HDI—it assigns different weights to the indicators that are part of the Knowledge conceptual dimension.<sup>28</sup> But the other two indices are quite different.

The GGGI is similar to the GEM and the GEI except that it breaks significantly with the equal-weight rule at the level of indicators. The GGGI's indicators are assigned weights according to their standard deviations relative to those of other indicators of the same conceptual dimension—higher weights are assigned to indicators with lower standard deviations and the weighting scheme calculated for the 2006 index is used in subsequent versions of the index (Hausmann et al. 2007a).<sup>29</sup> In turn, the SIGI uses the most-complicated aggregation procedure (Branisa et al. 2009). First, the developers of the SIGI conduct a dimensionality test on the indicators within each conceptual dimension and, finding that the Missing Women indicator did not correlate with the other indicators of the Physical Integrity conceptual dimension, placed the Missing Women indicator under a new

<sup>27</sup> As an example, 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. When a conceptual dimension is measured with only one indicator, as with the GEM's Political Participation and Decision-making dimension, no aggregation of indicators is necessary and the value of the relevant indicators—in this case the Parliamentary Seats indicator—is the value of the conceptual dimension.

<sup>28</sup> The Literacy Rate indicator is assigned double the weight of the combined three enrollment indicators.

<sup>29</sup> 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 et al. (2007a: 6).

conceptual dimension labeled Son Preference. Second, to calculate the value of the sub indices corresponding to each conceptual dimension, they use a factor-analytical technique and assign higher weights to indicators that correlate more.<sup>30</sup> Third, before aggregating the values of the subindices using an additive and unweighted aggregation rule, each subindex is rescaled using a non-linear (square function) transformation, which reduces the degree to which high levels of inequality on one conceptual dimension can be compensated with low levels of inequality on other conceptual dimensions.

The way index developers approach these choices has some merits. A plausible theoretical justification is offered or can be offered for many choices. The use of an additive aggregation rule is reasonable as all indices posit the existence of spheres that are neither interchangeable (gains in, say, the political sphere cannot make up or substitute for problems in, say, the economic sphere), nor prone to contamination (gains in one sphere are not overridden by problems in other spheres). Likewise, the rescaling using non-linear functions in the GDI and the SIGI is well justified (Anand and Sen 2003a: 140–41; Branisa et al. 2009: 7–8). But two weaknesses affect all five indices: one concerns the weighting scheme, and a second pertains to the testing of the selected aggregation rule against plausible alternatives.

The most common approach to weighting choices—used in the GDI, GEM, and GEI—is purely theory driven, meaning that weights are simply assigned by the index developers. And the selected choice, with only one exception, is to give equal weight to all indicators within a conceptual dimension and to apply equal weights across the conceptual dimensions. This weighting choice seems to be largely a default option used less out of reasoned analysis than a sense that agnosticism regarding all possible weighting schemes somehow constitutes a justification for assigning equal weights to indicators and conceptual dimensions.<sup>31</sup> In addition, index developers appear to be unaware of how the weights of indicators are affected by the formal weighting scheme that is adopted but also by the number of conceptual dimensions identified and the number of indicators selected to measure each conceptual dimension. The choices concerning the number of conceptual dimensions and indicators are never addressed, even though these two factors indirectly but strongly affect the weighting of indicators.<sup>32</sup> Indeed, we show that when these choices are factored into the weighting of the GDI's, the GEM's, and the GEI's indicators (see columns 2, 3 and 4 in Table 5), the actual weight of each indicator within each index is anything but equal and the same indicator has quite different weights in different indices. Ironically, the index developers that opt for a theory-driven weighting scheme pay little attention to the multiple choices that affect the weight of each indicator and hence offer very little theoretical justification for their choices.

The GGGI and SIGI exemplify a different approach to weighting choices. They assign an equal weight to each conceptual dimension. But, they combine this theory-driven choice with data-driven choices at the indicator level. This mixed approach is interesting and potentially

<sup>30</sup> At this stage, each subindex is rescaled by assigning the value 0 to the country with the best performance and 1 to the country with the worst performance.

<sup>31</sup> These index developers essentially follow Babbie's (1995: 171) advice that "items [should] be weighted equally unless there are compelling reasons for differential weighting. That is, the burden of proof should be on differential weighting; equal weighting should be the norm." Indeed, in the one exception to the equal-weighting option among theory-driven weighting—for the GDI's indicators of the Knowledge conceptual dimension—the choice of weights is explicitly justified on the basis of the distinction between stock and flow indicators (Jahan 2003: 156). But this is questionable methodological advice. Moreover, there are well-thought-out arguments against the default option of assigning equal weights. See, for example, Bardhan and Klasen (1999: 987–90, 994) on the GDI's weighting scheme.

<sup>32</sup> A related choice that affects the weights of indicators is the selection of extraneous and redundant indicators (on these indicators, see Sect. 2.2).



**Table 5** 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   |

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

fruitful. But it is important to note that data-driven choices entail assumptions and that the legitimacy of the assumptions used in the GGGI and SIGI is questionable. The GGGI's data-driven weighting scheme assigns higher weights to indicators with lower standard deviations (Hausmann et al. 2007a). But it is not clear, for example, why a ban on women holding seats in parliament in a certain country should be considered less of a problem simply because many other countries also have one. 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 an arbitrarily selected year to determine whether certain problems will be given more or less weight in the future.<sup>33</sup> When the weighting for indicators thus derived are combined with the theory-driven weights at the level

<sup>33</sup> To be consistent, given the justification offered by the creators of the GGGI, the empirically derived weights would be recalculated each year, using the data on all the years that were being compared.

of conceptual dimensions (see the last column in Table 5), the result is a weighting scheme that is open to many questions. In turn, the SIGI's data-driven weighting scheme—which assigns higher weights to indicators that correlate more—hinges on the assumption that because a problem (for example, early marriage) does not correlate as highly as two other problems (for example, polygamy and equal rights regarding inheritance) it is less of a problem. In sum, the weighting schemes used by the five gender indices rest on a shaky foundation.<sup>34</sup>

Last of all, and most generally, the developers of the five indices do little to present ideas about alternative ways in which they might have constructed their indices and to test the robustness of their chosen rules.<sup>35</sup> On the positive side, replicability of the process followed by each index allows for independent tests.<sup>36</sup> But when we conducted a test to assess the impact of weighting decisions, the results were mixed.

We recalculated each index using three broad alternative approaches to the choice of aggregation rule (see Table 6). One alternative, which we label as the “direct and indirect weighting” option, includes aggregation rules that assign variable weights to indicators through direct means—through the differential assignment of weights—but also via indirect means—that is, as a result of the way they are grouped under each conceptual dimension. Indeed, as shown above, the varying number of indicators per conceptual component yields a framing effect, whereby the conceptual framing of an index will result in indicators with different weights even if the conceptual dimensions are weighted equally. A second alternative, which we label the “indirect weighting” option, includes aggregation rules that directly assign an equal weight to each indicator but indirectly allow differential weighting of indicators by grouping the indicators under different conceptual components. And a third option, which we label the “no weighting” option, not only directly assigns an equal weight to each indicator but also avoids any indirect weighting through the framing effect.

The comparison of the mean value of each index when calculated using different aggregation rules is instructive. On the one hand, this test shows that the weighting choice used in some indices does not make much of a difference. Removing the GDI's theory-driven weighting of indicators does not introduce much change (the mean value goes from 0.665 to 0.659). Likewise, the data-driven weighting of the GGGI and the SIGI, if somewhat questionable, is of little consequence; the GGGI's mean value drops by 0.001 when this direct weighting of indicators is removed, and doing away with the direct weighting of the SIGI's indicators makes no statistically significant difference.<sup>37</sup> On the other hand, this test shows that the GEM and the SIGI would change considerably if all

<sup>34</sup> In addition to the problems mentioned in the text, several indices do not use the same weights for some indicators in all countries. Social Watch includes a country in the GEI even though data are missing on 4 of its 10 indicators, while the World Economic Forum includes a country in the GGGI even if 2 of the 14 indicators are missing. In those cases, the indicators that are used are reweighted—the weights of the missing indicators are redistributed across the available indicators—thus introducing a problem of comparability into the data.

<sup>35</sup> We found two partial exceptions. One concerns the GDI, in that the UNDP states that tests were conducted to assess the impact of assigning different weights to the HDI's indicators (Ul Haq 2003: 128). However, no results are reported and similar tests appear not to have been conducted with the GDI indicators proper. Another exception concerns SIGI (Branisa et al. 2009: 6); but here again no results are reported.

<sup>36</sup> The replicability of the aggregation process is ensured because the aggregation procedures are explicitly articulated and the underlying data on each indicator are made public by the index producers or are 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.

<sup>37</sup> Given this result, one wonders whether it would be wiser if the data driven weightings were simply dropped by these two indices and a simpler and easier to communicate choice were adopted.

**Table 6** The impact of alternative aggregation rules II: the aggregation of the values of all indicators

| Aggregation weighting options       | 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) | OECD, Social institutions and gender index (SIGI) |
|-------------------------------------|--|--|---|--|---|
| Direct and indirect weighting rules | <b>0.665</b>                                 | ...                                    | ...                                     | <b>0.674</b>   | <b>0.136</b>                                      |
| Indirect weighting rules            | 0.659  | <b>0.582</b>                           | <b>0.611</b>                            | 0.675  | 0.142   |
| No weighting rules                  | 0.673  | 0.648                                  | 0.623                                   | 0.654  | 0.209   |

A comparison of means. The set of indicators used in each index are maintained constant; the rules for weighting indicators directly and indirectly, through the way indicators are grouped under conceptual dimensions, are those used in the actual indices. The aggregation option actually used by each index is highlighted in bold. We followed the exact procedure used in each index, except for the population weighting element in the UNDP and Social Watch indices and the rescaling of subindex scales of the SIGI. Concerning missing data, for the SIGI we followed Social Watch's rule of including cases that had at least 6 of its 10 indicators; for the GGGI we adopted a laxer standard than the World Economic Forum, dropping a country only if it lacked one indicator for the Health and Survival conceptual dimension or two indicators for the other conceptual dimensions. The scales are 0.000–1.000 scales with higher values being indicative of a more positive situation (to facilitate the comparison, the SIGI scale was inverted). All differences of means are significant at a 0.00 level, except for the difference between the SIGI's actual aggregation rule and the indirect weighting rules option (which is significant at the 0.03 level) and the difference between the GGGI's actual aggregation rule and the indirect weighting rules option (which is not significant). The N for the GDI data is 120 countries, for the GEI 165, for the GEM 95, for the GGGI 116, and for the SIGI 71. Sources: Authors' calculations based on data on indicators, c. 2005; for more, see the data "Appendix"

their indicators were weighted equally. The conceptual framing used by both these indices, as with the other indices as well, relies on standard and defensible distinctions. Nonetheless, the greater importance of framing choices in the GEM and the SIGI relative to the other indices is a weakness.

In short, index developers have a responsibility to offer a justification for their choice of aggregation rule and also to consider alternative-but-feasible ways of constructing an index and to present the results of tests that assess the impact of these differences. Nonetheless, the developers of the reviewed indices have failed to shoulder the burden of performing sensitivity analyses and hence to consider evidence that could either strengthen or weaken the case for the validity of the proposed indices.

## 5 Overview and Implications

Measures of gender inequality, and gender indices in particular, are tools for raising awareness and they have become important in policy-agenda setting and decision-making. This is a welcome development. But it has also raised the stakes of measurement exercises and given urgency to doubts about the quality of gender measures. To respond to this need for a responsible use of data in discussions about gender matters, this article has provided a systematic and comprehensive assessment of the most visible gender indices. As shown, the methodological choices that go into the making of these indices are complex; therefore,

any quick summary assessment of these indices risks oversimplifying matters. But the complexity of the issues addressed in this article also makes an overview all the more necessary and valuable. Thus, by way of a conclusion, we first provide an overall assessment of the gender indices we have scrutinized, addressing the two questions posed in the introduction: What are the five reviewed indices a measure of? And, how valid are these indices? Thereafter, we draw some lessons from our assessment in the form of advice for users and producers of gender indices.

### 5.1 What Do These Indices Measure?

The five reviewed gender indices differ in their meanings, which are not always evident from the index's name or even the description offered by the index developers. The UNDP's GDI is the best-explained and most-easily understood index because of its focus on a well-theorized concept—development—and the wide dissemination of the human development index (HDI). As the UNDP states in the report that introduced the GDI, this index “measures the same basic capabilities as the HDI 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: 73). Less self evidently, to correctly interpret the GDI it is also important to recall that it is a general measure of gender inequality, which considers inequalities favoring men *or* women.<sup>38</sup> But the GDI stands out as a model of clarity and serves as reference point in an effort to understand the meanings of the other four indices.

The GEM has not been as well explained as the GDI. The explanation that the GEM focuses “on women's opportunities rather than their capabilities” (UNDP 2007: 360), and that it serves therefore as a supplement to the GDI, is not very helpful. After all, the GEM includes indicators that address results more than opportunities and, furthermore, it includes one of the same indicators used in the GDI. Explanations notwithstanding (UNDP 1995: 82), this contrast between opportunities and capabilities is somewhat puzzling. Moreover, since the GEM relies on the same general concept of gender equality as the GDI, the emphasis on women is misleading. Thus, the most obvious point to emphasize in interpreting the GEM is that it measures dimensions of the concept of human development related to economic and political power that were not included in the GDI and approaches these dimensions of human development from a gendered perspective.

The developers of the GEI and the GGGI do not explicitly state what concepts they intend to measure. That is, though these indices are advertised as measures of gender equality, their overarching concepts are not clearly formulated. The best one can do is to consider the conceptual dimensions of these two indices, infer that these conceptual dimensions are related to human development, as defined by the UNDP, and interpret the indices as measures of development that subsume (especially in the case of the GGGI) the indicators used in both the GDI and the GEM. But two key differences between the GEI and GGGI on the one hand, and the GDI on the other, need to be emphasized. First, as the creators of the GEI and GGGI clearly state (Social Watch 2005: 73; Hausmann et al. 2007a: 3), these indices are measures of the situation of women relative to men that do not take into consideration—in contrast to the GDI—the aggregate absolute situations of women and men. Second, the GEI and GGGI are measures of inequalities favoring men that—in contrast to the GDI—

<sup>38</sup> 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.

disregard inequalities that favor women.<sup>39</sup> Thus, in interpreting the GEI and GGGI, it is useful to see them as rather similar indices that are broader measures than the GDI, in that they cover more conceptual dimensions, but also narrower measures than the GDI, in that they focus only on relative positions and only on female disadvantages relative to men.

Finally, interpreting the OECD's SIGI presents a set of distinct problems. The SIGI intends to provide a measure that goes beyond the other indices and the overarching concept of social institutions is elaborated sufficiently to make it clear how SIGI relies on indicators that supplement those used by the other indexes (Jütting et al. 2008: 78). Nonetheless, some ambiguities cloud the interpretation of the SIGI. As with the GEI and GGGI, the SIGI does not address inequalities favoring women. But the SIGI includes relational measures based on an implicit comparison of the rights of women relative to men, and absolute measures of the rights of women that have no obvious male counterpart. And this lack of symmetry in the SIGI's indicators and indicator scales makes it unclear what is being measured. The SIGI can be seen, somewhat loosely, as a combined measure of female disadvantages relative to men concerning some basic rights and the fulfillment of rights distinctive to women.

## 5.2 How Valid Are These Indices?

Turning to the question of how well the five gender indices measure what they measure, some methodological strengths and weaknesses can be identified. Taken as a group, strengths include a reliance on well-established distinctions to differentiate among 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 7). Common weaknesses include an inadequate theorization of the index's overarching concept, the choice of indicators that do not cover the full meaning of the concept being measured or include extraneous or redundant indicators, the lack of justification for various choice made in the aggregation process, and the lack of sensitivity analyses that are essential to validation of measures. The important advances made in developing gender indices notwithstanding, the failure to meet some standard and basic methodological criteria cast doubts on the quality of these indices.

Beyond these shared features, the five indices differ in the extent to which they meet the criteria we highlight in our methodological framework. The UNDP's pioneering GDI stands out as the soundest index, its key problem being the lack of a clear differentiation with respect to the HDI. At the other extreme, the OECD's SIGI is a case of a measure that holds much promise but has been poorly executed. Finally, the other three indices fall in between these extremes, the lack of robustness of the UNDP's GEM aggregate data (see Tables 4 and 6) pointing to a relative advantage of the Social Watch's GEI and World Economic Forum's GGGI.

## 5.3 Implications for Data Users and Producers

The valuable contributions made by the developers of the five reviewed gender indices should be duly recognized. These indices constitute a resource for the study of

<sup>39</sup> Though Social Watch (2005: 77) suggests that the GEI 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. 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 et al. 2007a: 3).

**Table 7** Indices with gender-differentiated data IV: methodological strengths and weaknesses

| Index  | Assessment | Measuring instrument aspect  |   |   | From raw data on indicators to data on an index   |
|--|------------|--|---|---|---|
|  |            | Conceptual dimensions  | Indicators  | Indicator value assignment                        |   |
| UNDP, Gender-related development index (GDI) | Strength   | A well theorized concept<br>The introduction of well-established conceptual dimensions | Use of (relatively) reliable indicators   | Established and refined method of generating data | Well justified normalization of scales<br>Overall cogent justification of aggregation procedure<br>Replicability of the aggregation process |
|  | Weakness   | Dropping of conceptual dimensions due to a lack of available data                      | Narrow set of indicators  |   | Somewhat arbitrary weighting choices<br>Lack of sensitivity analyses  |
| UNDP, Gender empowerment measure (GEM)       | Strength   | The introduction of well-established conceptual dimensions                             | Use of (relatively) reliable indicators   | Established and refined method of generating data | Well justified normalization of scales<br>Replicability of the aggregation process  |
|  | Weakness   | The undertheorization of the overarching concept of empowerment                        | Selection of a (partly) extraneous indicator<br>Selection of a (partly) redundant indicator |   | Partial justification of aggregation process<br>Lack of robustness of the aggregate data<br>Lack of sensitivity analyses                    |

**Table 7** continued

| Index  | Assessment | Measuring instrument aspect   |  |   | From raw data on indicators to data on an index   |
|--|------------|---|--|---|---|
|  |            | Conceptual dimensions   | Indicators   | Indicator value assignment  |   |
| Social Watch, Gender equity index (GEI)              | Strength   | The introduction of well-established conceptual dimensions  | Use of (relatively) reliable indicators  | Established and refined method of generating data                       | Robustness of the aggregate data<br>Replicability of the aggregation process<br>Partial justification of aggregation process<br>Ad hoc reweighting compromises<br>comparability of the index<br>Lack of sensitivity analyses    |
|  | Weakness   | Relatively broad set of indicators<br>The lack of specification and theorization of the index's overarching concept | Selection of a (partly) redundant indicator  |   |   |
| World Economic Forum, Global gender gap index (GGGI) | Strength   | The introduction of well-established conceptual dimensions<br>Relatively broad set of indicators                    | Use of (relatively) reliable indicators  | No information on the validity and reliability of data of one indicator | Robustness of the aggregate data<br>Replicability of the aggregation process<br>Disputable justification of aggregation process<br>Ad hoc reweighting compromises<br>comparability of the index<br>Lack of sensitivity analyses |
|  | Weakness   | The lack of specification and theorization of the index's overarching concept                                       | Selection of a (partly) extraneous indicators<br>Selection of a (partly) redundant indicator |   |   |

**Table 7** continued

| Index   | Assessment | Measuring instrument aspect  | From raw data on indicators to raw data on indicators  | From raw data on indicators to data on an index  |
|---|------------|--|--|--|
|   |            | Conceptual dimensions  | Indicators   | Indicator scales   |
|   |            |  | Indicators   | Indicator value assignment   |
| OECD, Social institutions and gender index (SIGI) | Strength   | A moderately theorized concept<br>The introduction of well-established conceptual dimensions | Inclusion of not-readily-available indicators on key spheres of social life  |  |
|   | Weakness   |  | Poorly designed indicator scales: gender-disaggregated data are not systematically collected<br>Some vaguely formulated scales | No information on the validity and reliability of data on most indicators<br><br>Partial justification of aggregation process<br>Lack of robustness of the aggregate data<br>Only partial replicability of the aggregation process<br>Lack of sensitivity analyses |



development and a range of rights, which has allowed analysts to portray gender disparities and the rights of women in a more comprehensive and integrated manner than was possible prior to 1995. Having an index, especially one of global empirical scope, is better than having none, even if the index suffers from methodological weaknesses. But the availability of multiple indices—which could support different conclusions—also creates a problem for users, who now must offer a rationale for opting for one index over the others.

This is not an easy choice, but it is becoming an increasingly pressing choice. After all, confusion about the interpretation of gender indices has been common and has led to some resistance to embrace the use of gender indices (Shuller 2006: 173–76). Thus, in addition to standard concerns about measurement error and the appropriate uses of data,<sup>40</sup> data users should be cautious when interpreting these gender indices, should be cognizant that grasping the meaning of these indices is a more complicated matter than may seem at first sight, and recognize, as stressed in this article, that, as they measure different concepts, they are not interchangeable.

The shortcomings identified in this article also highlight the need to see the production of measures that incorporate a gender perspective as an evolving research agenda. This article has two central implications for this agenda. First, we emphasized the value of an explicit methodological framework in work on measurement. The value of such a framework has been recognized in discussions of measures of development (Booyesen 2002: 116–17). But current discussions of measures with gender-differentiated data do not rely on such a framework or focus only on some sources of measurement error. Yet, as we have shown, the meaning and validity of an index can be ascertained only if consideration is given to the full range of methodological choices highlighted in our framework. Thus, future research on data that incorporate a gender perspective should build more explicitly on a broad and integrated methodological framework.

Second, we showed the value of critical-but-constructive evaluations of existing measures to the challenge of developing better gender indices. In this regard, the assessment of gender indices in this article contains a host of ideas to guide future work on gender indices. Researchers frequently remain wedded to established measures and such a tendency is likely to hold regarding gender indices. Nonetheless, as we have showed, there is a need to develop better gender indices. And, by pinpointing both the valuable insights that should be incorporated in new indices and the problems that need to be rectified, we hope to have contributed to the collective effort to generate better gender measures.

**Acknowledgments** We thank Robert Harris for research assistance, and Jonathan Kulick and Svend-Erik Skaaning for comments.

## Appendix

See Table 8.

---

<sup>40</sup> Going beyond the methodological issues we addressed in this article, 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 et al. 2007a: 3, Social Watch 2005: 73, 2008a: 68–85). We would suggest that the data we currently have do not allow us to talk with confidence about trends over time.

**Table 8**

| Indicators and indices   | Data source          | Data location   |
|--|----------------------|---|
| UNDP, Gender-related development index (GDI)                                 | UNDP                 | UNDP, <i>Human Development Report 2007/2008. Fighting Climate Change: Human Solidarity in a Divided World</i> (New York: UNDP, 2007)    |
| UNDP, Gender empowerment measure (GEM)                                       | UNDP                 | UNDP, <i>Human Development Report 2007/2008. Fighting Climate Change: Human Solidarity in a Divided World</i> (New York: UNDP, 2007)    |
| Social Watch, Gender equity index (GEI)                                      | Social Watch         | Social Watch. <i>Social Watch Annual Report 2008</i> (Montevideo: Instituto del Tercer Mundo, 2008)                                     |
| World Economic Forum, Global gender gap index (GGGI)                         | World Economic Forum | Ricardo Hausmann, Laura D. Tyson, and Saadia Zahidi, <i>The Global Gender Gap Report 2007</i> (Switzerland: World Economic Forum, 2007) |
| OECD, Social institutions and gender index (SIGI)                            | OECD                 | <a href="http://genderindex.org/ranking/all">http://genderindex.org/ranking/all</a>   |
| Life expectancy (at birth)   | WHO                  | <a href="http://www.who.int/whosis/en/index.html">http://www.who.int/whosis/en/index.html</a>   |
| Healthy life expectancy  | WHO                  | <a href="http://www.who.int/whosis/en/index.html">http://www.who.int/whosis/en/index.html</a>   |
| Number of births (sex ratio)   | CIA                  | CIA, <i>The World Factbook 2006</i> (Washington, DC: Central Intelligence Agency, 2006).  |
| Literacy rate (for adults)   | UNDP                 | <a href="http://hdr.undp.org/en/statistics/">http://hdr.undp.org/en/statistics/</a>   |
| Literacy rate (for youths)   | UNESCO               | UNESCO, <i>EFA Global Monitoring Report 2006</i> (Paris: UNESCO, 2005)  |
| Primary level enrolment (net)  | UN                   | UN stats division, statistics and indicators on women and men   |
| Secondary level enrolment (net)  | UN                   | UN stats division, statistics and indicators on women and men   |
| Tertiary level enrolment (gross)   | UN                   | UN stats division, statistics and indicators on women and men   |
| Combined gross enrolment ratio for primary, secondary and tertiary education | UNDP                 | <a href="http://hdr.undp.org/en/statistics/">http://hdr.undp.org/en/statistics/</a>   |
| Earned income (PPP US\$)   | UNDP                 | <a href="http://hdr.undp.org/en/statistics/">http://hdr.undp.org/en/statistics/</a>   |
| Professional and technical positions   | UNDP                 | <a href="http://hdr.undp.org/en/statistics/">http://hdr.undp.org/en/statistics/</a>   |

**Table 8** continued

| Indicators and indices                                  | Data source          | Data location   |
|---|----------------------|---|
| Positions as legislators, senior officials and managers | UNDP                 | <a href="http://hdr.undp.org/en/statistics/">http://hdr.undp.org/en/statistics/</a>   |
| Labor force participation                               | UN                   | UN stats division, statistics and indicators on women and men   |
| Wage equality for similar work                          | World Economic Forum | Ricardo Hausmann, Laura D. Tyson, and Saadia Zahidi, <i>The Global Gender Gap Report 2006</i> (Switzerland: World Economic Forum, 2006) |
| Parliamentary seats                                     | UNDP                 | <a href="http://hdr.undp.org/en/statistics">http://hdr.undp.org/en/statistics</a>   |
| Ministerial level positions                             | UNDP                 | <a href="http://hdr.undp.org/en/statistics">http://hdr.undp.org/en/statistics</a>   |
| Heads of state  | World Economic Forum | Ricardo Hausmann, Laura D. Tyson, and Saadia Zahidi, <i>The Global Gender Gap Report 2006</i> (Switzerland: World Economic Forum, 2006) |

The data on all the indicators are for c. 2005

## References

- Abdullah, R. (2000). *A framework on indicators for action on women's health needs and rights after Beijing*. Kuala Lumpur: Asian-Pacific Resource and Research Centre for Women Malaysia.
- Anand, S., & Sen, A. (2003a). Human development index: Methodology and measurement. In S. Fukuda-Parr & A. K. S. Kumar (Eds.), *Readings in human development: Concepts, measures, and policies for a development* (pp. 138–51). Oxford: Oxford University Press. [Originally published as Human Development Report Office *Occasional Paper* No 12 (New York: UNDP, 1994)].
- Anand, S., & Sen, A. (2003b). Gender inequality in human development: Theories and measurement. In S. Fukuda-Parr & A. K. S. Kumar (Eds.), *Readings in human development: Concepts, measures, and policies for a development* (pp. 210–227). Oxford: Oxford University Press. [Originally published as Human Development Report Office *Occasional Paper* No 19. New York: UNDP, 1995].
- Babbie, E. (1995). *The practice of social research*. Belmont, CA: Wadsworth Publishing Company.
- Bardhan, K., & Klasen, S. (1999). UNDP's gender-related indices: A critical review. *World Development*, 27(6), 985–1010.
- Booyens, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research*, 59(2), 115–151.
- Branisa, B., Klasen, S., & Ziegler, M. (2009). Background paper: The construction of the social institutions and gender index (SIGI). University of Goettingen, Germany, March 6, 2009. Available at: <http://www.oecd.org/dataoecd/49/19/42295804.pdf>.
- Breitenbach, E., & Galligan, Y. (2006). Measuring gender equality: Reflecting on experiences and challenges in the UK and Ireland. *Policy and Politics*, 34(4), 597–614.
- Chant, S. (2006). Re-thinking the 'Feminization of Poverty' in relation to aggregate gender indices. *Journal of Human Development*, 7(2), 201–220.
- Charmes, J., & Wieringa, S. (2003). Measuring women's empowerment: An assessment of the gender-related development index and the gender empowerment measure. *Journal of Human Development*, 4(3), 419–435.
- Dijkstra, A. G. (2002). Revisiting UNDP's GDI and GEM: Towards an alternative. *Social Indicators Research*, 57(3), 301–338.
- Dijkstra, A. G. (2006). Towards a fresh start in measuring gender equality: A contribution to the debate. *Journal of Human Development*, 7(2), 275–283.
- Dijkstra, A. G., & Hanmer, L. (2000). Measuring socio-economic inequality: Towards an alternative to the UNDP gender-related development index. *Feminist Economics*, 6(2), 41–75.
- Forsythe, N., Korzeniewick, R. P., & Durrant, V. (1998). Gender inequalities, economic growth and structural adjustment: A longitudinal evaluation. In Paper presented to XXI conference of the Latin American studies association (LASA), Washington, September 24–26, 1998.
- Fukuda-Parr, S. (2003). Rescuing the human development concept from the HDI: Reflections on a New Agenda. In S. Fukuda-Parr & A. K. S. Kumar (Eds.), *Readings in human development: Concepts, measures, and policies for a development* (pp. 117–124). Oxford: Oxford University Press.
- Hausmann, R., Tyson, L. D., & Zahidi, S. (2006). *The global gender gap report 2006*. Switzerland: World Economic Forum.
- Hausmann, R., Tyson, L. D., & Zahidi, S. (2007a). The global gender gap index 2007. In R. Hausmann, L. D. Tyson, & S. Zahidi (Eds.), *The global gender gap report 2007* (pp. 3–27). Switzerland: World Economic Forum.
- Hausmann, R., Tyson, L. D., & Zahidi, S. (2007b). *The global gender gap report 2007*. Switzerland: World Economic Forum.
- Hausmann, R., Tyson, L. D., & Zahidi, S. (2008). *The global gender gap report 2008*. Switzerland: World Economic Forum.
- Jahan, S. (2003). Evolution of the human development index. In S. Fukuda-Parr & A. K. S. Kumar (Eds.), *Readings in human development: Concepts, measures, and policies for a development* (pp. 152–163). Oxford: Oxford University Press.
- Jütting, J. P., Morrisson, C., Dayton-Johnson, J., & Drechsler, D. (2006). *Measuring gender (In)Equality: Introducing the gender, institutions and development data base (GID)* (Paris: OECD Development Centre Working Papers 247, OECD Development Centre).
- Jütting, J. P., Morrisson, C., Dayton-Johnson, J., & Drechsler, D. (2008). Measuring gender (In)Equality: The OECD gender, institutions and development data base. *Journal of Human Development*, 9(1), 65–86.
- Kelly, L., Kennedy, L., & Horvath, M. (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.
- Klasen, S. (2006a). UNDP's gender-related measures: Some conceptual problems and possible solutions. *Journal of Human Development*, 7(2), 243–274.
- Klasen, S. (2006b). Gender-related indicators of well-being. In M. McGillivray & A. Shorrocks (Eds.), *Measuring human well-being* (pp. 167–192). London: Palgrave-Macmillan.
- Klasen, S. (Ed.). (2006c). Special issue: Revisiting the gender-related development index (GDI) and gender empowerment measure (GEM). *Journal of Human Development* 7(2).
- Morrisson, C., & Jütting, J. P. (2005). Women's discrimination in developing countries: A new data set for better policies. *World Development*, 33(7), 1065–1081.
- Nussbaum, M. C. (2000). *Women and human development*. Cambridge: Cambridge University Press.
- OECD Development Centre. (2009a). Country profiles: Gender equality and social indicators. Available at: <http://genderindex.org/countries>.
- OECD Development Centre. (2009b). Coding of the social institutions variables. Available at: [http://genderindex.org/sites/default/files/GID\\_variables.pdf](http://genderindex.org/sites/default/files/GID_variables.pdf).
- Permanyer, I. (2008). On the measurement of gender equality and gender-related development levels. *Journal of Human Development*, 9(1), 87–108.
- Permanyer, I. (2010). The measurement of multidimensional gender inequality: Continuing the debate. *Social Indicators Research*, 95(2), 181–198.
- Plantenga, J., Figueiredo, H., Remery, C., & Smith, M. (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, M., & Schwab, K. (2008). Executive opinion survey 2008. Available at: <http://www.kadin-indonesia.or.id/enm/images/dokumen/KADIN-117-2573-29022008.pdf>.
- Sen, A. (1999). *Development as freedom*. Oxford: Oxford University Press.
- Shuller, D. (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), 161–181.
- Social Watch. (2004). *Social Watch annual report 2004*. Montevideo: Instituto del Tercer Mundo.
- Social Watch. (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* (pp. 73–77). Montevideo: Instituto del Tercer Mundo.
- Social Watch. (2008a). *Social Watch annual report 2008*. Montevideo: Instituto del Tercer Mundo.
- Social Watch. (2008b). *Technical notes: The construction of the GEI*. Available at: [http://www.socialwatch.org/en/avancesyRetrocesos/IEG\\_2008/tablas/technicalNotes.html](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.
- Ul Haq, M. (2003). The birth of the human development index. In S. Fukuda-Parr, & A. K. S. Kumar (Eds.), *Readings in human development: Concepts, measures, and policies for a development* (pp. 127–37). Oxford: Oxford University Press.
- UNDP. (1990). Defining and measuring human development. In UNDP, *Human Development Report 1990* (pp. 9–16). New York: Oxford University Press.
- UNDP. (1995). Measuring gender inequality. In UNDP, *Human development report 1995. Gender and human development* (pp. 72–86). New York: UNDP and Oxford University Press.
- UNDP. (2006). *Measuring democratic governance: A framework for selecting pro-poor and gender sensitive indicators*. Oslo: UNDP Oslo Governance Centre.
- UNDP. (2007). Calculating the human development indices. In UNDP, *Human Development Report 2007/2008. Fighting climate change: human solidarity in a divided world* (pp. 355–61). New York: UNDP.
- UNECA (United Nations Economic Commission for Africa). (2004). *The African gender and development index*. Addis Ababa: 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).
- UNECLAC (Economic Commission for Latin America and the Caribbean). (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. In UNIFEM, *Progress of the world's women 2002 Volume 2: Gender equality and the millennium development goals* (pp. 55–61). New York: UNIFEM.
- UNRISD (United Nations Research Institute for Social Development). (2005). *Gender equality: Striving for justice in an unequal world*. Geneva: UNRISD.
- Unterhalter, E. (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, S. (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, H. (1997). *Patterns of gender discrimination: an examination of the UNDP's gender development index*. The Hague: Institute of Social Studies.
- Yinger, N., et al. (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.