

ELICITING, APPLYING AND EXPLORING MULTIDIMENSIONAL WELFARE WEIGHTS: EVIDENCE FROM THE FIELD[†]

By **LUCIO ESPOSITO***

School of International Development, University of East Anglia

AND

ENRICA CHIAPPERO-MARTINETTI*

Department of Political and Social Sciences, University of Pavia

By combining primary data on dimension importance collected in the field from three different samples and nationally representative survey data from the Dominican Republic, 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 framework. This poses important questions on the anatomy of dimension importance and on the use of weights in empirical analyses, and opens the door to what we call a ‘concordance paradox’ related to the very essence of the constructs of poverty and wellbeing. As a second contribution, we employ the sets of weights collected in the field to assess the trend of multidimensional poverty and wellbeing in the country. We find that the picking one set of weights or another is not a trivial choice, as they lead to opposite assessment results.

JEL Codes: I32, I31, D63, O12

Keywords. Multidimensional poverty, wellbeing, weighting schemes, fieldwork, between-subject design

[†] The authors thank Sarah Tustin, the editors, two anonymous referees and the participants of the IARIW-BOK conference for valuable comments which led to substantial improvement of our work. We would also like to thank the University of Pavia for funding the primary data collection this paper is based on.

* Corresponding author. Lucio Esposito, School of International Development, University of East Anglia, NR47TJ, Norwich, UK (lucio.esposito@uea.ac.uk).

*Enrica Chiappero-Martinetti, Department of Political and Social Sciences, University of Pavia, Strada Nuova 65 – 27100 Pavia, Italy (enrica.chiappero@unipv.it).

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 including 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.

While multidimensional evaluation enabled researchers to unveil aspects of poverty and wellbeing neglected by unidimensional monetary evaluation (Victor et al., 2014, Alkire et al., 2015, Trani et al., 2016), it also confronted them 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 the weights to be attached to them, 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 focus on the issue of dimension weights. We offer a twofold contribution on this issue 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 university 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

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). For recent contributions discussing the main theoretical and empirical aspects of multidimensional poverty see, respectively, Chakravarty and Chattopadhyay (2018) and Guio (2018).

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 in the literature 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 address this issue by running a questionnaire experiment with our university student sample (N=1,083). 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). This result indicates that people may value dimensions differently depending on the construct under consideration and therefore a blanket set of weights to be applied for any multidimensional evaluation may be inappropriate. In addition, our finding raises what we call a ‘concordance paradox’ which has meaningful implications for the conceptualisation of the notions of poverty and wellbeing, as will be discussed in the paper. The second offer of our paper relates to the debate as to whether the adoption of different weighting schemes produces qualitative differences in multidimensional evaluations or not. 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 three samples (i.e. the student, ‘expert’ and heterogeneous samples). Our results show that picking a certain set of weights rather than another is not a trivial choice, because different weighting schemes lead to opposite conclusions on the change in multidimensional poverty and wellbeing, and that simply adopting equal weights leads to the most optimistic assessment in both the cases of poverty and wellbeing.

The paper is structured in the following way. Section 2 critically 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 supporting the choice of the Budget Allocation Technique as the method for the derivation of dimension importance scores in the field. 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). 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 idea of taking people's preferences for different life domains into account is also at the centre of the equivalent-income approach, where a metric is derived by adjusting individuals' actual income figures on the basis of information on their achievements in other life domains as well as on their preferences (Decancq, Fleurbaey and Schokkaert, 2015a,b).

The introduction of weighting schemes in multidimensional evaluation, however, brings about operational as well as conceptual issues. Dimension importance scores can be given 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 – see, among others, Russell et al. (2006), Philip et al. (2009), Wu, Yang and Huang (2014) and Hsieh (2016). The results of Pasha (2017) challenge the statistical desirability of equal weighting. The economics literature has taken a slightly different angle, focussing in particular on 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 (2018) point out, these studies explore robustness to all possible sets of weights rather than 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 Decancq and Lugo's subclasses into the two broader categories of *direct* and *indirect* weights.

[place Table 1 here]

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. A less common approach is based on political constructivism and employs existing expression of people’s values such as Constitutions and other important pieces of legislation (Burchi, De Muro and Kollar, 2018). Given our interest in eliciting dimension importance scores directly from our respondents, in the following subsection we critically present the most common methods for the direct derivation of dimension importance scores.

2.b Direct approaches: a critical review and the case for the Budget-Allocation Technique

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 method 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 whilst 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

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’ method, 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).

weights have little sense on their own, and should actually be seen as having a relative character. 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 the 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 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 method 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 method enables researchers to overcome some 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 valued, so that the attribution of importance scores takes place simultaneously. Second, the amount of points to be allocated is fixed across subjects; this enables researchers 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 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 three-month fieldwork (February-May 2009), 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 (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

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 researchers to get around this problem directly at the stage of weights elicitation.

judgements from experts (Moldan and Billharz, 1997, and Mascherini and Hoskins, 2008).⁵ A description of the data collection strategy for each sample is in order.

Student sample (collected February-March 2009). 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, characterised by extremely low tuition fees making it one of the least elitist in the whole of Latin America (Liz, 2001). Around 29% of youth enrolls in tertiary education in the Dominican Republic and students at UASD have lower socioeconomic profiles compared to students in private universities (OECD, 2012). In spite of this, it should be kept in mind that our student sample may still have a relatively privileged background since around 1/3 of them has at least one parent with a university degree (against a national figure of university degree holders among the adult population of around 12%, López and Mejía, 2016) and that for poorer households even UASD fees may be inaccessible (ONE, 2013). 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 (response rate >96%). The development of the questionnaire benefited from inputs offered by academics in the School of Education at UASD and from a pilot with around 20 students.

In order to test heterogeneity in dimension importance scores across the poverty and the wellbeing constructs, twin versions of our questionnaire were produced and randomly allocated to our student sample 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*’. Between-group analysis of socio-demographic characteristics shows that the null hypothesis of a significant difference between the two

⁵ A similar method 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).

subsamples is rejected, confirming that randomization has worked (see Appendix 1). Views on dimension importance were elicited through the following question:

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 (collected February-May 2009). This sample consists of 309 adults interviewed face to face. While our resources did not allow us to pursue a formal strategy to achieve national representativeness, 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. Respondents aged from 18 to 79 (mean age is around 37), 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; around a fifth barely completed primary education and 12% have a university degree or higher. Almost half is catholic and 58% is employed. 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). Also for this sample the questionnaire development benefited from a pilot. After gathering information on a range of demographics, 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. 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 method was previously used by the authors in the context of a research on literacy in Mozambique (Esposito et al., 2011, 2015).

Expert sample (collected April 2009). 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 selected among those with a general development mission – i.e. avoiding organisations with a specific focus on education, health, housing or safety. They were first contacted by telephone and then visited in person. 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 sample.

3.2 Secondary data

Secondary micro-data were obtained from the 1997 and 2007 "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 as to keep individuals as the unit of analysis. Poverty thresholds were set up for each dimension – e.g. being illiterate or without any formal education, having three out of five poverty symptoms related to the housing condition such as lack of electricity, inadequate type of house or walls or sanitation, etc. Table

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.

A2 in Appendix 2 provides a more detailed description of variables, poverty thresholds and wellbeing scores assigned.

4. DIMENSION IMPORTANCE SCORES IN A POVERTY AND WELLBEING FRAMEWORK

4.1 Treatment effect

The distribution of dimension importance scores obtained in the experiment carried out with the student sample is graphically presented in Figure 1. For each dimension, two overlaying histograms are presented – with shaded and unshaded histograms referring to the wellbeing and the poverty questionnaire versions, respectively. The importance scores attributed to education are visibly more dispersed in the case of the poverty framework, while scores are more concentrated within the 20-40 range in the case of the wellbeing framework. Looking at health, wellbeing scores appear shifted to the right (higher values) while the opposite holds for housing – although to a lesser extent. There seems to be little difference instead between the two sets of scores in the case of safety. In Table 2 we present summary statistics as well as univariate tests for the significance of the treatment effect. These 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; it should be noted that the higher mean value for education in the poverty framework is the result of greater extreme values rather than a shift of the histograms across the board as in the case of health (see Figure 1). 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 – it is possible that education is also seen as factor enabling individuals to get out of poverty. 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); in particular, in the case of education the higher mean value for the

[place figure 1 here]

[place table 2 here]

While random allocation to treatment can be argued to control for potential confounders by design, so that univariate analysis would suffice, we run multivariate analysis to both test the significance of the treatment effect further controlling for a range of personal characteristics and family circumstances of our respondents. We employ a method which enables us to account for the interdependent nature of the dimension importance scores provided by each respondent. 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 imply that fewer tokens are available for the other dimensions. 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 score), taking into account their within-respondent interdependence. Formally, for n individuals, p outcome variables for each individual and q explanatory variables, we employ a system of equations

$$(1) \quad Y = \beta X + \varepsilon,$$

where Y is a $np \times 1$ vector of responses, X is a $np \times q$ matrix of explanatory variables, β is a q -dimensional vector of regression coefficients and ε_i is a $np \times 1$ error vector. The j^{th} outcome ($j = 1, \dots, p$) for the i^{th} individual, y_{ij} , is modelled through equations of the form

$$(2) \quad y_{ij} = \beta_j x_{ij} + \varepsilon_{ij}$$

where the error terms are expected to be correlated across equations, i.e. $E(\varepsilon_{ij} \varepsilon_{ij'} | X) \neq 0$ with $j \neq j'$.

Table 3 presents regression results for two specifications which differ in the set of explanatory variables used. 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 (choosing other baseline equations does not produce qualitative differences in our results, analysis available upon request). 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 which may be hypothesised to shape respondents' evaluations and affect the importance attached to the life domains we address. These variables range from information more specifically related to our four dimensions of interest – how far students are into 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 – to subjective socio-economic status variables – perceived family income and perceived relative standard of living. We also add two further variables which debriefing activities carried out during the pilot phase indicated as potentially related to views on dimension importance: when asked about the reasons for valuing highly a certain dimension, in some cases students referred to that dimension being a 'human right' or an 'urgent problem' in the country. In the questionnaire therefore, after students 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 and columns 5-7 (for Specification I and II, respectively) 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 (baseline equation). Columns 4 and 8 refer instead to the joint significance test, 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 (the 'poverty' version being the baseline category), and points to health being valued as more important in a wellbeing framework, whilst the opposite holds for education and housing – in line with the insights gained by the univariate analysis presented above. 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 the control variables mentioned above in Specification II, and are robust to alternative selections of explanatory variables. Our control variables do not show particular associations with respondents' views, nor do they seem to interact with the treatment dummy – the relative socio-demographic homogeneity of university students is likely to play a role in this. Variables related to being a human right and being the most urgent problem in the country are as expected positive and significant in the relevant equations. Finally, in order to avoid more stringent assumptions, in Specifications I and II we have maintained the same set of regressors for all outcome variables – so that, for example, health-related characteristics of our respondents are included also in equations relative to dimension importance scores attached to other dimensions. Results do not show qualitative differences if equations for the importance score of a certain dimension include only regressors closely related to that dimension (available upon request).

[place table 3 here]

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, while it is accepted that individuals hold different views on dimension importance, accounting for these individual-specific views in empirical analysis may conflict with the so-called dominance principle. Suppose that individuals or societies *A* and *B* have different preferences over health and education and a bi-dimensional index is used to compare their multidimensional wellbeing; the introduction of weights reflecting the different importance *A* and *B* attach to health and education may lead to *A*'s computed wellbeing to be lower than *B*'s even if *A* outperforms *B* in both health and education (see Fleurbaey and Trannoy, 2003, Brun and Tungodden, 2004 and Fleurbaey, 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 possibly having to accept that *B* is better off than *A*, or adopting a paternalistic approach where individual preferences are silenced and the dominance principle is preserved.

Consider now a 'concordance principle' stating that if individual or society *C* has more poverty than *D*, then she must also have less wellbeing than *D*. While this principle may appear as hardly questionable, it is easy to think of a situation where multidimensional evaluation would

bring about a deviation from this principle. 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 generally be greater than zero;⁸ at the same time, an array of wellbeing indices would quantify C 's overall wellbeing as greater than D 's given C 's high levels of achievements in all but one dimension. Clearly, in this case the deviation from the concordance principle originates in neglect for achievements above the poverty line in the conceptualisation and measurement of absolute poverty but not of wellbeing.

In order to mute this source of heterogeneity, think now about assessing multidimensional poverty and wellbeing of individuals or societies E and F who are both below the poverty lines in each of the two dimensions of interest. With a poverty line set at 10 for both health and education, achievement scores for E and F are respectively $(E_{edu}=3, E_{health}=4)$ for E and $(F_{edu}=5, F_{health}=2)$ for F . In addition, in order to rule out the role of the indices' functional forms, suppose that we assess both wellbeing and poverty through additive indices based on smooth linear functions – i.e. the sum of individual achievement scores and the sum of poverty gaps, respectively (results do not qualitatively change if a multiplicative form is used). In this way, not only are we comparing in terms of multidimensional poverty and wellbeing two individuals or groups who are both below the poverty line in all dimensions, but we are also doing this using well-behaved and (symmetrically) identical indices for poverty and wellbeing. It is clear that under equal weighting E and F will be deemed to be equal in terms of poverty and wellbeing, and that if health (education) were given more importance across the board then E (F) would be deemed to have less poverty and more wellbeing. Would we instead be ready to accept a 'discordance' between the poverty and wellbeing assessments, namely the conclusion that one individual between E or F has both more poverty and more wellbeing than the other?

As the following example shows, this conclusion is indeed possible if the importance attached to a certain dimension is allowed to vary not only across individuals but also across the constructs of poverty and wellbeing. Suppose E 's weights are $(Ew^P_{edu}=.4, Ew^P_{health}=.6)$ in the case of poverty and $(Ew^W_{edu}=.4, Ew^W_{health}=.6)$ in the case of wellbeing, while for F these are respectively $(Fw^P_{edu}=.5, Fw^P_{health}=.5)$ and $(Fw^W_{edu}=.6, Fw^W_{health}=.4)$. Using these importance

⁸ It should be noted that it may still be zero in the case of an overall poverty line defined in the multidimensional space rather than defining separate poverty lines for each dimension

scores as simple multiplicative weights, F would result as having both more poverty and more wellbeing than E :

$$E_{poverty} = (10 - 3) \times .4 + (10 - 4) \times .6 = 6.4 < F_{poverty} = (10 - 5) \times .5 + (10 - 2) \times .5 = 6.5$$

$$E_{wellbeing} = (3) \times .4 + (4) \times .6 = 3.6 < F_{wellbeing} = (5) \times .6 + (2) \times .4 = 3.8$$

It should be noted that this seemingly paradoxical conclusion arose in an illustration where we adopted very simple functional forms for the assessment of poverty and wellbeing, and where we employed sets of weights which are quantitatively very similar to one another. The seemingly paradoxical conclusion is, therefore, a possible outcome when we have reasons to believe that a difference in dimension importance scores between a poverty and a wellbeing framework (whether large or small) really exists. The existence of such difference is indeed the indication we derive from our questionnaire experiment. Statistical significance of our treatment effect indicates that the difference in dimension importance scores between a poverty and a wellbeing frameworks is not due to chance but it reflects a real feature of our respondents' views.

Should this seemingly paradoxical conclusion be accepted or rejected? The implication of accepting it would be that poverty and wellbeing would be seen not as two sides of the same coin but as two distinct phenomena. Poverty and wellbeing would be different aspects of an individual's condition, and more of one would not necessarily imply less of the other – an individual could have both more poverty and more wellbeing than another individual in a similar fashion as she can have more cholesterol and more eyesight. The strongest grounds to reject the seemingly paradoxical conclusion possibly reside in the idea that importance scores for multidimensional evaluation should not be taken as fixed, but they should be allowed to 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 view, 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. Since the achievements of E and F are below the poverty line, the appropriate sets of weights for them 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 to individuals below the poverty line, for whom only the only the construct of poverty would apply.

5. ASSESSING MULTIDIMENSIONAL POVERTY AND WELLBEING WITH DIFFERENT WEIGHTS

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 for us to obtain such a large sample of university students.

In Figure 2 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)⁹. More generally, it is useful to keep in mind that these two samples are likely to be more informative on the values held by higher rather than lower socioeconomic profiles. The views expressed by the ‘development experts’ show the largest gap between health and education on the one hand and housing and safety on the other; when asked to explain the reason for such a disparity, respondents often evoked the idea of health and education being central to the notion of human development. A final remark on the importance scores obtained is that if the scores for housing and safety are summed up, a pattern similar to the $\frac{1}{3}; \frac{1}{3}; \frac{1}{3}$ structure of the widely adopted Human Development Index (UNDP, 1990) and the Multidimensional Poverty Index (Alkire and Santos, 2010) emerges – with similar weights for an educational domain, a health domain and a ‘living conditions’ domain. This is clearly an ex-post judgement and in our case it may be the result of employing two highly valued dimensions (education and health) and two considerably less valued ones (housing and safety). Whether this importance scores structure would be observed also on the bases of different domains or of a larger number of domains is however unclear.

[place figure 2 here]

Moving to multidimensional evaluation, in Figure 3 we report the percentage change in multidimensional poverty and wellbeing¹⁰ in the Dominican Republic between 1997 and 2007 (upper panel) as well as variations in specific dimensions (lower panel). In both the cases 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. The

⁹ The scope for assessing the statistical significance of the differences in dimension importance scores across the three samples is limited, due to the difference in the data collection instruments and the sample size of the ‘expert’ sample. We however flag that *t*-tests on the heterogeneous sample scores where student importance scores are taken as the hypothesised values show that there is a statistical difference in the case of health and housing ($p < 0.0001$) but not in the case of education and safety.

¹⁰ Headcount ratios and linear additive indices are used for the evaluation of poverty. Similar results are obtained using different approaches such as multiplicative indices. As a limitation of our analysis, it should be considered that we are able to use two sets of weights (poverty-specific and wellbeing-specific) only with regard to our student sample and not for our heterogeneous and the ‘expert’ samples – this is because our incentivised experiment was carried out only with students.

variations in specific dimensions shed some light on this, with a poverty decrease in all dimensions except for health on the one hand and a wellbeing increase for education and housing but a decrease for health and safety. It is also interesting to note that, in both the cases of multidimensional poverty and multidimensional wellbeing, the rosier picture on the social development trend in the Dominican Republic is obtained by using equal weights. The simple adoption of equal weights, therefore, would have made us more optimistic than we probably should have been.

[place figure 3 here]

6. CONCLUSION

In this paper, we address two aspects related to dimension weights in multidimensional evaluation. The first moves from the consideration that although for decades various dimensions (education, health, etc.) have featured in indices representing different constructs (poverty, wellbeing, ‘development’, etc.), as yet there is no evidence on whether views on dimension importance would differ depending on the construct. We address this issue through a between-subject unincentivised experiment with random allocation to treatment, carried out with a large sample of university students in the Dominican Republic. We find that respondents’ views are significantly different if dimensions are presented as ‘dimensions of poverty’ or ‘dimensions of wellbeing’. We also show how this evidence can lead to what we label a ‘concordance paradox’, namely the possibly disturbing conclusion that one individual can be deemed to have at the same time more poverty and more wellbeing than another – even in the case that they are both below the poverty line in every dimension. As we argued, this paradox suggests alternative takes on the very essence of the concepts of ‘poverty’ and ‘wellbeing’, namely whether they should be thought of as two different phenomena or they may still be thought of as two sides of the same coin along a continuum with dimension weights changing along the achievement line. Further research is needed to fully expound the implications of our finding and of this seemingly paradoxical conclusion, as well as to clarify what these mean for the co-existence of the notions of poverty and wellbeing.

As a second offer of our paper, we test whether the use of alternative sets of weights brings about appreciable differences in poverty and wellbeing assessment. Using national survey data, we analyse the changes in multidimensional poverty and wellbeing in the Dominican Republic between 1997 and 2007 on the basis of equal weights and of the three sets of weights we collected in the field. We find that the use of different weights leads to opposite conclusions on

the trend of multidimensional poverty and wellbeing, making it difficult to answer the question “have poverty and wellbeing gone up or down?” This urges researchers to take seriously the issue of ‘who decides’ upon the relative importance of life domains in multidimensional evaluation and how to reach a consensus among different sectors of society.

REFERENCES

- Alkire, S. and M.E. Santos, “Acute multidimensional poverty: A new index for developing countries,” OPHI Working Paper Series No. 38, Oxford University, 2010.
- Alkire, S. and J. E. Foster, “Counting and multidimensional poverty measurement,” *Journal of Public Economics*, 95, 476-487, 2011a.
- , “Understandings and misunderstandings of multidimensional poverty measurement,” *Journal of Economic Inequality*, 9, 289-314, 2011b.
- Alkire, S., J.E. Foster, S. Seth, M.E. Santos, J.M. Roche and P. Ballon, *Multidimensional Poverty Measurement and Analysis*, Oxford University Press, Oxford, 2015.
- Angelini, V., D. Cavapozzi, L. Corazzini and O. Paccagnella, “Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases,” *Oxford Bulletin of Economics and Statistics* 76, 643–666, 2014.
- Athanassoglou, S., “Multidimensional welfare rankings under weight imprecision: a social choice perspective,” *Social Choice and Welfare*, 44, 719-744, 2015.
- Atkinson, A.B. and F. Bourguignon, “The comparison of multi-dimensioned distributions of economic status,” *Review of Economic Studies*, 49, 183-201, 1982.
- BCRD, Nueva canasta para medir la inflación IPC Base Diciembre 2010=100, Banco Central de la República Dominicana, Aviso 3 de Febrero 2011, Santo Domingo, 2011.
- Beegle, K., K. Himelein and M. Ravallion, “Frame of Reference Bias in Subjective Welfare Regressions,” *Journal of Economic Behaviours and Organization*, 81, 556-579, 2012.
- Belhadj, B. and M. Limam, “Unidimensional and multidimensional fuzzy poverty measures: new approach,” *Economic Modelling*, 29, 995–1002, 2012.
- Bossert, W., S.R. Chakravarty and C. D’Ambrosio, “Multidimensional poverty and material deprivation with discrete data,” *Review of Income and Wealth*, 59, 29-43, 2013.
- Bourguignon, F. and S.R. Chakravarty, “The measurement of multidimensional poverty,” *Journal of Economic Inequality*, 1, 25-49, 2003.
- Brown, M.B. and A.B. Forsythe, “Robust tests for the equality of variances,” *Journal of the American Statistical Association*, 69, 364–367, 1974.
- Brun, B.C. and B. Tungodden, “Non-welfaristic theories of justice: Is ‘the intersection approach’ a solution to the indexing impasse?,” *Social Choice and Welfare* 22, 49-60, 2004.
- Burchi, F., P. De Muro and E. Kollar, “Constructing Well-Being and Poverty Dimensions on Political Grounds,” *Social Indicators Research*, 137, 441-462, 2018.
- Cameron, A.C. and P.K. Trivedi, *Microeconometrics Using Stata*, Rev. ed. College Station, TX: Stata Press, 2010.

- Camfield, L. and D. Ruta, “‘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, 2007.
- Campbell, A., P.E. Converse and W.L. Rogers *The Quality of American Life: Perceptions, Evaluations, and Satisfaction*, Russel Sage, New York, 1976.
- Chakravarty, S., J. Deutsch and J. Silber, “On the Watts Multidimensional Poverty Index and its Decomposition,” *World Development* 36, 1067-1077, 2008.
- Chakravarty, S. and N. Chattopadhyay, "Multidimensional poverty and material deprivation: theoretical approaches," in C. D'Ambrosio ed., *The Handbook of Research on Economic and Social Wellbeing*, 153-170, Edward Elgar, London, 2018.
- Cherchye, L., E. Ooghe and T. Puyenbroeck, “Robust human development rankings,” *Journal of Economic Inequality*, 6, 287–321, 2008.
- Chiappero-Martinetti, E. and N. von Jacobi, “Light and shade of multidimensional indexes: how methodological choices impact on empirical results,” in Maggino F. and G. Nuvolati, eds., *Quality of life in Italy: research and reflections*, 69-106, Springer, 2012.
- Choo, E.W., B. Schoner and W.C. Wedley, “Interpretation of Criteria Weights in Multicriteria Decision Making,” *Computers & Industrial Engineering*, 37, 527-541, 1999.
- Crawford, G. and C. Williams, “A note on the analysis of subjective judgment matrices,” *Journal of Mathematical Psychology*, 29, 387–405, 1985.
- Dasgupta, P. and M. Weale, “On Measuring the Quality of Life,” *World Development*, 20, 119-131, 1992.
- Decancq, K. and M.A Lugo, “Weights in Multidimensional Indices of Well-Being: an Overview,” *Econometric Reviews*, 32, 7-34, 2013.
- Decancq, K., L. Van Ootegem and E. Verhofstadt, “What if we voted on the weights of a multidimensional well-being index? An illustration with Flemish data,” *Fiscal Studies*, 34, 315-332, 2013.
- Decancq, K., M. Fleurbaey and E. Schokkaert, “Inequality, Income, and Well-Being,” in A. B. Atkinson and F. Bourguignon, eds., *Handbook of Income Distribution*, 67-140, Elsevier, 2015a.
- , “Happiness, equivalent incomes, and respect for individual preferences,” *Economica*, 82, 1082-1106, 2015b.
- Donohue, C. and E. Biggs, “Monitoring socio-environmental change for sustainable development. Developing a multidimensional Livelihoods Index (MLI),” *Applied Geography*, 62, 391-403, 2015.
- Duclos, J.Y., D.E. Sahn and S.D. Younger, “Robust multidimensional poverty comparisons,” *Economic Journal*, 54, 943-968, 2006.
- Esposito, L. and E. Chiappero-Martinetti, “Multidimensional Poverty Measurement: Restricted and Unrestricted Hierarchy among Poverty Dimensions,” *Journal of Applied Economics*, 13, 181-204, 2010.
- Esposito, L., B. Kebede and B. Maddox, “Literacy practices and schooling: A case study from Mozambique,” *World Development*, 39, 1796-1807, 2011.
- , “The Value of Literacy Practices,” *Compare*, 45, 363-380, 2015.
- Feeny, S. and L. McDonald, “Vulnerability to Multidimensional Poverty: Findings from Households in Melanesia,” *Journal of Development Studies*, 3, 447–464, 2016.

- Ferrans, C.E. and M.J. Powers, "Quality of life index: development and psychometric properties," *Advances in nursing science*, 8, 15-24, 1985.
- Ferreira, F.H.G., "Poverty is multidimensional. But what are we going to do about it?," *Journal of Economic Inequality*, 9, 493-495, 2011.
- Ferreira, F.H.G. and M.A. Lugo, "Multidimensional poverty analysis: Looking for a middle ground," *The World Bank Research Observer*, 28, 220-235, 2013.
- Fleurbaey, M., "Social choice and the indexing dilemma," *Social Choice and Welfare*, 29, 633-648, 2007.
- Fleurbaey, M. and A. Trannoy, "The impossibility of a Paretian egalitarian," *Social Choice and Welfare*, 21, 243-263, 2003.
- Foster, J.E., M. McGillivray and S. Seth, "Composite Indices: Rank Robustness, Statistical Association, and Redundancy," *Econometric Reviews*, 32, 35-56, 2013.
- Glejser, H., L. Gevers, P. Lambot and J.A. Morales, "An econometric study of the variables determining inequality aversion among students," *European Economic Review*, 10, 173-188, 1977.
- Guio, A.-C., "Multidimensional poverty and material deprivation: empirical findings," in C. D'Ambrosio ed., *The Handbook of Research on Economic and Social Wellbeing*, 171-192, Edward Elgar, London, 2018.
- Halleröd, B., "The truly poor: direct and indirect consensual measurement of poverty in Sweden," *Journal of European Social Policy*, 5, 111-129, 1995.
- , "Deprivation and poverty: A comparative analysis of Sweden and Great Britain," *Acta Sociologica*, 39, 141-168, 1996.
- Hick, R., "Poverty as capability deprivation: Conceptualising and measuring poverty in contemporary Europe," *European Journal of Sociology*, 55, 295-323, 2014.
- Holland, P.W. and H. Wainer, *Differential Item Functioning*, Hillsdale, NJ, Lawrence Erlbaum Associates, 1993.
- Hsieh, C.M., "Domain Importance in Subjective Well-Being Measures," *Social Indicators Research*, 127, 777-792, 2016.
- Inglehart, R., "Value priorities, life satisfaction, and political dissatisfaction among western publics," *Comparative Studies in Sociology*, 1, 173-202, 1978.
- Jensen, M.C. and W.H. Meckling, "The Nature of Man," *Journal of Applied Corporate Finance*, 7(2), 4-19, 1994.
- Kahneman, D., A.B. Krueger, D. Schkade, N. Schwarz and A. Stone, "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method," *Science*, 306, 1776-1780, 2004.
- Kakwani, N. and J. Silber, *Quantitative approaches to multidimensional poverty measurement*, Palgrave Macmillan, London, 2008.
- Kaypten, A., J.P. Smith and A. van Soest, "Vignettes and Self-Reports of Work Disability in the U.S. and the Netherlands," *American Economic Review*, 97, 461-473, 2007.
- King, G., C.J.L. Murray, J. Salomon and A. Tandon, "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research," *American Political Science Review*, 98, 567-583, 2004.

- Kristensen, N. and E. Johansson, "New Evidence on Cross-Country Differences in Job Satisfaction Using Anchoring Vignettes," *Labour Economics*, 15, 96-117, 2008.
- Linton, M., P. Dieppe and A. Medina-Lara, "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, 2016.
- Liz, R.E., "Diagnóstico de la educación superior en la República Dominicana," *Theorethikos*, 5, 3-56, 2001.
- López, A. and R. Mejía, "Informe Educación Superior en Iberoamérica 2016: República Dominicana," Centro Interuniversitario de Desarrollo, Santiago, 2016.
- Maasoumi, E. and J. Racine, "A solution to aggregation and an application to multidimensional 'well-being' frontiers," *Journal of Econometrics*, 191, 374-383, 2016.
- Maslow, A.H., "A Theory of Human Motivation," *Psychological Review*, 50, 370-396, 1943.
- Massey P. et al. "Findings Toward a Multidimensional Measure of Adolescent Health Literacy," *American Journal of Health Behavior*, 37, 342-350, 2013.
- Markowski, C.A. and E.P. Markowski, "Conditions for the effectiveness of a preliminary test of variance," *American Statistician*, 44, 322-326, 1990.
- Mascherini, M. and B. Hoskins, "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, 2008.
- Mayer, S.E. and C. Jencks, "Poverty and the distribution of material hardship," *Journal of Human Resources*, 24, 88-114, 1989.
- Mitra, S., K. Jones, B. Vick, D. Brown, E. McGinn and M.J. Alexander, "Implementing a multidimensional poverty measure using mixed methods and a participatory framework," *Social indicators research*, 110, 1061-1081, 2013.
- Moldan, B. and S. Billharz, *Indicators of Sustainable Development*, John Wiley, Chichester, 1997.
- Morris, D., *Measuring the Conditions of the World Poor, the Physical Quality of Life Index*, Pergaman Press, New York, 1979.
- OECD, "La Educación Superior En La República Dominicana," Organization for Economic Development and Cooperation, Paris, 2012.
- ONE, "Boletín Mensual de la Oficina Nacional de Estadística," Panorama Estadístico - n.º 54, Oficina Nacional de Estadística, Santo Domingo, 2013.
- Pasha, A., "Regional Perspectives on the Multidimensional Poverty Index," *World Development*, 94, 268-285, 2017.
- Pattanaik P., S. Reddy and Y. Xu, "On measuring deprivation and living standards of societies in a multi-attribute framework," *Oxford Economic Papers*, 64, 43-56, 2012.
- Permanyer, I., "Assessing the Robustness of Composite Indices Rankings," *Review of Income and Wealth*, 57, 306-326, 2011.
- , "Uncertainty and Robustness in Composite Indices Rankings," *Oxford Economic Papers*, 64, 57-79, 2012.
- , "Assessing individuals' deprivation in a multidimensional framework," *Journal of Development Economics*, 109, 1-16, 2014.

- Philip, E.J., T.V. Merluzzi, A. Peterman and L.B. Cronk, "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, 2009.
- Ravallion, M., "On multidimensional indices of poverty," *Journal of Economic Inequality*, 9, 235-248, 2011.
- , "Troubling trade-offs in the Human Development Index," *Journal of Development Economics*, 99, 201-209, 2012.
- Rawls, J., *A Theory of Justice*, Harvard University Press, Cambridge MA, 1971.
- Royston, P., "An extension of Shapiro and Wilks's W test for normality to large samples," *Applied Statistics* 3, 115-124, 1982.
- , "Approximating the Shapiro-Wilk W-test for non-normality," *Statistics and Computing*, 2, 117-119, 1992.
- Russell, L.B., A.M. Hubley, A. Palepu and B.D. Zumbo, "Does weighting capture what's important? Revisiting subjective importance weighting with a quality of life measure," *Social Indicators Research*, 75, 141-167, 2006.
- Saaty, R.W., *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw-Hill, 1980.
- , "The analytic hierarchy process- what it is and how it is used," *Mathematical Modelling*, 9, 161-176, 1987.
- Saisana, M., Saltelli, A. and S. Tarantola, "Uncertainty and sensitivity analysis as tools for the quality assessment of composite indicators," *Journal of the Royal Statistical Society, Series A*, 168, 1-17, 2005.
- Salomon, J., A. Tandon and C.J.L. Murray, "Comparability of Self-Rated Health: Cross Section Multi-Country Survey Using Anchoring Vignettes," *British Medical Journal*, 328, 258-261, 2004.
- Schang, L., Y. Hynninen, A. Morton and A. Salo, "Developing robust composite measures of healthcare quality—Ranking intervals and dominance relations for Scottish Health Boards," *Social Science & Medicine*, 162, 59-67, 2016.
- Schenkerman, S., "Use and abuse of weights in multiple objective decisions support models," *Decision Science*, 2, 368-378, 1991.
- Sen, A.K., *Inequality Re-examined*, Oxford University Press, Oxford, 1992.
- Seth, S. and M. McGillivray, "Composite indices, alternative weights, and comparison robustness," *Social Choice and Welfare*, 52, 657-679, 2018.
- Shapiro, S.S. and M.B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, 52, 591-611, 1965.
- Stapleton, L. and G.D. Garrod, "Keeping things simple: Why the Human Development Index should not diverge from its equal weights assumption," *Social Indicators Research*, 84, 179-188, 2007.
- Trani, J.F., J. Kuhlberg, T. Cannings and D. Chakkal, "Multidimensional poverty in Afghanistan: who are the poorest of the poor?," *Oxford Development Studies*, 44, 220-245, 2016.
- Trauer, T. and A. Mackinnon, "Why are we weighting? The role of importance ratings in quality of life measurement," *Quality of Life Research*, 10, 579-85, 2001.

- Tsui, K.Y., “Multidimensional poverty indices,” *Social Choice and Welfare*, 19, 69-93, 2002.
- UNDP, *Human development report 1990. Concept and Measurement of Human Development*, Oxford University Press, New York, 1990.
- USAID, “Dominican Republic Country Development Cooperation Strategy,” FY 2014-2018, 2013 (<https://www.usaid.gov/dominican-republic/cdcs>).
- Victor B. et al., “Multidimensional Poverty in Rural Mozambique: A New Metric for Evaluating Public Health Interventions,” *PLoS ONE* 9: e108654, doi:10.1371/journal.pone.0108654, 2014
- Wu, C.H., C.T. Yang and L.N. Huang, “On the predictive effect of multidimensional importance-weighted quality of life scores on overall subjective well-being,” *Social indicators research*, 115, 933-943, 2014.
- Yalonetzky, G., “Conditions for the most robust multidimensional poverty comparisons using counting measures and ordinal variables,” *Social Choice and Welfare*, 43, 773-807, 2014.
- Zellner, A., “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias,” *Journal of the American Statistical Association*, 57, 348-368, 1962.
- , “Estimators for Seemingly Unrelated Regressions: Some Exact Finite Sample Results,” *Journal of the American Statistical Association*, 58, 977-992, 1963.
- Zheng, B. and C. Zheng, “Fuzzy ranking of human development: A proposal,” *Mathematical Social Sciences*, 78, 39-47, 2015.

TABLE 1.
DIRECT AND INDIRECT APPROACHES TO DERIVING WEIGHTS

<i>Direct</i>	<i>Indirect</i>
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

TABLE 2
STUDENT 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
	<i>p-values</i>	<i>0.0873^a; 0.0070^b</i>				
Health	Wellbeing	32.80194	10.70872	1	80	515
	Poverty	29.87662	10.24142	0	80	539
	<i>p-values</i>	<i>0.0000^a; 0.0062^b</i>				
Housing	Wellbeing	18.08544	7.895443	1	50	515
	Poverty	19.15492	8.560664	0	50	539
	<i>p-values</i>	<i>0.0393^a; 0.1675^b</i>				
Safety	Wellbeing	17.77184	6.8023	1	45	515
	Poverty	18.18646	7.614493	0	50	539
	<i>p-values</i>	<i>0.5052^a; 0.3482^b</i>				

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).

TABLE 3
MULTIVARIATE ANALYSIS - ZELLNER'S SEEMINGLY UNRELATED REGRESSIONS

	Specification I ^a				Specification II ^b			
	(1) <i>Edu</i>	(2) <i>Health</i>	(3) <i>Housing</i>	(4) Joint test (chi-2)	(5) <i>Edu</i>	(6) <i>Health</i>	(7) <i>Housing</i>	(8) Joint test (chi-2)
<i>Questionnaire version (treatment)</i>								
Wellbeing	-1.484** (0.715)	2.870*** (0.645)	-1.055** (0.511)	20.16***	-1.402** (0.687)	2.657*** (0.633)	-0.886* (0.515)	17.78***
<i>Demographics</i>								
Age	0.085 (0.068)	-0.060 (0.061)	0.033 (0.048)	3.96	-0.003 (0.072)	-0.022 (0.067)	0.058 (0.054)	1.77
Male	1.343 (0.848)	-0.427 (0.765)	-0.443 (0.607)	2.55	1.066 (0.823)	-0.254 (0.759)	-0.356 (0.617)	1.74
<i>Disciplines</i>								
Architecture	-1.145 (1.081)	0.641 (0.976)	0.916 (0.774)	35.83***	-0.107 (1.159)	0.984 (1.068)	0.213 (0.868)	21.49**
Medicine	-0.405 (1.038)	4.110*** (0.936)	-1.346* (0.743)		0.115 (1.008)	3.279*** (0.929)	-1.017 (0.755)	
Education	-0.488 (1.049)	-0.009 (0.946)	0.384 (0.751)		0.025 (1.057)	0.517 (0.974)	0.311 (0.792)	
<i>Dimension which most should be seen as a human right</i>								
Education					4.573*** (1.339)	-0.243 (1.234)	0.282 (1.003)	134.57***
Health					-1.901 (1.352)	5.391*** (1.246)	0.303 (1.013)	
Housing					-1.258 (1.753)	1.884 (1.615)	3.104** (1.313)	
<i>Dimension as the most urgent problem in the country</i>								
Education					4.225*** (0.873)	0.887 (0.804)	-2.443*** (0.654)	90.13***
Health					-1.479 (1.036)	3.956*** (0.955)	-1.082 (0.776)	
Housing					0.176 (1.248)	-0.160 (1.150)	2.444*** (0.935)	
Constant	30.637*** (1.689)	30.412*** (1.689)	18.478*** (1.340)		29.994*** (3.031)	24.083*** (2.792)	19.994*** (2.270)	
N	1,030	1,030	1,030		974	974	974	
Equation significance	0.0446	0.0000	0.0153		0.0000	0.0000	0.0000	
Breusch- Pagan test		0.0000				0.0000		

Notes.

^a: no control variables included other than those reported.

^b: additional control variables include 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) and subjective socio-economic status variables (perceived family income and perceived relative standard on living).

Figure. 1 Distribution of student importance scores

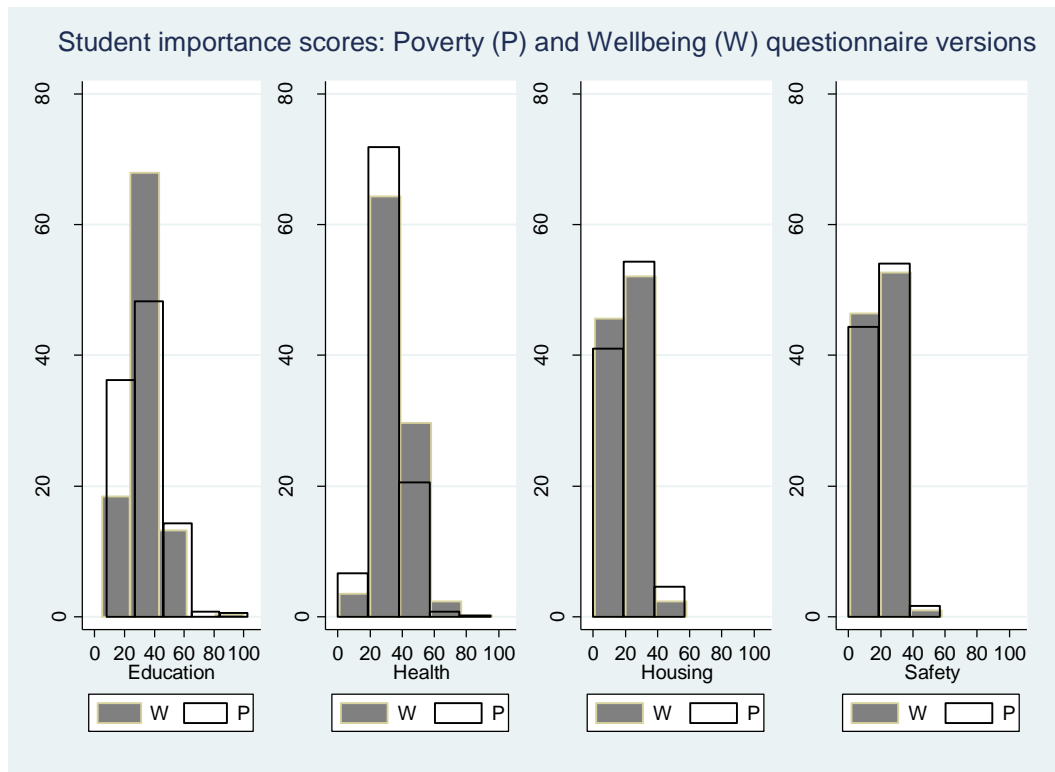


Figure 2. Dimension importance scores across samples
Notes: Importance scores for the heterogeneous sample are normalised to 100

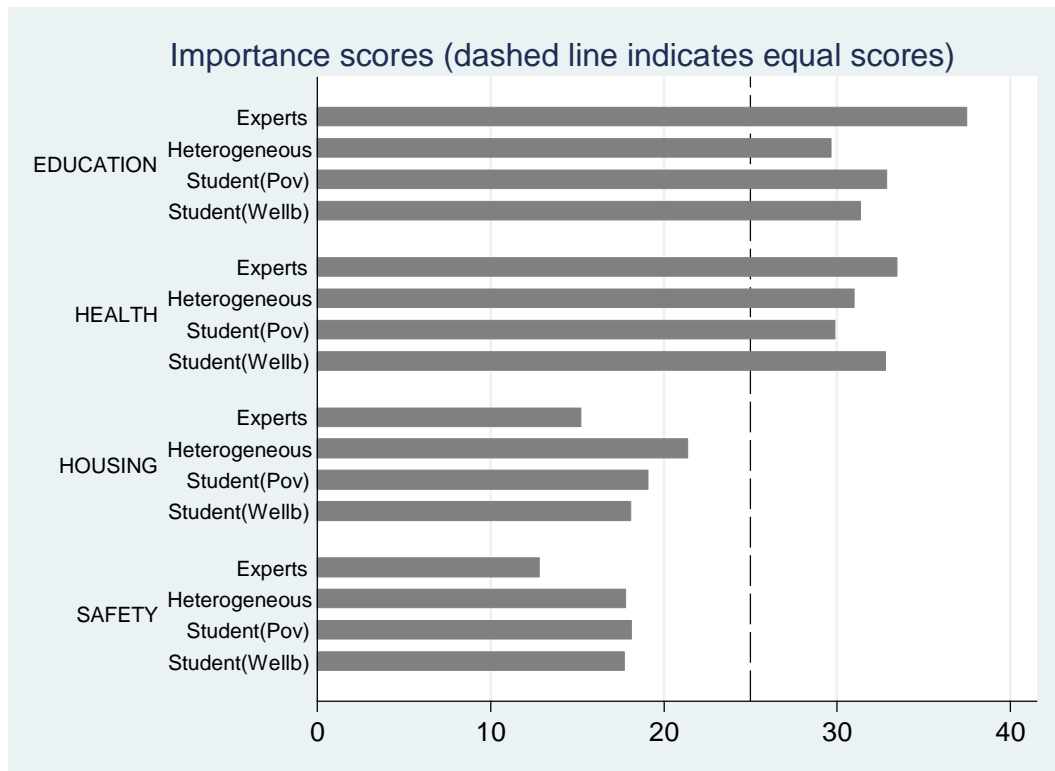
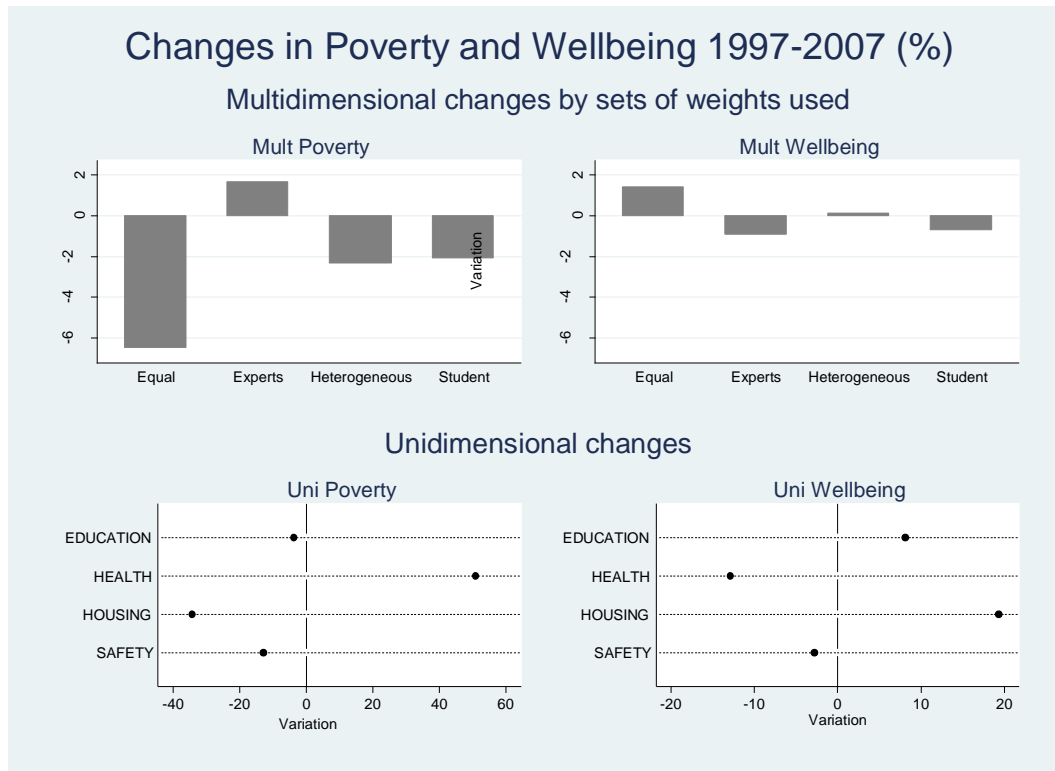


Figure 3. Change in multidimensional poverty and wellbeing 1997-2007



APPENDIX 1. RANDOMIZATION CHECK

Below are the tests for the null hypothesis of a significant difference between the two subsamples (one receiving the poverty version of our questionnaire and the other receiving the wellbeing version). Test of proportions, *t*-test and Wilcoxon Mann Whitney test performed depending on the nature of the variables.

TABLE A1.
REJECTION OF THE NULL HYPOTHESIS
OF A SIGNIFICANT DIFFERENCE BETWEEN THE TWO STUDENT SUBSAMPLES

<i>Binary variables</i>	<i>p-value (test of proportions)</i>
Female	.5942
Architecture degree	.8214
Law degree	.9404
Medicine degree	.8362
Education degree	.6037
Victim of home robbery	.2766
Victim of street robbery	.7253
Victim of assault	.2164
Victim of verbal threat	.5611
Home owner	.5085
Own experience of serious illness	.0898
Serious illness experienced by parents	.7645
Serious illness experienced by siblings	.3054
Serious illness experienced by relatives	.8319
Serious illness experienced by partner	.9307
Serious illness experienced by close friends	.3030
<i>Continuous variables</i>	<i>p-value (t-test)</i>
Age	.9229
Semester of study	.2283
<i>Ordinal variables</i>	<i>p-value (Wilcoxon-Mann-Whitney rank-sum test)</i>
Perceived family income ^a	.4065
Perceived relative standard of living ^b	.7596

^aPerceived family income on a 5-point Likert scale from 'Very low' to 'Very high'.

^bPerceived family standard of living compared to other families on a 5-point Likert scale from 'Much lower' to 'Much higher'.

APPENDIX 2. INDICATORS USED

TABLE A2.

DATA DESCRIPTION, POVERTY THRESHOLDS AND WELLBEING 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. illiterate 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)	<u>Indicator thresholds:</u> 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 =1 st quartile <u>Housing poverty threshold:</u> 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, wellbeing scores were assigned to people living in the same region (*estrato*) x_i ($i=1, \dots, 10$) 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

[Table A2. Continued]

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 nacional de ingresos y gastos de hogares”, Departamento de Encuesta, Oficina Nacional de Estadística, República 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”?;)