

Women's Empowerment in Tunisia: A Discrete Choice Experiment to Elicit Weights for a Multidimensional Index

Natalie Nairi Quinn*

13 March 2018

[PRELIMINARY DRAFT: DO NOT CITE OR CIRCULATE]

Abstract

Multidimensional indices are widely used to measure socioeconomic characteristics that are not directly measurable. The choice of dimensions, indicators and aggregation function determine the meaning and interpretation of a multidimensional index. Where the index appropriately reflects a population's values or preferences, I argue that those preferences should be elicited to inform the choice of dimensions, indicators and aggregation function. Standard approaches to preference elicitation are not well suited to implementation, in the context of a field survey in a developing country, for the large number of indicators typically identified for multidimensional indices. An innovative experimental design that addresses these issues was implemented to elicit perceptions of empowerment among women in Tunisia. Estimation of a parametric model generates weights for an index of Women's Empowerment that reflects the perceptions and values of the survey respondents. The conventional assumption of equal weights is decisively rejected. The empirical approach resolves, in this context, an issue that emerges in the methodological literature on multidimensional indices: how to weight correlated indicators without 'double counting' latent factors.

Keywords: Measurement, Discrete Choice Experiment, Multidimensional Index.

JEL codes: C43, D63, J16

*University of Oxford; natalie.quinn@qeh.ox.ac.uk. Acknowledgements: The data used in this study were collected as part of Oxfam GB's Effectiveness Review of the AMAL project supporting women's transformative leadership in Tunisia. The project was funded by the Swedish International Development Cooperation Agency (SIDA), implemented by Oxfam Novib in conjunction with three local partner organizations, and the Effectiveness Review was funded by Oxfam GB. Additional research costs were funded by St John's College, University of Oxford. Thanks to Simone Lombardini (Global Impact Evaluation Advisor at Oxfam GB) for synergistic collaboration; to Claire Hutchings (Head of Effectiveness and Evaluation at Oxfam GB) who supported the collaboration; to Zeineb Saidani for excellent survey management; and to workshop participants at QEH and CSAE, University of Oxford and Queen Mary, University of London, for helpful feedback and comments. Thanks also to Prof Mohammad-Salah Omri of St John's College for help with back translations and insights on the political context in Tunisia. University of Oxford CUREC reference ECONCIA16-17-004.

1 Introduction

There exists a demand for quantitative measures of many socioeconomic concepts that are not directly, or straightforwardly, measurable. Examples include poverty, human development, corruption and peace. In each case, one or more indices has been proposed that combine information from several imperfect proxies for the concept that is being measured. Such indices may be used to describe development, evaluate programme or policy effectiveness and may even impact resource allocation. Clear dangers arise if an index does not appropriately capture the concept that it purports to measure.

This paper addresses the problem of how multiple indicators, which are imperfect proxies for a concept of interest, may be combined to create a meaningful and informative measure. The key conceptual step is the recognition that any quantitative index represents an ordering of alternative profiles of the different indicators. If an ordering that corresponds appropriately to the concept of interest can be established, then a quantitative index may be constructed that represents that ordering.

In some contexts an individual's, or a group of individuals', values or preferences would be the appropriate ordering. The index constructed in this paper is a measure of empowerment for women in Tunisia;¹ the objective is to construct a measure that reflects the perception of empowerment among those same women whose empowerment is being evaluated. Their perceptions of empowerment were elicited through implementation of a discrete choice experiment (DCE) in the context of a quantitative survey.

This paper contributes to a literature on choosing weights for multidimensional indices; Decancq and Lugo (2013) is a useful review. Rodriguez-Takeuchi (2014) surveys alternative approaches to incorporate people's values in measures of development and identifies DCEs as a promising approach. Two studies (Watson, Sutton, Dibben, and Ryan, 2008; Decancq and Watson, 2015) have implemented DCEs with a view to establishing weights for a multidimensional or composite measure, in quite different contexts.² DCEs are widely used in health economics and market research and there is an extensive literature on methods and applications; Ryan, Gerard, and Amaya-Amaya (2007) and Johnson, Lancsar, Marshall, Kilambi, Mühlbacher, Regier, Bresnahan, Kanninen, and Bridges (2013) are useful surveys.

The paper proceeds as follows. The measurement framework and approach to perception elicitation are described in section 2. The DCE design and its implementation are described in section 3. The empirical strategy and preliminary results are presented in section 4, and section 5 concludes.

¹The study was carried out in collaboration with Oxfam GB, an international NGO that implements development projects in many developing countries.

²Watson, Sutton, Dibben, and Ryan (2008) is conceptually most similar but carried out in the UK, through a dedicated study, and aspects of the DCE design are questionable: respondents are asked to choose between implausible combinations of attributes, and the treatment of unspecified attributes is not clear. The application in Decancq and Watson (2015) is quite different as it establishes weights for an HDI-type index that is a composite of society-level rather than individual-level indicators.

2 Measurement Framework and Approach

This study was conceived in response to a desire to improve measurement methodologies from Oxfam GB. Women's rights is a core priority for Oxfam, and many Oxfam projects focus specifically on the empowerment of women. Oxfam is implementing a programme of quasi-experimental Effectiveness Reviews of a sample of projects, evaluating Women's Empowerment projects through a multidimensional measure of empowerment.

This study was carried out in the context of the Effectiveness Review for Oxfam's AMAL project in Tunisia, which supported "women's transformative leadership" in the period 2012–2015. Project beneficiaries "strengthen their confidence, knowledge of their rights and their campaigning and advocacy skills" to "become leaders who can work with their communities reclaim their right to political and civic participation and achieve long lasting positive change."

2.1 Existing Measurement Framework

Oxfam GB's established approach to the measurement of women's empowerment is documented by Bishop and Bowman (2014) and Lombardini, Bowman, and Garwood (2017). It builds on the *Women's Empowerment in Agriculture Index* (Alkire, Meinzen-Dick, Peterman, Quisumbing, Seymour, and Vaz, 2013), but modified in various ways. Oxfam's indices are context-specific, with empowerment characteristics, indicators and thresholds defined for each country or evaluation. Each index captures the proportion of indicators in which a woman is empowered, with no dichotomisation into 'empowered' and 'not empowered'. Oxfam has adopted a three-fold framework of empowerment, incorporating indicators that capture aspects of empowerment at the personal, relational and environmental levels. The latter includes social norms, policies and laws experienced by the woman.

In the established approach to index construction, the indicators for each of the three levels (personal, relational and environmental) are chosen through a qualitative exercise with stakeholders of the project being evaluated. Each of these indicators is given equal weight in the index.³ Oxfam GB has identified issues that deserve exploration (Lombardini, Bowman, and Garwood, 2017): there is no reason to expect all indicators chosen through the qualitative exercise to be 'equally important', while some indicators might be highly correlated and thus capturing the same underlying concept; does equal weighting lead to over-weighting for these indicators?

In the existing evaluation process, the data required to evaluate a project is gathered through a short (1 hour) quantitative survey of a moderate sample of project beneficiaries and a matched comparison group (with a total sample size of five to six hundred respondents). The survey is carried out on tablet computers, implemented using SurveyCTO software.

Many elements of the existing approach were retained in this study. The conceptual framework (three levels of empowerment) and the qualitative exercise to choose indicators were not

³It has been argued that this results in a certain degree of comparability across different countries and projects, as the index captures the 'proportion of indicators' in which a woman is empowered, regardless of what those indicators are. Cross-context comparability is a difficult issue which is not addressed in this study.

modified. The elicitation of perceptions was integrated into the existing quantitative survey; time and budget constraints meant that less than 10 minutes was available for this module.

2.2 Elicitation Approach

Given the practical constraints of implementation, through a short module in a quantitative survey in a developing country context, the approach needs to be straightforward for both the enumerator and the respondent. Discrete choice experiments are a stated choice method that are widely used in health economics and marketing to elicit preferences and values. The respondent is presented with a set⁴ of hypothetical scenarios (in this application, hypothetical women with different characteristics) and asked to express which she 'prefers' (in this application, considers to be more empowered). This task seemed straightforward enough to implement in the applied context, and feasibility was established through piloting.

Alternative approaches have been used to determine weights for multidimensional measures, but none are appropriate. Qualitative exercises (focus group discussions, budget allocation games) have an important role, for example choosing dimensions and indicators, but cannot determine meaningful aggregation weights (despite having been used in some applications, for example Oxfam's Humankind Index). Some authors advocate statistical approaches (frequencies, Principal Component Analysis, Multiple Correspondence Analysis) in which the weights are determined by relationships between indicators in the sample; there is no reason to expect such relationships to reflect the ordering that one would wish the measure to embody.

3 DCE Design and Implementation

3.1 Design: Indicators

The indicators incorporated in the index correspond to the attributes in DCE terminology. Following Oxfam's usual procedure, indicators of women's empowerment in Tunisia were established through a series of facilitated workshops and discussions with stakeholders in the project. 14 indicators were determined to be relevant for women's empowerment in Tunisia; they were categorised according to Oxfam's Empowerment framework (Lombardini, Bowman, and Garwood, 2017) as Personal (5 indicators), Relational (6 indicators) or Environmental (3 indicators). Each was allowed to take two levels⁵ In order to elicit meaningful information, the levels need to be specified as precisely as possible and also need to be precisely mapped to the survey responses.

⁴Most straightforwardly just two

⁵This is restrictive, but consistent with Oxfam's existing approach. The sample size available did not allow for enough power to explore richer options; I hope to be able to explore the possibilities in future work.

The five personal indicators chosen, and the levels for each:

Personal	Code	LOW	HIGH
Self Confidence and Self Esteem	P1SelfConf	 has low self-esteem and low self-confidence	 has high self-esteem and high self-confidence
Ability to make decisions for herself	P2Decision	 is not able to make decisions for herself	 is able to make decisions for herself
Recognising that all types of violence are not acceptable and overcoming the taboo on talking about it	P3Violence	 does not recognise that all types of violence are unacceptable, and is not able to talk about it	 recognises that all types of violence are unacceptable, and is able to talk about it
Awareness that collective action is more effective	P4AwareCol	 is not aware that collective action is more effective	 is aware that collective action is more effective
Ability to recognise problems (i.e. knowledge of women's rights)	P5KnowRght	 does not have knowledge of women's rights	 has knowledge of women's rights

The six relational indicators chosen, and the levels for each:

Relational	Code	LOW	HIGH
Ability to make decisions in the household	R1DecHouse	 is not able to make decisions within the household	 is able to make decisions within the household
Participation and ability to make decisions in the public sphere	R2DecPubli	 does not participate in, and is not able to make decisions in the public sphere	 participates in, and is not able to make decisions in the public sphere
Participation and ability to influence or make decisions in political sphere	R3DecPolit	 does not participate in, and is not able to make or influence decisions in the political sphere	 participates in, and is able to make or influence decisions in the political sphere
Taking action to stop violence	R4ActViole	 is not able to take effective action to end violence	 is able to take effective action to end violence
Independent income	R5Indeplnc	 does not have an independent source of income	 has an independent source of income
Control over resources in her household	R6ResourHH	 has no control over resources in her household	 has control over resources in her household

The three environmental indicators chosen, and the levels for each:

Environmental	Code	LOW	HIGH
Equality of Opportunity	E1EqualOpp	 lives in a community that does not ensure equal opportunities for women and men	 lives in a community that ensures equal opportunities for women and men
Social Norms	E2SocNorms	 lives in a patriarchal society	 lives in a non-patriarchal society
Legislative protection for women's rights	E3LegRight	 lives in a community where women's rights are not enshrined in law	 lives in a community where women's rights are enshrined in law

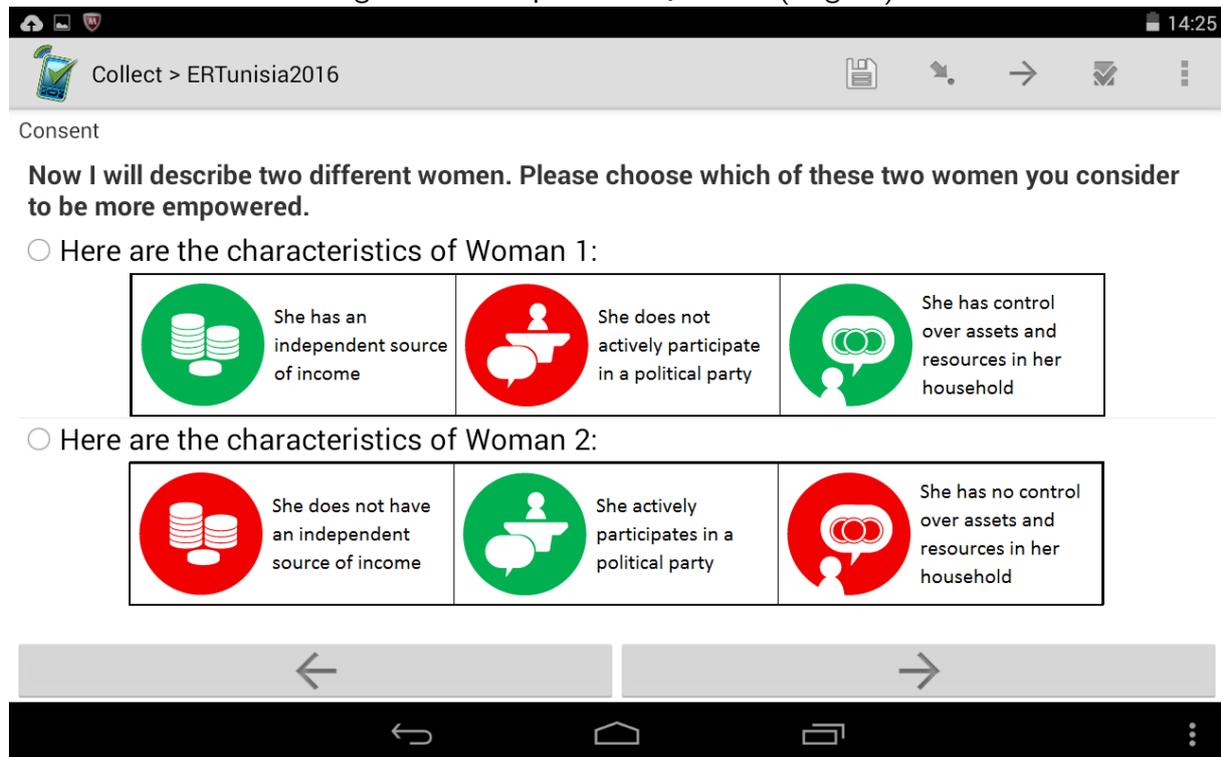
3.2 Design: Profiles and Questions

The standard DCE approach requires the specification of a level for each of the attributes. With 14 attributes (corresponding to the 14 indicators) the complexity would make field implementation in a developing country context infeasible: the cognitive demand on respondents would be excessive while the level of detail required would be impossible to present on the tablet computers.

Through piloting it was established that the most complex profiles feasible to implement specify just three of the 14 indicators. This meant that it was necessary to present respondents with choices between partial profiles in which just three indicators were specified. In order to simplify the task and to elicit the most useful information, all DCE questions required the respondent to compare a pair of profiles in which the *same* indicators were specified. Treatment of the unspecified indicators at the estimation stage is discussed in section 4 below.

An example 3-indicator question is illustrated in English and Arabic in Figures (1) and (2).

Figure 1: Example DCE Question (English).



3.3 Design: Allocation of Pairwise Comparisons

It was necessary to determine what combinations of indicators to use. There are 364 possible combinations of three indicators from 14; far too many to explore in a survey with just 10 minutes available with each of 500 respondents. An ad hoc design procedure was followed, with the DCE questions presented to each respondent falling into three sections.

In section A, each respondent was presented with one question in which they were asked to compare the simplest possible pair of profiles, in which just one indicator was specified.

Figure 2: Example DCE Question (Arabic).



This provided a straightforward introduction to the DCE questions for respondents. It also allows me to assess the extent to which the exercise was understood by respondents, and to determine whether respondents agree with the assumptions made about which of the levels for each indicator corresponds to greater empowerment. With 14 indicators, each of which could be presented high-level first or low-level first, there were 28 versions of section A. Each respondent was allocated one version uniformly at random.

Section B focussed on the six relational indicators as these were considered most pertinent to empowerment and to the project objectives. There are 20 possible combinations of three indicators from six; six of these combinations were chosen as illustrated in Figure (3) and each respondent was allocated one of these six combinations uniformly at random. The

Figure 3: Design: DCE B Indicator Combinations

1	 R1DecHouse	 R5IndepInc	 R3DecPolit
2	 R5IndepInc	 R3DecPolit	 R6ResourHH
3	 R3DecPolit	 R6ResourHH	 R2DecPubli
4	 R6ResourHH	 R2DecPubli	 R4ActViole
5	 R2DecPubli	 R4ActViole	 R1DecHouse
6	 R4ActViole	 R1DecHouse	 R5IndepInc

combinations were chosen to match political activity with both economic indicators and to

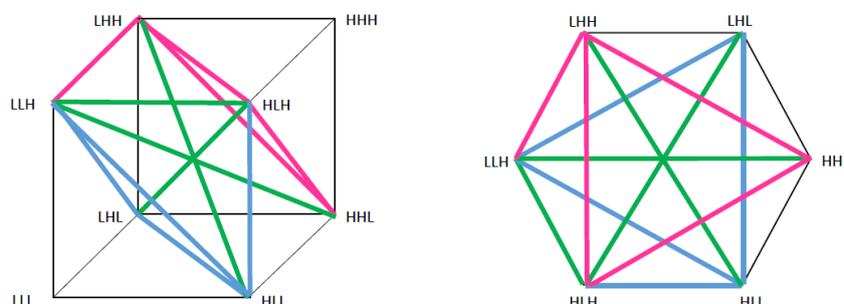
Table 1: The three groups of pairwise comparisons in Section B

		Design B1	Design B2	Design B3
Question 1	Woman 1:	H L H	L L H	L L H
	Woman 2:	H L L	L H H	H L H
Question 2	Woman 1:	H L L	L H H	H L H
	Woman 2:	L H L	H L H	L H L
Question 3	Woman 1:	L H L	H L H	H L L
	Woman 2:	L L H	H H L	L H H
Question 4	Woman 1:	L L H	H H L	H H L
	Woman 2:	H L L	L H H	L L H

match public and political activity, as it was considered essential to be able to determine the tradeoffs between these pairs of indicators. To the extent possible, implausible combinations were avoided to maximise the amount of meaningful data that could be collected.

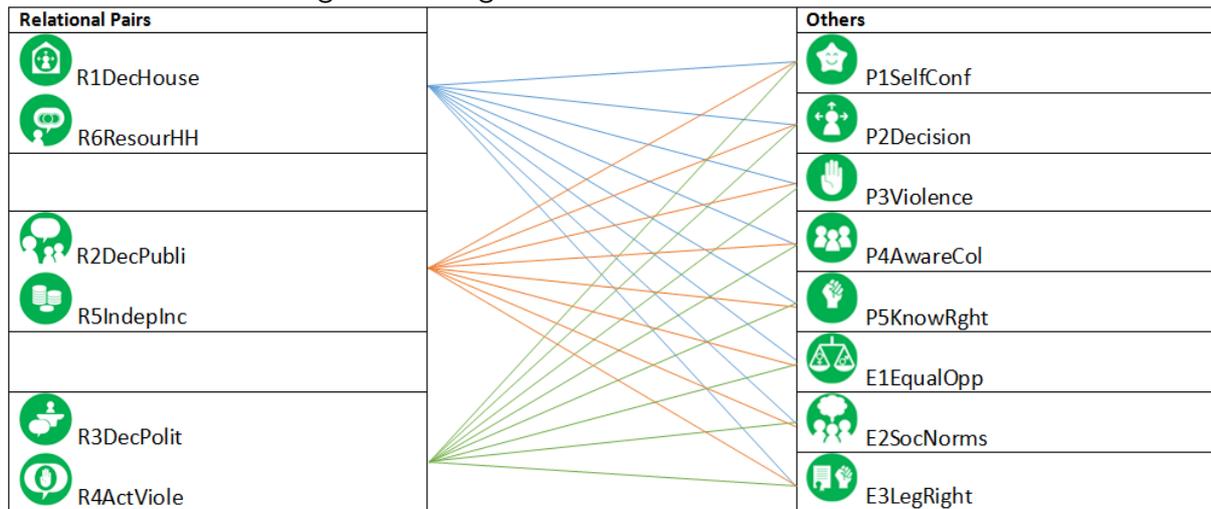
Given three indicators, each of which can take one of two levels (high or low), there are eight possible profiles among which there are 28 unique pairwise comparisons. Time constraints precluded asking each respondent to make all 28 pairwise comparisons for their allocated combination of indicators. However, the domain can be reduced substantially without losing information as given a clear empowerment ordering of levels for each indicator, many of the pairwise comparisons will not be informative. For example, if High corresponds to 'more empowered' for each indicator, a woman with the profile High-High-Low should unambiguously be considered more empowered than a woman with the profile High-Low-Low over the same three indicators. Removing the uninformative comparisons reduces the number of informative comparisons to 9. This was still too many to cover with each respondent, so the 9 pairs were divided into three groups of three (labelled 'Design B1' to 'Design B3'.) to which respondents were allocated uniformly at random. To each group was added (as the first question) one of the uninformative comparisons. This provided the most straightforward introduction to the three-indicator profile comparisons for the respondents. It also yielded further data that was useful for checks. The pairwise comparisons that constitute each of the three designs are documented in Table (1) and illustrated in Figure (4). Given six combinations of indicators and three designs, in total eighteen versions of section B were implemented.

Figure 4: The three groups of pairwise comparisons in Section B



In Section C, time constraints allowed two more pairwise comparison questions to be asked of each respondent, which were used to determine the role of the personal and environmental indicators (8 in total). These indicators were combined with relational indicators (the three pairs that had not been covered in Section B) and thus valued in relation to the relational

Figure 5: Design: DCE C Indicator Combinations



indicators. All 24 combinations of the 8 personal and environmental indicators with the three pairs of relational indicators were included (Figure 5); respondents were allocated two of these combinations uniformly at random and asked one of the 9 informative pairwise comparison questions for each combination allocated.

Implementation of this design in SurveyCTO was achieved by automating the code-writing; the SurveyCTO code as well as Imagemagick code for processing the illustrative image files was generated by running a Stata do-file as documented in Appendix A.

4 Empirical Strategy and Preliminary Results

4.1 Tests of Monotonicity and Consistency

Section A of the DCE (comparisons of one-indicator profiles) allows me to determine the extent to which respondents agreed with the assumed empowerment ordering of levels for each of the 14 indicators. Each respondent was allocated one of the 14 indicators, with Woman 1 allocated either the High level or the Low level of that indicator. Just 15 of the 505 respondents (2.97%) disagreed with the assumed ordering of the levels, from which I conclude with a high degree of confidence that the assumed ordering is reasonable in the study context.

There is no significant difference in 'error' rates across the 14 different indicators.

DCEversion	monotoneA		Total
	0	1	
vA1	42	1	43
vA10	43	0	43
vA11	34	0	34
vA12	25	2	27
vA13	35	0	35
vA14	34	2	36
vA2	36	0	36
vA3	38	1	39
vA4	25	1	26
vA5	39	2	41
vA6	43	0	43
vA7	34	2	36
vA8	32	2	34
vA9	30	2	32
Total	490	15	505

Pearson chi2 (13) = 12.2271 Pr = 0.509

However, the respondents were slightly more likely to make an 'erroneous' choice in favour of Woman 1, the difference in error rates being marginally significant.

DCEdesign	monotoneA		Total
	0	1	
dA1	243	11	254
dA2	247	4	251
Total	490	15	505

Pearson chi2 (1) = 3.2816 Pr = 0.070

For each respondent, the first question in Section B was an 'uninformative' comparison of three-indicator profiles. Given that the assumed ordering of indicator levels was confidently

verified through the Section A responses, this gives some sense of the impact of additional complexity of the three-indicator comparison on the rate of meaningful responses. Now 48 of the 505 respondents (9.5%) gave the 'wrong' response. This suggests that approximately 20% of the respondents may have failed to comprehend the exercise, or failed to give meaningful responses.

Of the three designs (groups of pairwise comparisons) implemented in Section B, Design B1 and Design B2 allow for the possibility that cyclical preferences could be expressed. Thus, for those respondents allocated B1 and B2 (325 in total) the consistency of their expressed empowerment ordering can be tested. Just 17 (5.32%) of these respondents expressed cyclical preferences. This is highly significantly less than the 25% of cycles that would have occurred if all respondents had made all choices entirely at random (there would be eight equally likely sets of responses, of which two are cyclical). Again this is consistent with approximately 20% of the respondents having failed to comprehend the exercise.

One-sample test of proportion		cycle: Number of obs = 325	
Variable	Mean	Std. Err.	[95% Conf. Interval]
cycle	.0523077	.0123502	.0281017 .0765137

p = proportion(cycle) z = -8.2306
 Ho: p = 0.25

Ha: p < 0.25 Pr(Z < z) = 0.0000	Ha: p != 0.25 Pr(Z > z) = 0.0000	Ha: p > 0.25 Pr(Z > z) = 1.0000
------------------------------------	---	------------------------------------

[Determine whether the same respondents gave non-monotone and cyclical responses.]

In conclusion, I am confident that the assumed ordering of levels was correct for all indicators, and that around 80% of the respondents comprehended the exercise and gave meaningful responses to the DCE questions.

4.2 Empirical Strategy: Preliminaries

The DCE design does not generate enough data to fully map every respondent's 'perception of empowerment' ordering, so non-parametric modelling is impossible, although it will be possible to examine some possibilities of heterogeneity. Furthermore, the existing measurement framework, an additive function of multiple binary indicators, is restrictive. But these issues have a common resolution, through estimation of a parametric model that reflects the restrictive structure.

4.3 Assumption of Equal Unspecified Attributes

If respondents' (homogeneous) empowerment-perception ordering may be represented by a linear combination of the 14 binary indicators then

$$E_i = \sum_{k=1}^{14} \beta_k x_{ik} + u_i$$

where i indexes a woman with level x_{ik} for indicator k and u_i is a random noise term ('random utility' model).

Each DCE question asks the respondent to choose which of two hypothetical women ($i = 1, 2$) she considers to be more empowered. The outcome $y = 1$ if woman 2 is chosen: if $E_2 > E_1$. Three indicators x_{ia} , x_{ib} and x_{ic} are specified for each woman. If we assume that u_1 and u_2 are independent and normally distributed then this yields the probit model

$$\Pr(y = 1) = \Phi(\beta_a(x_{2a} - x_{1a}) + \beta_b(x_{2b} - x_{1b}) + \beta_c(x_{2c} - x_{1c}))$$

Probit results for Section B (relational indicators only) are reported in Table (2); the Wald test decisively rejects equality of coefficients.

Variable	Coefficient	(Std. Err.)
DeltaR1	1.238	(0.070)
DeltaR2	0.801	(0.065)
DeltaR3	0.529	(0.062)
DeltaR4	1.343	(0.069)
DeltaR5	0.669	(0.063)
DeltaR6	0.982	(0.065)
<hr/>		
N	2020	
Log-likelihood	-1098.803	
$\chi^2_{(6)}$	489.326	

Probit results for all indicators (Section B and C data pooled) are reported in Table (3); again, the Wald test decisively rejects equality of coefficients.

4.4 Empirical Strategy: Imputed Unspecified Attributes

Implicit in the model above is the assumption that unspecified characteristics are the same for the two women. This is implausible. It is reasonable to assume that respondents will form a full, if uncertain, impression of the characteristics of the women from the partial information provided (that is, the indicators specified). This process of forming impressions is modelled as follows:

- Estimate conditional expectations from the survey data on respondents' characteristics:
 $\hat{x}_{ik} = \mathbb{E}(x_k | x_{ia}, x_{ib}, x_{ic})$.
- The probit model becomes

$$\Pr(y = 1) = \Phi\left(\sum_{k=a,b,c} \beta_k(x_{2k} - x_{1k}) + \sum_{k \neq a,b,c} \beta_k(\hat{x}_{2k} - \hat{x}_{1k})\right)$$

Table 3: All Indicators

Variable	Coefficient	(Std. Err.)
DeltaP1	1.129	(0.137)
DeltaP2	1.678	(0.162)
DeltaP3	1.788	(0.171)
DeltaP4	1.457	(0.146)
DeltaP5	1.585	(0.160)
DeltaR1	1.089	(0.059)
DeltaR2	0.710	(0.056)
DeltaR3	0.421	(0.054)
DeltaR4	1.208	(0.058)
DeltaR5	0.549	(0.055)
DeltaR6	0.855	(0.056)
DeltaE1	1.481	(0.170)
DeltaE2	1.576	(0.149)
DeltaE3	1.623	(0.157)
<hr/>		
N	3030	
Log-likelihood	-1644.304	
$\chi^2_{(14)}$	716.336	

The sampling errors will be complicated, firstly because sampling weights are not all equal (as the comparison group is weighted by propensity score matching strategy), and because of the 2-stage estimation procedure in which the conditional expectations are estimated for the unspecified attributes. I therefore obtained sampling errors using the bootstrap method, drawing the bootstrap sample after the determination of matching weights but before predicting the unspecified attributes.

4.4.1 DCE B Results: Relational Indicators

	Point estimates	Bootstrapped standard errors (2000 reps)	p-val for 1-sided test >0	sig >0	Standardised weights
R1DecHouse	0.89217	0.12603331	0.000	***	0.24
R2DecPubli	0.59010	0.13841629	0.000	***	0.16
R3DecPolit	0.04583	0.15126212	0.381		0.01
R4ActViole	1.30619	0.11798273	0.000	***	0.35
R5IndepInc	0.49285	0.10828052	0.000	***	0.13
R6ResourHH	0.43457	0.15209296	0.002	***	0.12
sum	3.76169969				1.00

*Note: observations weighted according to PSM (Kernel) weights
Unspecified indicators replaced by 'average' predictions.*

4.4.2 DCE B and C Results: All Indicators

	Point estimates	Bootstrapped standard errors (2000 reps)	p-val for 1-sided test >0	sig >0	Normalised weight
P1SelfConf	0.613045	0.22987108	0.004	***	0.13
P2Decision	0.389128	0.46846119	0.203		0.08
P3Violence	0.322185	0.45340666	0.239		0.07
P4AwareCol	0.820084	0.28510612	0.002	***	0.17
P5KnowRight	0.548244	0.25091311	0.014	**	0.12
R1DecHouse	0.365095	0.14096479	0.005	***	0.08
R2DecPubli	0.259113	0.12401418	0.018	**	0.05
R3DecPolit	-0.727567	0.17027044	1.000		-0.15
R4ActViole	0.756538	0.14103433	0.000	***	0.16
R5IndepInc	-0.001275	0.15303224	0.503		0.00
R6ResourHH	0.265203	0.14249888	0.031	**	0.06
E1EqualOpp	0.548256	0.29326314	0.031	**	0.12
E2SocNorms	0.160448	0.29996789	0.296		0.03
E2LegRight	0.397902	0.24261324	0.050	*	0.08
sum	4.7163996			sum	1.00

*Note: observations weighted according to PSM (Kernel) weights
Unspecified indicators replaced by 'average' predictions.*

4.4.3 DCE B and C Results: Dropping R3 and E1

The two indicators that took negative coefficients are dropped and the model re-estimated to generate the final weights for the index of women's empowerment.

	Point estimates	Bootstrapped standard errors (2000 reps)	p-val for 1-sided test >0	sig >0	Standardised weights
P1SelfConf	0.458045	0.2263660	0.022	**	0.11
P2Decision	0.515385	0.4103474	0.105		0.12
P3Violence	0.294105	0.4134369	0.238		0.07
P4AwareCol	0.814055	0.3037404	0.004	***	0.19
P5KnowRight	0.416416	0.2780255	0.067	*	0.10
R1DecHouse	0.400357	0.1256636	0.001	***	0.09
R2DecPubli	0.034743	0.1090942	0.375		0.01
R4ActViole	0.772841	0.1152889	0.000	***	0.18
R5IndepInc	0.081329	0.1189473	0.247		0.02
R6ResourHH	0.177230	0.1215467	0.072	*	0.04
E2SocNorms	0.067238	0.3240582	0.418		0.02
E2LegRight	0.256472	0.2073969	0.108		0.06
sum	4.288214			sum	1.00

*Note: observations weighted according to PSM (Kernel) weights
Unspecified indicators replaced by 'average' predictions.*

4.4.4 Pairwise tests for significant differences

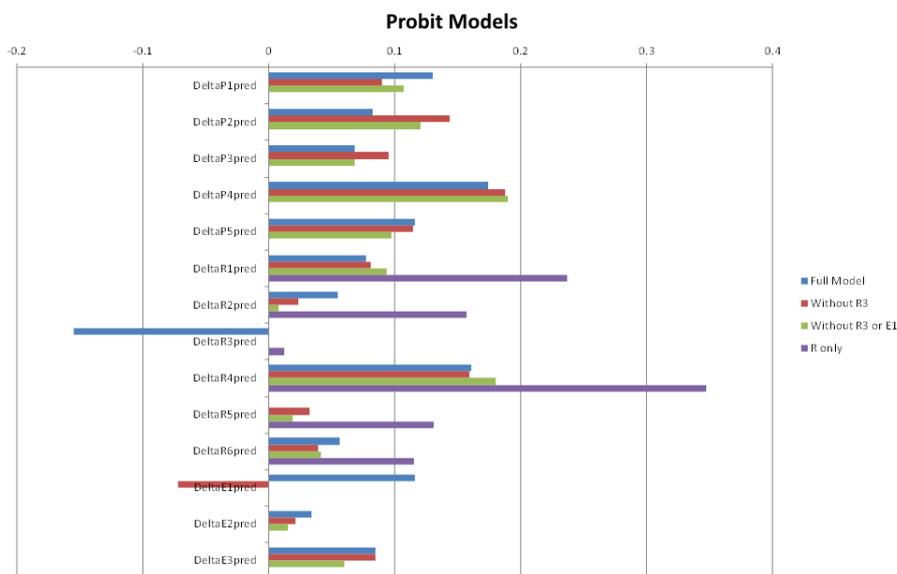
If the estimated coefficients on the different indicators are significantly different from one another then we can confidently conclude that an equal-weighted index is inappropriate to measure women's empowerment in the study context. Several of the coefficients are indeed significantly different from one another; it remains to determine whether this result is robust to multiple comparisons adjustments.

*10%, **5%, ***1%		Indicator	P1SelfConf	P2Decision	P3Violence	P4AwareCc	P5KnowRg	R1DecHou	R2DecPubl	R4ActViole	R5IndepInc	R6ResourH	E2SocNorm	E2LegRight
Indicator	Point est	SE	0.4580	0.5154	0.2941	0.8141	0.4164	0.4004	0.0347	0.7728	0.0813	0.1772	0.0672	0.2565
P1SelfConf	0.4580	0.2264												
P2Decision	0.5154	0.4103						*						
P3Violence	0.2941	0.4134												
P4AwareCol	0.8141	0.3037							**		**	*	*	
P5KnowRght	0.4164	0.2780												
R1DecHouse	0.4004	0.1257							**	**	*			
R2DecPubli	0.0347	0.1091	*			**	**	**	***					
R4ActViole	0.7728	0.1153		**		**	**	***	***	***	***	***	**	**
R5IndepInc	0.0813	0.1189					*							
R6ResourHH	0.1772	0.1215			*					***	***	***		
E2SocNorms	0.0672	0.3241			*					**	**	**		
E2LegRight	0.2565	0.2074								**	**	**		

number of tests:	66	proportion
number significant at 10%	13	0.197
number significant at 5%	9	0.136
number significant at 1%	3	0.045

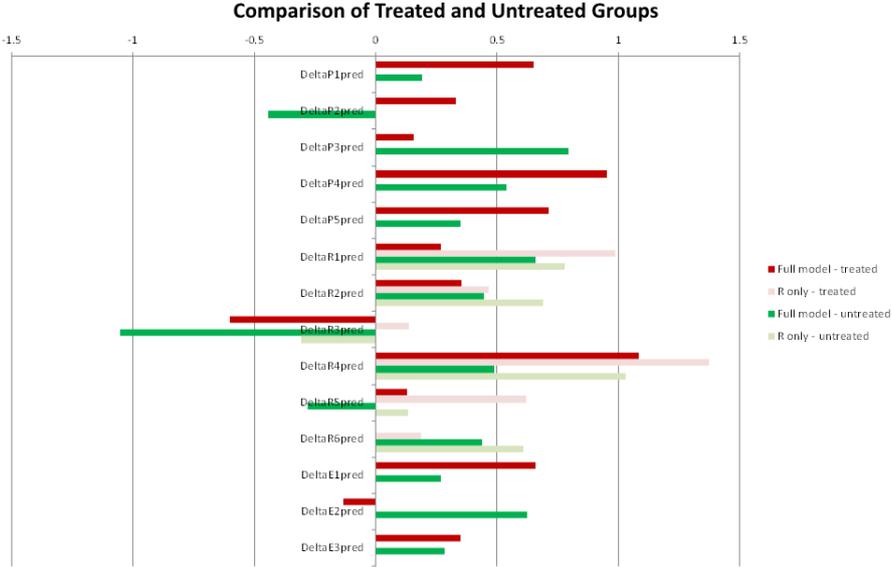
4.4.5 Comparison of Weights

The estimated coefficients are normalised (to add to 1) to give the weights for the index of women's empowerment.



4.4.6 Effect of Programme on Perceptions

Comparison of weights obtained for the beneficiary and (weighted) comparison groups suggests that the programme may have had an impact on perceptions of empowerment. These results should be interpreted with caution, as their validity will depend on the validity of the quasi-experimental identification which is beyond the scope of this paper.



5 Concluding Remarks

This study describes a pilot of a method to elicit values and perceptions to inform the weighting of a multidimensional index. Even in the small pilot sample, the conventional assumption of equally weighting different indicators of empowerment is decisively rejected. Further analysis is necessary to give firmer foundations for the interpretation of the results. Further work in larger samples would allow the restrictive assumptions to be relaxed, including homogeneity of preferences and the restriction of each indicator to just two levels.

References

- ALKIRE, S., R. MEINZEN-DICK, A. PETERMAN, A. QUISUMBING, G. SEYMOUR, AND A. VAZ (2013): "The womens empowerment in agriculture index," *World Development*, 52, 71–91.
- BISHOP, D., AND K. BOWMAN (2014): "Still learning: a critical reflection on three years of measuring women's empowerment in Oxfam," *Gender & Development*, 22(2), 253–269.
- DECANCO, K., AND M. A. LUGO (2013): "Weights in multidimensional indices of wellbeing: An overview," *Econometric Reviews*, 32(1), 7–34.
- DECANCO, K., AND V. WATSON (2015): "Eliciting preferences based weights for a multidimensional wellbeing index.," Presented at 14th Journes LAGV Conference on Public Economics, Aix-en-Provence, France.
- JOHNSON, F. R., E. LANCSAR, D. MARSHALL, V. KILAMBI, A. MÜHLBACHER, D. A. REGIER, B. W. BRESNAHAN, B. KANNINEN, AND J. F. BRIDGES (2013): "Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force," *Value in Health*, 16(1), 3–13.
- LOMBARDINI, S., K. BOWMAN, AND R. GARWOOD (2017): "A How ToGuide To Measuring Women's Empowerment: Sharing experience from Oxfams impact evaluations," Discussion paper, Oxfam GB.
- RODRIGUEZ-TAKEUCHI, L. (2014): "Incorporating peoples values in development: Weighting alternatives," Discussion paper, Overseas Development Institute.
- RYAN, M., K. GERARD, AND M. AMAYA-AMAYA (2007): *Using discrete choice experiments to value health and health care*, vol. 11. Springer Science & Business Media.
- WATSON, V., M. SUTTON, C. DIBBEN, AND M. RYAN (2008): "Deriving weights for the index of multiple deprivation based on societal preferences: The application of a discrete choice experiment.," Discussion Paper 24, OPHI Working Paper, Oxford, UK.

A Implementation

Implementing the DCE in SurveyCTO required long and complex code that would have been impossible to write and test in the tight timescale of the Effectiveness Review process, besides the risk of errors. The coding was automated by use of a Stata do-file; 500 lines of Stata code, taking the DCE design described above as input, output 1000 lines of Imagemagick code (for the image processing) and 5000 lines of SurveyCTO code. The coding process is illustrated in the flow diagram below.

