Measuring Women's Economic Empowerment in Financial Sector Deepening and Financial Inclusion programmes

Guidance Note

November 2023

Gareth Davies &
Joanna Ledgerwood

TANDEM

Acknowledgements

This Guidance Note is based on a set of guidance materials originally produced by the authors for Financial Sector Deepening Kenya (FSDK) and Access to Finance Rwanda (AFR). The Guidance Note builds on the seminal work of Naila Kabeer as well as the excellent guidance published by the Abdul Latif Jameel Poverty Action Lab (J-PAL) and the International Center for Research on Women (ICRW), among other sources. The authors would also like to thank Sabine Garbarino for her invaluable comments.

1 Overview

For many programmes working in Financial Sector Deepening (FSD) or financial inclusion programmes, women and Women's Economic Empowerment (WEE) are increasingly at the centre of programme strategies and Theories of Change. This makes it imperative that programmes are able to report credible and robust results on WEE, and that monitoring and evaluation systems generate timely and useful data and insights that can be used by staff to continuously improve the WEE impacts achieved. Better gender data, including on WEE, can also be effective in highlighting gender gaps and driving programme and partner commitment to delivering results for women and girls.

The measurement of WEE is widely acknowledged to be complex and challenging. Conceptually, WEE can be defined in different ways, and what counts as 'empowerment' can vary across contexts, regions, and individuals. Robustly measuring subjective elements of empowerment can be tricky, with direct survey questions often subject to reporting bias. Features of the data collection process itself – such as the gender of the enumerator or the time and place of the interview or survey – can also introduce measurement bias if not considered carefully.[1]

Given these complexities, this Guidance Note presents a framework and methodology for measuring the WEE outcomes of FSD projects and interventions. The Note can be used for both monitoring and evaluation purposes. The Note also provides useful guidance for undertaking research exploring concepts relating to empowerment (or any research where collecting the perspectives and experiences of women as well as men is important).

The rest of the Note is organised as follows:

- Section 2 presents a simple framework for conceptualising WEE. Key concepts in empowerment such as 'access' and 'agency' are introduced and defined, with illustrative examples from FSD programming. This section provides the conceptual foundation for the remainder of the Note.
- Section 3 presents a five-step methodology for measuring WEE. For any given project or intervention, the first step is to map the empowerment pathway. In the second step, indicators are defined for each element of the empowerment pathway. The third step involves developing a measurement plan which sets out the data collection methods, sources, samples, timing, and responsibilities for each indicator. The fourth step covers the actual data collection process, including the design of data collection instruments, pre-testing, enumerator recruitment, and so on. In the final step, data is analysed and synthesised and used to both report progress and results to donors and other stakeholders and to inform programme decision-making.
- Section 4 provides a list of external resources. Examples of survey questions and tools for measuring WEE are also included in the Annex.

[1] For an excellent summary of the challenges in measuring WEE, see: Glennerster, R. et al. "A practical guide to measuring women's and girls' empowerment in impact evaluations". J-PAL.

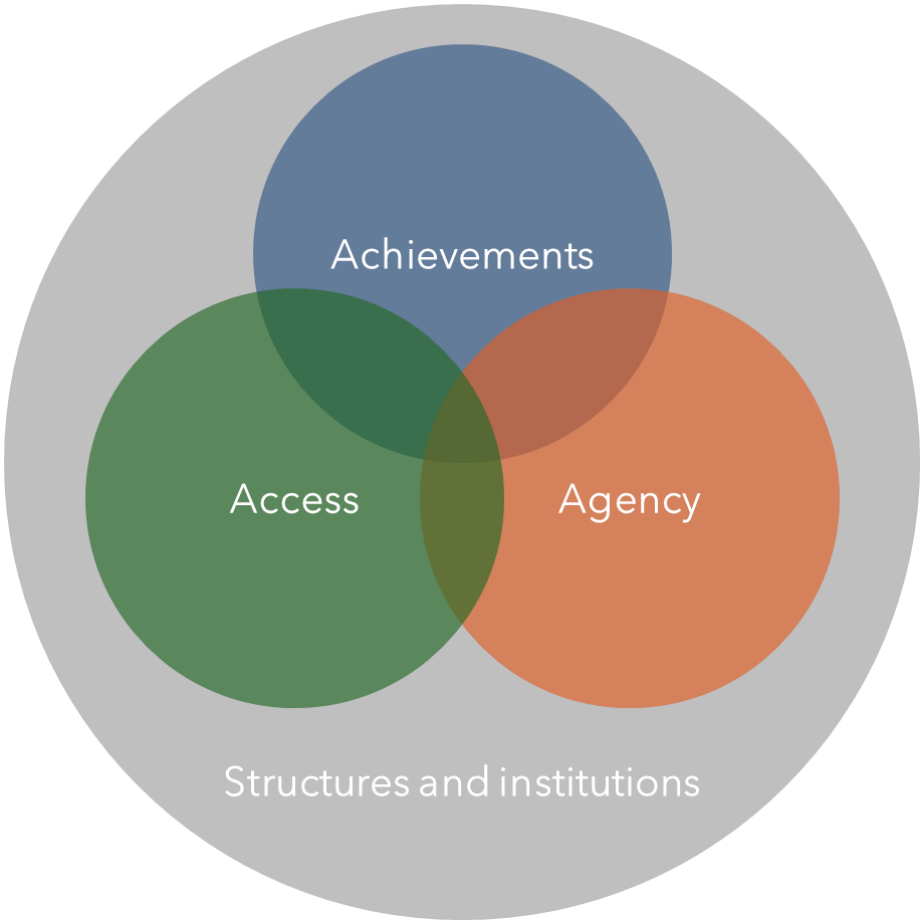
2 WEE Conceptual Framework

This section presents a simple framework for conceptualising WEE. Key concepts in empowerment such as 'access' and 'agency' are introduced and defined, with illustrative examples from FSD programming. This section builds on the extensive literature on WEE, including Kabeer's seminal 1999 article on empowerment[2] and more recent guidance from International Centre for Research on Women (ICRW).[3]

The 3A model of empowerment: Access, Agency, and Achievements

An economically empowered woman can be defined as a woman who is able to achieve the economic and non-economic goals she has set for herself (and her family). Rather than seeing empowerment as a binary outcome – 'empowered' or 'disempowered' – empowerment is best thought of as a continuum: the more a woman is able to set her own goals and achieve those goals, the more empowered she is. This definition of empowerment involves three interrelated concepts: *access*, *agency*, and *achievements* – see Figure 1.

Figure 1: the 3A model of empowerment



Each element of the model is discussed further below.

[2] Kabeer, N. (1999). "Resources, Agency, Achievements: Reflections on the Measurement of Women's Empowerment". Development and Change, 30 (3): 435-464.
[3] Golla, A. M. et al (2018). "Understanding & Measuring Women's Economic Empowerment: Definition, Framework & Indicators". ICRW.

To achieve their goals, women require *access* to certain resources. This includes access to goods, services, and assets, as well as more intangible resources such as access to information and networks, and human and social capital. In the context of FSD programming, examples might include:

- Access to formal and informal financial services;
- Access to technology and other enabling products and services, such as mobile phones;
- Access to sources of knowledge, skills, and information that help build an individual's numeracy, literacy, and financial capabilities;
- Access to assets, such as land, which can serve as collateral for loan products;
- Access to social networks and social capital, for example through savings groups.

To be empowered, women also need the *agency* to set and pursue their own goals and make their own choices. This includes active participation in decision-making processes and the power women have vis-à-vis other decision-makers (such as husbands and partners), as well as the ability and confidence women enjoy to set their own goals. One important factor in agency is the division of family unpaid care responsibilities, including housework, meal preparation, childcare, care of elderly relatives, and so on. Women with lower levels of agency often shoulder a disproportionate share of the unpaid care burden within a household, which in turn reduces their ability to set other goals and to pursue other options outside the home. Other important features of agency include:

- Mobility, both physically – such as being able to leave the home without the permission of a male family member – and economically – such as being free to enter occupations that may be seen as 'male occupations';
- Voice, self-confidence, and leadership;
- Financial independence and autonomy, such as having control over one's own savings;
- Freedom from violence, shame, or stigma that might otherwise restrict the real set of choices available.

Achievements considers the extent to which women are able to achieve the goals they have set for themselves and, as a result, enjoy meaningful improvements in well-being and life outcomes. In terms of economic empowerment, this could include goals such as starting a new business or growing an existing business, earning higher income, or securing employment in a desired occupation. Achievements can also include the non-economic benefits that result from these economic gains, such as improved health and education outcomes, as well as subjective benefits such as feelings of wellbeing, life satisfaction, and self-esteem.

Note that there is not necessarily a simple linear pathway from access to agency to achievements. Each of the three As can interact in different ways and there can be both positive and negative feedback loops between the three different concepts. For example, to access a financial service such as an e-wallet, a woman will first need a minimal level of *agency* such as the ability to decide to open an account, the ability to negotiate successfully with her husband (if she first needs her husband's permission), the mobility to leave the home and meet with the agent, and so on. She will also need *access* to other resources such as a mobile phone, information about the product, and the requisite levels of literacy and financial capabilities, which in turn may require additional starting levels of *agency*. If the woman is able to *access* the financial service, this may further boost her *agency*, for example by providing a safe place to save, thereby increasing her financial autonomy. It may also allow her to *achieve* other goals, for example helping her to start a business. This may contribute to further gains in

her *agency* if, for example, it further improves her self-confidence and negotiating power within the household.

There is also the potential for negative feedback loops or trade-offs. For example, *access* to a financial service that allows a woman to expand her business may increase her income (*achievements*) but also result in increased working hours which, without changes in the division of unpaid care (*agency*), will lead to increased time poverty and potentially lower life satisfaction overall. Similarly, gaining employment in a job that is informal, insecure, and with poor health and safety may not feel empowering and result in worse life outcomes on some metrics.

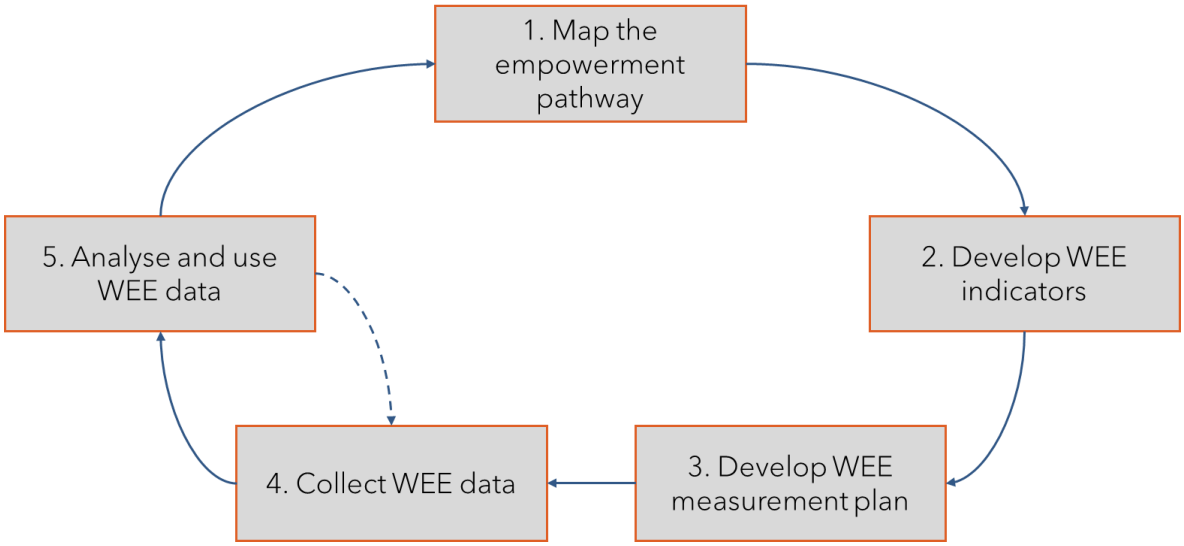
Underpinning the access, agency, and achievements of individual women are a broader set of structural factors and institutions. This includes cultural, social, political, and economic institutions, both formal and informal, that influence which resources women can access and the levels of agency they can exercise. For example, there may be a set of gendered social norms that favour men over women in the inheritance of land. This may be further reinforced and reflected in the laws around inheritance and land ownership, and in the cultural practices of community leaders, judges, and so on. This in turn shapes women's control over land and the ability to *access* collateralised loans, as well as their financial autonomy and decision-making power in the household (*agency*).

The strength and influence of these structural factors can vary across time and space and can interact with individual circumstances and characteristics in various ways. For example, women with higher levels of education and better access to legal services may be better able to challenge discriminatory informal norms around inheritance. In general, woman (and men) may seek to circumvent or directly challenge institutions that create barriers to the achievement of their goals, thereby contributing to societal change. However, depending on the context, this can carry the risk of backlash or social sanctions for the individuals involved.

3 Five-step Methodology for Measuring WEE

This section presents a five-step methodology for measuring WEE outcomes for a given FSD project or intervention, building on the conceptual framework presented in Section 2. The five steps are summarised in Figure 2 below.

Figure 2: five-step methodology for measuring WEE



Each of the steps are described in more detail below. References for additional external guidance can be found in Section 4.

3.1 Step 1: map the empowerment pathway

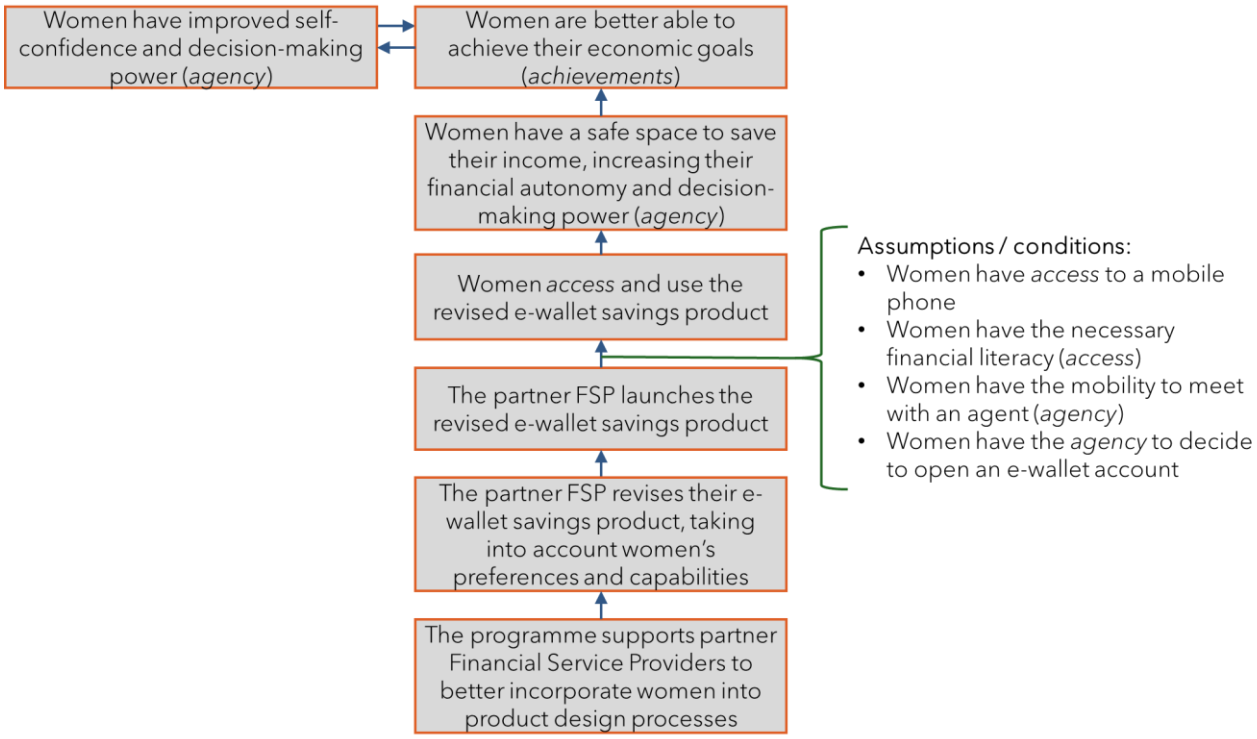
The first step is to map out the causal pathway illustrating how the intervention or project is expected to contribute to improved WEE outcomes. From a project design perspective, this step provides a systematic process for project teams to set-out the specific WEE results they hope to achieve as well as the causal mechanisms involved. It also provides an opportunity to systematically record the theories or hypotheses underpinning these causal mechanisms and the assumptions or conditions required for these causal mechanisms to hold, and to test these theories and assumptions against the available evidence and data. This is important for targeted interventions with the primary objective of empowering women, and for mainstreaming where implementers want to ensure results for women without WEE necessarily being the main objective. From a measurement perspective, this process also provides the project-specific framing for subsequent WEE data collection and analysis. Given that WEE is a broad and complex concept, this helps to focus data collection on those aspects of WEE that are most relevant to the given project or intervention.

The pathway should incorporate the three As of the WEE conceptual framework presented in Section 2. As noted in Section 2, WEE is not a linear process. The sequencing of changes in access and agency may therefore vary by project (although achievements will typically come after improvements in access and/or agency). The pathway may also include various feedback loops between the three As.

Note that the three As can appear as outcome boxes in the empowerment pathway and as assumptions underlying the causal logic (if they are preconditions needed for the causal logic to hold). To build on the example used in Section 2, the empowerment pathway for an

intervention around digital financial services might state that the launch of a revised e-wallet savings product is expected to increase women's *access* to financial services, which in turn increases her *agency* (because she now has a safe, private space to save thereby increasing her financial autonomy, and by allowing her to better spend her income in accordance with her own preferences), which in turn increases her *achievements* (because she can better achieve her economic goals, such as starting or growing a business). Underpinning this will be a set of assumptions about pre-existing levels of access and agency that need to be in-place if the pathway is to hold, such as access to a mobile phone, minimum levels of financial capability and decision-making power, and so on. This example is illustrated in Figure 3.

Figure 3: hypothetical example of an empowerment pathway for a digital financial services intervention[4]



The same thought-process can be used to identify project risks in terms of barriers to the delivery of positive WEE outcome and the potential for unintended negative WEE outcomes. For example, as noted in Section 2, the expansion of a business can lead to an increase in women's time poverty if not accompanied by a redistribution of unpaid care burdens. Documenting the assumptions underpinning the empowerment pathway also helps to identify potential barriers to economic empowerment (should the assumptions fail to hold in practice). For example, in a context where many women have low access to basic education, some women may not have the requisite financial capabilities to access an e-wallet. From a project design perspective, this can help project staff to think of ways in which the product can be modified in order to minimise these barriers, for example by working with the financial service provider to ensure that the product interface is as simple as possible or offers an interactive voice response platform to address barriers related to literacy or visual impairments, and that agents are trained to provide the necessary customer support. From a monitoring perspective,

[4] Note that in the interest of space, the assumptions or conditions have only been enumerated for one link in the empowerment pathway. In practice, assumptions or conditions should be developed for each link in the pathway.

this process helps to identify key assumptions to be tested and potential unintended negative outcomes that should be monitored for. It can also help to identify people and questions to investigate in the monitoring plan and data collection instruments, such as monitoring husbands' attitudes towards their wives accessing financial services like e-wallets (Step 4).

The empowerment pathway should be evidence-based, drawing on primary and secondary data and research. This can include evidence from previous programme interventions, interventions from other FSD programmes, and research and evidence from other countries that share similar contexts. From a strategy perspective, this can help to focus programme resources on projects and interventions that the evidence suggests will have the greatest likelihood of delivering strong, positive WEE outcomes.

Note that if a Results Chain exists for the project or intervention, the empowerment pathway can be incorporated into the Results Chain.[5] If no Results Chain exists, a separate empowerment pathway can be developed.

Learning Questions

Having constructed the empowerment pathway, project and monitoring staff have the option of developing a set of Learning Questions. These are key questions that staff and partners want answered through the monitoring and evaluation process. Learning Questions can be developed systematically for each link in the empowerment pathway, or might focus on specific linkages where the evidence base is currently weak. Following the hypothetical example developed above, Learning Questions for the e-wallet intervention might include:

- What are the key barriers for women accessing digital financial services like e-wallets, and how do these vary by characteristics – such as levels of education – and contexts – such as rural versus urban settings?
- To what extent does access to digital financial services like e-wallets contribute to improvements in women's agency and what are the causal mechanisms?

Learning Questions such as these can provide useful prompts when developing measurement plans (Step 3), Terms of Reference for evaluations or impact assessments, or when analysing and synthesising data (Step 5).

Participatory, self-defined empowerment outcomes

When mapping out the potential WEE outcomes for a given project or intervention, several WEE guidance notes recommend using participatory approaches whereby the women targeted by an intervention are themselves able to define what would count as a positive empowerment outcome.[6] This recognises the fact that what counts as 'empowerment' can vary across cultures, regions, and individuals, and avoids the imposition of an outsider's view of what women should want. It also helps to understand how women themselves would trade off different outcomes, such as extended working hours that increase income but may reduce leisure time and increase time poverty. One useful technique is the Participative Ranking Methodology (PRM). This involves focus group participants identifying and discussing a number of different possible responses to a question posed by the moderator, then ranking the responses in order of importance. By analysing responses across a number of groups, the

[5] For more guidance on constructing Results Chains, see the Donor Committee for Enterprise Development (DCED) Standard for Results Measurement (www.enterprise-development.org/implementing-the-dced-standard/).
[6] See, for example, Glennerster, R. et al., op. cit. and Anand, M. et al, (2014). "Practical Tools and Frameworks for Measuring Agency in Women's Economic Empowerment". The SEEP Network.

most frequently cited responses can be identified and used to develop locally-grounded WEE outcomes and indicators. However, as noted by Glennerster et al., "we also need to keep in mind that women's preferences may reflect society's views about gender rather than their own true preferences. Even though women's preferences are an important component of empowerment, measuring preferences alone may not always fully reflect women's ability to make a meaningful choice."

Project Example: participatory, self-defined empowerment outcomes

"The **Arab Women's Enterprise Fund (AWEF)** developed a tailored qualitative research process based on Focus Group Discussions (FGDs) conducted at the start of the project. It allowed for the collection of specific agency characteristics and pathways directly from female participants. To structure this process, an imaginary woman was introduced to the group. This woman was constructed as a 'typical case' with whom participants could easily identify. The facilitator told the story of the woman as she progressed through an intervention but stopped short of providing any outcomes. Participants were asked to continue the woman's story and describe the agency-oriented outcomes she would achieve. Once a sufficient number of agency characteristics were identified, participants voted individually on which are most important and likely to occur to the imaginary woman, resulting in a continuum of indicators that can be used for quantitative agency measurement. During this process, participants were also encouraged to share their own individual stories."

Reproduced from Anand, M. et al, (2014). "Practical Tools and Frameworks for Measuring Agency in Women's Economic Empowerment". The SEEP Network.

3.2 Step 2: develop WEE indicators

The next step is to develop a set of WEE indicators; at least one indicator should be developed for each box in the empowerment pathway. This will allow project teams to track progress and test the causal theories and assumptions underpinning the empowerment pathway. Should the WEE outcomes not materialise as expected, the team will then be able to pinpoint where the empowerment pathway is breaking down. By measuring *agency* as well as *achievements* it will also be possible to assess whether any positive achievements observed actually reflect women's own priorities and choices (the true measure of empowerment).

Indicators can be quantitative (numerical) or qualitative (descriptive); in general, it is good practice to combine quantitative and qualitative indicators. This is especially so in the case of WEE given the complexity and nuance involved in measuring empowerment outcomes and the subjectivity of many of the underlying concepts (such as self-confidence, wellbeing, and self-esteem). Qualitative indicators can also provide a rich source of data regarding the barriers facing women in accessing financial services and in achieving their goals, data which can be used by project teams and partners to adapt and improve project design and delivery (see Step 5).

Indicators can be defined to measure either the 'breadth' or the 'depth' of change. In the case of WEE, typically we want to measure change in both dimensions. For example, if a digital financial service project is expected to contribute to an increase in women's financial autonomy, an indicator of breadth might be: 'the number of women reporting an increase in savings that they control'. An indicator of depth might be: 'the average value of savings that women control', or 'qualitative changes in the extent to which women feel financially autonomous'. Combining breadth and depth indicators allows project staff to assess both how many women have been reached, and how substantive or consequential the outcome changes are for the women concerned.

As noted above, WEE is a complex term and it is not possible to measure every aspect of empowerment for every project or intervention; indicators should therefore focus on the specific aspects of access, agency, and achievements that the project or intervention is expected to bring about (as identified in Step 1). This also means that while a programme might use some common or standard WEE indicators (to aid comparability and aggregation across projects), these should be complemented by project-specific indicators tailored to the particular empowerment pathway and relevant to the project context.

Given the real potential for unintended WEE outcomes – such as increased time poverty – and the trade-offs many women have to negotiate – such as taking a job that provides increased income but that increases the risk of sexual harassment – the set of indicators defined needs to capture these different potential outcomes (to the extent that project M&E resources allow). For example, an indicator relating to the number of women gaining formal employment might be complemented by a qualitative indicator relating to job quality.

Whilst it is important to attempt to capture the subjective elements of WEE, these types of indicators can be difficult to measure and subject to reporting bias and should therefore be complemented with more 'objective' indicators. This is particularly the case with indicators relating to gender norms which can be subject to social desirability bias whereby respondents provide answers that they think the enumerator wants to hear or that are more in-line with what is perceived to be socially acceptable behaviour (rather than the answers that best reflect their reality). Some people may also be uncomfortable speaking freely about certain WEE-related topics. It is therefore good practice, where possible, to combine subjective indicators with objective or proxy indicators – see the box below for an example.

Project Example: combining 'subjective' and 'objective' indicators

In a randomised evaluation of commitment savings accounts in the Philippines, the evaluators identified women's influence in household spending decisions as their main outcome of interest. They collected data on two indicators:

1) An index of women and men's responses about who decides in nine common household spending decisions, and
2) Household expenditures on what respondents identified as typically 'male' or 'female' goods.

This allowed the evaluators to triangulate the more subjective indicator on household decision-making with a more objective measure of the actual outcome of the decision-making process in terms of actual household expenditures. However, this approach only works if women and men report different spending preferences, and individual preferences can be subject to change over time.

See Ashraf, N. (2010): "Female Empowerment: Impact of a Commitment Savings Product in the Philippines". World Development 38 (3): 333-344. https://doi.org/10.1016/j.orlddev.2009.05.010

All WEE indicators should be SMART:

- Specific: indicators are clearly defined and unambiguous, reducing the scope for subjective interpretation.
- Measurable: indicators can be measured or assessed (either by numbers or a description); indicators are within the means of the programme to measure. Proxies can be used when the underlying change is too complex or costly to measure given programme resources.
- Attributable (or 'contribution-sensitive'): indicators are tightly defined around the change the project is trying to catalyse; indicators are not so broad that there are many

external factors that could cause changes in the indicator such that the contribution of the project to changes in the indicator cannot be plausibly established.
- Relevant: indicators are appropriate and applicable to the box in the empowerment pathway being measured; indicators will provide useful information for management decision-making.
- Time-bound: when relating to a flow (e.g. number of women accessing a new or revised financial service) rather than a stock, indicators are defined over a specific timeframe (e.g. number of women accessing a new or revised financial service in the last 12 months).

By the end of Step 2, the set of indicators defined should cover each of the boxes in the empowerment pathway and provide a means for answering the key Learning Questions identified in Step 1. Indicators can also be developed to monitor and test any key assumptions or potential unintended negative outcomes.

Sex-disaggregation of indicators

In some cases a programme will already have defined indicators that are relevant to the empowerment pathway developed in Step 1. For example, at the access level, indicators may already exist for the number of individuals accessing and using new or revised financial services, or the number of individuals reporting increased income. Also, often programmes will want to measure certain outcomes for both women and men. In these cases, indicators should be sex-disaggregated.

Whilst this is straightforward for indicators relating to individuals (such as 'the number of people accessing a financial service in the last 12 months'), sex-disaggregation of indicators relating to households, enterprises, or other units or organisations is more complicated. Afterall, a unit such as an enterprise does not have a sex, only people do. One seemingly obvious solution is to look at the sex of the enterprise owner and disaggregate all enterprise-related indicators by Women-Owned Enterprises. However, beyond micro enterprises and sole-traders, an enterprise may have multiple owners, potentially of different sexes. Even in a small family enterprise primarily operated by a woman, her husband may exercise control over key business decisions. Conversely, an enterprise formally 'owned' by the husband may in fact be primarily run and controlled by the wife. Similar issues arise for household-related indicators, where sex disaggregation is usually based on who 'heads' the household. However, households designated as 'male-headed' can hide a full spectrum of decision-making power enjoyed by female household members. For this reason, when it comes to measuring WEE, it is best to define indicators relating to individuals rather than units. This means the black box of 'the household' or 'the enterprise' can be unpacked, giving a truer picture of the agency and achievements of women.

Similar issues arise when trying to define indicators relating to income. Measuring income at a household level, while useful from a wider impact perspective, does not provide much insight regarding WEE outcomes as it glosses over intra-household decision-making. Instead, income could be measured at the individual level. However, care is needed when defining the indicator as *receipt* of any additional income may not translate into *control* of that income, or into realising the *benefits* of the additional income. Project staff therefore need to carefully consider which concept of income they are trying to capture and also consider adding indicators relating to decision-making and/or household expenditure (see box above).

Beyond sex-disaggregation, indicators can be disaggregated by other characteristics such as age, marital status, education level, location (rural vs. urban), disability and so on. This builds on the idea of intersectionality: 'women' are not a homogenous block, and what works for

women with certain characteristics might not work as well for women with completely different characteristics. Which disaggregations will be most useful can be informed by the project assumptions identified in Step 1. For example, if a key condition or assumption is that women possess a minimal level of literacy and numeracy to access a new digital financial service, disaggregating an indicator by education level will be a useful means of testing this assumption and for identifying the threshold level of education women need. If the project includes objectives related to disability inclusion, specialised tools such as those by the Washington Group on Disability Statistics may be useful. These additional levels of disaggregation allow more nuanced insights to be developed during the data analysis stage (see Step 5).

Example WEE indicators

The table below provides some example WEE indicators for each of the three As in the conceptual framework.[7]

Table 1: example WEE indicators

Elements of WEE	Example indicator
Access	Number of women accessing and using a new or revised financial service
	Number of women with access to a mobile phone
	Average financial literacy scores of women (based on a standardised financial literacy test)
Agency	Number of women reporting improvements in household decision-making power
	Qualitative improvements in women's reported voice, self-confidence, and leadership
	Number of households where household expenditure more closely matches women's preferences
	Average number of hours spent by women per day on household chores and unpaid care responsibilities
Achievements	Number of women reporting improvements in their ability to achieve their economic goals
	Qualitative changes in the extent to which women report being able to achieve their goals
	Number of women reporting improvements in their feeling of wellbeing and life satisfaction (using standardised psychometric scales[8])

[7] Note that indicators are not necessarily the same as survey questions. For example, for the indicator 'Number of women reporting improvements in household decision-making power', directly asking women in a single question whether they have seen improvements in their household decision-making power is unlikely to yield meaningful results. Instead, the indicator could, for example, be calculated based on the responses to a set of more specific, relatable questions regarding different types of concrete decisions. See Section 3.4 for more guidance on designing WEE survey questions.

[8] See Section 3.4 for examples.

3.3 Step 3: develop a WEE measurement plan

Having developed WEE indicators for each box in the empowerment pathway, the next step is to develop a measurement plan. For each indicator, the measurement plan summarises:

- Any indicator *definitions / calculations*, including definitions for key terms used in the indicators (e.g. 'usage' defined as using the financial service at least once in the previous quarter) and the calculations used to populate the indicator (e.g. 'increase' calculated as a percentage increase from the baseline);
- The *data collection method(s)* to be used to populate the indicator (e.g. observation, interviews, Focus Group Discussions, surveys, etc.);
- The *sources / samples* used to populate the indicator, including the sampling frame and sampling method;
- Whether *baseline* data collection is required and the timing of any baseline data collection;
- The *frequency / timing* of post-baseline data collection;
- The *responsibility* for collecting and verifying the data (or for overseeing data collection by a third party).

The basic format of a measurement plan is reproduced in Table 2 below.

Table 2: project or intervention measurement plan

Empowerment pathway box	Indicator	Definitions / calculations	Data collection method	Sources / samples	Baseline data collection	Frequency / timing	Responsibility
Box 1…	Indicator 1.1…						
	Indicator 1.2…						
Box 2…							

Underpinning the detailed measurement plan should also be an attribution or contribution methodology, detailing how the programme will assess the attribution or contribution of programme activities to the observed changes in WEE outcome (versus other external factors).[9]

Each of these elements of the measurement plan are considered below, focusing on what is unique about measuring WEE.[10]

Data collection methods

For each WEE indicator, the Measurement Plan specifies which data collection method will be used to populate the indicator. A variety of measurement tools are available including: observation, interviews, Focus Group Discussions (FGDs), surveys, and so on. Some tools are more suited to collect either qualitative or quantitative data. For example, FGDs are usually used to collect qualitative data whereas surveys are primarily used to collect quantitative data (although some qualitative data can also be collected through open-ended survey questions, and scoring within FGDs can yield quantitative data).

[9] For more on attribution and contribution methodologies, see Sen, N. (2021). "Measuring Attributable Change: Implementation Guidelines for the DCED Standard". DCED.
[10] For more general guidance on measurement planning, see the DCED Standard for Results Measurement (www.enterprise-development.org/measuring-results-the-dced-standard/).

The choice of data collection method depends on the type of changes being measured, the robustness required, the type of data needed (quantitative or qualitative), and the budget available. Subject to project M&E resources, triangulating methods will provide a more robust and insightful set of data than just relying on one or two methods. This is especially so in the case of WEE given that common data collection methods like surveys and interviews can be subject to reporting and measurement bias. If possible, combining these tools with techniques such as direct observation, games, or experimental vignettes can provide more robust findings (see below). These other tools can also be helpful for collecting data on elements of WEE that are hard to measure, such as attitudes towards gender norms that respondents might not report honestly or that they are not even aware of (e.g. subconscious gender bias).

Including qualitative techniques alongside quantitative methods is also important given the complex and nuanced nature of the empowerment process and the real risk of causing unintended negative WEE outcomes which need to be monitored for (such as increased time poverty or sexual harassment). Qualitative data is also useful for identifying WEE barriers and informing adaptations to project design (see Step 5).

Some of the main data collection methods for measuring WEE are discussed further below.

Direct observation. This involves directly observing and recording behaviours and actions of individuals and organisations in the field. For example, project staff might directly observe savings group meetings and record various quantitative and qualitative observations such as how well-run the meetings are, how many people were in attendance, and whether and how many of the steps in the group's constitution were followed. Direct observation has the advantage of providing more objective measures of phenomenon of interest. For example, the number of times women speak in community meetings, and how long they speak, may provide a more objective measure of 'voice' and 'self-confidence' (to complement more subjective self-reported data from surveys and interviews). However, it may not be possible or may be too costly to observe many of the decision-making processes and other events that a programme may be interested in.

Secondary data. This involves obtaining data already collected by market actors, stakeholders, and other development programmes. Client data collected by partner financial service providers is one important example. The great advantage of secondary data is that, as it is already being collected by third parties, it costs very little for the programme to obtain such data. It can also be collected at a scale that would be prohibitively expensive for a donor programme to duplicate. However, a big disadvantage is that, because the data is primarily being collected for someone else's needs, it may not be as robust as we would like, and may not cover all of the things we are interested in. This is particularly so in the case of WEE: beyond *access*, few financial service providers will have an interest in, or the capabilities to, collect data on women's *agency* and *achievements*. Even at the access level, many financial service providers do not routinely collect accurate sex-disaggregated client data – see box.

Challenges in collecting sex-disaggregated data from financial service providers

Financial Sector Deepening and financial inclusion programmes often expect and rely on partner financial service providers to collect and share sex-disaggregated client data. However, across the world many financial service providers do not routinely collect sex-disaggregated data and, if they do, it can often be inaccurate.

For example, a study by the GBA and McKinsey found only 55% of banks interviewed reported having the capability to disaggregate customer data by sex.[11] Similarly, at the 2015 AFI Global Policy Forum, half of respondents stated that the capacity to disaggregate by sex was a big challenge. This included restricted functionality of basic 'off the shelf' Management Information System (MIS) software. They also cited the challenge of assigning a 'sex' to an enterprise – such as what counts as a Women-Owned Enterprise (discussed in Step 2) – and how to treat joint accounts. Participants also reported problems with the quality of data: staff were not always properly trained to understand the field categorisations in the MIS. Cultural and social barriers can also affect the quality of data. For example, HBL in Pakistan performed a random sampling exercise of its female deposit portfolio in 2014. It found that the actual female control of accounts tagged as 'woman-owned' in rural areas was only around 50%.[12]

As a first step, programmes should include a requirement for financial service providers to collect and report on sex-disaggregated client data (at least as it relates to the programme support provided) in MoUs and partnership agreements. However, recognising that some financial service providers may not have the systems and capabilities to do so, additional financial and/or technical support may be required to enable financial service providers to upgrade their data systems. Longer term, donor programmes can also work with regulators, policy-makers, and other stakeholders to encourage or mandate the collection of sex-disaggregated data across the financial sector.[13]

Where secondary data is available, consideration needs to be given to the reliability and relevance of the data. This might include an assessment of the methodology used in the collection of national statistics, or an assessment of the internal quality assurance processes partners use to verify their internal data.

Semi-structured interviews. Semi-structured interviews are loosely structured interviews, shaped around a pre-defined set of open-ended questions. They resemble a conversation, allowing a free flow of ideas and information. Interviewers frame follow-up questions spontaneously, probe for information, and take notes (which are elaborated on and organised after the interview). Because they are loosely structured, they are ideal for identifying unintended consequences of an intervention. They are also ideally suited to collecting nuanced data and insights into concepts around *agency* and the more subjective elements of *achievements*. Because they are conducted one-on-one (versus FGDs which are conducted in a group setting), they are also a good tool for collecting data on more sensitive aspects of WEE. However, the findings from semi-structured interviews are not always easily generalisable, and the small sample sizes means the findings may not be as robust as surveys. Subject to the available M&E resources, it is generally preferable therefore to combine interviews with surveys.

Focus Group Discussions (FGDs). FGDs are similar to in-depth interviews except that they are conducted with a group rather than an individual. They are primarily used to collect qualitative

[11] "How Banks Can Profit from the Multi-Trillion Dollar Female Economy". Global Banking Alliance for Women (2014).
[12] "The 2015 Afi Global Policy Forum Report". Alliance for Financial Inclusion (2015).
[13] For more ideas on how organisations can help close the gender data gap, see: "Transforming the Data Landscape: Solutions to Close Gender Data Gaps". Data 2X (2022).

data. They are useful when it is anticipated that a group dynamic will be useful to elicit fuller and more complete answers from respondents. Participants are guided by a moderator or facilitator who introduces topics for discussion and helps the group to participate in a lively and natural discussion amongst themselves. The strength of FGDs relies on allowing the participants to agree or disagree with each other so as to provide an insight into how a group thinks about an issue, and to highlight any inconsistencies and variation that exists in a particular group in terms of beliefs, experiences, and practices. Compared to interviews, however, FGDs may not be suitable for exploring more sensitive WEE topics. There is also a risk that one or two higher-status (more empowered) individuals dominate the discussion, leading to biased data.

Surveys. Surveys are used to gather (predominantly) quantitative information from a large number of respondents (although small amounts of qualitative information can also be gathered using open-ended questions). The great advantage of surveys is that they allow the collection of statistically robust data (assuming the questionnaire is well designed and biases are not introduced in the administration of the survey – see Step 4). Because a standardised set of questions is used, a larger number of respondents can be reached than through more time-intensive forms of data collection like semi-structured interviews. On the flip side, however, surveys can be quite reductive in the data collected, missing some of the important nuances and insights provided by interviews and FGDs. As noted above, they can also be quite limited in measuring certain elements of WEE, such as attitudes to gender norms.

Experimental vignettes. Vignettes are short descriptions of hypothetical situations. They are commonly used in questionnaires and in interview guides as a way of asking questions about concepts that may not be easy to grasp if asked directly. *Experimental* vignettes involve changing a key detail in the story, with each version assigned randomly to respondents to test whether the change makes any different to the responses given. In the case of WEE, changing the sex of the protagonist in the vignette, for example, would provide an insight into hard-to-measure things like subconscious gender bias. For example, to assess the effectiveness of gender training for bank loan staff, two versions of a loan application could be constructed, identical but for the sex of the fictitious applicant. Randomly assigning these two versions of the loan application to a sample of bank loan staff, before and after the training, could provide insights into whether the training had been effective in reducing gender bias.

Project Example: using experimental vignettes to examine unconscious gender bias

One evaluation of a local governance project in India looked at whether exposure to female leaders in Indian village councils changed perceptions about women's effectiveness as leaders. As part of the survey, researchers played a short recording of a speech by a local leader responding to a complaint from a villager. Respondents were randomly assigned to hear the recording spoken by a man or woman. After the speech was over, they were asked to rate the leader's effectiveness. The vignette allowed researchers to measure whether there was a subconscious bias that led people to rate female leaders as relatively less effective. The study found that exposure to a female leader through the policy that reserved village council head positions for women reduced men's bias against female leaders.

See Beaman, L. et al (2009): "Powerful Women: Does Exposure Reduce Bias?." The Quarterly Journal of Economics 124, no. 4: 1497-1540.

Games. This involves constructing an artificial scenario in which an individual or group of individuals are asked to make choices or decisions. They are designed to reveal something about how participants might act in the real world. For example, the case in the box below shows how a game involving small amounts of real money was used to measure the bargaining power of women vis-à-vis their husbands. Games are particularly useful for complementing

survey or interview questions that may be subject to reporting bias or that cover issues that are difficult for respondents to conceptualise or provide answers to, such as household power dynamics. However, games need to be carefully designed to avoid the bias introduced by the fact that people are being observed while playing the game, and how people act in the game may not reflect how they actually behave in real life situations or where the stakes are higher.

Project Example: using games to examine differences in spending preferences

In a randomised evaluation in Kenya, Simone Schaner examined how differences in spending preferences between husbands and wives affected demand for savings accounts. The researchers used a game to measure intra-household bargaining power to test whether the savings account programmes had different impacts on women with more or less bargaining power. At the end of the survey, husbands and wives, who were being surveyed separately, were asked to divide a small cash prize between themselves and their spouse. Each spouse recorded his or her allocation separately on cards and placed the amount they allocated to themselves in their tin and the amount allocated to the spouse in the spouse's tin. Then they came together to decide how to allocate the prize money between them and recorded it on cards, which they added to each of their tins. To ensure respondents' privacy, the husband and wife also each added an additional envelope to each of their tins with a randomly selected amount of money. The husband and wife then chose one card from each of their tins and were immediately given the cash amount allocated to them on the card they chose.

This game allowed researchers to identify women with relatively low- or high-bargaining power and test whether the impact of the intervention was different for women with different levels of bargaining power.

See Schaner, S. (2017): "The Cost of Convenience? Transaction Costs, Bargaining Power, and Savings Account Use in Kenya". Journal of Human Resources 52 (4): 919-945. https://doi.org/10.3368/jhr.52.4.0815-7350R1

Sources and sampling

Having determined the data collection methods, the next step is to sketch out the data sources and sampling approach. For methods such as interviews and FGDs, this means planning how many interviews or FGDs to conduct and how participants will be selected. For surveys, thought needs to be given to the sample size and sampling methodology. In each case, this involves thinking through who possesses the information we need, and whose voices and perceptions we need to capture. Note that from a WEE perspective this does not mean only interviewing or surveying women. For example, to measure changes in intra-household decision-making it is often useful to interview or survey both women and men in the household. This will allow answers to be triangulated and for differences in perspectives and attitudes to be identified.

There is no magic number for how many interviews or FGDs to conduct, or how many surveys to administer. For interviews and FGDs it is often not until the interviews and FGDs are being conducted that it is possible to determine exactly how many to do. If the time and resources allow, it is good practice to repeat them until they no longer provide new insights. For surveys, the exact sample size to be used depends on the level of robustness required, which in turn depends upon the size of the population and the amount of variability in responses (measured as the relative standard error). A larger sample size increases the robustness of the survey, but costs more to administer.[14] If questions relating to WEE outcomes are being included as part

[14] Note that increases in the size of the population above 10,000 do not materially affect the size of the sample required to achieve a given level of robustness. Robustness is measured in terms of the confidence interval and the confidence level. As a rough guide, with a relative standard error of 0.5

of a larger mixed-sex survey, it is important to ensure that women are adequately represented in the sample (for example by using random stratified sampling whereby the number of women in the sample is proportional to the number of women in the overall population)[15]. Depending on the attribution or contribution methodology, the sample may also need to include a treatment and control group.

Timing and frequency of data collection

For each indicator, the Measurement Plan specifies when and how often the indicator will be collected. In general, to calculate the change in an indicator, two observations are required: a baseline, which records the value of the indicator before the effects of the intervention are felt, and a second observation taken after the effects of the intervention are first felt. Often further observations are taken to record trends over time and to monitor sustainability. Note that when conducting a baseline assessment, 'before the effects of the intervention are felt' does not necessarily mean 'before the start of an intervention'. For example, it may take many months of working with a partner before they launch a new or revised financial service, in which case it is not necessary to conduct a baseline WEE assessment right at the start of the intervention. In fact, it may be impossible to do so if it is not possible to identify the treatment group until they have actually accessed the service. In some instances it might be possible to collect before and after data in one data collection exercise by asking respondents to recall the baseline situation. However, care should be taken to ensure the recall periods are realistic for respondents (see also Step 4).

When thinking about the timing of WEE data collection, thought needs to be given to the timeframes over which change is realistically expected to happen. In terms of the 3A model, access level changes will typically happen the quickest. Changes to agency and achievements can take longer to manifest, depending on the nature of the intervention and the empowerment pathway (see Step 1). These timeframes need to be factored into the Measurement Plan: if agency and achievement level changes are measured too soon, and without further rounds of data collection, important WEE outcomes might be missed. In some cases it might be necessary to continue to monitor changes a year or more after the end of the project or intervention.

Although the full WEE outcomes of a project may unfold over a long time period, projects should look for early signs of positive or negative change, to inform on-going project adaptation and decision-making. If WEE outcomes are only measured at the end of a project, or after a project has closed, this limits the ability of project teams to adapt their strategies and tactics in response to what is and is not working on-the-ground. The measurement plan should therefore include data collection exercises that will allow teams to test early on key assumptions underlying the empowerment pathway and to check for any potential negative outcomes. For example, FGDs and small surveys with female clients of a new or revised financial service, carried out within a few months of the product launch, may reveal useful insights into women's experiences of accessing and using the new product (including any barriers encountered), what they intend to use the product for, and whether they anticipate the product to contribute to any changes to intra-household dynamics. These insights can then be used by the project team to co-develop modifications to the product design or distribution strategy before the product is rolled-out further.

---

and population of 10,000, a sample size of 96 will give a +/-10% confidence interval with a confidence level of 95%.

[15] Depending on the indicator disaggregation, it might also be necessary to identify additional strata such as rural versus urban populations.

3.4 Step 4: collect WEE data

Having developed the Measurement Plan, the next step is to design the individual data collection instruments (such as survey tools and questionnaires, or interview guides) and collect the data. From a WEE perspective, getting this step right is critical given the potential for reporting or measurement bias when collecting WEE data.

Designing data collection instruments

For surveys, semi-structured interviews, and FGDs, a well-designed questionnaire or interview / FGD guide is important in minimising biases and measurement error. As discussed in Step 3, questions relating to WEE are prone to bias and error. For example, people may be reluctant to openly discuss or answer questions relating to sensitive gender topics; answers can also be subject to social desirability bias. Poorly designed surveys can also lead to the minimisation of women's roles and voices, for example by forcing a binary choice between 'male-headed' and 'female-headed' households or enterprises. Involving WEE experts from within or outside the programme when developing data collection instruments is recommended.

Glennerster et al. identify five good-practice principles for constructing survey (and interview) questions that minimise potential bias and error:

- **Specific**. Each question should only ask one thing at a time.
- **Neutral**. The wording of the question should not bias respondents to give a particular answer one way or another.
- **Understandable**. All survey questions should be relatively easy to comprehend by anyone in the sample. This means avoiding unfamiliar terms or concepts that are not clearly defined in the survey prompt. In the context of WEE, questions should typically avoid abstract concepts like 'empowerment' or 'agency'. For example, instead of asking 'are you empowered to make decisions around accessing financial services?', a better question might be: 'who in your household decides whether to open a new bank account or other financial service?'.
- **Clearly framed**. Good questions have clear boundaries and well-defined timeframes. If the responses are multiple choice, the list of possible answers should be mutually exclusive and collectively exhaustive. Any question asking a respondent to recall something from the past should include a clear timeframe and this recall period should be short enough that the respondent can remember accurately.
- **Relevant**. It is important to be respectful of study participants' time, so all the survey, interview, or FGD questions should measure indicators that will actually be used by the programme. Questions should also be relevant to the local context.[16]

Various techniques can be used to put the respondent at ease and increase their willingness to answer potentially sensitive questions. A good way to start a survey or interview is with simple questions such as basic demographic information. This is useful for disaggregating the analysis (see Step 5); it also helps the enumerator to build rapport with the respondent before proceeding to more sensitive topics. Another technique is to frame questions indirectly through the use of vignettes or hypothetical scenarios. For example, instead of asking a direct question about the respondent's own life, the question could be framed by asking: 'for someone similar to you in your community…'. Vignettes can also be used to ask for people's opinions about a fictional situation, the answers to which reveal something about their real-life opinions and viewpoints. To elicit insightful data, it is important that these vignettes are grounded in the local context and therefore realistic and relatable for respondents (see the

[16] Adapted from Glennerster et al., op. cit.

Annex for examples). Similarly, when asking about things like decision-making processes, experience suggests that asking specific questions about specific decision-making processes provide more reliable data than more general questions. For example, rather than asking a single general question such as 'who in your household makes decision about finances?', consider developing a series of more specific questions such as: 'who in your household decides whether to open a savings account (if you have one)?', 'who in your household decides how much money to save each month?', 'if you have taken a loan in the last 6 months, who in your household decided to take the loan', etc. As noted by Glennerster and Walsh (2017), asking more specific questions about a concrete scenario tailored to the choices women care most about in the study context may be easier for respondents to answer accurately, and may reveal more about whether they can make meaningful choices that matter to them.[17]

Project Example: measuring psychological components of empowerment

The *agency* and *achievements* element of the 3A empowerment framework includes various subjective or psychological aspects such as feelings of wellbeing, self-esteem, and self-confidence. There are various existing survey tools and questionnaires from psychology that can be used to measure these aspects of WEE.

For example, the CARE evaluation of a financial inclusion and empowerment project in Uganda used two different psychometric scales as a measure of empowerment (and disempowerment). The Pearlin Mastery Scale asks respondents to rate responses from 1 (strongly disagree) to 7 (strongly agree) to questions such as:

- 'There's no way I can solve the problems I have'
- 'Sometimes I feel I am being pushed here and there in life'
- 'I have little control over things that happen to me'
- 'I can do anything when I put my mind to it'

Similarly, the Kessler Phycological Distress Scale (K10) asks for responses rated on a scale from 'none of the time' to 'all of the time' to questions such as:

- 'During the last 30 days, about how often did you feel tired out for no good reason?'
- 'During the last 30 days, about how often did you feel hopeless?'
- 'During the last 30 days, about how often did you feel depressed?'

The two scales were used in before and after surveys of three different groups: a treatment group that accessed a Village Savings and Loan Association (VSLA) plus mobile wallet for women; a second treatment group that accessed the VSLA, mobile wallet, plus a course of seven household counselling sessions; and a control group. The evaluation found that for married women, there was a significant positive improvement against both psychometric scales for the second treatment group but not the first treatment group (vis-à-vis the control group). The CARE study complemented these psychometric scales with more 'objective' proxy indicators such as the redirection of male income from alcohol toward family goals.

Another example of a psychometric scale used in WEE project evaluations is the Relative Autonomy Index, which provides a measure of the extent to which people feel they have the capacity to achieve their own goals, or are at the fate of external factors like social norms or coercion.

Given the standardised nature of these scales and the questions they are based on, a risk is that they are not well suited to the project's local context or easily understandable or answerable for the survey population. It is therefore important to pre-test the questions before they are deployed in the full survey (see below).

See Scott, L. (2020). "Digital Subwallets and Household Dialogues: Final Report". CARE.

[17] See: Glennerster, R., and Walsh, C. (2017). "Is It Time to Re-Think How We Measure Women's Household Decision-Making Power?" J-PAL.

Pre-testing data collection instruments

Before embarking on a large-scale survey exercise it is important to pre-test the data collection instruments in locations similar to the ones where the full survey will take place. This is especially important for WEE data collection given the often complex and abstract nature of the concepts we are ultimately interested in, the sensitivity of some topics, and the potential for reporting bias. Pre-testing is useful for checking things such as:

- **Whether respondents understand the questions and find them easy to answer.** For example, some terms may not translate well into local languages or be confusing for respondents.
- **Whether the survey is too long.** Respondents can become fatigued during long surveys and their answers may become less accurate towards the end of the survey.
- **Whether recall periods are realistic.** People may struggle to accurately answer questions about decisions or behaviours that occurred too long ago.
- **Whether the questions are sensitive enough to capture variation.** For example, if asking questions to assess a respondent's level of financial capabilities, the questions should not be too easy such that every respondent gets them right, or too hard such that every respondent gets them wrong.
- **Whether the timing and location of the interviews is right.** For example, while it may be more convenient for the enumerator to interview female workers at their place of work, respondents may be too busy to answer the questions. Thought also needs to be given to women's unpaid care burdens. For example, female retailers may tend to open their shops latter then men due to childcare responsibilities – this needs to be taken into account if planning to interview women at their shops (see also below).

To gather respondents' feedback on the survey instrument itself, two techniques can be used:

1. **Respondent debriefings**, whereby enumerators implement the survey and then collect feedback and comments from respondents at the end of the survey;
2. **Cognitive interviews**, whereby respondents describe their reactions and thought processes in addition to answering the survey question.

Based on the findings of the pre-test, revisions may be required to the data collection instrument and/or the planned data collection process.

Data collection processes

During WEE data collection, careful consideration needs to be given to enumerator training and recruitment and the where and when of data collection. All of these considerations have a gender aspect which have the potential to bias results or introduce measurement error.

Enumerator recruitment. When collecting WEE data from respondents it is good practice to match the gender of the enumerator with the gender of the respondent. This is especially true of more sensitive aspects of WEE and in contexts where gender norms restrict the socially acceptable interactions between women and men. At the same time, male researchers are often required to engage with (largely male) community leaders when approval for community-based research is sought. To build rapport and put the respondent at ease, it can also be helpful if the enumerator is from the same region and ethnicity as the respondent, and should also be able to speak any local languages. Because the identity of the enumerator can potentially affect the answers given by respondents, ideally the same enumerators should be used for the treatment and control group, and in baseline and repeat surveys.

Where to conduct interviews. When collecting potentially sensitive WEE data, whether the respondent is alone or within earshot of other family members or members of the community can affect the truthfulness of their answers. Ideally interviews should be conducted alone. In some contexts, this may require the consent of a family member such as a husband (or parents if interviewing young women). A consent script, explaining the purpose of the research and guaranteeing the anonymity of the respondent, can help to alleviate any concerns. Interviews can also be conducted within sight, but out of earshot, of the husband or parents. If it was not possible to perform the interview in private, this should be recorded in the survey. Even if the respondent is alone they may be reluctant to talk openly about certain issues. One solution is to allow respondents to directly input their answers into a tablet or smart-phone. For respondents with low levels of literacy, technology such as Audio Computer-Assisted Self-Interview platforms can be used.

If respondents are required to travel, as can be the case for FGDs, there is also a need to consider if this is safe for women and how to avoid introducing any sampling biases (for example, by excluding women who are less mobile or have greater unpaid care burdens). Other considerations include the provision of child-care for smaller children and whether the participation of older children is appropriate.

When to conduct interviews. As noted above, the time of day when interviews are conducted can potentially bias the survey results. For example, female retailers with childcare responsibilities may open their shops later in the morning. Conducting interviews at a marketplace early in the morning may therefore result in a sample that is bias towards women without children or who do not have the same unpaid care burdens. The time of year can also potentially affect survey responses. For example, women in rural households may have different workloads at different times of year, depending on the agricultural cycle. When conducting panel surveys, it is therefore important to consider whether each survey round should be conducted during the same month each year.

Ethics and safeguarding. It is always important to maintain high standards of ethics when conducting primary research. The following standards should be observed:

- Voluntary participation. Participation should be voluntary and free from external pressure. Information should not be withheld from prospective participants that might affect their willingness to provide data. All participants should have a right to withdraw and withdraw any data already provided at any point without fear of penalty.
- Informed consent. Data enumerators should provide a clear statement of intent to inform participants how information and data obtained will be used, processed, shared, and disposed of, prior to obtaining consent.
- Child protection. Informed consent should be obtained from parents or caregivers and from children themselves. Children should not be required to participate even if their parents provide consent.
- Treatment of participants. Data enumerators must be aware of differences in culture, local customs, religious beliefs and practices, personal interaction and gender roles, disability, age and ethnicity, and should be mindful of the potential implications of these differences when collecting data.
- Confidentiality. Data enumerators must respect people's right to provide information in confidence, and must ensure that sensitive information cannot be traced to its source. They should also inform participants about the scope and limits of confidentiality.
- Data security. Data collectors should guard confidential material and personal information by the proper use of passwords and other security measures. There is a duty to state how data will be stored, shared, archived and (if necessary) disposed.

Given the sometimes sensitive nature of WEE data collection, it is even more imperative that high ethical standards are maintained. In some cases this may require the development of additional ethical and safeguarding guidelines. This is especially true if collecting data on Gender Based Violence. This includes carefully wording questions to avoid unnecessary distress, checking the national laws for the mandatory reporting of violence, and considering what safeguarding and referral processes to put in place.[18]

3.5 Step 5: analyse and use WEE data

The final step is to analyse and use the WEE data collected.

Interpreting and analysing data is a vital step in ensuring that the programme generates useful learning and insights around WEE which supports evidence-based decision-making, allowing project staff to respond in a dynamic way to what is and is not working on the ground. Data on its own carries no meaning: for data to become knowledge and learning it must be analysed, interpreted, and put into context. Data analysis includes:

- Calculation of **summary statistics** – such as totals and averages,
- **Outlier analysis** – to identify interesting positive or negative outcomes,
- **Trend analysis** – to examine changes over time, and
- **Correlation analysis** – to examine relationships between variables.

The data from different tools (especially quantitative and qualitative tools) should be triangulated and synthesised to build up a more complete picture of the change process. Beyond the planned data collection activities from the Measurement Plan, additional field observations and tacit information collected by project teams and consultants can also be included in the data analysis process.

If any project-level learning questions were identified at the start of the project (see Step 1), the data analysis process should aim to answer these questions. Data analysis can also be used to answer the general learning questions: what works and why, for whom, in what circumstances. Building on the idea of intersectionality, this recognises that women are not a homogenous group: what works for some women might not work well for other women with different characteristics or in different circumstances. For example, analysis of the *access* data may reveal lower usage of a financial service among rural versus urban women. There may be clues in the existing data, particularly qualitative data, to explain why this is the case. It may be, for example, that rural women have to travel further to access the service, or that they tend to have lower financial literacy and are therefore are less confident at using the service. Sometimes further data collection may be required to differentiate between competing hypotheses (represented by the dotted arrow in Figure 2).

By measuring change in each box in the empowerment pathway it should be possible to validate the assumptions underlying the pathway, monitor progress in achieving WEE outcomes, and identify any WEE barriers or unintended negative outcomes. By plotting data and findings against the empowerment pathway it should be possible to identify whether the project or intervention is on track to deliver the desired WEE outcomes and, if not, to pinpoint where the causal logic is breaking down. It may be that certain assumptions are not holding in practice, or that women face additional barriers that were not previously considered. As noted above, further data may be required to narrow down the range of possible explanations. Data collection is therefore seen as part of a dynamic problem-solving process, not a rigid exercise

[18] For more in-depth guidance on measuring domestic violence and intimate partner violence, see Annex 1 of the WHO Multi-country Study on Women's Health.

designed only to record programme successes. Being curious is an important aspect of analysing and using WEE data.

The process of generating lessons and insights from data can be a collaborative process involving project partners. For example, towards the end of the pilot phase of an innovation project, project staff may facilitate an insights and lesson learning workshop – involving the project team, partner counterparts, and external consultants – to analyse the pilot data, share observations, identify any issues, generate hypotheses, plan additional data collection (if required), and brainstorm solutions. In some cases a partner financial service provider may have tested multiple different versions or calibrations of a new model or innovation. Data should be analysed to understand which version is most effective (which may vary across contexts and client segments). The process of collecting and discussing gender data can in itself challenge prejudice or gender stereotypes and facilitate stronger partner engagement with WEE objectives.

Project teams are also encouraged to hold internal learning workshops where they can share initial data and insights from their project(s), and brainstorm hypotheses and potential solutions with colleagues. More formal review processes can also be used to promote intra- and inter-project learning. For example, the learning from one innovation project might well be relevant for other innovation projects also aiming to improve WEE outcomes. Creating diverse project teams can increase the effectiveness of these sessions – different team members will bring different ideas, perspectives, knowledge, and experiences, and can help to challenge each other's assumptions and biases.

To complete the cycle, project staff should use the insights and lessons generated to improve the intervention, modifying or dropping what does not work, and scaling-up what does. This might involve, for example, working with project partners to modify the design of a new financial service so that it is more accessible to women (which may in turn require modifications to the support package provided by the programme). In some cases it might be necessary to pause or exit an intervention if, for example, serious negative WEE outcomes are identified. The empowerment pathway should also be regularly reviewed and updated to reflect the latest learning and insights (thereby completing the cycle in Figure 2).

4 Resources

This section lists useful external resources for conceptualising and measuring WEE.

WEE concepts and frameworks

Kabeer, N. (1999). "Resources, Agency, Achievements: Reflections on the Measurement of Women's Empowerment". Development and Change, 30 (3): 435-464. https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-7660.00125

Golla, A. M. et al (2018). "Understanding & Measuring Women's Economic Empowerment: Definition, Framework & Indicators". International Center for Research on Women. www.icrw.org/publications/understanding-and-measuring-womens-economic-empowerment/

Measuring WEE

Anand, M. et al, (2014). "Practical Tools and Frameworks for Measuring Agency in Women's Economic Empowerment". The SEEP Network. https://seepnetwork.org/files/galleries/2019-WEE-MeasuringWomensAgency-_EN-DIGITAL.pdf

Badiee, S. et al (2022). "Transforming the Data Landscape: Solutions to Close Gender Data Gaps". Data 2X. https://data2x.org/resource-center/transforming-the-data-landscape-solutions-to-close-gender-data-gaps/

Buvinic, M., and Furst-Nichols, R. (2013). "Measuring Women's Economic Empowerment: Companion to A Roadmap for Promoting Women's Economic Empowerment." United Nations Foundation and Exxon Mobil Foundation. http://womeneconroadmap.org/sites/default/files/Measuring%20Womens%20Econ%20Emp_FINAL_06_09_15.pdf

Calder, R. et al (2020). "Measurement of Women's Economic Empowerment: WOW Helpdesk Guidance Note No. 2". Work and Opportunities for Women. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/895765/Guidance-Measurement-Womans-Economic-Empowerment2.pdf

Donald, A. et el (2020). "Measuring Women's Agency". Feminist Economics 26(3):200-26. https://openknowledge.worldbank.org/handle/10986/35282

Glennerster, R. et al. "A practical guide to measuring women's and girls' empowerment in impact evaluations". J-PAL.

Glennerster, R., and Walsh, C. (2017). "Is It Time to Re-Think How We Measure Women's Household Decision-Making Power?" J-PAL. www.povertyactionlab.org/blog/9-6-17/it-time-rethink-how-wemeasure-women's-household-decision-making-power-impact

Lombardini, S. et al. (2017). "A 'How To' Guide to Measuring Women's Empowerment". Oxfam GB. https://policy-practice.oxfam.org/resources/a-how-to-guide-to-measuring-womens-empowerment-sharing-experience-from-oxfams-i-620271/

Qualitative data collection (general guidance)

Ager, A, et al (2010). "Participative Ranking Methodology: A Brief Guide: Version 1.1." New York, NY: Program on Forced Migration & Health, Mailman School of Public Health, Columbia University. https://doi.org/10.13140/RG.2.2.34356.45448

Mack, N. et al (2005). "Qualitative Research Methods: A Data Collector's Field Guide." Research Triangle Park: Family Health International. www.fhi360.org/sites/default/files/media/documents/Qualitative%20Research%20Methods%20-%20A%20Data%20Collector%27s%20Field%20Guide.pdf

Raworth, K. et al. (2012). "Conducting Semi-Structured Interviews." Oxfam. https://policy-practice.oxfam.org.uk/publications/conducting-semi-structuredinterviews-252993

Researching Gender Based Violence

García-Moreno, C. et al (2005). "WHO Multi-Country Study on Women's Health and Domestic Violence against Women: Summary Report of Initial Results on Prevalence, Health Outcomes and Women's Responses". Geneva: World Health Organization, 101. www.who.int/reproductivehealth/publications/violence/24159358X/en/

Innovations for Poverty Action (2018). "The Safe and Ethical Conduct of Violence Research: Guidance for IPA Staff and Researchers". Poverty Action. www.poverty-action.org/publication/ipv-ethical-guidance

ANNEX: example survey questions

Annan, J. et al (2016). "The Returns to Microenterprise Support among the Ultrapoor: A Field Experiment in Postwar Uganda." American Economic Journal: Applied Economics 8(2): 35-64.

When you have small amounts of money, such as 500 or 2,000 shillings, can you decide how to spend it on your own? (Yes / No)

Follow up with:

- When an expensive item like a bicycle or a cow is purchased by the household, is your opinion listened to in the decision of what to buy?
- If you have some money you have earned, can you use it to purchase clothing for yourself or children without asking the permission of anyone else?
- Are you allowed to buy and sell things in the market without asking the permission of your partner?

If money is available, who in your household decides whether to pay school fees for a relative from your side of the family? (You primarily / You with someone else / Someone else without you)

Follow up with:

- If money is available, who in your household decides whether to purchase items like a radio or a paraffin lamp?

If you have money that you have earned, can you refuse to give some to your partner if he/she wishes to purchase alcohol? (Often / Sometimes / Rarely / Never)

Do you agree that a wife has a right to buy and sell things in the market without asking the permission of her husband? (Yes / No / Don't know)

If a wife has earned some money, does she have the right to buy clothing for herself or her children without asking the permission of her husband? (Yes / No / Don't know)

Almas, I. et al (2015). "Measuring and Changing Control: Women's Empowerment and Targeted Transfers." National Bureau of Economic Research Working Paper No. 21717.

In the last 2 weeks, did you and your spouse argue about managing money? (Yes / No / Not applicable / Don't know)

Who in the household usually decides how much money to be spent on food? (Wife / Husband / Together / Not applicable / Don't know)

Follow up with:

- Who in the household usually decides about the financial administration?

Imagine the following household composed of a wife, a husband and three children. The wife is 40 years old and her husband is 43 years old. The three children are aged 5, 10 and 14. Both wife and husband have been unemployed in the last 2 years and have been receiving Financial Assistance. Today, the wife receives X Macedonian Denars (MKD) from her parents to

help the family. Who do you think should decide what to do with that amount? (Wife / Husband / Together / Not applicable / Don't know)

In the following questions you will be facing different scenarios in which you will have to choose between two alternatives, A or B. You cannot choose both. If you choose A it means you prefer alternative A to alternative B. Which of these two alternative options do you prefer?

- 550 MKD paid to you (A) or 600 MKD paid to your partner (B)?
- 500 MKD paid to you (A) or 600 MKD paid to your partner (B)?
- 400 MKD paid to you (A) or 600 MKD paid to your partner (B)?
- 300 MKD paid to you (A) or 600 MKD paid to your partner (B)?

Björkman Nyqvist, M. et al (2017). "Mothers Care More, but Fathers Decide: Educating Parents about Child Health in Uganda". American Economic Review, 107(5): 496-500

To what degree do you agree with these statements? (Strongly agree / Agree / Disagree / Strongly disagree / Not applicable)

- When women get rights they are taking rights away from men
- Gender equality, meaning that women and men are equal, has come far enough already
- A wife should obey her husband, even if she disagrees.
- It is important for a man to show his wife/partner who is the boss.
- It is the job of men to be leaders, not women
- A woman should be able to choose her own friends, even if her husband disapproves
- A man should decide how to spend his free time on his own
- A woman should decide how to spend her free time on her own
- If a woman has power in the household, it means she is taking power away from her husband
- A husband and wife can share power
- Women's opinions are valuable and should always be considered when household decisions are made

Green, D. et al (2018). "Silence Begets Violence: A Mass Media Experiment to Prevent Violence against Women in Rural Uganda". Working Paper.

In some of the other villages we have visited, (some) people think that a man has good reason to hit his wife if she disobeys him, while (other) people in those communities do not think this is a good reason to hit one's wife. In your community, do people think a man has a good reason to hit his wife if she disobeys him? (Yes / No / Don't know / Refuse to answer)

Thinking of the opinions held by people in your community, do most people think that a man has a good reason to hit his wife if she does not complete her household work to his satisfaction? (Yes / No / Don't know / Refuse to answer)

Karlan, D. (2017). "Impact of Savings Groups on the Lives of the Poor". Proceedings of the National Academy of Sciences

How would you rank your ability to do these activities on a scale of 1 to 10?

- Run your own business
- Identify business opportunities to start up a new business
- Obtain credit to start up a new business or expand an existing business
- Save in order to invest in future business opportunities
- Make sure that your employees get the work done properly
- Manage financial accounts
- Bargain to obtain cheap prices when you are buying anything for the business
- Bargain to obtain high prices when you are selling
- Protect your business assets from harm by others
- Collecting the money someone owes you

How true is the following statement, on a scale of 1-10?

- While doing any task, it is important for me to do it better than others
- If I start working in a task, I definitely see the end of it no matter how difficult it is
- If I have the chance, I would make a good leader
- I want to be a respectful person in my village
- I do not care what others think about my success or failure
- I am in control of what happens in my life
- I often make plans for the future
- I believe that my future is determined by luck no matter how hard I work

Lucia, R. et al (2020). "Measuring And Understanding Unpaid Care And Domestic Work: Household Care Survey". Oxfam.

	1 *Primary care*	*2* *Secondary care*		*3* *Supervision care*	*4* *Supervision care*
	What were you doing yesterday from [TIME]? See codes below	**What else were you doing at the same time?** 0 = Nothing else See codes below		**Were you responsible for looking after a child (<18 years) during that hour?** 0 = No 1 = Yes 99 = There is no child in my household	**Were you responsible for looking after a dependent adult during that hour?** 0 = No 1 = Yes 99 = There is no dependent adult in my household
04am – 05am	|__|__|	|__|__|		|__|	|__|
05am – 06am	|__|__|	|__|__|		|__|	|__|
…	|__|__|	|__|__|		|__|	|__|

https://policy-practice.oxfam.org/resources/measuring-and-understanding-unpaid-care-and-domestic-work-household-care-survey-621082/

Website
www.tandem-development.com

Contacts
gareth.davies@tandem-development.com

Address
Tandem Development International Ltd,
56 Leman Street,
London,
United Kingdom,
E1 8EU

TANDEM