Oxford Poverty and Human Development Initiative (OPHI)
Oxford Department of International Development
University of Oxford

OPHI RESEARCH IN PROGRESS SERIES 67a

The Participatory Index of Women's Empowerment: Development and an application in Tunisia

Natalie Naïri Quinn* and Simone Lombardini**

September 2023

Abstract
In this paper we develop the Participatory Index of Women's Empowerment, an innovative measurement tool that reflects its subjects' own perceptions of empowerment. Participatory measurement is a response to the paradoxical potential for measurement of empowerment to disempower. A simple stated choice experiment allows participants to implicitly reveal the trade-offs that they make between different indicators of empowerment. This permits participatory determination of the relative weights for each indicator in a composite index, through estimation of a random utility model. We demonstrate the implementation of PIWE through a pilot application in the context of a quasi-experimental impact evaluation of an Oxfam project in Tunisia. Despite a relatively small sample size, we can reject the hypothesis that participants' perceptions of empowerment are consistent with equal weights. We find that the project had a significant positive impact on participants' empowerment and find suggestive evidence of impact on their perceptions of empowerment.

Keywords: Women's empowerment; participatory methods; multidimensional measurement; discrete choice experiment

JEL classification: B54, C43

---

1. Introduction

There is a broad consensus in the development sector around the importance of the fifth Sustainable Development Goal, 'to achieve gender equality and empower all women and girls', as adopted by the United Nations General Assembly in 2015. Many development agencies explicitly target women's equality and empowerment (UN Women, 2014). Consequently, development practitioners and researchers face the challenge of how to define and measure the intangible concept 'empowerment' in order to evaluate and inform policies.

In this paper we develop the Participatory Index of Women's Empowerment (PIWE), an innovative measurement tool that reflects its subjects' own perceptions of empowerment. We propose PIWE as a response to the concern, highlighted by Taylor (2000), Aslanbeigui et al. (2010) and Raj (2020), that external or expert measurement of empowerment embodies *power over* dynamics and can thus, paradoxically, perpetuate disempowerment. We agree with Bridges (2001) that engaging through a participatory approach is a more appropriate response than disengagement.

Earlier participatory approaches have focussed on choice of indicators of empowerment, often through qualitative exercises with small groups of participants. PIWE brings participation at scale, collectively eliciting from women whose empowerment is being measured both the indicators of empowerment that for them are most important and the relative weights that should be assigned to those indicators in an index of empowerment.

The paper proceeds as follows. In the next section, we examine the case for participatory measurement. In section three we review measurement approaches proposed and applied in the existing literature, identifying a gap in the availability of participatory methods for assigning weights to indicators. In section four we develop PIWE to fill that gap. In section five we demonstrate, through a pilot application in the evaluation of an Oxfam project in Tunisia, that PIWE may be operationalised in a quantitative study by embedding a simple stated choice exercise into the same survey questionnaire that is used to measure respondents' indicators of empowerment. We present the results of our analysis to generate weights for the index and its application to evaluation of the project in section six. We discuss our results in section seven, highlighting practical and conceptual considerations that emerged from our pilot implementation of PIWE.

2. The case for participatory measurement of empowerment

In this section, we examine the case for participatory measurement of empowerment. We argue that adoption of a participatory approach goes some way toward addressing the inherent paradox of external measurement of empowerment, while acknowledging the challenge posed by internalisation of disempowering values.

2.1 Definitions of empowerment

Empowerment was defined by Kabeer (1999), building on the concepts of agency and capability introduced by Sen (1985a, 1985b), as 'the expansion in people's ability to make strategic life choices in a context where this ability was previously denied to them' (p. 437). She recognises that individual agency is necessary but not sufficient to achieve this, emphasising that 'structural inequalities cannot be addressed by individuals alone' (p. 457).

A similar point was emphasised by both Narayan (2002, 2005) and Alsop et al. (2006), who explicitly recognise that *agency* – the 'ability to make purposeful choices' – may be constrained by *opportunity structure* – 'those aspects of the institutional context within which actors operate that influence their ability to transform agency into action' (Alsop et al., p. 10). This is reflected in Narayan's 2002 definition of empowerment as 'the expansion of assets and capabilities of poor people to participate in, negotiate with, influence, control, and hold accountable institutions that affect their lives' (p. xviii).

The typology of power developed by Rowlands (1995) provides a complementary framework. She distinguishes between *power over* ('controlling power'), *power to* ('generative power, … [creating] new possibilities and actions without domination'), *power with* (emerging from collective action) and *power from within* (based in 'self-acceptance and self-respect', extending 'to respect for and acceptance of others as equals'), identifying aspects of empowerment of all forms (Rowlands, 1997, p.13). VeneKlasen and Miller (2002) combine these 'expressions of power' with the experiences of grassroots activists to explore empowerment as a 'process and the result of the process' to 'address the negative forms and results of *power over*' (pp. 45, 53).

In practice, a variety of definitions of women's empowerment have been used; in their 2007 survey, Ibrahim and Alkire identified 29 definitions in use at that time. And yet, as Malhotra and Schuler (2005) argue, despite diversity of emphasis and terminology in the academic literature, there is in fact a broad consensus on the concept of empowerment. However, it is a complex concept, and not one that is directly measurable.

2.2 The paradox of externally mediated or measured empowerment

Giving an example from a project in Kenya, Rowlands (1995) discusses the potential for outsiders, as mediators of empowerment in a development context, to engage in ways that are ineffective or even actively disempower 'local people' (p. 105). This reflects the inherent problems of 'mediated empowerment' discussed by Rocha (1997, pp. 36–37), which we interpret as perpetuating disempowering *power over*.

Reflecting on empirical studies of the impact of microcredit on women's empowerment in Bangladesh, Aslanbeigui et al. (2010) observe that '[t]he empowered woman is conceived as a construct, an artifact of specialists' (p.191). Discussing the activities of these specialists or 'experts', they conclude that, paradoxically, '[e]mpowerment, therefore, is mainly a consequence of what is *done to* women as opposed to what they do on their own behalf' (p. 191, our emphasis).

These concerns apply just as much to the measurement of empowerment as its conceptualisation and promotion. Writing from a 'Southern' development practitioner's perspective, Taylor (2000) highlights the potential for the external measurement of empowerment to disempower, arguing that 'measurement of empowerment must not become something that the more powerful do to the less powerful' (p. 12). A similar concern about the 'power over' dynamic maintained by 'expert' measurement of empowerment is expressed from a public health perspective by Raj (2020, p.2).

2.3 Participatory measurement as reconciler and another paradox

Bridges (2001), in his treatment of the ethics of outsider research, examines the claim that *any* research conducted by outsiders into the experience of disempowered communities is intrinsically disempowering. Bridges recognises the potential for reinforcement of disempowerment through outsider research but argues that to avoid it entirely is 'not, any more than the paternalism of the powerful, the route to a more just society' (p. 382). He explores the character of participatory research approaches that have potential to reconcile this tension.

We might conclude that the adoption of appropriate participatory approaches could resolve the inherently disempowering aspect of external measurement of empowerment. A further paradox emerges, however. As Kabeer (1999) discusses, illustrating with examples from India, social norms play a fundamental role in disempowerment and it is natural for the disempowered to internalise values that justify their subordinate status. Khader (2011) explores in depth the role of these internalised values, or 'inappropriately adaptive preferences (IAPs)', which she sees as central to the concept of empowerment, 'the process of overcoming […] IAPs' (pp. 171, 176). While acknowledging some problematic implications, Kabeer concludes that women's empowerment

must be evaluated from an 'external normative standpoint' (p. 458), undermining the case for participatory measurement.

2.4 Conclusion

Despite the phenomenon of internalisation of disempowering values, we argue that there nevertheless remains an important role for participatory approaches to the measurement of empowerment. The extent of internalisation will be context-specific and thus a matter for empirical assessment, to which participatory approaches can contribute. Even Khader (2011) argues for the important role of subjective data in identifying and elucidating divergences between subjective and external perspectives. Furthermore, some participatory approaches may be more robust to internalisation, so careful choice of approach is important.

3. Existing measurement approaches

In this section we review existing approaches to measurement of empowerment, focussed mainly but not exclusively on women's empowerment. We do not attempt to survey the literature comprehensively, but focus on composite indices that aggregate multiple indicators, which have become established as the typical measurement approach. We discuss several examples, highlighting their measurement properties and the extent to which participatory approaches contributed to their development. This enables us to identify a gap in the availability of participatory methods for assigning weights to indicators.

Broader reviews of the recent literature include a critical review by Richardson (2018), a typology of concepts by Gram et al. (2019) and a bibliometric analysis by Priya et al. (2021).

3.1 Empowerment as a complex multidimensional concept

Narayan (2005) highlighted the necessity of measuring empowerment in order for it to be taken seriously in development policy making and programming. In moving from abstract definition to concrete attempts to measure it, many authors have recognised and have attempted to structure its complexity. We outline here some key contributions, highlighting their implications for the choice of measurement approach.

3.1.1 Multidimensionality of empowerment

Kabeer (1999) identified three *dimensions* of empowerment: resources (material, human and social), agency and achievements. The multidimensionality of empowerment has been recognised broadly in the literature, although different authors have used a variety of terms to describe it.

Malhotra and Mather (1997) emphasise 'that power is multilocational and exists in multiple domains' (p. 604), demonstrating empirically that women who are empowered in one area of life may not be empowered in others.

Alsop et al. (2006) recognise that a person's empowerment may vary across different *domains* (state, market and society) and at different *levels*: macro (national or state level), intermediary (familiar but not experienced daily) and local (the 'immediate vicinity of a person's everyday life') (p. 21).

Lombardini et al. (2017) frame the process of empowerment as involving changes that take place at three levels: *personal* (within the person), *relational* (encompassing relationships within the household and in the community), and *environmental* (encompassing both the formal institutional context and informal characteristics of the wider society).[1] This structure is reminiscent of Sen's (1985a) typology of *personal, social* and *environmental* conversion factors, through which goods and services are related to functionings and capabilities in the capability approach.[2] Lombardini et al. also map the typology of power developed by Rowlands (1995) to their framework, stating that changes in *power within* and *power to* take place at personal level, while changes in *power with* and *power over* take place at relational level. We observe that the mapping may, in practice, be more complex; for example, changes in *power to* may result from changes taking place at environmental level.

3.1.2 Multiplicity of indicators

Kabeer (1999) discusses indicators to measure empowerment in each of her three dimensions, arguing that their meanings, and thus their validity as measures of empowerment, are context-specific and determined by interrelationships between the three dimensions. An extensive literature on indicators of empowerment has since emerged: many indicators have been developed to measure empowerment in different domains and at different levels, through individual and household surveys.

---

1 Note that these levels do not map precisely to the levels identified by Alsop et al. (2006): for example, changes within the person may impact on empowerment at any of Alsop et al.'s levels, and some changes at Alsop et al.'s intermediary level may be relational, pertaining to relationships in the community, while others may be environmental.

2 See also Robeyns (2005) for a clear exposition of the role of these conversion factors in the capability approach.

For example, Ibrahim and Alkire (2007) proposed a shortlist of indicators that address many of the conceptual and practical issues that Narayan (2005) had identified and argued should be considered in the measurement of empowerment. Ibrahim and Alkire focus on agency rather than opportunity structure, while emphasising international comparability. Conversely, Kabeer (1999) and Malhotra and Schuler (2005) emphasise the importance of context: behaviours or attributes that might signify empowerment in one context may have different meanings elsewhere, so appropriate indicators may vary across contexts.

More recently, Glennerster et al. (2018, Appendix 1) comprehensively documented and categorised, across several domains, indicators of women's empowerment that have been used in recent empirical research. Glennerster et al. also emphasise the value of non-survey instruments, including direct observation, games, vignettes, implicit association tests and biomarkers, to assess aspects of empowerment that are difficult to capture through surveys.

3.2 Implications for measurement

The multidimensionality of empowerment and multiplicity of its indicators in different domains and at different levels pose particular issues for its measurement and analysis. It is natural to aggregate multiple indicators into a composite index; this approach has become conventional for measurement of empowerment in most policy contexts. Similarly, many empirical academic studies have adopted as their outcome of interest a composite index of empowerment that aggregates indicators across several dimensions, not necessarily those identified by Kabeer (1999).

A typical example of the composite indices of empowerment adopted in empirical studies is the Women's Empowerment Index (henceforth OxWEI), described in Lombardini et al. (2017) and Bishop and Bowman (2014), which was developed for Oxfam's Effectiveness Reviews (Hutchings, 2014). Acknowledging the context-specificity of empowerment, in any implementation of OxWEI a number of individual-level binary indicators (indexed $j = 1, 2, \ldots, J$) are identified, through preliminary qualitative research, to reflect the characteristics of an empowered woman in the evaluation context. Each indicator is categorised by level (personal, relational or environmental) and assigned a weight w_j. The empowerment score attained by individual i is then

$$E_i = \sum_{j=1}^{J} w_j x_{ij} \tag{1}$$

where $x_{ij} = 1$ if she achieves indicator j and $x_{ij} = 0$ if she does not. We observe that this index embodies an implicit definition of individual empowerment. Women with different profiles of

achievements in its constituent indicators will be assessed as more, or less, empowered, while relative indicator weights represent the implicit trade-offs between different indicators.

Other indices with a similar sum-of-binary-indicators structure include the Women's Empowerment in Agriculture Index or WEAI (Alkire et al., 2013; strictly its uncensored 5DE component), the project-level (pro-)WEAI (Malapit et al., 2019) and the Women's Empowerment in Livestock Index or WELI (Galiè et al., 2019). These and other examples are discussed in more detail in section 3.4 below.

It is important to acknowledge that adoption and implementation of a composite index does not resolve all challenges around the measurement of empowerment. Malhotra and Schuler (2005) recognise that indices and scales that aggregate multiple indicators are needed, but caution that the use of such indices and scales may obscure differential effects on, and effects of, particular aspects of empowerment. Alsop et al. (2006) also discuss, in chapter 3, the use of composite indices to combine different indicators, emphasising the need for caution when choosing indicators and their weights. More recently, Gram et al. (2019) have argued for conceptual clarity and cautioned against the arbitrary aggregation of indicators that measure distinct notions of empowerment.

3.3 Participatory measurement

Participatory approaches have been extensively applied to the selection of indicators of empowerment. Bishop and Bowman (2014) recognised the importance of involving people affected by a project in its evaluation. Consequently, wherever possible project participants contribute to the identification of context-specific characteristics of an empowered woman in the OxWEI construction process (Lombardini et al., 2017). Other examples include the identification of indicators for the evaluation process of a grass-roots social movement in Bangladesh documented by Jupp et al. (2010) and the choice of indicators and dimensions for the WELI (Galiè et al., 2019).

In each of these cases, participation was achieved through qualitative activities including focus group discussions and other participatory rural appraisal methods. As a qualitative exercise, selection of indicators is well-suited to such approaches. It is important to work with a sufficiently representative sample of participants to achieve saturation (Bowen, 2008), being conscious that self-selection of participants who already have the knowledge, self-confidence and time to participate may impede this, but statistical representativeness is not necessary.

But choice of indicators is just one aspect of the implicit definition of empowerment embodied in an index; the assignment of weights to indicators is equally important. In most applications of

OxWEI, all indicators have been assigned equal weights. An interesting exception is Vigneri and Lombardini (2017), whose implementation of OxWEI in Mali is the only previous study that we are aware of that has involved those whose empowerment is being measured in the assignment of weights. Vigneri and Lombardini conducted an exercise with 17 workshop participants who were asked to identify and rank the three indicators, from 20 that they had collectively identified, that they considered to capture the most important dimensions of girls' empowerment in the project context. Each participant allocated three points to the most important, two points to the second and one point to the third, and weights were assigned to indicators in proportion to the total number of points allocated. This approach was similar to that taken – in a different context and on a much larger scale, with 1800 participants across six countries – by Wisor et al. (2014) to assign weights to dimensions of the Individual Deprivation Measure (IDM). The 15 (of 25 candidate) dimensions receiving the highest average rank were selected for the IDM and divided into three tiers of five; arbitrary relative weights of 1.5, 1 and 0.5 respectively were assigned to the dimensions in these tiers.

However, there is no reason to expect ranking exercises to generate weights that reflect trade-offs between indicators consistent with participants' perceptions of empowerment. Furthermore, we argue that, as a quantitative exercise, the participants in a participatory weighting exercise should be representative of those whose empowerment is being evaluated. In section four below, we propose an alternative approach that addresses these issues.

3.4 Example composite indices

Having focussed on OxWEI as our primary example above, we outline here several other notable examples of composite indices of empowerment that have been proposed and applied in the literature. We highlight the definitions and measurement properties of these indices, as well as the extent to which participatory approaches have contributed to their development and application.

3.4.1 Combination of population-aggregate indicators

Early composite indices such as the Gender Empowerment Measure (GEM) published by the United Nations Development Programme (UNDP, 1995)[3] or the Gender Gap Index published by the World Economic Forum (Hausmann et al., 2006) combine information about several dimensions of empowerment $d = 1, 2, \ldots, D$ measured at population-aggregate level; in the case

---

3 Now superseded by the Gender Inequality Index (UNDP 2019).

of GEM, economic, political and professional participation of women. The typical structure of such indices is

$$E = \sum_{d=1}^{D} w_d X_d \qquad (2)$$

where the components X_d are the population-aggregate indicators and the weights w_d are typically equal across dimensions. For example, in the case of the GEM, the components are 'equally distributed equivalent percentages' that penalise gender inequality by taking the population-share weighted harmonic mean of the male and female achievements (Klasen, 2006, p.153).

The choice of indicators for such indices and the way in which they are combined has typically been expert-led and at least partly determined by data availability (UNDP, 1995). We are not aware of any examples in which participatory approaches have been implemented.

As in the GEM example, the indicators are typically not simple population averages, which means that the order of population and dimension aggregation cannot be interchanged. This has two important implications. Firstly, the index cannot be used for individual-level impact evaluation. Secondly, the index is not subgroup consistent over population subgroups (Foster and Shorrocks, 1991), which means that it is entirely possible for the index to rise in a subgroup of the population and remain unchanged for the rest of the population, but fall overall. This may be reasonable for an index of gender *inequality*, but is not what one might expect of an index of *empowerment* as such.

3.4.2 Black economic empowerment in the South African wine industry

A study that implemented a participatory approach is the evaluation of black economic empowerment in the wine industry in South Africa by Janssens et al. (2006). Their unit of analysis was the individual (survey respondent), although they reported results as sample averages. Their initial choice of dimensions and indicators was informed by the empowerment literature together with qualitative participatory activities (focus groups and a workshop) in which workers, managers and experts were represented. Their aggregation of indicators within each dimension was linear, with weights determined by statistical methods (exploratory and confirmatory factor analysis) rather than participatory or external normative approaches. Janssens et al. did not aggregate respondents' scores across different dimensions.

3.4.3 Evaluation of a social movement in Bangladesh

The evaluation process of a grass-roots social movement in Bangladesh documented by Jupp et al. (2010) has participatory elements at both the index-development stage and evaluation stage. Initial participatory activities with a sample of local groups utilised a variety of Participatory Rural

Appraisal (PRA) tools, 'mood meter, well-being analysis, scoring, network mapping, timelines, flow diagrams, drawings and drama' (p. 45). These generated 8,000 statements of empowerment, which were distilled by the research team, using qualitative analysis approaches, into 132 indicators spanning four categories or dimensions and three levels of 'developmental progression' (p. 50).

The assignment of weights w_j to the indicators was, however, based on external 'recognition that certain indicators have greater value than others' (Jupp et al., 2010, p. 54). The assignment of weights to indicators reflects the dimensions and levels, but the structure of the empowerment score does not otherwise reflect this multidimensionality. The indicators apply at the group level, so the empowerment score for group g has the form

$$E_g = \sum_{j=1}^{132} w_j X_{gj}, \tag{3}$$

where X_{gj} is group g's achievement of indicator j. Groups self-assess their achievement of each of the indicators.

3.4.4 The Women's Empowerment in Agriculture Index

The Women's Empowerment in Agriculture Index (WEAI) was developed by Alkire et al. (2013) to measure empowerment of women in an agricultural context. It is an adaptation of the subgroup consistent Adjusted Headcount Ratio M_0, which was developed by Alkire and Foster (2011) to measure multidimensional poverty. Indicators of empowerment in different dimensions are aggregated at individual level, before aggregation of individual scores to assess empowerment at community level. The resulting index is decomposable by population subgroups[4] including subnational region, age and social group, as well as by indicator.

The WEAI combines 10 binary individual-level indicators of empowerment, agency and inclusion (indexed $d = 1,2,\dots,10$) in five domains identified by The United States Agency for International Development (USAID) as priorities for the assessment of agricultural programmes. Alkire et al. (2013) sought comparability over time and space in their choice of indicators; there was no use of participatory approaches in the development of the index. Weights w_d were assigned to each indicator such that weights were equal and summed to ⅕ within each of the five domains, summing to 1 in total. Each individual i, achieving outcome x_{id} in indicator d (where $x_{id} = 1$ if

---

4 Note that the 5DE and GPI sub-indices described below are decomposable, but the WEAI itself, which combines both, is not, as 5DE is estimated for all women in a population while GPI is estimated only for households including adults of both genders.

the individual is empowered and $x_{id} = 0$ if disempowered in that indicator), is assigned an *empowerment score*[5]

$$e_i = \sum_{d=1}^{10} w_d x_{id}.$$ (4)

All individuals achieving $e_i \geq 0.8$ are considered empowered, so the empowerment score is censored above with $e_i^c = 1$ if $e_i \geq 0.8$, while $e_i^c = e_i$ if $e_i < 0.8$. The WEAI 'five domains of empowerment' (5DE) index is then

$$5\text{DE} = \frac{1}{N} \sum_{i=1}^{N} e_i^c$$ (5)

where $i = 1, 2, \ldots, N$ indexes the women in the sample, that is, the average censored empowerment score for women in the sample. The WEAI itself combines the 5DE index, weighted 90%, with a 10%-weighted Gender Parity Index (GPI) that measures the divergence in empowerment between women and men in the same household.

As the 5DE index (though not the WEAI itself) is the average of an individual-level index, it could be straightforwardly utilised for impact evaluation, with the censored empowerment score e_i^c as the individual outcome of interest. However, we argue that unless there is an objective reason to do so, discarding information through censoring is not appropriate in the context of impact evaluation; it would be more appropriate to use the uncensored empowerment score (equation 4) as the outcome of interest. The natural wish to focus on outcomes and impacts for the less-empowered would be better achieved through heterogeneity analysis of the treatment effect.

The creators of the WEAI acknowledged that the index should be adapted, by revision of choice of indicators, to measure women's empowerment in different contexts (Alkire et al., 2013). For example, Ghali et al. (2018) document their adaptation and implementation in Tunisia, while Gupta et al. (2019) document their adaptation and implementation in India. Other adaptations, some detailed below, simplify the structure of the WEAI by focussing on the 5DE sub-index alone.

3.4.5 The pro-WEAI

The project-level (or pro-) WEAI (Malapit et al., 2019) retains the full structure of the WEAI, but modifies its domains to align with Rowlands' (1995) typology of power. It was developed in

---

[5] We modify the WEAI notation to present in 'positive' empowerment terms, to make clear the relationship with other indices. Alkire et al. (2013), pursuing consistency with Alkire and Foster (2011), presented WEAI using 'inadequacy' notation.

collaboration with a portfolio of agricultural development projects with explicit women's empowerment goals. The indicators and survey instrument were developed through a participatory process with project implementers, although not project participants.

3.4.6 The Women's Empowerment in Nutrition Index

The Women's Empowerment in Nutrition Index (WENI), developed by Narayanan et al. (2019), adapts the 5DE structure of the WEAI (equation 5) without implementing the GPI component or an equivalent of the WEAI itself. The creators of the WENI cross-cut domains of nutrition following United Nations Children's Fund (UNICEF, 1990) with the dimensions of empowerment identified by Kabeer (1999).

A participatory approach was adopted to develop the list of candidate indicators for the WENI. Focus group discussions and individual interviews were conducted with women in rural South Asia, to identify barriers and opportunities that they face regarding nutrition (Narayanan et al., 2019). The WENI creators then combined statistical data reduction techniques and normative considerations to choose indicators within each of the domain-dimensions.

3.4.7 The Women's Empowerment in Livestock Index

Women's Empowerment in Livestock Index or WELI (Galiè et al., 2019) is also based on the WEAI, but simplifies the index structure further by omitting the censoring of the individual empowerment score, so that the (individual-level) empowerment index has the same structure as the WEAI empowerment score (equation 4). As we argue above, this approach is particularly appropriate for impact evaluation.

A participatory approach contributed to the choice of indicators and dimensions for the WELI, which was informed by a preliminary qualitative study with women and men in livestock communities as well as community leaders and experts at community, national and international level. In total, 16 indicators of empowerment were chosen, which are weighted equally within each of six dimensions.

4. The Participatory Index of Women's Empowerment

In the previous section, we identified a gap in the availability of participatory methods for assigning weights to indicators such that the resulting index reflects trade-offs between indicators consistent with participants' perceptions of empowerment. In this section, we develop our Participatory Index of Women's Empowerment to fill that gap. We propose a stated choice exercise to elicit participants' orderings of profiles of empowerment indicators and show how indicator weights can be estimated from the choice data, to generate the participatory index.

4.1 Relationship between empowerment indicator weights and orderings of profiles

Our starting point is the observation that any individual-level composite index of empowerment represents an ordering of alternative profiles of achievements in the various indicators of empowerment that comprise the index. By 'ordering' we mean a specification, for each profile, of which other profiles are considered more, less, or equivalently empowered. Given an index of the form equation (1), the choice of weights determines the ordering represented, and thus the definition of empowerment embodied by the index.

Turning this around, if we were able to observe the ordering of alternative profiles of empowerment indicators perceived by those whose empowerment is being evaluated, we could attempt to determine weights w_j^p for each indicator j such that the index

$$PIWE_i = \sum_{j=1}^{J} w_j^p x_{ij} \qquad (6)$$

represents the perceived ordering.[6] This is our Participatory Index of Women's Empowerment (PIWE). In order to implement PIWE, we will need both a method to observe or elicit the perceived empowerment ordering and a method to determine weights w_j^p such that the index represents the observed empowerment ordering.

4.2 Eliciting perceived empowerment ordering of profiles

Following Watson et al. (2019), who apply alternative preference-based methods to explore the robustness of the expert-based weights used in the UK's Index of Multiple Deprivation (IMD),

---

6 While, with a finite number of indicators that may take a finite number of values, we can be sure that there exists some function of the indicators that represents the ordering, there is no guarantee that that function will be additive. In fact, existence of an additive representation requires separability of the ordering across indicators, which need not be the case: the extent to which one empowerment indicator is traded off with a second might depend on the value of a third. We are relaxed about this, as estimation will yield the best additive approximation to the 'true' representation, which seems appropriate if we seek an additive index.

we suggest that a discrete choice experiment (DCE) be implemented to elicit participants' perceived empowerment ordering. A DCE is a stated choice exercise in which participants are invited to express preferences amongst alternatives specified in terms of the discrete *levels* attained in each of several *attributes* (Ryan et al., 2008; Hensher et al., 2015).[7] To implement PIWE, the alternatives will be hypothetical women, described in terms of their achievements (*levels*) in each of several candidate indicators (*attributes*) of empowerment. We will demonstrate in our pilot application (section five) that it is possible to achieve participation at scale in a quantitative study by embedding a DCE into the same survey questionnaire that is used to measure participants' indicators of empowerment.

4.3 Recovering indicator weights

Having implemented a DCE, the PIWE weights may be recovered by application of standard discrete choice methods. Consider a participant i who expresses which of two hypothetical women 1 and 2 she considers more empowered. $PIWE_1 = \sum_{j=1}^{J} w_j^p x_{1j}$ and $PIWE_2 = \sum_{j=1}^{J} w_j^p x_{2j}$, where x_{1j} and x_{2j} are the hypothetical women's achievements in the various indicators; the weights w_j^p are as yet unknown. Let participant i's evaluation of 1 and 2's empowerment be E_{1i} and E_{2i} respectively, such that $E_{2i} > E_{1i}$ if she considers 2 more empowered. It is possible that E_{1i} and E_{2i} diverge from $PIWE_1$ and $PIWE_2$, so $E_{1i} = PIWE_1 + \varepsilon_{1i}$ and $E_{2i} = PIWE_2 + \varepsilon_{2i}$. $E_{2i} > E_{1i}$ if $\varepsilon_{1i} - \varepsilon_{2i} < PIWE_2 - PIWE_1 = \sum_{j=1}^{J} w_j^p (x_{2j} - x_{1j})$ so the probability that i will identify hypothetical woman 2 as more empowered is

$$\mathbb{P}(i \text{ chooses 2 from } \{1,2\}) = F\left(\sum_{j=1}^{J} w_j^p (x_{2j} - x_{1j})\right) \tag{7}$$

where F is the cumulative distribution function of $\varepsilon_{1i} - \varepsilon_{2i}$. This is a straightforward linear random utility model (Greene, 2012), whose parameters are the participatory weights w_j^p.

4.4 Contrast with earlier methods

The key difference between PIWE and the ranking exercises applied in previous studies is that in PIWE the participants rank hypothetical women described by their profile of achievements in several empowerment indicators, rather than ranking the indicators (or dimensions) directly. By

---

7 DCEs have been widely applied in marketing, transport and health economics, as surveyed in Ryan et al. (2008) and Hensher et al. (2015). More recently, political scientists and sociologists have implemented DCEs in surveys to explore voter preferences and social attitudes (Hainmueller et al., 2014; Liebe et al., 2020). Decancq and Watson (2019) implement a DCE to explore weights for the Human Development Index.

ranking profiles, participants implicitly reveal the trade-offs that they make between the different indicators. It is essential to elicit these trade-offs as they are what the indicator weights in an index represent. Direct ranking of indicators cannot reveal the trade-offs among them, and so cannot yield weights for the index.

5. Application of PIWE in Tunisia: methods

In this section we describe the study context and methods that we adopted to operationalise PIWE in our pilot application. We achieved this by embedding a simple stated choice exercise into the same survey questionnaire that was used to measure respondents' indicators of empowerment. As well as documenting the experimental design, we describe the empirical methods that we use to estimate the PIWE indicator weights.

5.1 Study context and sample

We piloted the implementation of PIWE in the context of an Effectiveness Review, conducted by Oxfam GB in Tunisia in November 2016, to assess the impact of the project *AMAL: Supporting Women's Transformative Leadership* on women's empowerment (Lombardini, 2018). This project started in 2012, following the democratic transition of 2011, with the objectives of increasing women's awareness of their political and socio-economic rights and supporting women to play a more active role in the political and socio-economic life of their community and country. It was implemented by three organisations in Tunisia: the League of Tunisian Women Voters (LET), the Tunisian Association of Democratic Women (ATFD), and the Association of Tunisian Women for Research and Development (AFTURD).

Tunisian households (Ghali et al., 2018). Similarly, the 39% with formal sector employment was substantially higher than in the representative survey.[8]

As an ex-post impact evaluation, the Effectiveness Review followed a quasi-experimental propensity score matching (PSM) approach. A comparison group of 290 non-project-participant women was sampled from matched communities. Sampling mimicked the targeting process employed by the project, in an attempt to minimise both observable and unobservable differences between the project participant and comparison groups. Propensity for project participation was estimated as a function of recalled baseline and fixed characteristics. The sample was restricted to

---

[8] In the representative survey 1% of rural women and 5% of urban women in the household head/spouse sample and 32% of the youth sample had completed tertiary education, while the proportion engaged in wage work were 8%, 14% and 19% respectively.

the region of common support of the propensity score distribution, comprising 213 project participants and 285 non-participants. Non-participants were assigned matching weights (henceforth 'PSM weights') determined by application of a Gaussian kernel. Impact evaluation methods and findings are reported in detail by Lombardini (2018); the present study focusses on the implementation of PIWE, in the context of the Effectiveness Review.

5.2 Choice and measurement of indicators of empowerment

As an Oxfam study, the Effectiveness Review adopted the OxWEI approach to measuring women's empowerment; we integrated PIWE with the OxWEI index-building process described by Lombardini et al. (2017).

The authors facilitated a preliminary workshop with Oxfam Tunisia programme staff, staff from the implementing organisations and a Tunisian expert in empowerment and gender. Workshop participants identified characteristics of an empowered woman in the evaluation context. Logistical constraints precluded the representation of study participants in this workshop. While this was a limitation of our pilot implementation, our validity checks show that its impact was limited (section 6.1). We discuss more extensively in section 7.1 its implications for our results and what would have been an ideal approach.

Through brainstorming, discussion and consensus-building, the workshop participants established a list of 14 indicators of empowerment together with, for each, an icon to visually represent it and brief descriptors of a woman who achieves it and a woman who does not, all summarised in Table 1. They also categorised each indicator as representing empowerment at the personal, relational or environmental level.

Building on the workshop results and in collaboration with the Tunisian expert, we developed a questionnaire to capture each of the 14 indicators. Questions drew on established survey tools including the Rosenberg Self-Esteem Scale, the Demographic and Health Surveys (DHS) toolkit questionnaire, and the WEAI questionnaires (Alkire et al., 2013), as well as the accumulated experience described by Lombardini et al. (2017). The measurement of each indicator is detailed in Lombardini (2018), Appendix 1.

Table 1. Indicators of empowerment, icons and descriptors.

Indicator	Code	Icon	Descriptor (not achieved)	Descriptor (achieved)
Personal				
Self confidence and self esteem	P1		She feels she doesn't have many good qualities	She feels that she has a number of good qualities
Ability to make decisions for herself	P2		She cannot personally take decisions regarding herself	She can personally take decisions regarding herself
Recognising that violence is not acceptable	P3		She considers it acceptable for a man to beat his wife	She considers it unacceptable for a man to beat his wife
Awareness that collective action is effective	P4		She does not believe that acting as a group is effective to solve issues	She believes that acting as a group is effective to solve issues
Knowledge and awareness of women's rights	P5		She does not think that men and women should have the same rights	She thinks that men and women should have the same rights
Relational				
Ability to make decisions in the household	R1		She is not able to make decisions within the household	She is able to make decisions within the household
Participation and ability to make decisions in the public sphere	R2		She does not participate in civil society and associations	She actively participates in civil society and associations
Participation and ability to influence or make decisions in the political sphere	R3		She does not actively participate in political parties	She actively participates in political parties
Taking action to stop violence	R4		She experiences violence and does not report it	In cases of experience of violence, she is able to report it
Independent income	R5		She does not have an independent source of income	She has an independent source of income
Control over resources in her household	R6		She has no control over assets and resources in her household	She has control over assets and resources in her household
Environmental				
Equality of opportunity	E1		She lives in a community that does not allow women to have equal political opportunities with men	She lives in a community that ensures that women have equal political opportunities with men
Social norms	E2		She lives in a society that does not allow her to be free	She lives in a society that allows her to be free
Legislative protection for women's rights	E3		She lives in a society where women's rights are not enshrined in law	She lives in a society where women's rights are enshrined in law

Notes: Indicator definitions and descriptors developed, and icons identified, in preliminary workshop in Tunis, November 2016.

5.3 Discrete choice experiment questions

In section 4.2 we proposed the implementation of a discrete choice experiment (DCE) to elicit perceived empowerment orderings and thus determine indicator weights for PIWE. In the

evaluation of AMAL in Tunisia, we embedded this DCE within the same questionnaire used to capture study participants' achievement of the indicators of empowerment.

To minimise cognitive demand, the choice set for each DCE question comprised just two alternative hypothetical women, with different profiles of indicators of empowerment. The study participant was asked to identify which of these hypothetical women she considered to be more empowered. Careful phrasing of this question was important, to ensure that we captured perceptions of *empowerment* rather than any other concept. The DCE question texts, translations and back translations are documented in Appendix A. The phrase 'القدرة باش تكون فاعلة' used to translate 'empowered' into Tunisian Arabic may be literally translated into English as 'capable of action'. It reflects the phrase 'active citizen', which has become a stock phrase in Tunisian Arabic, entering the vocabulary of democratisation and reflecting a transition in perceptions of the individual from passive subject to active citizen (M.-S. Omri, personal communication, May 30, 2017). While it is possible that the phrase may not have been understood in that way by less-educated study participants, we note that typical education levels of study participants are relatively high.

5.4 Discrete choice experiment design

The choice of which profiles to present, in which choice sets, to which study participants, comprised the *design* of the DCE. Our design was constrained by the fact that variation (or even specification) of all 14 indicators of empowerment for each hypothetical woman risked cognitive overload for study participants; Arentze et al. (2003) demonstrated in their transport-choice study in South Africa that specification of more than three attributes had negative consequences for data quality. Therefore, we presented *partial* profiles (Chrzan, 2010) in which at most three indicators of empowerment were specified for both hypothetical women within each choice set. This precluded the estimation of interaction effects, but, as the goal of the exercise was to estimate weights for an additive index, we considered this an acceptable limitation. An example DCE question is illustrated in Figure 1; its Tunisian Arabic implementation is illustrated in Figure B2.

Another constraint was logistical: with limited module duration, we could ask only seven DCE questions with each study participant. We therefore created a 'blocked' design, allocating different sets of questions randomly to different study participants. With 14 indicators of empowerment, $\binom{14}{3} = 364$ combinations of three indicators may be selected, while given that each of the three indicators may be achieved, or not, there are $2^3 \times 364 = 2912$ potential hypothetical profiles.

With just seven questions per study participant, it was impossible to explore all these hypothetical profiles, even allocating different 'blocks' of questions to different study participants.

There is an active literature on efficient design of DCEs, surveyed comprehensively in Chapter 6 of Hensher et al. (2015). The objective is typically to maximise the statistical power that can be obtained for a given sample size; results are contingent on specification (including parameter values) of the choice model to be estimated. However, implementation of an efficient design can jeopardise identification if the model is mis-specified (Hainmueller et al., 2014), and the literature on efficient design for DCEs with partial profiles is less-developed; contributions include Kessels et al. (2011, 2015) and Großmann (2019). Given the exploratory nature of our study, we implemented a modified version of Chrzan's (2010) rotational design algorithm, which is not necessarily efficient, but minimises attribute (indicator) presence correlations and is thus relatively robust to alternative 'true model' specification. We structured the DCE module in three parts, which we now describe in more detail.

Figure 1. Example DCE question
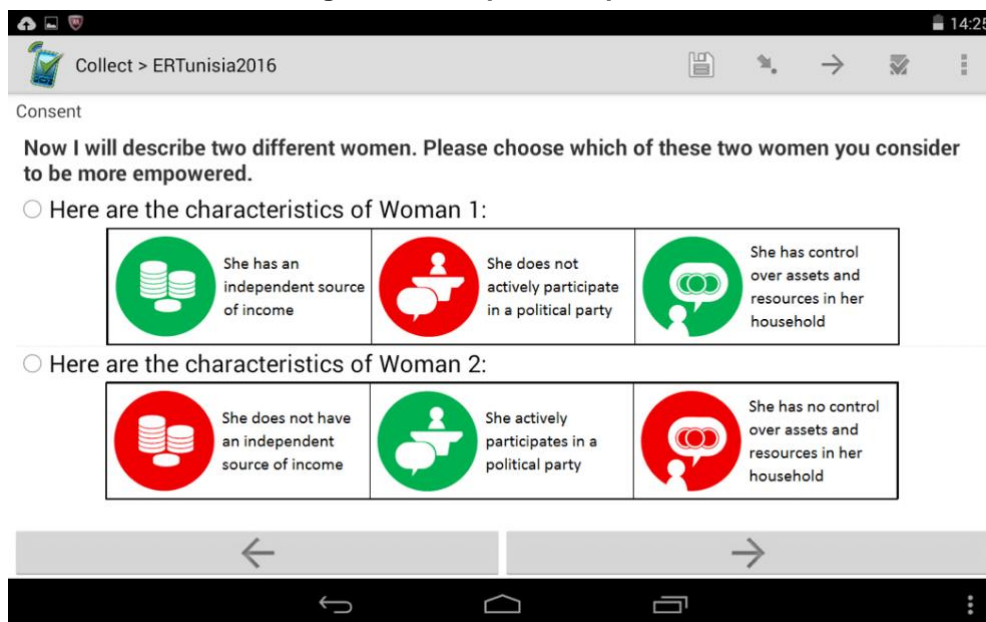
5.4.1 DCE part A

Part A of the DCE contained one question, in which only one indicator of empowerment was specified; one of the hypothetical women achieved this indicator, while the other did not. This very straightforward question was included to aid study participants' comprehension of the activity and to provide a fully participatory check of whether they do indeed consider each indicator a valid characteristic of empowerment. The 14 indicators were allocated to study participants at random and the ordering of the hypothetical women was also randomised, to check possible order effects.

5.4.2 DCE part B

Part B of the DCE focussed on the six relational indicators, identified by participants in the preliminary workshop as particularly targeted by project activities. Implementing Chrzan's (2010) rotational design algorithm, six overlapping combinations of three indicators were selected, illustrated in Figure 2 and detailed in Table B3. Each study participant was assigned one of these combinations at random.

There are eight (2^3) possible profiles of achievements in three indicators. For example, representing achievement with 1 and non-achievement with 0, the hypothetical woman with profile (101) achieves the first and third indicators but not the second. Eight profiles pair to form 28 choice sets. However, all but nine of these pairs are uninformative; for example, the profile (101) is unambiguously more empowered than (001). Time constraints only permitted four questions in Part B, so the nine informative pairs were divided into three blocks of three, to each of which was added one uninformative pair as a further check of comprehension and indicator validity.

Figure 2. Part B rotational design
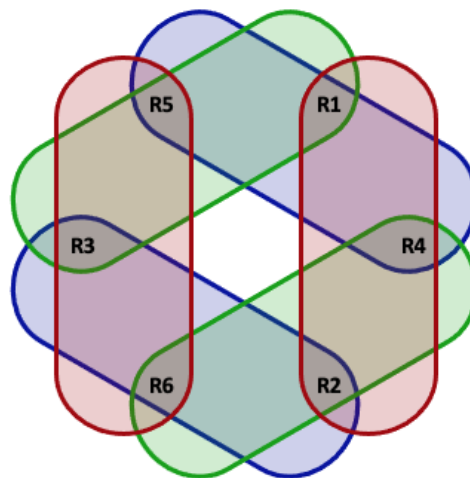
These question blocks are documented in Table 2. In each block, the first question was the uninformative pair, giving a second straightforward exercise (with the Part A question) to aid study participants' comprehension and check the validity of each indicator. The question blocks are also illustrated in Figure B2, in which it can be seen that blocks 1 and 2 permit cyclic choices, providing a check of consistency.

Table 2. Choice set blocks in part B of the DCE module

Hypothetical Profile	Block 1						Block 2						Block 3					
	Woman 1			Woman 2			Woman 1			Woman 2			Woman 1			Woman 2		
Indicator	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Question 1	1	0	**1**	1	0	**0**	0	**0**	1	0	**1**	1	**0**	0	1	**1**	0	1
Question 2	**1**	**0**	0	**0**	**1**	0	**0**	**1**	1	**1**	**0**	1	**1**	**0**	1	**0**	**1**	0
Question 3	0	**1**	0	0	**0**	1	1	**0**	1	1	**1**	0	**1**	0	0	**0**	1	1
Question 4	**0**	0	1	**1**	0	**0**	1	1	0	**0**	1	1	**1**	1	0	**0**	0	1

Notes: 1 represents achievement of the relevant indicator by the hypothetical woman; 0 represents non-achievement. Indicators that vary within choice sets are highlighted in boldface.

5.4.3 DCE part C

Part C (two questions) combined relational indicators with a personal or environmental indicator. Three pairs of the relational indicators were not matched in any of the part B combinations; these pairs were each matched with each of the personal and environmental indicators (24 combinations in total), detailed in Table B1. Each study participant was assigned two of these combinations at random. In each case, one question was asked, the choice sets randomly selected from the nine informative choice sets described under part B and documented in Table 2.

5.5 Discrete choice experiment implementation

The survey platform used for Oxfam Effectiveness Review questionnaires is SurveyCTO, so it was necessary to implement the DCE module in this platform. For each DCE question, we needed to display up to three (of 28) indicator texts and icons, for each of the two hypothetical women amongst whom the study participant was asked to choose. Unfortunately, SurveyCTO is not sufficiently flexible to allow us to implement this directly. However, it was possible to display a single image to illustrate each option, so we prepared image files that combined the indicator texts and icons for each of the hypothetical women in each of the DCE questions. An example of these image files is illustrated in Figure 3.

Figure 3. Example image file
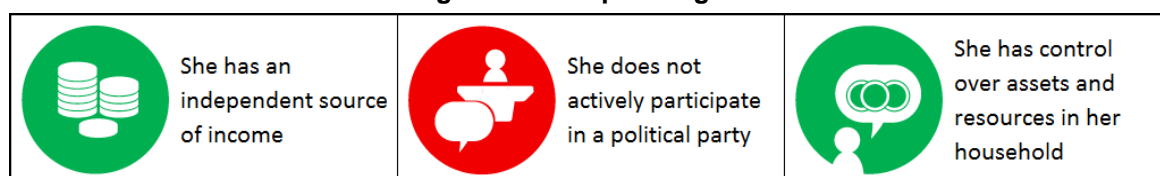
The DCE design required us to represent 208 different hypothetical women (two achievement profiles for each of the 14 indicators in part A, plus six achievement profiles for each of the six indicator combinations in part B and each of the 24 indicator combinations in in part C).

Considering both English and Arabic, this required 416 image files in total. Manually preparing and naming each of these files, as well as preparing the SurveyCTO code to randomise among the profile pairs included in the design and call the correct image files, would have been prohibitively time-consuming and prone to error. We therefore automated this process, writing a Stata do-file whose outputs were the SurveyCTO survey and choices sheets for the DCE module as well as ImageMagick batch files containing the commands needed to compile the text and icons for each hypothetical woman in both English and Tunisian Arabic. The workflow for our automated implementation process is documented in more detail in Appendix B.2.

5.6 Empirical methods

5.6.1 Estimation of indicator weights

As described in section 4.3, the PIWE weights may be recovered through estimation of a random utility model (eq. 7). If the error term $\varepsilon_{1i} - \varepsilon_{2i}$ is normally distributed we may re-scale the coefficients so that $\varepsilon_{1i} - \varepsilon_{2i} \sim N(0,1)$ and this becomes the probit model,

$$\mathbb{P}(i \text{ chooses } 2 \text{ from } \{1,2\}) = \Phi\left(\sum_{j=1}^{14} \beta_j(x_{2j} - x_{1j})\right) = \Phi\left(\sum_{j=1}^{14} \beta_j \Delta x_j\right), \tag{8}$$

which may be estimated by maximum likelihood. We weight observations by the PSM weights, so the results are representative for the sample of project participants (subject to the common support restriction). The PIWE weights are then

$$w_j^p = \frac{\beta_j}{\sum_{k=1}^{14} \beta_k}. \tag{9}$$

5.6.2 Imputation of unspecified indicators

Our partial profile design, adopted to minimise cognitive demand, allowed us to specify only three of the 14 indicators of empowerment for the hypothetical women in each DCE question. This complicates estimation of the model. We were reluctant to direct study participants to assume that the hypothetical women were identical in characteristics other than those specified, as is standard practice with partial profiles (Chrzan, 2010), or to impose this assumption in our analysis of the data. If – as is likely – indicators of empowerment are, in the participants' experience, highly correlated, they would perceive the directed assumption as implausible. Therefore, we did not make any reference to unspecified characteristics but allowed participants to make their own inferences on the basis of the information provided.

In a different context (a UK study of medical screening preferences), Ryan et al. (2009) find qualitative evidence that DCE respondents do indeed make inferences about unspecified characteristics of the hypothetical alternatives that they are asked to choose amongst. There are significant positive pairwise correlations among most of the fourteen indicators for our study participants, recorded in Table 3, and it is not plausible that study participants would assume otherwise; it is reasonable to assume that they make implicit inferences about characteristics of the hypothetical women other than those specified. Failing to account for this inference-making process when estimating the binary choice model is likely to result in biased estimates for the model parameters and thus the index weights. Unfortunately, we have not found any prior studies that model such an inference-making process. We therefore implemented a heuristic approach, recognising that identification of the optimal method is an area that warrants future research.

Table 3. Empirical pairwise correlations among the indicators of empowerment

	P1	P2	P3	P4	P5	R1	R2	R3	R4	R5	R6	E1	E2	E3
P1	1.00													
P2	0.23**	1.00												
P3	0.30**	0.42**	1.00											
P4	0.22**		0.23**	1.00										
P5	0.17	0.20**	0.26**		1.00									
R1		0.43**	0.23**		0.17*	1.00								
R2	0.19*	0.24**	0.23**	0.17		0.28**	1.00							
R3	0.27**	0.24**	0.26**	0.18*		0.29**	0.46**	1.00						
R4			0.20**		0.22**				1.00					
R5	0.16	0.18*	0.18*		0.18*	0.35**	0.24**	0.26**		1.00				
R6	0.19*	0.35**	0.26**			0.54**	0.33**	0.39**		0.36**	1.00			
E1	0.20**	0.34**	0.34**	0.17	0.22**	0.30**	0.38**	0.46**	0.19*	0.28**	0.29**	1.00		
E2	0.27**	0.30**	0.28**	0.27**		0.24**		0.21**			0.21**		1.00	
E3														1.00

Notes: Pairwise correlations between indicators in the empirical data (N=498), PSM weights applied. Values reported when p<0.05; * p<0.01; ** p<0.001; Sidak multiple-comparisons adjustment applied.

To represent study participants' inference-making process we utilised the empirical joint distribution of their measured empowerment indicators. We assume that study participants (implicitly) impute the unspecified indicators for the hypothetical women as their conditional expectations (conditional on the specified indicators), and estimate these conditional expectations using their sample counterparts (conditional averages). We then use these sample counterparts to represent the unspecified indicators when estimating our binary choice models.

Consider the sample of study participants $i = 1,2, \dots, N$, assigned PSM weights $w_1, w_2, \dots, w_N$ and whose *observed* attainments of the fourteen binary indicators are $a_{ij} \in \{0,1\}$ for each $j = 1,2, \dots, 14$.

We will need the (PSM-weighted) sample averages of the indicator attainments, conditional on combinations of three specified indicators. Let the three specified indicators be labelled $\{a, b, c\} \subset \{1, 2, \dots, 14\}$. Given specified indicator values x_a, x_b, x_c, the weighted conditional average value of indicator $j \notin \{a, b, c\}$ in the sample is

$$\bar{x}_j(a, b, c | x_a, x_b, x_c) = \frac{\sum_{i=1}^{N} a_{ij} w_i I(a_{ia} = x_a, a_{ib} = x_b, a_{ic} = x_c)}{\sum_{i=1}^{N} w_i I(a_{ia} = x_a, a_{ib} = x_b, a_{ic} = x_c)}, \tag{10}$$

where the indicator function $I(a_{ia} = x_a, a_{ib} = x_b, a_{ic} = x_c) = 1$ for individuals i whose attainments of indicators a, b and c match the specified values, and 0 otherwise.

Now consider a hypothetical woman $v \in \{1, 2\}$ presented to study participant i in DCE question q. Her (hypothetical) attainments $x_{va}^{iq}, x_{vb}^{iq}, x_{vc}^{iq}$ in three indicators $\{a^{iq}, b^{iq}, c^{iq}\} \subset \{1, 2, \dots, 14\}$ are specified, while her attainments x_{vj}^{iq} in all other indicators $j \notin \{a^{iq}, b^{iq}, c^{iq}\}$ are not specified. Let $s_{vj}^{iq} \in \{0, 1\}$ be a specification indicator such that $s_{va}^{iq} = s_{vb}^{iq} = s_{vc}^{iq} = 1$ and $s_{vj}^{iq} = 0$ for $j \notin \{a^{iq}, b^{iq}, c^{iq}\}$.

The hypothetical woman's specified and imputed attainments, which provide the right-hand side variables for the estimated binary choice model (equation 8), are then, for each for each $j = 1, 2, \dots, 14$,

$$\tilde{x}_{vj}^{iq} = s_{vj}^{iq} x_{vj}^{iq} + (1 - s_{vj}^{iq}) \bar{x}_j(a^{iq}, b^{iq}, c^{iq} | x_{va}^{iq}, x_{vb}^{iq}, x_{vc}^{iq}). \tag{11}$$

5.6.3 Standard errors

The imputation of explanatory variables invalidates usual approaches to estimation of standard errors, so we implement a bootstrap procedure to obtain standard errors. As we re-sample from the sample of individuals rather than DCE question observations, the bootstrap accounts for within-individual clustering as well as the imputation of unspecified characteristics (Cameron et al., 2008).

5.6.4 Heterogeneity of perceptions of empowerment

We may be interested in heterogeneity of perceptions of empowerment across different subgroups of the study participants; in particular, in this impact evaluation context, between project participants and non-participants. In order to assess such heterogeneity, we introduce interaction terms with a project participation indicator P_i. The probit model then becomes

$$\mathbb{P}(i \text{ chooses } 2 \text{ from } \{1, 2\}) = \Phi\left(\sum_{j=1}^{14} (\gamma_j \Delta x_j + \delta_j \Delta x_j P_i)\right), \tag{12}$$

where the coefficients γ_j reflect the perceptions of the project non-participants, while the coefficients δ_j reflect any divergence in perceptions of empowerment between project participants and non-participants.

6. Application of PIWE in Tunisia: results

In this section we report our analysis of the data collected in our pilot application. We document and discuss the results of the validity and consistency checks, the estimation of weights for the index, the application of the index as the outcome variable of interest in the evaluation of the AMAL project and conclude by assessing the impact that the project may have had on participants' perceptions of empowerment.

6.1 Validity and consistency checks

We designed the DCE to incorporate several checks of the validity of the exercise. Such checks have been interpreted by many implementers of DCEs as tests of the 'rationality' of the respondents (Lancsar and Louviere, 2006) and in some studies respondents who 'fail' such tests have been excluded from the sample for analysis. We concur with Lancsar and Louviere that 'failure', in most cases, is consistent with entirely rational preferences, while even the case of transitivity failures may be absorbed by the error term of the random utility model. We interpret the checks as tests of the validity of the measurement exercise rather than the rationality of the study participants, and do not exclude any study participants from the analysed data on the basis of the checks. As the checks are independent of the impact evaluation, we do not restrict the sample to the region of common support, nor do we use the PSM weights.

Study participants were not directly represented in the workshop at which the indicators of empowerment were identified, discussed in section 5.2, so an important first check is whether or not participants agree that each, individually, is indeed a characteristic of empowerment. The single question in part A and the first question in part B of the DCE module are both directly informative.

Results from these checks are reported in Table 4. Study participants overwhelmingly responded that the hypothetical woman achieving the indicator was the more empowered, with no more than 7.4% disagreement for any indicator and just 3% disagreement across all 14 indicators for the single-indicator question in part A. There was some variation across indicators, with the lowest agreement (at 92.6%) for equality of opportunity (E1). There was weak evidence of an order effect, with $p = 0.070$ for the test of equal proportions, suggesting that participants were slightly more likely to identify woman 1 as the more empowered even if woman 2 achieved the indicator.

Overall, this reassures us that the study participants agreed with the workshop participants that each is indeed an indicator of empowerment.

Table 4. Checks on individual indicators as characteristics of empowerment

		Part A		Part B	
		P(agree)	N	P(agree)	N
Self confidence and self esteem	P1	0.977	43		
Ability to make decisions for herself	P2	1.000	36		
Recognising that violence is not acceptable	P3	0.974	39		
Awareness that collective action is more effective	P4	0.962	26		
Knowledge and awareness of women's rights	P5	0.951	41		
Ability to make decisions in the household	R1	1.000	43	0.951	82
Participation and ability to make decisions in the public sphere	R2	0.944	36	0.931	72
Participation […] influence or make decisions in the political sphere	R3	0.941	34	0.812	85
Taking action to stop violence	R4	0.938	32	0.901	101
Independent income	R5	1.000	43	0.902	82
Control over resources in her household	R6	1.000	34	0.940	83
Equality of opportunity	E1	0.926	27		
Social norms	E2	1.000	35		
Legislative protection for women's rights	E3	0.944	36		
Woman 1 achieves indicator (pooled)		0.984	251	0.935	170
Woman 2 achieves indicator (pooled)		0.957	254	0.890	335
All (pooled)		0.970	505	0.905	505

Notes: P(agree) indicates the proportion of participants whose response to the question confirmed the indicator as a characteristic of empowerment. N is the relevant sample size, which varies as indicators (part A) and indicator combinations (part B) were assigned to participants at random. P-values for tests of equal proportions, whether hypothetical woman 1 or 2 achieves the indicator, are 0.070 (part A) and 0.098 (part B).

Similar results, also reported in Table 4, were obtained with the first question in part B, in which the hypothetical women differed in their attainment of only one of the three specified indicators. Here the proportion disagreeing was higher at 9.5% across all six relational indicators. This may result from the greater complexity of the three-indicator comparisons, resulting in greater error variance in the random utility model (equation 8) or simply from cases where the participant did not have a strong opinion so the error term dominated. A worst-case interpretation is that around 19% of the participants had not understood the exercise and responded effectively at random, but even in this case we can be confident that over 80% of respondents understood. There is some evidence of variation across indicators ($p = 0.030$ for the F-test of equality), with the greatest disagreement of 18.8% for political participation (R3) as a characteristic of empowerment. Again, there is weak evidence of an order effect, with $p = 0.098$ for the test of equal proportions, which highlights the importance of indicator-achievement balance across hypothetical woman 1 and woman 2 in the DCE design.

Blocks 1 and 2 of part B of the DCE, allocated to 325 of the study participants, admit the possibility of an intransitive (inconsistent) cyclic response in which profile A is considered more empowered than profile B, profile B is considered more empowered than profile C, and yet profile C is considered more empowered than profile A. If all study participants had made choices entirely at random there would have been eight equally likely patterns of response to questions 2-3 in part B, two of which (25%) are cyclic. In practice, 17 of the 325 participants (5.23%) expressed cyclic responses; we strongly reject the hypothesis of 25% ($p = 0.000$). The point estimate of 5.23% is consistent with 21% of respondents choosing effectively at random, very close to the 19% that explains the part B indicator checks above. Again, this is consistent with around 80% of the study participants understanding the exercise and expressing consistent choices.

Overall, the validity and consistency checks reassure us that the DCE elicited useful information about participants' perceptions of empowerment. Study participants agree that each indicator represents a characteristic of empowerment, while 80% or more fully understood the DCE exercise and expressed consistent choices. We do not exclude any participants from the sample as a result of these checks, maintaining our fully-participatory approach and allowing the choices expressed even by those who did not have strong opinions or did not fully understand the exercise to contribute to the results. Inconsistencies both within and across study participants are absorbed by the error term in our random utility model (eq. 7).

6.2 Estimation of PIWE weights

We pool study participants' responses to questions in parts B and C of the DCE module to estimate a probit model (eq. 8) and thus recover the PIWE weights w_j^p (eq. 9). Table 5 reports coefficient estimates for four versions of the model, together with tests of the hypothesis of equal coefficients.[9] We also report the proportion of study participants' choices that are correctly predicted by each model (and thus any resulting index). For each model this is substantially greater than the 59.7% of study participants' choices correctly predicted by an equally-weighted index.

Model (1) maintains the assumption that non-specified characteristics are equal for the hypothetical women within each choice set (no imputation of non-specified characteristics). Models (2) – (4) report coefficient estimates with imputation of unspecified characteristics as
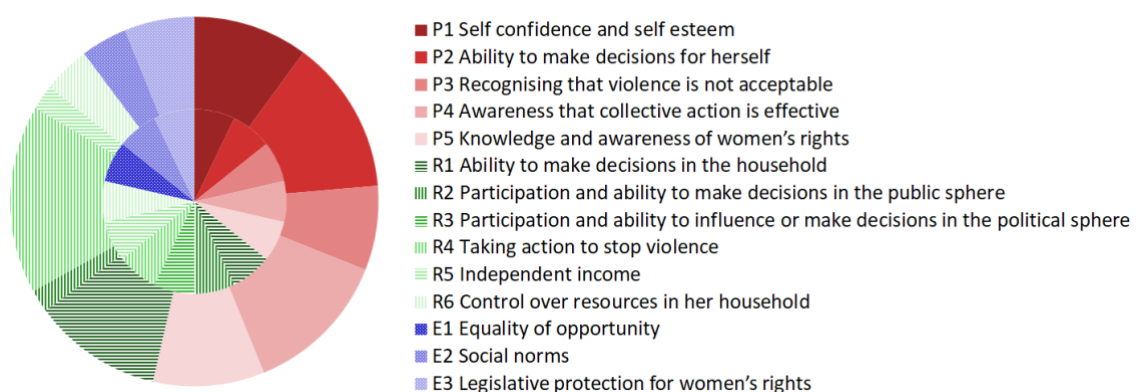
---

9 More coefficients are significantly different from zero; at p<0.05 all 14 in model (1), 7 of 14 in model (2), 5 of 14 in model (3) although only 2 of 14 in model (4). We do not report these tests, as the hypothesis of equality is more relevant to the study.

described in section 3.4. In models (3) and (4), indicators that yield negative coefficients are sequentially excluded, to ensure that all PIWE weights are non-negative.

We chose model (4), in which unspecified characteristics are imputed and coefficients constrained to be non-negative, to provide the weights for PIWE, which are obtained by re-scaling the estimated coefficients to sum to 100% (eq. 9). The PIWE weights are illustrated and contrasted with OxWEI (equal) weights in Figure 4. This demonstrates that the total weight (54% solid red) ascribed to the personal indicators is substantially greater than it would have been had they been allocated equal weights (36%), while the total weights allocated to the relational and, especially, environmental indicators (36% striped green and 10% dotted blue respectively) are substantially less than under equal weights (43% and 21% respectively). Only three of the coefficients of model (4) were individually significantly different from equal at the 5% significance level; interestingly these are all coefficients on relational indicators, reflecting the greater power for the relational indicators that arises from the greater effective sample size, as the part B DCE questions specified only relational indicators. Participation in the public sphere (R2) and independent income (R5) receive significantly smaller weights than they would have received in an equally weighted index. Conversely, taking action to stop violence (R4) receives a significantly greater weight. The weights assigned to decision making (P2) and awareness of collective action (P4) are relatively high, but not significantly different from equal weights.

Figure 4. Comparison of PIWE (outer ring) and OxWEI (inner circle) weights.

Legend:
- P1 Self confidence and self esteem
- P2 Ability to make decisions for herself
- P3 Recognising that violence is not acceptable
- P4 Awareness that collective action is effective
- P5 Knowledge and awareness of women's rights
- R1 Ability to make decisions in the household
- R2 Participation and ability to make decisions in the public sphere
- R3 Participation and ability to influence or make decisions in the political sphere
- R4 Taking action to stop violence
- R5 Independent income
- R6 Control over resources in her household
- E1 Equality of opportunity
- E2 Social norms
- E3 Legislative protection for women's rights

Having estimated the PIWE weights, we may compute the PIWE and OxWEI (equally-weighted) scores for the study participants on the basis of their empirical achievement of all indicators. We find a strong but not perfect correlation of 0.93 between the two indices. The relationship between the indices is illustrated in Figure 5, which demonstrates that individuals' PIWE and OxWEI scores differ by up to 29 percentage points.

Table 5. Estimation of probit coefficients and PIWE weights

		(1) Choice	(2) Choice	(3) Choice	(4) Choice	PIWE weights
Self confidence and self esteem	P1	1.104 (0.150)	0.726 (0.246)	0.381 (0.259)	0.391 (0.267)	10.10%
Ability to make decisions for herself	P2	1.637* (0.209)	0.733 (0.429)	0.735 (0.426)	0.524 (0.427)	13.53%
Recognising that violence is not acceptable	P3	1.857*** (0.193)	0.136 (0.417)	0.575 (0.387)	0.286 (0.379)	7.39%
Awareness that collective action is effective	P4	1.546 (0.168)	0.677 (0.286)	0.665 (0.314)	0.496 (0.325)	12.81%
Knowledge and awareness of women's rights	P5	1.553 (0.185)	0.578 (0.270)	0.549 (0.314)	0.380 (0.330)	9.79%
Ability to make decisions in the household	R1	1.200 (0.073)	0.340 (0.137)	0.422 (0.133)	0.440 (0.133)	11.35%
Participation and ability to make decisions in the public sphere	R2	0.758*** (0.060)	0.266 (0.130)	0.222 (0.126)	0.076*** (0.113)	1.97%
Participation […] influence or make decisions in the political sphere	R3	0.428*** (0.062)	-0.863*** (0.168)			0.00%
Taking action to stop violence	R4	1.306 (0.077)	0.592** (0.138)	0.623*** (0.122)	0.640*** (0.119)	16.50%
Independent income	R5	0.616*** (0.062)	0.040*** (0.142)	0.158 (0.126)	0.085** (0.122)	2.19%
Control over resources in her household	R6	0.898*** (0.066)	0.215 (0.133)	0.180 (0.122)	0.161 (0.121)	4.15%
Equality of opportunity	E1	1.605 (0.218)	0.472 (0.256)	-0.622*** (0.215)		0.00%
Social norms	E2	1.769*** (0.179)	0.315 (0.312)	0.067 (0.362)	0.158 (0.367)	4.08%
Legislative protection for women's rights	E3	1.648* (0.202)	0.333 (0.237)	0.463 (0.227)	0.238 (0.221)	6.14%
Unspecified indicators imputed		NO	YES	YES	YES	
Sum of coefficients		17.925	4.559	4.418	3.876	100%
Coefficient values if equal		1.280	0.326	0.340	0.323	8.33%
Proportion of choices correctly predicted		0.735	0.694	0.693	0.672	67.2%
Observations		2988	2988	2988	2988	

Notes: All models restricted to common support (N=498; six DCE questions and thus observations per study participant), PSM weights applied, standard errors in parentheses. Model (1): no imputation of non-specified indicators; standard errors clustered at respondent level. Models (2) – (4): imputation of non-specified indicators; standard errors obtained through bootstrap re-sampling of respondents (2000 repetitions). Hypotheses of equal rather than zero coefficients reported; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5. Comparison of PIWE and OxWEI scores.
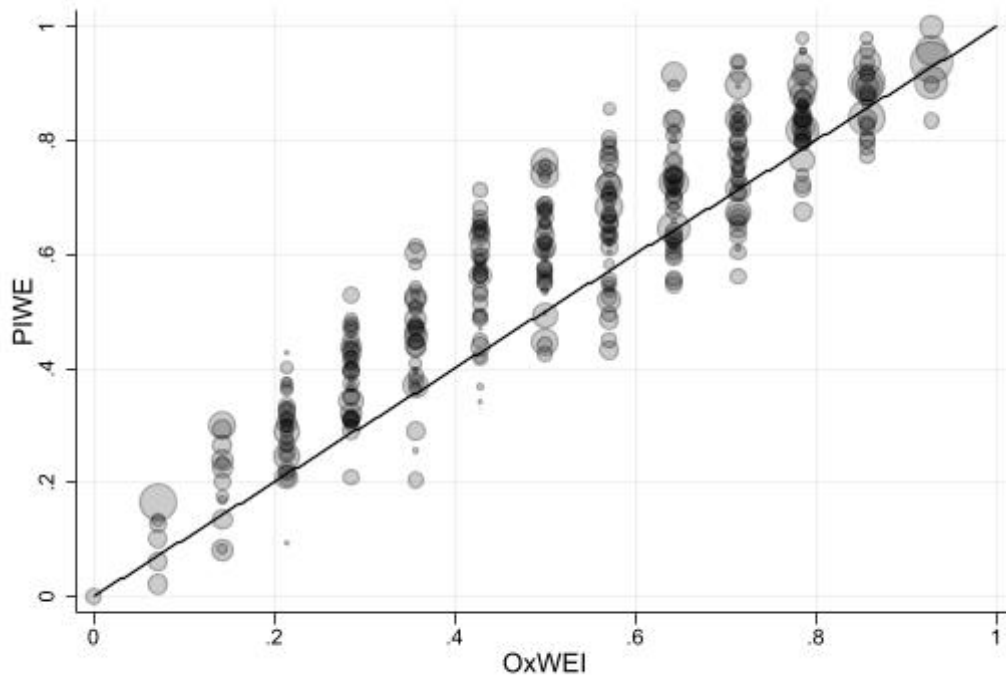
Source: authors' calculations. Marker size represents aggregate PSM weight for each OxWEI/PIWE combination; solid line is the line of equality.

6.3 Application to evaluation of AMAL

6.3.1 Impact on empowerment

This pilot implementation of PIWE was conducted in the context of the quasi-experimental impact evaluation of the Oxfam project AMAL in Tunisia, reported in detail by Lombardini (2018). To illustrate its application, we report the key estimates of impact in Table 6. According to both PIWE and the equally-weighted OxWEI, the project had a small but significant positive impact on the empowerment of project participants. As the PIWE reflects the study participants' collective perception of empowerment, we may conclude that the project had a positive impact on empowerment as perceived by the study participants.

Table 6. Impact of AMAL project on women's empowerment

	(1) OxWEI (equal weights)	(2) PIWE
Project participants' mean	0.578	0.641
Matched non-participants' mean	0.513	0.595
Difference in means (ATT)	0.065***	0.047**
	(0.023)	(0.023)
Observations	498	498

Notes: Sample restricted to common support (N=498 of which 214 are project participants); matched non-participant sample constructed by application of a Gaussian kernel to estimated propensity scores. Standard errors in parentheses. Hypotheses of zero difference between project participants' mean score and matched non-participants' mean score reported; $* p < 0.1$, $** p < 0.05$, $*** p < 0.01$.

6.3.2 Impact on perceptions of empowerment

It is possible that project participation had an impact on perceptions of empowerment. To assess the extent to which this is a relevant concern in our study, we re-estimate the probit model with project participation interaction terms (eq. 12). We impute the unspecified indicators and retain all indicators in the model, so this amounts to a decomposition of model (2) in Table 2. Results are reported in Table 7.

The only interaction term coefficient significantly different from zero is that on the ability to make decisions for herself (P2), which is substantially higher for the project participants than non-participants. This suggests that the project may have had an impact on the perception of this indicator as an important characteristic of empowerment among project participants.

The standard errors are relatively large for the personal and environmental indicators, so the lack of significant differences for these indicators (other than P2) cannot be taken as indicative of no impact. However, the standard errors are smaller for the relational indicators, and it is interesting to observe that we do not find a significant impact of the project on the weight that the respondents attach to any of these indicators.

Table 7. Estimation of probit model with project participation interactions

		γ_j (Coefficient for non-participants)	δ_j (Interaction)
Self confidence and self esteem	P1	0.583 (0.257)	0.327 (0.290)
Ability to make decisions for herself	P2	0.044 (0.517)	1.134** (0.555)
Recognising that violence is not acceptable	P3	0.315 (0.482)	-0.323 (0.513)
Awareness that collective action is effective	P4	0.502 (0.339)	0.309 (0.342)
Knowledge and awareness of women's rights	P5	0.636 (0.309)	-0.054 (0.316)
Ability to make decisions in the household	R1	0.477 (0.155)	-0.210 (0.175)
Participation and ability to make decisions in the public sphere	R2	0.260 (0.142)	0.005 (0.149)
Participation and ability to influence or make decisions in the political sphere	R3	-0.775*** (0.189)	-0.219 (0.214)
Taking action to stop violence	R4	0.532 (0.166)	0.113 (0.180)
Independent income	R5	0.038* (0.155)	-0.008 (0.151)
Control over resources in her household	R6	0.305 (0.066)	-0.162 (0.163)
Equality of opportunity	E1	0.435 (0.306)	0.200 (0.394)
Social norms	E2	0.467 (0.349)	-0.193 (0.345)
Legislative protection for women's rights	E3	0.283 (0.283)	0.070 (0.304)
Sum of coefficients		4.100	
Coefficient values if equal		0.293	
Observations			2988

Notes: Restricted to common support (N=498; six DCE questions and thus observations per study participant), PSM weights applied. Imputation of non-specified indicators; standard errors (in parentheses) obtained through bootstrap re-sampling of respondents (2000 repetitions). Hypotheses of equal rather than zero coefficients reported for γ_j; hypotheses of zero coefficients reported for interaction term coefficients δ_j; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

7. Discussion

In this final session we discuss our results in more depth, highlighting practical and conceptual considerations that emerged from our pilot implementation of PIWE.

7.1 Implementation of PIWE

We reflect here on several practical and conceptual considerations that emerged in our pilot implementation.

7.1.1 Indicator choice

The DCE allowed study participants to collectively determine the weights assigned to a candidate list of indicators. Rejection of candidate indicators would have been possible (an indicator not considered relevant by the participants is assigned low weight), but there is no mechanism for participants to propose extra indicators at the quantitative survey stage.

In our pilot implementation, the validity checks reassured us that the participants considered all 14 candidates to be valid indicators of empowerment. However, it is possible that indicators considered important by the participants had been overlooked. With an extensive candidate list, it is likely that included indicators served as proxies for important excluded indicators, minimising the effect on impact evaluation results but complicating interpretation of the estimated weights. In future applications, we recommend that study participants should be represented in any preliminary qualitative exercise to identify the candidate list of indicators.

7.1.2 Experimental design

The need to develop and implement a bespoke DCE design in the context of a tight evaluation timeframe was challenging. Our decision to focus attention on a subset of indicators resulted in an inefficient experimental design that limited our power to estimate precise weights for some of the PIWE indicators. With limited sample size and survey time availability, there is scope to benefit in future applications from the implementation of more efficient experimental designs. As the literature on efficient designs for partial profile DCEs that are robust to model mis-specification is underdeveloped, this is a priority for future research.

7.1.3 Heterogeneity

It is likely that study participants' perceptions of empowerment, in particular the trade-offs that they make between different indicators, will be heterogeneous. In our model (eq. 7) this heterogeneity is absorbed in the error term $\varepsilon_{1i} - \varepsilon_{2i}$, while our balanced experimental design and

pooled estimation approach ensure that the estimated participatory weights represent an impartial average over participants' perceptions.

With sufficiently large sample size, it is possible to empirically assess the extent of heterogeneity by including interaction terms with subgroup dummies (eq. 12). We implemented this for project participants and non-participants (Table 6).

7.1.4 Functional form

We have imposed the assumption that PIWE is additive in the indicators, while in our pilot implementation we imposed the further assumption that all indicators are binary. These assumptions align with other indices in the literature, discussed in sections 3.2 and 3.4, but need not align with study participants' perceptions of empowerment. Our experimental design and estimation approach (described in section 5) ensure that PIWE is the best fit to participants' perceptions given these constraints; in future applications with sufficiently large sample sizes it may be possible to relax these restrictive assumptions.

7.1.5 Comparability

Like all context-specific and participatory approaches to measuring empowerment, PIWE has no guarantee of comparability across different contexts or implementations. As we discuss in section 2, there are strong arguments in favour of participatory approaches, while given the context-specificity of indicators of empowerment, apparently comparable indices (such as WEAI) are not in practice straightforwardly comparable across different contexts. Where it is necessary to make comparisons across different contexts, we suggest that researchers may wish to explore the extent to which context-specific indices align, and whether findings are robust to the choice of index.

7.1.6 Subjectivity and internalisation of disempowering norms

It is important to minimise the distortions that can emerge in subjective measurement (Bertrand and Mullainathan, 2001; Jahedi and Méndez, 2014). This is one reason why we do not directly ask study participants to assess their own empowerment. DCE questions are relatively straightforward, minimising cognitive demands that can distort subjective data; Arentze et al. (2003) find that data quality in a transport-choice DCE in South Africa did not vary with respondent literacy. Distortion through social desirability may be minimised by assuring participants of anonymity; in a literate context, they could record their responses without involvement of an enumerator. In some cases, participants may not have a strong attitude about which of the hypothetical women is more empowered and so may respond essentially at random. This is absorbed by the error term in the

random utility model and, appropriately, would tend to reduce the coefficients and thus weights on the relevant indicators.

Internalisation of disempowering norms poses a particular issue for subjective or participatory measurement of empowerment. Its extent will vary across contexts, while its impact may be partly mitigated by DCE question phrasing. Participants are asked which of the hypothetical women is 'more empowered' rather than 'better', while the DCE elicits trade-offs between indicators rather than value judgements about levels of empowerment. Nevertheless, it is important to be conscious of this issue. Where it is a particular concern, the PIWE exercise could be repeated with community activists or even empowerment experts familiar with the context, to explore the extent of divergence between participants' and activists' or experts' perceptions of empowerment.

7.1.7 Instability of perceptions

It is quite possible that the process of empowerment impacts perceptions of empowerment. This is an empirical question that PIWE can help to answer, through comparison of the index for different groups or at different times. It is a particular issue when PIWE is implemented for impact evaluation, as in our application. With a relatively small sample size and thus low power, we choose not to develop separate PIWEs for the project participants and non-participants. Had we been able to do so, an important element of the impact evaluation would have been to assess robustness of the results to the choice of index. We recommend that, where possible, divergence in project participants' and non-participants' perceptions, and robustness of evaluation to any divergence, be assessed.

7.2 Empirical methods and results

In the absence of an established method to model the inferences that study participants draw about unspecified indicators, we implemented an ad hoc imputation method. Two aspects of our empirical results suggest that this method was not optimal, indicating that further methodological research is needed to develop a better approach.

Firstly, the explanatory power of models (2) – (4), in which we impute the unspecified indicators, is actually slightly lower, at 67 – 69% of study participants' choices correctly predicted, than model (1) in which we do not impute, at 74%.

Secondly, despite study participants demonstrating through the validity checks that they consider all indicators individually to be indicators of empowerment, when we estimated model (2) to obtain the PIWE weights, we found that the coefficient on one indicator (R3, political participation) was

significantly negative. Some coefficient estimates were unstable on elimination of this indicator; the coefficient on a second indicator (E1, equality of opportunity) became negative and we eliminated it also. This coefficient instability may result from the correlations between indicators introduced by the imputation, so that included indicators proxy for excluded indicators. The diminution in the explanatory power is marginal, so we remain confident that the aggregate index (PIWE) reflects study participants' collective perceptions of empowerment. However, the instability makes us reluctant to ascribe direct meaning to individual indicator coefficient values or PIWE weights.

In our impact evaluation results, we observe with interest that the average PIWE scores are greater for both project participants and non-participants than the equally-weighted OxWEI scores. This suggests that the weighting of indicators in PIWE aligns more strongly with those indicators that are more commonly achieved by the study participants, perhaps suggesting a salience effect in their perceptions of empowerment as expressed through the DCE.

7.3 Concluding remarks

We have demonstrated, through a pilot implementation in the context of an impact evaluation in Tunisia, that it is possible to achieve participation at scale when measuring women's empowerment. The resulting Participatory Index of Women's Empowerment reflects the collective perceptions of empowerment of the participating women. While there remains scope for methodological improvement, in particular through development of efficient experimental designs and better imputation approaches, we hope that this measurement tool will prove a useful addition to the portfolio of methods available to researchers and practitioners who seek to fully involve the communities that they work with.

References

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

Alkire, S., Meinzen-Dick, R., Peterman, A., Quisumbing, A., Seymour, G., and Vaz, A. (2013). The Women's Empowerment in Agriculture Index. *World Development, 52*, 71–91.

Alsop, R., Bertelsen, M., and Holland, J. (2006). *Empowerment in practice: From analysis to implementation*. World Bank.

Arentze, T., Borgers, A., Timmermans, H., and DelMistro, R. (2003). Transport stated choice responses: effects of task complexity, presentation format and literacy. *Transportation Research Part E: Logistics and Transportation Review*, *39*(3), 229–244.

Aslanbeigui, N., Oakes, G., and Uddin, N. (2010). Assessing microcredit in Bangladesh: A critique of the concept of empowerment. *Review of Political Economy*, *22*(2), 181–204.

Bertrand, M., and Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *American Economic Review*, *91*(2), 67–72.

Bishop, D., and Bowman, K. (2014). Still learning: a critical reflection on three years of measuring women's empowerment in Oxfam. *Gender and Development, 22*(2), 253–269.

Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: a research note. *Qualitative Research*, *8*(1), 137–152.

Bridges, D. (2001). The ethics of outsider research. *Journal of Philosophy of Education*, *35*(3), 371–386.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414-427.

Chrzan, K. (2010). Using partial profile choice experiments to handle large numbers of attributes. *International Journal of Market Research*, *52*(6), 827–840.

Decancq, K., and Watson, V. (2019, July 3–5). *Eliciting weights for the Human Development Index with a discrete choice experiment* [Paper presentation]. Eighth ECINEQ Meeting, Paris, France.

Foster, J. E., and Shorrocks, A. F. (1991). Subgroup consistent poverty indices. *Econometrica*, *59*(3), 687–709.

Galiè, A., Teufel, N., Korir, L., Baltenweck, I., Girard, A. W., Dominguez-Salas, P., and Yount, K. M. (2019). The women's empowerment in livestock index. *Social Indicators Research*, *142*(2), 799–825.

Ghali, S., Goedhuys, M., Grimm, M., Meysonnat, A., Nillesen, E., and Zouari, S. (2018). *Women's and youth empowerment in rural Tunisia: An assessment using the Women's Empowerment in Agriculture Index (WEAI)* (Taqeem Impact Report Series No. 11). International Labour Office.

Glennerster, R., Walsh, C., and Diaz-Martin, L. (2018). *A practical guide to measuring women's and girl's empowerment in impact evaluations*. Abdul Latif Jameel Poverty Action Lab.

Gram, L., Morrison, J., and Skordis-Worrall, J. (2019). Organising concepts of 'women's empowerment' for measurement: a typology. *Social Indicators Research*, *143*(3), 1349–1376.

Greene, W. (2012). *Econometric analysis* (7th ed.). Prentice Hall.

Großmann, H. (2019). A practical approach to designing partial-profile choice experiments with two alternatives for estimating main effects and interactions of many two-level attributes. *Journal of Choice Modelling*, *32*, 100136.

Gupta, S., Vemireddy, V., Singh, D., and Pingali, P. (2019). Adapting the Women's Empowerment in Agriculture Index to specific country context: Insights and critiques from fieldwork in India. *Global Food Security*, *23*, 245–255.

Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, *22*(1), 1–30.

Hausmann, R., Tyson, L. D., and Zahidi, S. (2006). *The global gender gap report 2006*. World Economic Forum. https://www.weforum.org/reports/global-gender-gap-report-2006

Hensher, D., Rose, J., and Greene, W. (2015). *Applied choice analysis* (2nd ed.). Cambridge University Press.

Hutchings, C. (2014). Balancing accountability and learning: A review of Oxfam GB's Global Performance Framework. *Journal of Development Effectiveness*, *6*(4), 425–435.

Ibrahim, S., and Alkire, S. (2007). Agency and empowerment: A proposal for internationally comparable indicators. *Oxford Development Studies*, 35(4), 379–403.

Jahedi, S., and Méndez, F. (2014). On the advantages and disadvantages of subjective measures. *Journal of Economic Behavior and Organization*, *98*, 97–114.

Janssens, W., Sefoko, N., van Rooyen, J., and Bostyn, F. (2006). Measuring perceived black economic empowerment in the South African wine industry. *Agrekon*, *45*(4), 381–405.

Jupp, D., Ibn Ali, S., and Barahona, C. (2010). *Measuring empowerment? Ask them: Quantifying qualitative outcomes from people's own analysis* (Sida Studies in Evaluation 2010:1). Swedish International Development Cooperation Agency.

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change*, *30*(3), 435–464.

Kessels, R., Jones, B., and Goos, P. (2011). Bayesian optimal designs for discrete choice experiments with partial profiles. *Journal of Choice Modelling*, *4*(3), 52–74.

Kessels, R., Jones, B., and Goos, P. (2015). An improved two-stage variance balance approach for constructing partial profile designs for discrete choice experiments. *Applied Stochastic Models in Business and Industry*, *31*(5), 626–648.

Khader, S. J. (2011). *Adaptive preferences and women's empowerment*. Oxford University Press.

Klasen, S. (2006). Guest editor's introduction. *Journal of Human Development*, *7*(2), 145–159.

Lancsar, E., and Louviere, J. (2006). Deleting 'irrational' responses from discrete choice experiments: A case of investigating or imposing preferences? *Health Economics*, *15*(8), 797–811.

Liebe, U., Moumouni, I. M., Bigler, C., Ingabire, C., and Bieri, S. (2020). Using factorial survey experiments to measure attitudes, social norms, and fairness concerns in developing countries. *Sociological Methods and Research 49*(1), 161–192.

Lombardini, S. (2018). *Women's empowerment in Tunisia: Impact evaluation of the project 'AMAL: Supporting Women's Transformative Leadership' in Tunisia'* (Effectiveness Review Series 2016/17). Oxfam GB.

Lombardini, S., Bowman, K., and Garwood, R. (2017). *A 'how to' guide to measuring women's empowerment: Sharing experience from Oxfam's impact evaluations*. Oxfam GB.

Malapit, H., Quisumbing, A., Meinzen-Dick, R., Seymour, G., Martinez, E. M., Heckert, J., Rubin, D., Vaz, A., Yount, K. M., and GAAP2 Study Team (2019). Development of the project-level Women's Empowerment in Agriculture Index (pro-WEAI). *World Development*, *122*, 675–692.

Malhotra, A., and Mather, M. (1997). Do schooling and work empower women in developing countries? Gender and domestic decisions in Sri Lanka. *Sociological Forum 12*(4), 599–630.

Malhotra, A., and Schuler, S. R. (2005). Women's empowerment as a variable in international development. In D. Narayan (Ed.), *Measuring empowerment: Cross-disciplinary perspectives* (pp. 71–88). World Bank.

Narayan, D. (Ed.). (2002). *Empowerment and poverty reduction: A sourcebook.* World Bank.

Narayan, D. (2005). Conceptual framework and methodological challenges. In D. Narayan (Ed.), *Measuring empowerment: Cross-disciplinary perspectives* (pp. 3–38). World Bank.

Narayanan, S., Lentz, E., Fontana, M., De, A., and Kulkarni, B. (2019). Developing the Women's Empowerment in Nutrition Index in two states of India. *Food Policy, 89*, 101780.

Priya, P., Venkatesh, A., and Shukla, A. (2021). Two decades of theorising and measuring women's empowerment: Literature review and future research agenda. *Women's Studies International Forum, 87,* 102495.

Raj, A. (2020). Women and girls' health agency: Operationalizing the "Can-Act-Resist" of the women's empowerment process. *EClinicalMedicine, 20*, 100287.

Richardson, R. A. (2018). Measuring women's empowerment: A critical review of current practices and recommendations for researchers. *Social Indicators Research, 137*(2), 539–557.

Robeyns, I. (2005). The capability approach: a theoretical survey. *Journal of Human Development, 6*(1), 93–117.

Rocha, E. M. (1997). A ladder of empowerment. *Journal of Planning Education and Research, 17*(1), 31–44.

Rowlands, J. (1995). Empowerment examined. *Development in Practice, 5*(2), 101–107.

Rowlands, J. (1997). *Questioning empowerment: Working with women in Honduras.* Oxfam UK and Ireland.

Ryan, M., Gerard, K., and Amaya-Amaya, M. (Eds.). (2008). *Using discrete choice experiments to value health and health care.* Springer.

Ryan, M., Watson, V., and Entwistle, V. (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health Economics, 18*(3), 321–336.

Sen, A. (1985a). *Commodities and capabilities.* North-Holland.

Sen, A. (1985b). Well-being, agency and freedom: The Dewey lectures 1984. *The Journal of Philosophy, 82*(4), 169–221.

Taylor, J. (2000). *So now they are going to measure empowerment!.* Community Development Resource Association.
https://www.cdra.org.za/uploads/1/1/1/6/111664/so_now_they_are_going_to_measure_emp owerment_james_taylor_2000.doc

United Nations Children's Fund. (1990). *Strategy for improved nutrition of children and women in developing countries.* (UNICEF Policy Review E/ICEF/1990/ L.6).

United Nations Development Programme. (1995). *Human development report 1995: Gender and human development.* Oxford University Press.

United Nations Development Programme. (2019). *Technical notes to the Human Development Report 2019.*

United Nations General Assembly. (2015). *Resolution 70/1. Transforming our world: The 2030 agenda for sustainable development* (UNGA Res A/RES/70/1).

UN Women. (2014). *Gender mainstreaming in development programming* (Guidance Note).

VeneKlasen, L., and Miller, V. (2002). *A new weave of power, people and politics: The action guide for advocacy and citizen participation.* World Neighbours.

Vigneri, M., and Lombardini, S. (2017). *Women's empowerment in Mali: Impact evaluation of the educational project: 'Girls CAN-Promoting Secondary Education in West Africa'* (Effectiveness Review Series 2015/16). Oxfam GB.

Watson, V., Dibben, C., Cox, M., Atherton, I., Sutton, M., and Ryan, M. (2019). Testing the expert based weights used in the UK's Index of Multiple Deprivation (IMD) against three preference-based methods. *Social Indicators Research*, *144*(3), 1055–1074.

Wisor, S., Bessell, S., Castillo, F., Crawford, J., Donaghue, K., Hunt, J., Jaggar, A., Liu, A. and Pogge, T. (2014). *The Individual Deprivation Measure: A gender-sensitive approach to poverty measurement.* Individual Deprivation Measure.

Appendices

In these appendices we document several details of the implementation of the discrete choice experiment (DCE): the translation and back translation of the DCE question and indicator texts, some details of the DCE experimental design and its implementation in SurveyCTO.

A. Translation of DCE questions

Careful phrasing of the DCE questions was important, to ensure that we captured perceptions of *empowerment* rather than any other concept. Back translation allows us to check the quality of the translation into Tunisian Arabic and to be aware of any issues that may have arisen. The back translations of the DCE questions are reported in Table A1.

Table A1. Translation and back translation of DCE question texts.

English	Tunisian Arabic	Back translation
Please choose which of these two women you consider to be more empowered	من فضلك اختار شكوني المرا إلي تعتبرها الأكثر عندها القدرة باش تكون فاعلة	Please choose which woman you consider more able of being an actor/ capable of action
Now I will describe two different women. Please choose which of these two women you consider to be more empowered.	توة باش نوصفلك زوز نسا مختلفين وانتي باش تختار أناهي منهم المرا إلي تعتبرها الأكثر عندها القدرة باش تكون فاعلة	Now I'm going to describe to you two different women and you choose which of them you consider more capable of action.
Here are the characteristics of Woman 1:	هاذي صفات المرا1 :	These are the features/ attributes of Woman 1
Here are the characteristics of Woman 2:	هاذي صفات المرا2 :	These are the features/ attributes of Woman 2

Interesting observations that emerged during the back translation of the DCE question texts are discussed in section 5.3 of the main text. The back translations of the indicator texts, with observations that emerged, are reported in Table A2.

Table A2. Translation and back translation of DCE indicator texts.

Indicator	Code	Icon	English	Tunisian Arabic	Back translation
Self confidence and self esteem	P1		She feels that she doesn't have many good qualities	تحس انها ماعندهاش برشا صفات باهية	You feel she doesn't have a lot of good qualities/ attributes
			She feels that she has a number of good qualities	حس انهآعندها برشا صفات باهية	You feel she has a lot of good qualities / attributes
Back translator's note: spelling mistake (bold) in 'achieved' text.					
Author observation: translation error in grammatical person.					
Ability to make decisions for herself	P2		She cannot independently take decisions regarding herself	ماهيش قادرة تاخذ قرارات متعلقة باها بصفة مستقلة	She is not capable of taking decisions related to herself in an independent way.
			She can independently take decisions regarding herself	قادرة تاخذ قرارات متعلقة باها بصفة مستقلة	She is capable of taking decisions related to herself in an independent way.

Indicator	Code	Icon	English	Tunisian Arabic	Back translation
Recognising that violence is not acceptable	P3		She considers it acceptable for a man to beat his wife	تشوفها حاجة مقبولة وقتلي الراجل يضرب مرتو	She finds it/ sees it as acceptable that a husband beats his wife.
			She considers it unacceptable for a man to beat his wife	ما تشوفهاش حاجة مقبولة وقتلي الراجل يضرب مرتو	She does not see it as something acceptable that a husband beats his wife.
Awareness that collective action is effective	P4		She does not believe that acting as a group is effective to solve issues	ما تأمنش أنو التصرف مع مجموعة يكون فعال لحل المشاكل	She does not believe that acting in a group is effective in solving problems.
			She believes that acting as a group is effective to solve issues	تأمن أنو التصرف مع مجموعة يكون فعال لحل المشاكل	She does believe that acting in a group is effective in solving problems.

Back translator's note: alternative translation 'working with a group' (for 'acting in a group')

Indicator	Code	Icon	English	Tunisian Arabic	Back translation
Knowledge and awareness of women's rights	P5		She does not think that men and women should have the same rights	ما تعتبرش أنو الرجال والنساء لازم أن يكون عندهم نفس الحقوق	She does not consider that men and women must have the same rights.
			She thinks that men and women should have the same rights	تعتبر أنو الرجال والنساء لازم أن يكون عندهم نفس الحقوق	She does consider that men and women must have the same rights.
Ability to make decisions in the household	R1		She is not able to make decisions within her household	ماهيش قادرة باش تاخو قرارات في الدار	She is incapable of taking/unable to take decisions at home/in the home.
			She is able to make decisions within her household	قادرة باش تاخو قرارات في الدار	She is capable of taking/able to take decisions at home/in the home.

Back translator's note: the word representing 'able/capable of' is the same used in R4 (achieved) below.

Indicator	Code	Icon	English	Tunisian Arabic	Back translation
Participation and ability to make decisions in the public sphere	R2		She does not participate in civil society and associations	ما تشاركش في المجتمع المدني والجمعيات	She does not participate in civil society [*activities*] and associations [*formal organisations*].
			She actively participates in civil society and associations	ناشطة في المجتمع المدني والجمعيات	She is active in civil society [*activities*] and associations [*formal organisations*].

Back translator's note: a different verb used in 'not achieved' and 'achieved' texts. 'Civil society' indicates activities, while 'associations' indicates formal organisations.
Author observation: the verb inconsistency reflects slight inconsistency in the English phrasing.

Indicator	Code	Icon	English	Tunisian Arabic	Back translation
Participation and ability to influence or make decisions in the political sphere	R3		She does not actively participate in a political party	ماهيش منخرطة ولا ناشطة في حزب سياسي	She is not a member of or active in a political party.
			She actively participates in a political party	ناشطة في حزب سياسي	She is active in a political party.

Back translator's note: the word 'ناشطة' translates as 'activist'.

Indicator	Code	Icon	English	Tunisian Arabic	Back translation
Taking action to stop violence	R4		She experiences violence and does not report it	تتعرض للعنف و ماتشكيش	She is subjected to violence, but does not complain.
			If she experiences violence she is able to report it	إذاكان تتعرض للعنف قادره باش تشكي	If she is subjected to violence, she is able to complain.

Back translator's note: the word representing 'able' is the same used in R1 (achieved) above. 'Complain' could be interpreted as talking to authorities or simply 'talking to someone'.

Indicator	Code	Icon	English	Tunisian Arabic	Back translation
Independent income	R5		She does not have an independent source of income	ماعندهاش مورد رزق مستقل	She doesn't have an independent source of income
			She has an independent source of income	عندها مورد رزق مستقل	She has an independent source of income
Control over resources in her household	R6		She has no control over assets and resources in her household	ماعندهاش السيطرة على الممتلكات والموارد في دارە	She doesn't have control over belongings and resources in her home/household
			She has control over assets and resources in her household	عندها السيطرة على الممتلكات والموارد في دارها	She has control over belongings and resources in her home/ household
Equality of opportunity	E1		She lives in a community that does not allow women to have equal political opportunities as men	تعيش في مجتمع ما يسمحش للمرا باش يكون عندها نفس الفرص السياسية اللي يتمتع بها الراجل	She lives in a society which does not permit a woman to have the same political opportunities that men enjoy.
			She lives in a community that ensures that women have equal political opportunities as men	تعيش في مجتمع يسمح للمرا باش يكون عندها نفس الفرص السياسية اللي يتمتع بها الراجل	She lives in a society which allows a woman to have the same political opportunities that men enjoy.
Social norms	E2		She lives in a society which does not allow her to be free	تعيش في مجتمع مايخليهاش تكون حرة	She lives in a society which does not allow her to be free.
			She lives in a society which allows her to be free	تعيش في مجتمع يخلاها تكون حرة	She lives in a society which allows her to be free.
Legislative protection for women's rights	E3		She lives in a society where women's rights are not enshrined in law	تعيش في مجتمع حيث حقوق المرأة ماهيش منصوصة في القانون	She lives in a society where the rights of women are not enshrined/ expressed in law.
			She lives in a society where women's rights are enshrined in law	تعيش في مجتمع حيث حقوق المرأة منصوصة في القانون	She lives in a society where the rights of women are enshrined/ expressed in law.

Back translator's note: these [E3] texts are quite erudite, using formal Arabic.

B. DCE design

The Tunisian Arabic implementation of a DCE question is illustrated in Figure B1. The 3-indicator combinations included in parts B and C of the DCE module are documented in Table B1. The question blocks utilised in parts B and C, documented in Table 2 in the main text, are illustrated in Figure B2. This demonstrates clearly that blocks 1 and 2 permit cyclic choices, which allows us to implement a check of consistency.

Figure B1. Example DCE question (Tunisian Arabic).

Figure B2. Choice sets and their blocks in part B of the DCE module.
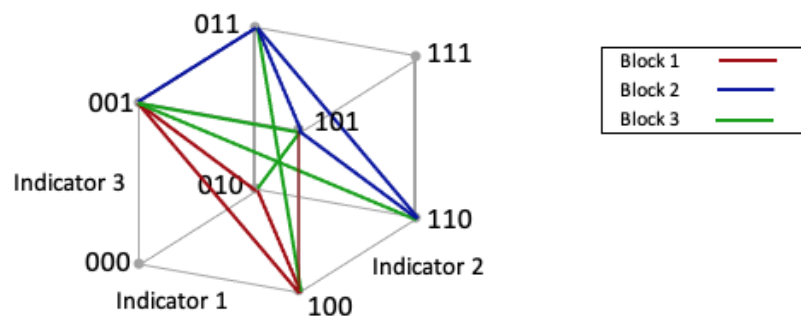
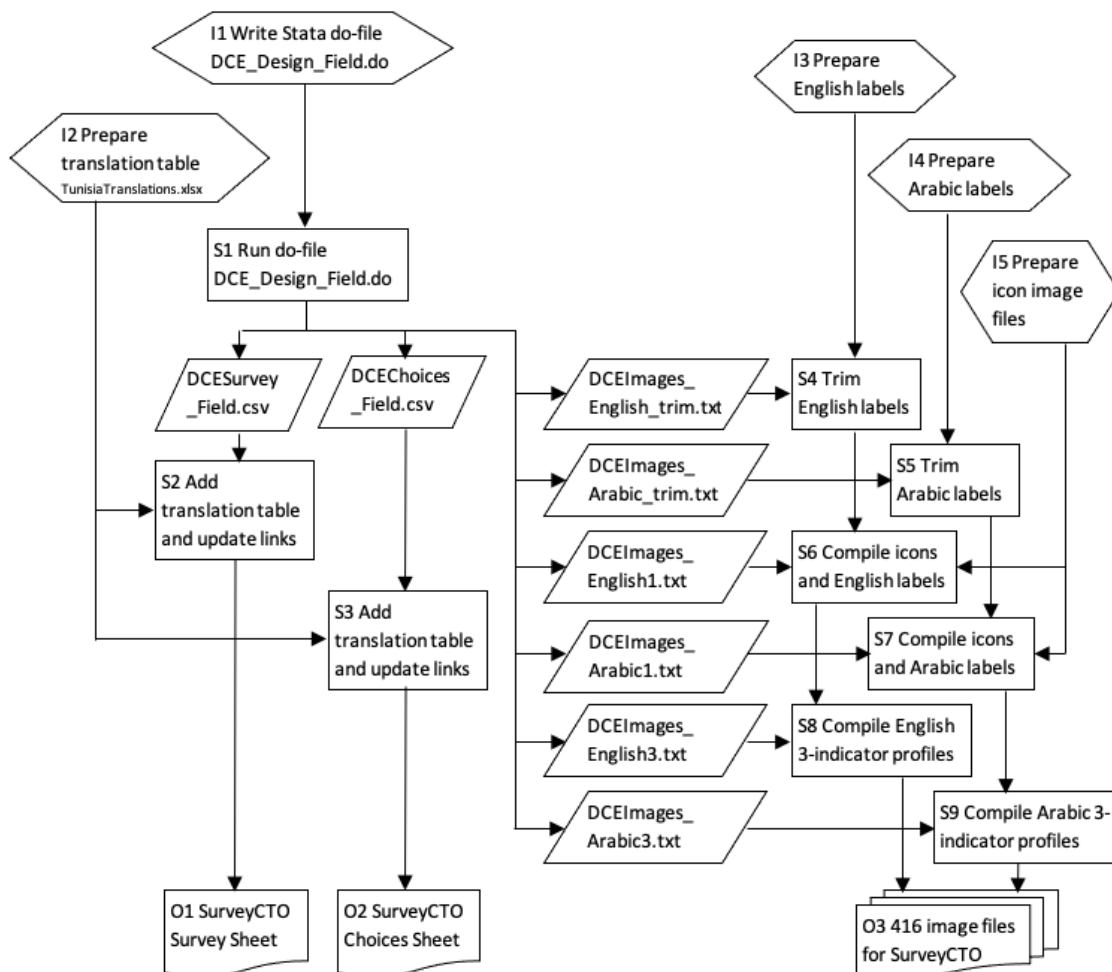**Table B1. 3-indicator combinations in parts B and C of the DCE module**

		Part B	Part C
Personal Indicators			
Self confidence and self esteem	P1		✓ ✓ ✓
Ability to make decisions for herself	P2		✓ ✓ ✓
Recognising that violence is not acceptable	P3		✓ ✓ ✓
Awareness that collective action is effective	P4		✓ ✓ ✓
Knowledge and awareness of women's rights	P5		✓ ✓ ✓
Relational Indicators			
Ability to make decisions in the household	R1	✓	✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓
Participation and ability to make decisions in the public sphere	R2	✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓ ✓
Participation and ability to influence or make decisions in the political sphere	R3	✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓ ✓
Taking action to stop violence	R4	✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓
Independent income	R5	✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓ ✓
Control over resources in her household	R6	✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓ ✓
Environmental Indicators			
Equality of opportunity	E1		✓ ✓ ✓
Social norms	E2		✓ ✓ ✓
Legislative protection for women's rights	E3		✓ ✓ ✓

Notes: Columns represent indicator combinations.

C. Implementation in SurveyCTO

As described in the main text, we automated the process to prepare and name the compiled indicator text and icon files to illustrate each hypothetical woman as well as preparing the SurveyCTO code to randomise among the profile pairs and call the correct image files. This was achieved by writing a Stata do-file (of approximately 500 lines) whose outputs were the SurveyCTO survey and choices sheets for the DCE module (almost 5,000 lines in total), incorporating the image filenames, as well as six ImageMagick batch files (approximately 1,000 lines in total). These batch files contained the commands needed to compile the text and icons for each hypothetical woman in both English and Tunisian Arabic. Unfortunately, it was not possible for the Stata do-file to output the Arabic translations of the question and choice texts, so we coded it to output Excel vlookup commands and created a lookup table, then manually ran the commands to pull in the Arabic translations. The workflow for the SurveyCTO implementation process is detailed in Figure C1 and illustrated in Table C1.

Figure C1. Workflow chart for preparation of the DCE question image files.

Table C1. Detailed SurveyCTO implementation process.

Inputs

Item	Filename	Type	Description
I1	DCE_Design_Field.do	Stata do-file	Generates the intermediate outputs (SurveyCTO sheets and ImageMagick commands)
I2	TunisiaTranslations.xlsx	Excel file	Lookup table with the Tunisian Arabic translations for the SurveyCTO notes, questions and choice labels
I3	TunisiaTexts.zip	28 .png files	English labels (196x151 pixels) for each level (achieved and not) of the 14 indicators
I4	TunisiaTexts.zip	28 .png files	Tunisian Arabic labels (196x151 pixels) for each level (achieved and not) of the 14 indicators
I5	TunisiaIcons.zip	28 .png files	Icon image files (132x132 pixels) for each level (achieved and not) of the 14 indicators

Process Steps

S1	Run 'DCE_Design_Field.do' (**I1**) in Stata to generate eight intermediate outputs:		

Item	Filename	Type	Description
S1a	DCESurvey_Field.csv	SurveyCTO	Preliminary survey sheet
S1b	DCEChoices_Field.csv	SurveyCTO	Preliminary choices sheet
S1c	DCEImages_English_trim.txt	ImageMagick	Commands to trim the borders of the English labels
S1d	DCEImages_Arabic_trim.txt	ImageMagick	Commands to trim the borders of the Arabic labels
S1e	DCEImages_English1.txt	ImageMagick	Commands to combine icon images with English labels
S1f	DCEImages_Arabic1.txt	ImageMagick	Commands to combine icon images with Arabic labels
S1g	DCEImages_English3.txt	ImageMagick	Commands to compile the three English-labelled icons representing the specified characteristics of each of the hypothetical women presented
S1h	DCEImages_Arabic3.txt	ImageMagick	Commands to compile the three Arabic-labelled icons representing the specified characteristics of each of the hypothetical women presented

S2	Add translation table sheet (**I2**) to 'DCESurvey_Field.csv' (**S1a**) in Excel and update the links on the survey sheet to pull in the Arabic translations of notes and question texts. Copy and paste values to fix translations, delete the translation table and save as an Excel spreadsheet (**O1**).
S3	Add translation table sheet (**I2**) to 'DCEChoices_Field.csv' (**S1b**) in Excel and update the links on the choices sheet to pull in the Arabic translations of choice texts. Copy and paste values to fix translations, delete the translation table and save as an Excel spreadsheet (**O2**).
S4	Run 'DCEImages_English_trim.txt' to trim the English label text image files.
S5	Run 'DCEImages_English_trim.txt' to trim the Arabic label text image files.
S6	Run 'DCEImages_English1.txt' to combine the English label texts with the relevant icon image files.
S7	Run 'DCEImages_Arabic1.txt' to combine the Arabic label texts with the relevant icon image files.
S8	Run 'DCEImages_English3.txt' to compile the three specified indicators for each of the hypothetical women presented in Part B and Part C of the DCE.
S9	Run 'DCEImages_Arabic3.txt' to compile the three specified indicators for each of the hypothetical women presented in Part B and Part C of the DCE.

Outputs

Item	Filename	Type	Description
O1	DCESurvey_Field.xlsx	SurveyCTO	Survey code for the DCE module
O2	DCEChoices_Field.xlsx	SurveyCTO	Choice code for the DCE module
O3	Various	416 .png files	Icons and indicator texts in both English and Arabic for each of the hypothetical women presented in the DCE.