IARIW-Bank of Korea Conference "Beyond GDP: Experiences and Challenges in the Measurement of Economic Well-being," Seoul, Korea, April 26-28, 2017

Eliciting, applying and exploring multidimensional welfare weights: evidence from the field

Lucio Esposito (University of East Anglia, United Kingdom) and Enrica Chiappero-Martinetti (University of Pavia, Italy)

Paper prepared for the IARIW-Bank of Korea Conference

Seoul, Korea, April 26-28, 2017

Session 2A: Multidimensional Well-being

Time: Wednesday, April 26, 2017 [Morning]

Eliciting, applying and exploring multidimensional welfare weights: evidence from the field

Lucio Esposito[♣] and Enrica Chiappero-Martinetti[♦]

University of East Anglia and University of Pavia

Abstract

By combining primary data on dimension importance collected in the field from three different samples (a total of 1,402 subjects) and nationally representative survey data, we offer a twofold contribution. The first one comes from an unincentivised questionnaire experiment, where the significance of the treatment effect shows that life domains are valued differently in a poverty vs a wellbeing frameworks. This opens the door to what we call a 'concordance paradox', and poses important questions not only on the anatomy of dimension importance but also on the essence of the notions of poverty and wellbeing. Our second offer relates to the so-called 'weight or not to weight' debate – i.e. to the issue of whether alternative sets of weights lead to qualitative differences in empirical analyses. On the basis of the sets of weights we derive in the field (from a student sample, a 'development experts' sample a more heterogeneous sample), we find that they do. Depending on which set of dimension importance scores is employed, opposite conclusions are reached on the trend of multidimensional poverty and wellbeing in the Dominican Republic.

1. Introduction

Researchers from a variety of disciplines in the social and medical sciences are increasingly interested in the multidimensional evaluation of human achievements or deprivations, the underlying phenomenon of interest being poverty, wellbeing, capabilities, quality of life, health, literacy, etc. – see Esposito, Kebede and Maddox (2011), Massey et al. (2013), Hick (2014), Alkire et al. (2015), Donohue and Biggs (2015), Feeny and McDonald (2016) and Schang et al (2016). The array of aspects of human life being taken into examination is extremely wide; for example, the interdisciplinary review by Linton et al. (2016), which focuses on the concept of wellbeing and does not cover age-specific or condition-specific measures, identifies as many as 196 dimensions being used in the literature.

The reason for the growing interest in a multidimensional approach is that monetary metrics may fail to offer an accurate picture of people's concrete living conditions. Monetary metrics do quantify an important means for human flourishing, but the extent to which this means does translate into

[♣] University of East Anglia and IUSS Pavia. Corresponding author, lucio.esposito@uea.ac.uk.
[♦] University of Pavia and IUSS Pavia.

ends depends on a number of variables including individual conversion factors, the existence of efficient markets and the public provision of goods and services (e.g. health or education). While monetary and non-monetary indicators can be closely interrelated (Merz and Rathjen 2014 and Callander and Schofield 2015), the literature has documented that the latter are often able to unveil aspects of deprivation which the former neglects (Alkire and Santos 2014 and Trani et al. 2016). Multidimensional evaluation also proved useful to highlight gender-based or caste-based deprivation inequalities (Alkire and Seth 2015 and Rogan 2016), to illustrate social consequences of economic crises (Stoeffler et al 2016) or to identify the poorest of the poor in global as well as very localised contexts (Alkire et al 2015 and Vasquez, Cabieses and Tunstall 2016). In addition, multidimensional evaluation has not been confined to developing contexts and is increasingly employed for analysis of high-income countries as well – see Coromaldi and Zoli (2012), Nowak and Scheicher (2014), Wagle (2014), Betti et al (2015) and Hick (2016). Finally, multidimensional indicators have been argued to be able to provide valuable information for policy (Victor et al 2014 and Angulo, Díaz and Pardo 2016), although this view has been questioned on pragmatic and ontological grounds – see Ravallion (2011) and Michener (2015).

There is no such thing as a free lunch though, and the additional informational content provided by multidimensional measurement comes with increased technical complexity and possibly greater scope for arbitrariness – with regard to, for example, desirable functional forms, aggregation procedures, the choice of the relevant dimensions and of their relative importance, etc. In the past decade, a number of contributions have significantly increased our command over the technical difficulties behind a multidimensional approach to poverty and wellbeing measurement.[1] While this body of work has brought us a long way from the initial contributions of Morris (1979), Atkinson and Bourguignon (1982), UNDP (1990) and Dasgupta and Weale (1992), the field of multidimensional evaluation still presents a number of challenges and hosts heated debates – e.g. the 'single index approach' vs 'dashboard approach' debate, see Alkire and Foster (2011b), Ferreira (2011), Ravallion (2011) and Ferreira and Lugo (2013).

In this paper we take the literature on multidimensional evaluation forward with respect to the issue of dimension importance. In particular, we offer a twofold contribution by combining nationally representative survey data from the Dominican Republic and primary data on dimension importance personally collected in the field by one of the authors – the primary data amounting to 1,402 observations and comprising a student sample, a sample of local 'development experts' and a sample of respondents who are more heterogeneous in terms of socio-economic characteristics. Our first offer stems from the following consideration. While it often occurs that a certain dimension (e.g. education) features in the measurement of different constructs (e.g. 'poverty', 'wellbeing', 'development', etc.), there is no evidence as to whether the public would attach different importance to the dimension depending on which construct it refers to – i.e. depending on whether it is intended 'as a dimension of poverty' or 'as a dimension of wellbeing'. We tackle this issue by running a questionnaire experiment with 1,083 university students at the national university in the capital city. Random allocation of a 'poverty' and a 'wellbeing' questionnaire versions does produce a significant difference in the importance attached to the dimensions we consider in our study (education, health, housing and personal safety). Whilst this result may seem *prima facie* innocuous, it raises what we call a 'concordance paradox' and it bears substantial implications for the conceptualisation of the notions of poverty and wellbeing, which we discuss in the paper. The second offer of our paper relates to the 'weight or not to weight' debate, that is, the issue of whether

[1] See, inter alia, Tsui (2002), Bourguignon and Chakravarty (2003), Duclos, Sahn and Younger (2006), Kakwani and Silber (2008), Chakravarty, Deutsch and Silber (2008), Alkire and Santos (2010) and Alkire and Foster (2011a,b), Belhadj and Limam (2012), Pattanaik, Reddy and Xu (2012), Ravallion (2012), Bossert, Chakravarty and D'Ambrosio (2013), Decancq and Lugo (2013), Seth (2013), Permanyer (2014), Yalonetzky (2014) and Maasoumi and Racine (2016).

adopting different weighting schemes produces appreciable differences in empirical analyses. We estimate multidimensional poverty and wellbeing in the Dominican Republic using national household surveys from 1997 and 2007, and employing equal weights as well as the sets of weights elicited from our different samples (i.e. the student, 'expert' and heterogeneous samples). Our results show that picking a set of weights or another is not a trivial choice, because different weighting schemes lead to opposite conclusions on the change in multidimensional poverty and wellbeing.

The paper is structured in the following way. Section 2 reviews the literature on multidimensional weights and on the main approaches to the derivation of dimension importance scores, with a focus on what we call direct approaches – those where importance scores originate from explicit questions posed to the respondent about the value of the chosen dimensions. This section provides a framework to introduce the methodology we used to derive dimension importance scores in the field, namely the Budget Allocation Technique. Section 3 presents the primary data collection strategy for each of the three samples and briefly describes the secondary data used in the assessment of multidimensional poverty and wellbeing. Sections 4 and 5 present the first and the second sets of results, respectively. Section 6 concludes.

2. Setting domains importance: direct approaches and the Budget Allocation Technique

2.a Adopting dimension weights in multidimensional analyses

The issue of heterogeneity in dimension importance in multidimensional analyses has been addressed since the work of Campbell et al. (1976) and Inglehart (1978). Interestingly, the issue was raised also by Rawls (1971), who in his influential Theory of Justice notes that the selection of an appropriate wellbeing index is faced with the choice of the relative weights to be attached to life domains. The idea that more important dimensions should play a larger role in a composite index of individual achievements or deprivations has a straightforward conceptual appeal and has long been advocated by a number of scholars – e.g. Ferrans and Powers (1985), Mayer and Jencks (1989) and Sen (1992). The central point is that if an individual or a society attaches little importance to a life domain then attainments in that domain should be somehow deflated vis-a`-vis those in highly valued domains.

The introduction of weighting schemes in multidimensional evaluation, however, brings about operational as well as conceptual issues. Dimension importance scores can be attached different meanings (e.g. substitution rates, relative contribution to overall value, scaling factors, discriminating power, etc.) and this can affect the weighting system's operational effectiveness, dependence on measurement units, as well as their suitability to a certain aggregation strategy – see Crawford and Williams (1985), Schenkerman (1991) and Choo, Schoner and Wedley (1999). Trauer and Mackinnon (2001) criticise the use of dimension importance scores on the grounds that they may bring about interpretational difficulties and undesirable psychometric properties, and that weighted indices provide little gain in empirical exercises. Stapleton and Garrod (2007) suggest that if the added value of using unequal weights is negligible (i.e. it brings about little difference in empirical assessment), then the use of indices based on equal weights should be preferred on the basis of Occam's Razor – a principle which rejects unnecessary complexity in the name of parsimony.

The quality of life literature has focussed on the question of whether indicators based on equal or unequal weights better predict outcome variables of interest, with the dilemma of whether 'to weight or not to weight' not being set yet – see, among others, Russell et al (2006), Philip et al (2009), Wu, Yang and Huang (2014) and Hsieh (2016). The economics literature has been more

interested in the robustness or sensitivity of distribution rankings to the use of alternative weighting schemes – Saisana, Saltelli and Tarantola (2005) Cherchye, Ooghe and Puyenbroeck (2008), Permanyer (2011, 2012), Foster, McGillivray and Seth (2013), Athanassoglou (2015) and Zheng and Zheng (2015). As Seth and McGillivray (2016) point out, these studies explore robustness to all possible sets of weights rather to concentrating on a more restricted set of weights selected according to criteria deemed desirable. Examples of papers which carried out multidimensional poverty analyses on the basis of selected weights, derived from preferences expressed by the population under study, are Esposito and Chiappero-Martinetti (2010), Chiappero-Martinetti and von Jacobi (2012), Decancq, Van Ootegem and Verhofstadt (2013) and Mitra et al. (2013).

There is a plethora of approaches to setting weights. Decancq and Lugo (2013) carefully review the literature on the derivation of weights in multidimensional evaluation and develop a useful taxonomy. They categorise approaches to deriving weights in three classes (data-driven, normative and hybrid weights), each in turn divided into subclasses. *Data-driven* weights depend on the actual distribution of achievements in society and do not rely on value judgements on the perceived importance of the dimensions. Conversely, *normative* weights are solely based on value judgements and are independent of how well society is doing in the domains of interest. *Hybrid* weights result from a combination of the two sources of information. In Table 1 below, we rearrange Decanc and Lugo's subclasses into the two broader categories of *direct* and *indirect* weights.

Table 1. Direct and indirect approaches to deriving weights

Direct	*Indirect*
Equal/arbitrary[a]	Frequency[c]
Expert opinion[a]	Statistical[c]
Self-stated[b]	Most favourable[c]
	Price-based[a]
	Hedonic[b]

The classification according to Decancq and Lugo (2013) is as follows:
[a] Normative – based on value judgements
[b] Hybrid – based on both value judgements and actual achievements
[c] Data driven – based on actual achievements

What distinguishes *direct* from *indirect* approaches is whether dimension importance scores are expressed through a direct judgement on dimensions' value (whoever makes this judgement) or they are indirectly inferred by the researcher – regardless of whether objective variables (e.g. achievements), or subjective variables (e.g. happiness) are used for this purpose. *Indirect* approaches are not only those defined as 'data-driven' by Decancq and Lugo (2013), but also the subcategories they label as 'hedonic' and 'price based'. In hedonic approaches weights are based on coefficients resulting from econometric models where achievement in the different domains are explanatory variables for self-reported happiness/life satisfaction; price-based approaches consider prices as the base to build a weighting system in that these would reflect revealed preferences in society – on this see also Ferreira and Lugo (2013). Looking at Table 1, *direct* approaches comprise equal/arbitrary, expert opinion and self-stated. These three subgroups differ in the source of the value judgements on dimension importance – respectively, the researcher(s) carrying out the analysis, a set of experts and a sample of the population. In other words, the difference relies on *whose* opinions should be taken into account for the derivation dimension importance scores. Given our interest in eliciting dimension importance scores directly from our respondents, in the following subsection, we focus on the most common methodologies for the direct derivation of dimension importance scores.

2.b Direct approaches to eliciting weights and the case for the Budget-Allocation Technique

With respect to the question of *how* to elicit dimension importance within the 'direct' category, a variety of methodologies have been adopted. In what follows, we present the most widely used among them. We make the case for the desirability of the so-called Budget Allocation Technique, in particular in cases where respondents have low levels of formal education.

Ordered Scale Valuation – Likert Scales. Respondents provide dimension importance scores by rating the dimensions along a numerical or a verbally described scale (i.e. from '1 to 10' or from 'not at all important' to 'extremely important').[2] Although this methodology is widely used, it is found unsatisfactory for a number of reasons. Firstly, it is well known that scores chosen by respondents on ordinal scales are affected by individual- or cultural-specific scale biases – see Holland and Wainer (1993) and Kahneman et al. (2004).[3] It follows that due to idiosyncratic factors some respondents may choose values lower down and others higher up the scale whist not genuinely differing in the importance attached to the dimensions. Secondly, respondents rate the importance of dimensions one after another, with the consequence that each importance score is provided in isolation with little reference to the whole picture. This appears inappropriate in multidimensional welfare evaluation, where the selected dimensions are assumed to jointly represent the phenomenon under study. In this framework, single dimension weights have little sense on their own, and should actually be seen as having a relative character. This is most evident when domain-specific indicators are combined into a multidimensional additive index: as illustrated by Decancq and Lugo (2013) dimension weights directly affect the marginal rate of substitution among dimensions, that is, the amount of achievement in a certain dimension that an individual would give up for an extra unit of achievement in another dimension. Thirdly, at a more practical level, along the exercise respondents are unlikely to be able to remember the exact score attributed to previous dimensions, with scope for inaccurate cardinal content of reported scores and false dimension rankings.

Perceived Status of Necessity. Respondents are asked to state which item/dimension represents a necessity and which does not. Larger weights are attributed to dimensions which are more widely identified as necessity. For example, Halleröd (1995, 1996) derives weights by computing the proportion of respondents regarding a certain item/dimension as a necessity. Two major drawbacks affect this procedure. First, the attribution of dimension importance scores has to rely on individual interpretations of the notion of need or necessity. Hence, interpersonal heterogeneity in value judgements on dimension importance is likely to be biased by different views of what constitutes a necessity. This concern acquires further strength in the light of stances dismissing the significance of the concept of necessity altogether (see, for example, Jensen and Meckling, 1994). Second, at an individual level, in this approach dimension evaluation is limited to a dichotomous partition into 'needs' and 'non-needs', neglecting a graduation in their importance which has been advocated by psychological literature since Maslow (1943).

Analytic Hierarchy Process. Dimension importance is gathered by presenting the respondent with a series of pair-wise comparisons. Each time the respondent first chooses which dimension is the most important and next states 'by how much' on a scale from 1 to *k*. A value of 1 is chosen if the

[2] It should be noted that the two options have different implications in data analysis; if the scale is numerical then the resulting variable can (to some extent) be considered interval (hence, for example, entered as it is as explanatory variable in econometric models); if, instead, the different scale points are described verbally then the variable will be only ordinal.

[3] A number of contributions have attempted to correct for this making use of the 'anchoring vignettes' methodology, where personal valuations are set against a standard in order to increase interpersonal comparability – see King et al. (2004), Salomon et al. (2004), Kaypten et al. (2007), Kristensen and Johansson (2008), Angelini et al. (2014) and Beegle et al. (2009).

two dimensions are deemed to be equally important, while a value of *k* if one is *k*-times as important as the other. Responses are inserted in a matrix and relative weights are computed using an eigenvector technique. This methodology has been developed by Saaty (1980, 1987) in the field of multi-attribute decision-making. The main advantage of this technique is that complex decisions/comparisons are decomposed into more easily manageable sub-problems. However, some of the drawbacks mentioned earlier apply here as well; for example, binary comparisons are carried out with little reference to the whole spectrum of dimensions to be evaluated. A limitation which is specific to the Analytic Hierarchy Process is the fact that elicited differentials in dimension importance are bound to be in the form of exact multiples (a dimension can only be deemed to have double, triple, etc. importance than another one).

Budget Allocation Technique. Respondents are invited to distribute a fixed budget of 'points' to different dimensions according to the importance attached to them, with more points allocated to the dimensions more highly valued. The adoption of this methodology enables to overcome most of the problems highlighted above. Three features of the Budget Allocation Technique emerge as particularly valuable. First, the respondent is presented at once with the whole array of dimensions to be valuated; it follows that the attribution of importance scores takes place simultaneously. Second, the amount of points to be allocated is fixed across subjects; this enables to circumvent the problem of individual scale biases.[4] Third, since a point allocated to a certain dimension implies that less points are available for the other dimensions, the Budget Allocation Technique is able to present the respondent with explicit trade-offs among dimensions; this feature appears extremely desirable when dimension importance scores are to be used in the development of an aggregate multidimensional index, where, as mentioned above, weights affect directly the marginal rate of substitution among dimensions.

3. Data

3.1 Elicitation of dimension importance scores in the field

The collection of primary data on dimension importance took place in the Dominican Republic during a three-month fieldwork, and was carried out entirely by one of the authors without the use of enumerators. The Budget Allocation Technique was adopted to elicit views on the importance of four dimensions (*education*, *health*, *housing* and *personal safety*). The choice of these dimensions was driven by the existence of secondary data at a national level which could be used to estimate multidimensional poverty and wellbeing in the country, as well as by the relevance these dimensions have in the country's public opinion and political discourse – for example, they all feature as key points in the National Development Strategy Plan of Dominican Republic 2010-2030, which represents a roadmap of the country's development priorities (see USAID, 2013). Dimension importance scores were elicited from a threefold sample – university students, local 'development experts' and a sample of respondents with highly heterogeneous socio-economic profiles. This represents a novelty, since the use of the Budget Allocation Technique has been typically restricted to the elicitation of value judgements from experts (Moldan and Billharz, 1997, and Mascherini and Hoskins, 2008).[5] A detailed description of the data collection strategy for each sample is in order.

[4] In order to correct for the problem of individual scale bias in Ordered Scale Valuation, Hsieh (2004) suggests to normalise the importance score of each dimension by the sum of the importance scores across all domains. Rather than this *ex post* solution, the Budget Allocation Technique enables to get round this problem directly at the stage of weights elicitation.

[5] A similar methodology was adopted by Esposito et al (2011, 2015) with a sample with low educational attainments. A slightly different approach was followed by Camfield and Ruta (2007).

Student sample. The idea of seeking university students' views on normative questions is a long used approach in economics (e.g. Glejser et al 1977). We approached 1,083 undergraduate students in the Universidad Autonoma de Santo Domingo (UASD), the public national university, in the main campus in the capital city. We selected the disciplines of study of our respondents in a way that would enable us to explore potential discipline-specific biases in the importance given to our four dimensions of interest. It follows that in our sample we have students from the following disciplines: Education (251 respondents), Medicine (255), Architecture (269) and Law (308). Students were administered a written questionnaire in sessions supervised by one of the authors during lecture time, typically the first or the last 20 minutes of a teaching session. The development of the questionnaire benefited from inputs offered by academics in the School of Education at UASD. In addition, the questionnaire was piloted with a sample of around 20 students, who gave useful feedback on their understanding of the questions posed and on ways of improving the wording. In the final survey, response rate was around 96% and random post-questionnaire interviews carried out with respondents reassured about the understanding of the questions posed. Views on dimension importance were elicited through the following question (our translation from the original version in Spanish):

> We would like to ask your view about the importance of the 4 dimensions mentioned above. Please assign a number from 1 to 100 to each dimension according to the importance you personally think they have, making sure that those values sum up to 100:
>
> - Education: ……..
> - Health: ……..
> - Housing: ……..
> - Personal Safety: ……..

Heterogeneous sample. This sample consists of 309 adults interviewed face to face by one of the authors. Our resources did not allow us to pursue a formal strategy to achieve national representativeness; still, in our data collection we strived to achieve substantial demographic, socio-economic and geographic heterogeneity. Interviews took place across two urban and two rural locations (the two main cities, Santo Domingo and Santiago de Los Caballeros, and two rural areas in the North and South-East of the country). Respondents aged from 18 to 79, 53% were female and number of children ranged from 0 to 13. Educational levels ranged from as little as 0 years of schooling (11 respondents) to postgraduate degrees (5 respondents), with mean and median around 10 years of schooling; respondents with a university degree were around 10%, against a national figure of about 9% (ONE, 2013). The variation in the standard of living of our respondents was also rather wide; in terms of durable goods possessed, 26 respondents owned both a computer and air conditioning while 41 owned neither a fridge nor a washing machine, and personal income ranged from 900 to 70,000 Dominican Pesos per month – the national absolute poverty line being 2,601.75 Dominican Pesos (BCRD, 2011).

As was the case for the student sample, the development of the interview schedule and of the flashcard used for the valuation exercise benefited from piloting and from inputs provided by academics in the School of Education at UASD. The interviews lasted around 30 minutes and consisted of two parts. The first part was a customary structured interview where information on a range of demographics was collected. The second part was a Budget Allocation activity eliciting information on dimension importance. Respondents were presented with a flashcard where each quadrant showed a pictorial representation of the four dimensions under study, and were asked to apportion 40 tokens among them according to the importance they attributed to each dimension. The pictorial representations of the four dimensions were produced by a local painter in order to obtain contextually meaningful images and were also subjected to piloting. Respondents were clearly and repeatedly explained what dimension each quadrant was meant to represent, so that

these would merely serve as mnemonic support; in addition, to minimise potential biases induced by the pictorial representations themselves, no colour images were used and each quadrant included two different illustrations of the dimension (e.g. drawings of medicines and a of hospital in the case of health, of books and of a teacher with a student in the case of education, etc.). It was evident during the pilot that for respondents with no or very little formal education abstract questions about dimension importance were unintelligible, while the physical allocation of tokens enabled them to express their views – a similar methodology was previously used by the authors in the context of a research on literacy in Mozambique (Esposito et al 2011, 2015).

Expert sample. Our third sample is made of 10 local development agencies and committees,[6] which were chosen to seek the views of local 'development experts'. The organisations were first contacted by telephone and then visited in person, and were selected among those with a general development mission – i.e. avoiding organisations with a specific focus on education, health, housing or safety. The semi-structured interviews lasted on average 25 minutes and were carried out in the organisations' premises with the chief or deputy chief of the organisation. The level of education of the interviewees made it easy to elicit their views on dimension importance in an abstract way through a question similar to the one posed to our student subsample – i.e. they were asked to state the importance of the four dimensions from 0 to 100 in a way that the sum would amount to 100.

3.2 Secondary data

The empirical analysis has been conducted on 1997 and 2007 micro-data of the "Encuesta de hogares de propósitos múltiples (Enhogar)", a large nationally representative survey on individuals and households living conditions (involving, approximately, 19.000 individuals in 1997 and 30.000 in 2007 corresponding to, respectively, 4500 and 8300 households).[7] For each of education, health, housing and personal safety, indicators were created on the basis of the following criteria: i) deprivation/achievement was measured in a gradual manner and not as a "switch-on, switch-off" condition; ii) indicators were normalized along a scale ranging from zero (best condition/no deprivation) to one (worst condition/full deprivation); iii) indicators are monotonic as an increase in their value means an increase in terms of deprivation. These criteria have been applied to all dimensions described by means of ordinal or categorical variables, the only exception being health, which was dichotomous. Information related to education, health and security were available at individual level, whereas housing variables were collected at the household level and assigned to each household's member so to keep individuals as unit of analysis. Poverty thresholds were set up for each dimension assuming a high level of hardship: in particularly, individuals were deemed to be poor if they were illiterate or without any formal education; in case of disease or negative health occurrences in the past month; if three out of five poverty symptoms related to the housing condition occurred such as lack of electricity, inadequate type of house or walls or sanitation, or overcrowded housing conditions. Table A1 in the appendix provides a more detailed description of variables, poverty thresholds and well-being scores assigned.

[6] The organisations involved in the survey are the following: 1) Acción social de promoción humana campesina, 2) Asociación de San José de las matas prodesarrollo de la comunidad, 3) Asociación para el desarrollo de Santiago, 4) Consejo comunitario de desarrollo de la sierra, 5) Consejo para el desarrollo estrategico de la ciudad y el municipio de Santiago, 6) Fundación de desarrollo comunitario, 7) Consejo comunitario de Santiago, 8) Fundación comunidad y acción, 9) Junta pro-desarrollo y bienestar del Limon y lugares aledaños, 10) Fundación hogar hacia una mejor calidad de vida.

[7] The 1997 dataset has been integrated by Census Data for supplementing the lack of information related to personal safety.

4. Dimension importance scores in a poverty and wellbeing framework

4.1 Treatment effect

In order to test heterogeneity in dimension importance scores across the poverty and the wellbeing constructs, twin versions of our questionnaire were produced and allocated through a between-subject design – each student was presented with only one version. The two versions are identical in everything except that in one the text reads as "Poverty is a multidimensional phenomenon… We are interested in your opinion about the importance of the following poverty dimensions…", while in the other version the word 'poverty' is replaced with the word 'wellbeing'. In each classroom, the questionnaires were dispensed in a chessboard-like distribution of the two versions so that students were randomly allocated to treatment. Between-group analysis shows that the null hypothesis of a significant difference between the two subsamples is rejected for all socio-demographic characteristics (available upon request).

We first explore the significance of the treatment effect through univariate analysis. As can be seen in Table 2, univariate statistical tests show that there is a statistically significant difference in the dimension importance scores attached to education ($p<0.1$), health ($p<0.01$) and housing ($p<0.05$) across the two versions of the questionnaire, with a sizable difference in the case of health; there is no statistical difference instead in the case of safety. In addition, it is interesting to note that in a poverty framework the most valued dimension is education while in a wellbeing framework the most valued dimension is health. Lastly, in the case of education and health, also the dispersion of dimension importance scores is statistically different across the two questionnaire versions (significant higher dispersion of importance scores in the poverty version for education and in the wellbeing version for health).

Tab 2. Dimension importance scores and statistical tests for treatment effect

		Mean	St. Deviation	Min	Max	N
Education	Wellbeing	31.37961	10.86869	5	97	515
	Poverty	32.88404	12.14756	8	100	539
	p-values	*0.0873*[a]; *0.0070*[b]				
Health	Wellbeing	32.80194	10.70872	1	80	515
	Poverty	29.87662	10.24142	0	80	539
	p-values	*0.0000*[a]; *0.0062*[b]				
Housing	Wellbeing	18.08544	7.895443	1	50	515
	Poverty	19.15492	8.560664	0	50	539
	p-values	*0.0393*[a]; *0.1675*[b]				
Safety	Wellbeing	17.77184	6.8023	1	45	515
	Poverty	18.18646	7.614493	0	50	539
	p-values	*0.5052*[a]; *0.3482*[b]				

a, b: between group Wilcoxon-Mann-Whitney test and Robust Variance test, respectively. These tests are used due to the non-normality of underlying distributions (p=0.0000, Shapiro-Wilks normality test, for all four dimensions) hence the inability to use the more common t-test and F-test –see Shapiro and Wilk (1965), Brown and Forsythe (1974) Royston (1982, 1992), and Markowski and Markowski (1990).

While random allocation to treatment can be argued to control for potential confounders by design, so that univariate analysis would suffice, we now run multivariate analysis to both test the significance of the treatment effect further and to explore the association between views on dimension importance and additional characteristics of our respondents. We employ a methodology which enables us to account for the interdependent nature of our dimension importance scores. Since in the Budget Allocation Technique the total number of tokens is fixed, the importance score attributed to each dimension is related to the importance scores attributed to the others – more tokens placed on one dimension automatically mean fewer tokens available for the others. Therefore, the significance of the treatment effect and the predictive role of other variables are explored through Zellner's Seemingly Unrelated Regressions (SUREG) – see Zellner (1962, 1963) and Cameron and Trivedi (2010). The model simultaneously estimates one equation per each dependent variable (dimension importance scores), so that the interdependence of the dependent variables is taken into account. One regression is omitted because of linear dependence and serves as a baseline; we omit the regression for safety, hence estimated coefficients of each regression are to be interpreted as relative to those for safety.

Table 3 presents regression results for two specifications differing in the number of explanatory variables used. Specification I (columns 1-4) includes only the treatment dummy (i.e. questionnaire version), gender, age and discipline of study of the respondent, while Specification II (columns 5-8) includes several additional regressors. These range from general background information (parents' education, perceived family income and perceived relative standard on living) to information potentially related our dimensions of interest – how far students are in their tertiary education (semester of study), own and family experience of illness, whether the student's family owns their house and variables accounting for episodes of robbery, burglary and physical threat suffered. We also add two further variables which debriefing activities carried out during the pilot phase indicated as potentially related to dimension importance; to derive these variables, after respondents have stated the importance of the four dimensions they are asked to select which one dimension according to them should be considered as a human right and which one is the most urgent problem in the country. Columns 1-3 refer to Specification I and provide coefficients and significance of predictors for the importance attributed to health, education and housing, respectively, relative to the role of that variable as an explanatory variable for the importance attributed to safety; the same applies to columns 5-7 in the case of Specification II. Columns 4 and 8 refer instead to a joint significance tests, which determines whether a certain explanatory variable plays a statistically significant role in the equations jointly considered.

For both specifications, general statistics on the validity of the model confirm the reliability of our estimations. In particular, all equations are highly significant (for all of them, $p<0.05$ in Specification I and $p<0.01$ in Specification II) and the Breusch-Pagan test, as expected, rejects the null hypothesis of no correlation among the error terms in the estimated equations hence confirming the value in using a SUREG model rather than simple ordinary least squares. Looking at Specification I, the role of the treatment effect is confirmed; the dummy variable for the 'wellbeing' version of the questionnaire is highly significant in every equation and points to health being valued as more important in a wellbeing framework, whilst the opposite holds for education and housing. Demographics such as age and gender are not significant. The joint test for discipline of study suggests the existence of a discipline-specific pattern, but looking at individual variables this appears to be driven mainly by medical students attributing notable importance to health. The above results do not change after the inclusion of an array of control variables in Specification II, and are robust to alternative specifications.

Table 3: Treatment effect and other predictors - Zellner's seemingly unrelated regressions

	Specification I[a]				***Specification II[b]***			
	(1) *Edu*	(2) *Health*	(3) *Housing*	(4) Joint test (chi-2)	(5) *Edu*	(6) *Health*	(7) *Housing*	(8) Joint test (chi-2)
Questionnaire version (treatment)								
Wellbeing	-1.484**	2.870***	-1.055**	20.16***	-1.402**	2.657***	-0.886*	17.78***
	(0.715)	(0.645)	(0.511)		(0.687)	(0.633)	(0.515)	
Demographics								
Age	0.085	-0.060	0.033	3.96	-.003	-0.022	0.058	1.77
	(0.068)	(0.061)	(0.048)		(0.072)	(0.067)	(0.054)	
Male	1.343	-0.427	-0.443	2.55	1.066	-0.254	-.356	1.74
	(0.848)	(0.765)	(0.607)		(0.823)	(0.759)	(0.617)	
Disciplines								
Architecture	-1.145	0.641	0.916	35.83***	-0.107	0.984	0.213	21.49**
	(1.081)	(0.976)	(0.774)		(1.159)	(1.068)	(0.868)	
Medicine	-0.405	4.110***	-1.346*		0.115	3.279***	-1.017	
	(1.038)	(0.936)	(0.743)		(1.008)	(0.929)	(0.755)	
Education	-0.488	-0.009	0.384		0.025	0.517	0.311	
	(1.049)	(0.946)	(0.751)		(1.057)	(0.974)	(0.792)	
Dimension which most should be seen as a human right								
Education					4.573***	-0.243	0.282	134.57***
					(1.339)	(1.234)	(1.003)	
Health					-1.901	5.391***	0.303	
					(1.352)	(1.246)	(1.013)	
Housing					-1.258	1.884	3.104**	
					(1.753)	(1.615)	(1.313)	
Dimension as the most urgent problem in the country								
Education					4.225***	0.887	-2.443***	90.13***
					(0.873)	(0.804)	(0.654)	
Health					-1.479	3.956***	-1.082	
					(1.036)	(0.955)	(0.776)	
Housing					0.176	-0.160	2.444***	
					(1.248)	(1.150)	(0.935)	
Constant	30.637***	30.412***	18.478***		29.994***	24.083***	19.994***	
	(1.689)	(1.689)	(1.340)		(3.031)	(2.792)	(2.270)	
Observations	1,030	1,030	1,030		974	974	974	
Equation significance	0.0446	0.0000	0.0153		0.0000	0.0000	0.0001	
Breusch-Pagan test		0.0000				0.0000		

Notes.
[a]: no control variables included other than those reported.
[b]: additional control variables include general demographics (parents' education, perceived family income and perceived relative standard on living) and dimension-specific indicators (semester of study, own and family experience of illness, whether the student's family owns their house and indicators accounting for episodes of robbery, burglary and physical threat).

4.2 Another dimension importance paradox?

The literature has shown that the introduction of people's individual preferences in multidimensional evaluation can lead to paradoxical results. For example, accounting for individual-specific views on dimension importance may conflict with the so-called dominance principle. Suppose that individuals A and B have different preferences over health and education and a bi-dimensional index is used to compare their multidimensional wellbeing; A may be deemed

to be worse off than B even if she outperforms B in both health and education (see Fleurbaey and Trannoy, 2003, Brun and Tungodden, 2004 and Fleurbay, 2007). This means that if A and B have genuinely different preferences, we are confronted with the dilemma of either accounting for this difference and accept that B is better off, or adopting a paternalistic approach where individual preferences are silenced and the dominance principle is savaged.

Consider a 'concordance principle' stating that if individual C has more poverty than individual D, then she must also have less wellbeing than D. While this principle may appear as hardly questionable, in multidimensional evaluation it is easy to think of a situation where a deviation from this principle may be, at least to some extent, admissible. Think of a situation where C is poor in one dimension and extremely well-off in the others, while D is barely above the poverty line in all dimensions. In such a situation, D's poverty would be zero while C's poverty would be greater than zero; at the same time, an array of wellbeing indices would quantify C's wellbeing as greater than D's. Clearly, in this case the deviation from the concordance principle originates in poverty indices' neglect for achievements above the poverty line.

To rule out the role of poverty thresholds, think now about assessing multidimensional poverty and wellbeing of individuals E and F, who are both below the poverty lines in each of the two dimensions of interest – with poverty lines set at 10 days of physical functioning per fortnight for health and 10 years of schooling for education, individual achievements are (E_{edu}=8, E_{health}=8) for E and (F_{edu}=7, F_{health}=9) for F. In addition, in order to rule out the role of the indices' functional forms, suppose that in our assessment exercise in the case of both poverty and wellbeing, and for both dimensions, we use simple indices based on continuous and linear functions; in other words, these indices have positive (negative) first order derivative for wellbeing (poverty) and null second order derivative – examples of these functions are the poverty gap for poverty and a linear function of achievements for wellbeing. In this way, not only are we comparing in terms of multidimensional poverty and wellbeing two individuals who are both below the poverty line, but we are also doing this using (symmetrically) identical indices for poverty and wellbeing. Would we be ready to accept the conclusion that one between E or F has both more poverty and more wellbeing than the other?

It is clear that if no weight is used then E and F will be deemed to be equal in terms of both poverty and wellbeing. However, suppose that we use dimension importance scores as simple multiplicative factors of a weighted average – not necessarily individual-specific dimension importance scores as we saw in the case of dominance paradox, but simply the average dimension importance score, so that the same weights are applied to E and F. Then, any weighing system where, as was the case for our student sample, one dimension is deemed more important than the other in a poverty framework but less important in a wellbeing framework, or vice versa, will lead to the conclusion that one between E and F has both more poverty and more wellbeing than the other.

It should be clarified that this seemingly paradoxical conclusion is obtained regardless of whether the difference in dimension weights between a poverty and a wellbeing framework is large or minuscule. The seemingly paradoxical conclusion is, therefore, a possible outcome of the evaluation exercise whenever we have reason to believe that a difference in dimension importance scores between a poverty and a wellbeing framework really exists. This is indeed the evidence we find in our questionnaire experiment. Looking at our results one could well argue that, in terms of magnitude, this difference is appreciable in the case of health, worth-mentioning in the case of education and housing, and small in the case of safety. However, beyond a subjective judgement on magnitudes is the fact that statistical significance indicates that the difference in dimension importance scores between a poverty and a wellbeing frameworks is not due to chance but it is real, and it truly reflects respondents' views.

We see two possible ways of dealing with the possibility of the paradoxical conclusion. The first one is to accept the paradoxical conclusion as true. The implication of this is that poverty and wellbeing should be seen not as two sides of the same coin, but as two distinct phenomena. Along this take on the matter, between two poor individuals one can have both more poverty and more wellbeing than the other in the same fashion as she can have more cholesterol and more eyesight. The second one is to reject the paradoxical conclusion. In our view, the strongest grounds allowing the researcher to do this is to hypothesise that importance scores may not be fixed for a certain dimension, but they may change along the achievement line (e.g. a dimension may be very important at lower levels of achievement but become less important at higher achievement levels). Along this interpretation, the difference in the importance scores given by our respondents across the poverty and wellbeing frameworks could be made sense of by thinking that the former would apply to low achievements while the latter to high achievements. In this way, the appropriate sets of weights in the case of E and F would be those provided in the poverty framework and the paradox would disappear. A difficulty with this interpretation is, however, that it may jeopardise the applicability of the concept of wellbeing (different from that of poverty) to individuals below the poverty line and may therefore lead to the notion that below the poverty line only the construct of poverty applies.

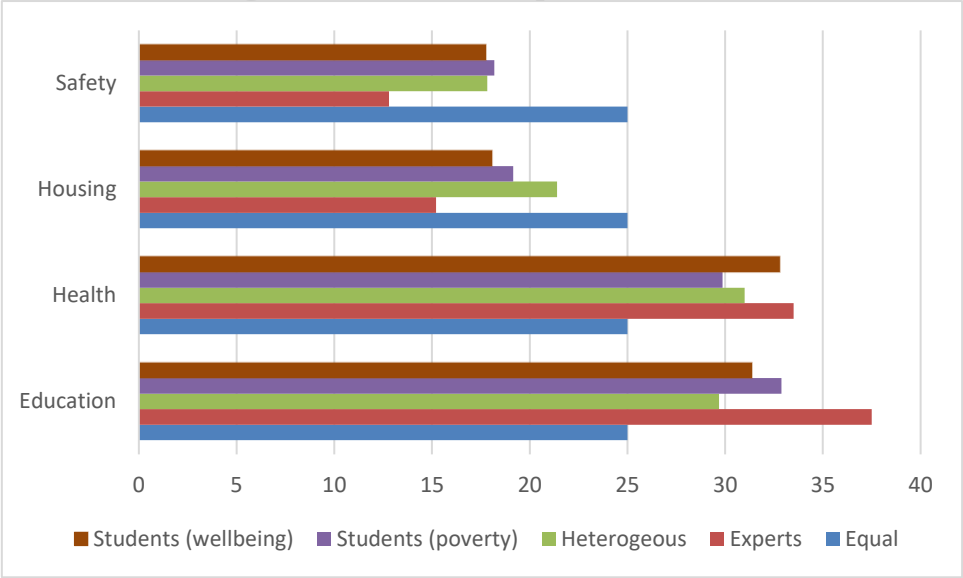
5. Multidimensional poverty and wellbeing in the Dominican Republic

Our second aim in this paper is to explore whether the use of alternative sets of weights brings about appreciable differences in the assessment of multidimensional poverty and wellbeing. Before presenting our evidence, we clarify that we are not interested in studying which sets or ranges of weights, among all the theoretically possible ones, produce qualitatively different empirical results. Rather, we want to explore whether qualitatively different results are produced by *specific* sets of weights, namely those we collected in the field – which are non-paternalistic and contextually relevant to the country whose poverty and wellbeing are studied. An additional remark regards the limitations affecting the sets of weights we elicited in the field. It should be clear that, while we believe that our fieldwork enabled us to produce meaningful views on dimension importance, the derived sets of weights are not statistically representative of the student, 'expert' and adult populations in the country, given the non-probabilistic nature of our samples. In addition, while the different sets of importance scores can be seen as comparable because they were all collected using the Budget Allocation Technique, at the same time this comparability encounters some limits given that this approach was implemented following different procedures. The reason for this was again opportunity and resource constraints – the only procedure viable for all samples was the one followed for our heterogeneous sample, but the resources needed for this would have made it impossible to obtain such a large sample of university students.

In Figure 1 we illustrate the different weighting schemes to be used in our empirical analysis – the average values attributed by our samples to the four dimensions. It appears clear how the set of equal weights brings about an overestimation of the low-valued dimensions (housing and safety) and an underestimation of the high-valued ones (health and education). Among our respondents, the lowest value to education is given by the heterogeneous sample – which is also the group with the lowest average level of formal education. The higher level of education of students and development experts suggests a relationship between educational attainment and value attached to education; this idea is reinforced by the results from univariate and multivariate analyses of the heterogeneous sample data, where respondents' years of schooling are strong predictors of the value attached to education (results available upon request). The views expressed by development experts show the largest gap between health and education on the one hand and housing and safety on the

other; when asked to motivate the reason for such a disparity, respondents often evoked the idea of health and education being central to the notion of human development.

Figure 1. Dimension importance scores



Notes: Importance scores for the heterogeneous sample are normalised to 100

Moving to multidimensional evaluation, in Figures 2 and 3 we report the percentage change in multidimensional poverty[8] and wellbeing, respectively, in the Dominican Republic between 1997 and 2007. In both the case of multidimensional poverty and multidimensional wellbeing, opposite conclusions are reached depending on which set of weight is used. A negative variation in poverty (poverty decrease), is obtained if the analysis is carried out using the dimension importance scores suggested by the heterogeneous sample, by the student subsample having received the poverty version of the questionnaire, or giving equal importance to the four dimensions; by contrast, the adoption of the dimension importance scores provided by the expert sample suggest an increase in multidimensional poverty. The evidence on multidimensional wellbeing is even more mixed, with two sets of weights indicating a positive trend and two indicating a negative trend. It is also interesting to notice that, in both the cases of multidimensional poverty and multidimensional wellbeing, the rosiest picture on the social development trend in the Dominican Republic is obtained by using equal weights.

[8] For each dimension, headcount ratios are used for the evaluation of poverty.

Figure 2. Change in multidimensional poverty 1997-2007 by sets of weights used (%)

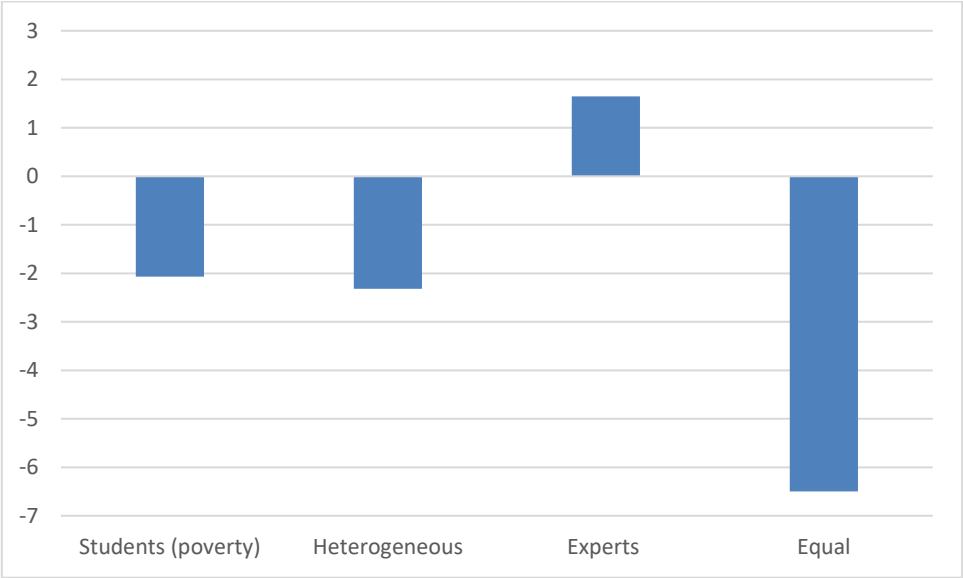
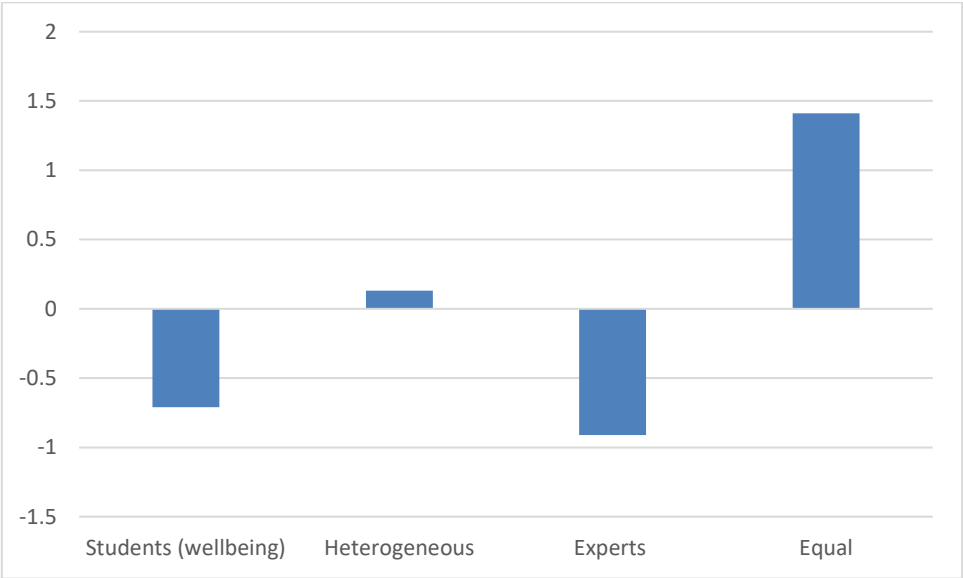


Figure 3. Change in multidimensional wellbeing 1997-2007 by sets of weights used (%)



6. Conclusions

In this paper, we tackle the issue of weighting in multidimensional evaluation from different angles. If one dimension is believed to be more important than another, should this difference in importance be accounted for when a multidimensional index is built? We believe that the answer is yes, no matter whether the use of equal weights or alternative sets of unequal weights leads to large or minor differences in empirical analyses. As philosopher Carveth Read (1914) argued, "It is better to be vaguely right than precisely wrong" (Ch. 2, Section 2.b.iii). If a dimension counts little, it should not play the same role as a dimension which is believed to be paramount to a certain phenomenon (poverty, wellbeing, etc.). However, should this conceptual line of reasoning not convince the sceptic, the empirical evidence provided in this paper shows that the choice of which sets of weights is used is not a trivial one. We collect primary data on the importance of four life domains such as education, health, housing and personal safety from a threefold sample in the Dominican Republic. We employ these three sets of weights for the assessment of the trend of multidimensional poverty in the country, and find that opposite conclusions are reached depending on which set of weights is used. Clearly, there is no pretension that these sets of weights are the 'true' views of the student, society and expert categories in the Dominican Republic since our means did not allow the pursuit of representative samples. These weights are, however, meaningful to the country whose poverty is analysed since derived from an intensive three-month fieldwork carried out with the greatest care by one of the authors. That the use of these weights leads to a reversal in the multidimensional poverty ranking does invite researchers to take seriously the issue of 'who decides' upon the relative importance of life domains in multidimensional evaluation.

Another offer of our paper relates to the determinants of views on dimension importance. A randomly selected half of our sample of university students is presented with a questionnaire framed in terms of poverty, while the other half receives a script which is identical except that it is framed in terms of wellbeing. The treatment effect is strongly significant, meaning that the set of weights attached to the four dimensions is significantly different across the two questionnaire versions. In other words, the four dimensions receive a different set of importance scores according to whether these are presented as 'dimensions of poverty' or 'dimensions of wellbeing'. We show how this evidence can lead to what we called a 'concordance paradox', namely the possibly disturbing conclusion that between two individuals who are below the poverty line in every dimension, one can be deemed to have at the same time more poverty and more wellbeing than the other. We argue that there can be two ways out of this apparently paradoxical conclusion. The conclusion can be accepted, with the implication of poverty and wellbeing needing to be conceptualised as entirely different phenomena, so that one can have both more poverty and more wellbeing than the other in the same fashion as she can have more cholesterol and more eyesight. Or it can be rejected, hypothesising dimensional weights to be function of achievements – a conclusion which is probably less novel than we may think, given that the literature on sour grapes began two and a half millennia ago. This conclusion, however, poses questions on the relevance of the construct of wellbeing for individuals below the poverty line.

References

Alkire, S. and M.E. Santos (2010), Acute multidimensional poverty: A new index for developing countries, OPHI Working Paper Series No. 38, Oxford University.

Alkire, S., and J. Foster (2011a), Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95(7-8), 476-487.

Alkire, S. and J. Foster (2011b), Understandings and misunderstandings of multidimensional poverty measurement, *Journal of Economic Inequality*, 9(2), 289-314.

Alkire, S., Foster, J., Seth,S., Santos, M.E., Roche, J.M. and P. Ballon (2015), *Multidimensional Poverty Measurement and Analysis*, Oxford: Oxford University Press.

Alkire, S., and Seth, S. (2015), Multidimensional poverty reduction in India between 1999 and 2006: Where and how?. *World Development, 72*, 93-108

Angelini, V., Cavapozzi, D., Corazzini, L. and O. Paccagnella (2014), Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases, *Oxford Bullettin of Economics and* Statistics 76(5), 643–666.

Angulo, R., Díaz Y. and R. Pardo (2016), The Colombian Multidimensional Poverty Index: Measuring Poverty in a Public Policy Context, *Social Indicators Research*,127(1), 1–38.

Athanassoglou, S. (2015). Multidimensional welfare rankings under weight imprecision: a social choice perspective. *Social Choice and Welfare*, 44(4), 719-744.

Atkinson, A. B., and F. Bourguignon (1982), The comparison of multi-dimensioned distributions of economic status, *Review of Economic Studies*, 49, 183-201.

BCRD (2011), Nueva canasta para medir la inflación IPC Base Diciembre 2010=100, Banco Central de la República Dominicana, Aviso 3 de Febrero 2011, Santo Domingo

Beegle, K., Himelein, K. and M. Ravallion (2012), Frame of Reference Bias in Subjective Welfare Regressions, *Journal of Economic Behaviours and Organization*, 81, 556-579.

Belhadj, B. and M. Limam (2012), Unidimensional and multidimensional fuzzy poverty measures: New approach, *Economic Modelling*, 29(4), 995–1002.

Betti, G., Gagliardi, F., Lemmi A. and V. Verma (2015), Comparative measures of multidimensional deprivation in the European Union, *Empirical Economics*, 49(3), 1071-1100.

Bossert, W., Chakravarty, S. R. and C. D'Ambrosio (2013), Multidimensional poverty and material deprivation with discrete data, *Review of Income and Wealth*, 59(1), 29-43.

Bourguignon, F. and S.R. Chakravarty (2003), The measurement of multidimensional poverty, *Journal of Economic Inequality*, 1, 25-49.

Brown, M. B., and A.B. Forsythe (1974), Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69, 364–367.

Brun, B. C. and B. Tungodden (2004), Non-welfaristic theories of justice: Is "the intersection approach" a solution to the indexing impasse?, *Social Choice and Welfare* 22, 49-60.

Callander, E. and D. Schofield (2015). Multidimensional poverty and health status as a predictor of chronic income poverty. *Health Economics*, 24(12), 1638-1643.

Cameron, A. C., and P.K. Trivedi (2010), *Microeconometrics Using Stata*, Rev. ed. College Station, TX: Stata Press.

Camfield, L. and D. Ruta (2007), 'Translation Is Not Enough': Using the Global Person Generated Index (Gpgi) to Assess Individual Quality of Life in Bangladesh, Thailand, and Ethiopia, *Quality of Life Research*, 16, 1039-51.

Campbell, A., Converse, P.E. and W.L. Rogers (1976), *The Quality of American Life: Perceptions, Evaluations, and Satisfactions*, Russel Sage, New York.

Chakravarty, S., Deutsch, J. and J. Silber, (2008), On the Watts Multidimensional Poverty Index and its Decomposition, *World Development* 36(6): 1067-1077.

Cherchye, L., Ooghe, E., and T. Puyenbroeck (2008) Robust human development rankings, *Journal of Economic Inequality*, 6, 287–321.

Chiappero-Martinetti, E. and N. von Jacobi (2012), Light and shade of multidimensional indexes: how methodological choices impact on empirical results, in "*Quality of life in Italy: research and reflections*" in Maggino F. and G. Nuvolati, eds. Springer, 69-106.

Choo, E.W., Schoner, B. and W.C. Wedley (1999), Interpretation of Criteria Weights in Multicriteria Decision Making, *Computers & Industrial Engineering*, 37, 527-541.

Coromaldi, M. and M. Zoli (2012), Deriving Multidimensional Poverty Indicators: Methodological Issues and an Empirical Analysis for Italy, *Social Indicators Research*, 107,(1), 37-54.

Crawford, G. and C. Williams (1985), A note on the analysis of subjective judgment matrices, *Journal of Mathematical Psychology*, 29, 4, 387–405

Dasgupta, P., and M. Weale (1992), On Measuring the Quality of Life, *World Development*, 20(1) 119-131.

Decanq, K. and M.A Lugo (2013), Weights in Multidimensional Indices of Well-Being: an Overview, *Econometric Reviews*, 32, 7-34.

Decancq, K., Van Ootegem, L., and E.Verhofstadt (2013), What if we voted on the weights of a multidimensional well-being index? An illustration with Flemish data, *Fiscal Studies*, 34(3), 315-332.

Donohue, C. and E. Biggs (2015), Monitoring socio-environmental change for sustainable development. Developing a multidimensional Livelihoods Index (MLI), *Applied Geography*, 62 391-403.

Duclos, J.Y, Sahn, D.E. and S.D. Younger (2006), Robust multidimensional poverty comparisons, *Economic Journal* 54, 943-968.

Esposito, L. and E. Chiappero-Martinetti (2010) Multidimensional Poverty Measurement: Restricted and Unrestricted Hierarchy among Poverty Dimensions, *Journal of Applied Economics*, 13, 181-204.

Esposito, L., Kebede, B. and B. Maddox (2011), Literacy practices and schooling: A case study from Mozambique, *World Development*, 39, 1796-1807.

Esposito, L., Kebede, B. and B. Maddox (2015), The Value of Literacy Practices, *Compare*, 45(3), 363-380.

Feeny, S. and L. McDonald (2016) Vulnerability to Multidimensional Poverty: Findings from Households in Melanesia, *Journal of Development Studies*, 3, 447–464.

Ferrans, C. E. and M.J. Powers (1985), Quality of life index: development and psychometric properties. *Advances in nursing science*, 8(1), 15-24.

Ferreira, F. H. G. (2011), Poverty is multidimensional. But what are we going to do about it? *Journal of Economic Inequality*, 9(3): 493-495.

Ferreira, F. H. G. and M.A. Lugo (2013), Multidimensional poverty analysis: Looking for a middle ground, *The World Bank* Research Observer, 28(2), 220-235.

Fleurbaey, M. (2007). Social choice and the indexing dilemma. *Social Choice and Welfare*, 29(4), 633-648.

Fleurbaey, M. and A. Trannoy (2003), The impossibility of a Paretian egalitarian, *Social Choice and Welfare*, 21(2), 243-263.

Foster, J. E., McGillivray, M. and S. Seth (2013), Composite Indices: Rank Robustness, Statistical Association, and Redundancy, *Econometric Reviews*, 32(1), 35-56.

Glejser, H., Gevers, L., Lambot, P., & Morales, J. A. (1977). An econometric study of the variables determining inequalitty aversion among students. *European Economic Review*, 10(2), 173-188

Halleröd, B. (1995), The truly poor: direct and indirect consensual measurement of poverty in Sweden, *Journal of European Social Policy*, 5(2), 111-129.

Halleröd, B. (1996), Deprivation and poverty: A comparative analysis of Sweden and Great Britain, *Acta Sociologica*, 39(2), 141-168.

Hick, R. (2014), Poverty as capability deprivation: Conceptualising and measuring poverty in contemporary Europe, *European Journal of Sociology* 55(3), 295–323.

Hick, R. (2016), Material poverty and multiple deprivation in the UK: The distinctiveness of multidimensional assessment, *Journal of Public Policy*, 36(1), 277 – 308.

Holland, P. W. and H. Wainer (1993), *Differential Item Functioning*, Hillsdale, NJ, Lawrence Erlbaum Associates.

Hsieh, C. M. (2016), Domain Importance in Subjective Well-Being Measures. *Social Indicators Research*, 127(2), 777-792.

Inglehart, R. (1978), Value priorities, life satisfaction, and political dissatisfaction among western publics, *Comparative Studies in Sociology.* 1, 173–202.

Jensen, M. C. and W.H. Meckling (1994), The Nature of Man, *Journal of Applied Corporate Finance*, 7(2), 4-19.

Kahneman, D., Krueger, A.B., Schkade, D., Schwarz, N. and A. Stone (2004),A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method, *Science*, 306, 1776-1780.

Kakwani, N. and J. Silber, eds., (2008), *Quantitative approaches to multidimensional poverty measurement*, London, Palgrave Macmillan.

Kaypten, A., Smith, J. P., and A. van Soest, (2007), Vignettes and Self-Reports of Work Disability in the U.S. and the Netherlands, *American Economic Review*, 97, 461-473.

King, G., Murray, C. J. L., Salomon, J. and A. Tandon (2004), Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research, *American Political Science Review*, 98, 567-583.

Kristensen, N. and E. Johansson (2008), New Evidence on Cross-Country Differences in Job Satisfaction Using Anchoring Vignettes, *Labour Economics*, 15, 96-117.

Linton, M., Dieppe, P. and A. Medina-Lara (2016), Review of 99 self-report measures for assessing well-being in adults: exploring dimensions of well-being and developments over time, *BMJ Open*;6:e010641; doi: 10.1136/bmjopen-2015-010641.

Maslow, A. H. (1943), A Theory of Human Motivation. *Psychological Review*, 50, 370-396.

Massey P. et al. (2013), Findings Toward a Multidimensional Measure of Adolescent Health Literacy, American Journal of Health Behavior, 37(3), 342-350.

Markowski, C. A., and E. P. Markowski (1990), Conditions for the effectiveness of a preliminary test of variance, *American Statistician*, 44, 322–326.

Mascherini, M. and B. Hoskins (2008), *Retrieving expert opinion on weights for the Active Citizenship Composite Indicator,* European Commission, Joint Research Centre, Institute for Social Protection and Security of the Citizen.

Mayer, S.E. and C. Jencks (1989), Poverty and the distribution of material hardship, *Journal of Human Resources*, 24(1),88-114.

Merz, J. and T. Rathjen (2014), Time And Income Poverty: An Interdependent Multidimensional Poverty Approach With German Time Use Diary Data, *Review of Income and Wealth*, 60: 450–479.

Maasoumi, E. and J. Racine (2016), A solution to aggregation and an application to multidimensional 'well-being' frontiers, *Journal of Econometrics*, 191(2), 374-383.

Michener, G. (2015), Policy Evaluation via Composite Indexes: Qualitative Lessons from International Transparency Policy Indexes, *World Development*, 74(C), 184-196.

Mitra, S., Jones, K., Vick, B., Brown, D., McGinn, E., an M.J. Alexander (2013), Implementing a multidimensional poverty measure using mixed methods and a participatory framework, *Social indicators research*, 110(3), 1061-1081.

Moldan, B. and S. Billharz (1997), *Indicators of Sustainable Development*. Chichester, John Wiley.

Morris, D. (1979), *Measuring the Conditions of the World Poor, the Physical Quality of Life Index*, Pergaman Press, New York.

Nowak D. and C. Scheicher (2014), Considering the extremely poor: Multidimensional poverty measurement for Germany, Discussion Papers in Econometrics and Statistics, No. 02/14, University of Cologne, Institute of Econometrics and Statistics.

Pattanaik P., Reddy S. and Y. Xu (2012), On measuring deprivation and living standards of societies in a multi-attribute framework. *Oxford Economic Papers*. 64, 43–56.

Permanyer, I. (2011), Assessing the Robustness of Composite Indices Rankings, *Review of Income and Wealth*, 57(2), 306–326.

Permanyer, I. (2012), Uncertainty and Robustness in Composite Indices Rankings, *Oxford Economic Papers*, 64(1), 57–79.

Permanyer, I. (2014), Assessing individuals' deprivation in a multidimensional framework, *Journal of Development Economics*, 109, 1–16.

Philip, E. J., Merluzzi, T. V., Peterman, A., and L. B. Cronk (2009), Measuring accuracy in assessing patient's quality of life: to weight or not to weight domains of quality of life, *Quality of Life Research*, 18, 775-782

Ravallion, M. (2011), On multidimensional indices of poverty, *Journal of Economic Inequality*, 9(2),235-248.

Ravaillon, M. (2012), Troubling trade-offs in the Human Development Index, *Journal of Development Economics*, 99(2), 201-209.

Rawls, J. (1971), *A Theory of Justice, Cambridge*, MA, Harvard University Press.

Read, C. (1914), *Logic: Deductive and Inductive*, London, England, Simpkin, Marshall, Hamilton, Kent & Co. LTD.

Rogan, M. (2016), Gender and Multidimensional Poverty in South Africa: Applying the Global Multidimensional Poverty Index (MPI), *Social Indicators Research*, 126(3), 987–1006.

Royston, P. (1982), An extension of Shapiro and Wilks's W test for normality to large samples, *Applied Statistics* 3, 115–124.

Royston, P. (1992), Approximating the Shapiro–Wilk W-test for non-normality, *Statistics and Computing*, 2,117–119.

Russell, L. B., Hubley, A. M., Palepu, A., and B. D. Zumbo (2006), Does weighting capture what's important? Revisiting subjective importance weighting with a quality of life measure, *Social Indicators Research*, 75, 141-167.

Saaty, R. W (1980), *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw-Hill.

Saaty, R. W. (1987),The analytic hierarchy process- what it is and how it is used, *Mathematical Modelling*, 9(3-5), 161-176.

Saisana, M., Saltelli, A., and Tarantola, S. (2005), Uncertainty and sensitivity analysis as tools for the quality assessment of composite indicators, *Journal of the Royal Statistical Society*, Series A, 168, 1–17.

Salomon, J., Tandon, A. and C. J. L. Murray (2004), Comparability of Self-Rated Health: Cross Section Multi-Country Survey Using Anchoring Vignettes, *British Medical Journal*, 328, 258-261.

Schang, L., Hynninen, Y., Morton, A., & Salo, A. (2016). Developing robust composite measures of healthcare quality–Ranking intervals and dominance relations for Scottish Health Boards. *Social Science & Medicine*, *162*, 59-67

Schenkerman, S. (1991), Use and abuse of weights in multiple objective decisions support models, *Decision Science*, 2, 368-378.

Sen, A.K. (1992), *Inequality Re-examined*, Oxford University Press, Oxford.

Seth, S. and M. McGillivray (2016), Composite indices, alternative weights, and comparison robustness, OPHI Working Paper 106, University of Oxford.

Shapiro, S. S., and M. B. Wilk (1965), An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591–611.

Stapleton, L. and G.D. Garrod (2007), Keeping things simple: Why the Human Development Index should not diverge from its equal weights assumption, *Social Indicators Research*, 84, 179-188.

Stoeffler, Q., Alwang, J., Mills, B. and N. Taruvinga (2016), Multidimensional Poverty in Crisis: Lessons from Zimbabwe, *Journal of Development Studies,* 52(3), 428-446.

Trani, J.F., Kuhlberg, J., Cannings, T. and D. Chakkal (2016), Multidimensional poverty in Afghanistan: who are the poorest of the poor?, *Oxford Development Studies*, 44(2), 220-245.

Trauer, T. and A. Mackinnon (2001), Why are we weighting? The role of importance ratings in quality of life measurement, *Quality of Life Research*, 10(7), 579-85.

Tsui, K. Y. (2002), Multidimensional poverty indices, *Social Choice and Welfare*, 19, 69-93.

UNDP (1990). *Human development report 1990. Concept and Measurement of Human Development*, New York: Oxford University Press.

USAID (2013), Dominican Republic Country Development Cooperation Strategy, FY 2014-2018 (https://www.usaid.gov/dominican-republic/cdcs).

Vasquez, A., Cabieses ,B. and H. Tunstall (2016), Where Are Socioeconomically Deprived Immigrants Located in Chile? A Spatial Analysis of Census Data Using an Index of Multiple Deprivation from the Last Three Decades (1992-2012), PLOS ONE 11(1): e0146047, doi:10.1371/journal.pone.0146047.

Victor B., et al. (2014), Multidimensional Poverty in Rural Mozambique: A New Metric for Evaluating Public Health Interventions, PLoS ONE 9(9): e108654, doi:10.1371/journal.pone.0108654

Wagle, U. R., (2014), The Counting-Based Measurement of Multidimensional Poverty: The Focus on Economic Resources, Inner Capabilities, and Relational Resources in the United States, *Social Indicators Research*, 115(1), 223-240.

Wu, C. H., Yang, C. T., and L.N. Huang, L. N. (2014), On the predictive effect of multidimensional importance-weighted quality of life scores on overall subjective well-being, *Social indicators research*, 115(3), 933-943.

Yalonetzky, G., (2014), Conditions for the most robust multidimensional poverty comparisons using counting measures and ordinal variables, Social Choice and Welfare, 43(4), 773-807.

Zellner, A. (1962), An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias, *Journal of the American Statistical Association*, 57, 348-368.

Zellner, A. (1963), Estimators for Seemingly Unrelated Regressions: Some Exact Finite Sample Results, *Journal of the American Statistical Association*, 58, 977-992.

Zheng, B., and C. Zheng, (2015). Fuzzy ranking of human development: A proposal, *Mathematical Social Sciences*, 78, 39-47.

Appendix

Table A1. Data description, poverty thresholds and well-being scores, Dominican Republic 1997 and 2007

DIMENSIONS	INDICATOR(S)	TYPES OF VARIABLES	DESCRIPTION	WELLBEING SCORES	POVERTY LINES (Z)	NO. OF OBSERV.
EDUCATION	Highest level of education attained	Ordinal	1. illitterate 2. read&writing but no formal edu 3. primary school (basic) 4. high school (middle) 5. univ degree or doctorate	0 (min wb) .25 .50 .75 1 (max wb)	$Z \leq 2$	19,083 (1997) 21,578 (2007)
HEALTH	Presence/absence of a disease or negative health occurrences in the past month	Dichotomous	1. health problems 2. no health problems	0 (min wb) 1 (max wb)	Z=1	19,083 (1997) 30,969 (2007)
HOUSING	Housing conditions	Categorical	1. Type of housing 2. Walls 3. Electricity 4. Sanitation 5. Overcrowding index (no of adults/no. of bedrooms)	count # of poverty symptoms 0= 5 sympt. (min wb) .2=4 sympt. .4=3 sympt. .6=2 sympt. .8=1 sympt. 1=0 sympt. (max wb)	Indicator thresholds: Z_1=shanty or building house or house shared with workplace/shop Z_2=pasteboard or wood or palm leaf Z_3=no electricity or polluting source of energy (i.e. kerosene) Z_4=outhouse or private cesspit Z_5=1st quartile Housing poverty threshold: 3 out of 5 symptoms	16,937 (1997) 31,369 (2007)
PERSONAL SAFETY	Feeling insecure in the neighborhood where people live (*)	Categorical	1. very safe 2. safe 3. quite safe 4. unsafe 5. very unsafe	0 (min security) (°) .2 .4 .6 .8 1 (max security)	Z= mean value (1997=.540) (2007=.525)	19,103 (1997) 31,609 (2007)

Notes:

(*) Individual micro-data integrated by district data on the perception of personal security in the neighborhood

(°) in order to facilitate a time comparison, well-being scores were assigned to people living in the same region (*estrato)* x_i (i=1,…10) on the basis of the standard deviation and the national mean values (μ) observed in the two years. Namely, a zero value (worst security condition) is assigned if the observed value in region *i* (x_i) was larger than -2sd from the national mean; 0.2 if it was

included between two and one sd below the mean; 0.4 for values comprised between -1 sd and the national mean: 0.6 if was included between +1 sd and the national mean; 0.8 if between two and one sd above the mean and finally a value equal to one (best security condition) was assigned if the observed value x_i was 2 or more sd above the mean.

Data sources:

Education, Health and Housing: 1997 and 2007 Microdata “Encuesta national de ingresos y gastos de hogares”, Departemento de Encuesta, Oficina Nacinonal de Estadistica, Republica Dominicana.

Personal Security: microdata by district of the "Encuesta de hogares de propósitos múltiples (enhogar)" 2007; estimated value for 1997 based on the variation of personal security 2007-2005 and perception on changes between 2005-2000 (Q: “Desde el 2005 (2000), considera su barrio igual, mas o menos seguro que antes”?;)