

DEVELOPING AND WEIGHING OUTCOME MEASURES TO ASSESS THE IMPACT OF ECONOMIC STRENGTHENING INTERVENTIONS

HES Research Dialogues: Methods Brief III

Marija Pantelic and Janina Steinert

Introduction

Good evidence on the impact of an intervention depends on the quality of the outcome measures used. If outcome measures are not carefully selected and adapted, the conclusions and findings drawn from a study can be biased. For example, if there is a discrepancy between what the researchers intend to ask in a survey and how study participants understand the questions, it becomes challenging to interpret findings. This brief provides a simple set of guidelines for adapting or developing adequate outcome measures for an impact evaluation of Household Economic Strengthening (HES) interventions.

Development and Assessment of Outcome Measures

Before designing an impact study, it is essential to conduct a thorough search to identify key

relevant measures that are already being used by related studies and projects. It is recommended that project managers and program developers identify systematic reviews of measurements on a given subject.¹ For example, to assess whether a project has affected decreases in HIV-related stigma, you may consult an existing systematic review of illness-related stigma scales. As the review has already assessed the quality of existing stigma scales, it will make it easier for you to decide on which scale to use (Stevlink et al. 2012). If no systematic review is available, you can identify measures from individual studies assessing the outcome of your interest. Given that many studies might be ongoing and not yet published, it is additionally advised to consult experts in the field to refer you to existing outcome measures. They might also be able to share information and insights that have not been published – for example, how certain measurements worked in their practical experience. Once you have identified one or more measure, it is essential to determine its quality as well as its relevance to your population of interest.

ASSESSING THE QUALITY OF EXISTING MEASURES

Not all existing measures will be adequate to measure your outcome of interest. Below is a set of criteria to consider before deciding whether or not to use the existing measures that you have identified:

- **Reliability:** There are different types of reliability but one common type is **internal consistency**, which evaluates whether

¹ For more information on systematic reviews, see: Methods Brief I.

various questions in a measurement relate to the same underlying construct (e.g. housing quality, ownership of durables, nutrition and health should reflect the underlying construct of household wealth). One way to check this is to read prior research that provides evidence of correlations between indicators of interest. Sometimes researchers will report “Cronbach’s α ” as a measurement of internal consistency or correlation between indicators. In general, values of >0.7 should be considered acceptable.

- **Validity: Content validity** aims to assess whether a set of questions measure what they intend to measure – for more guidance on content validity see back translation and cognitive interviews in the following section. **Criterion validity** should assess whether, and to what extent, a measure correlates with a theoretically expected outcome. For example, we might want to assess whether a measure of poverty correlates with food insecurity. Or whether responses to an adherence measure correlate with the respondents’ viral loads? Correlation in the expected direction (positive or negative) would indicate high criterion validity. By contrast, **face validity** refers to the extent to which a measurement tool is *subjectively* viewed by non-experts as accurately measuring what it is designed to measure.

ASSESSING THE MEASUREMENT’S APPLICABILITY TO YOUR POPULATION OF INTEREST

Important questions to consider here are:

- How many studies have used this measure with **your population of interest and in**

your cultural/ geographic setting? If a measure has been used only with a specific population (e.g. adult populations in Western countries), dissimilar from your population of interest, this might suggest that the measure is not the most appropriate for your evaluation.

- Was the population of interest involved in **designing or adapting the measurement?**

If the answer to either question is *no*, researchers may **adapt an existing outcome measure to their specific context or population**. Below is a set of steps that help to ensure that the measure is comprehensive and valid to your target population.

1. **Back translation:** After your measurement questions have been translated into the language of your target population, an independent interpreter should translate these back into English (or original language). This should highlight any inconsistencies between the original questions and the back-translated questions and therefore ensure that the intended meaning is retained (Brislin 1970).
2. **Cognitive interviewing** is a qualitative method that is commonly used to uncover inconsistencies between what the measurement items are meant to ask and the way in which members of the target population interpret these items. Cognitive interviewing involves probing of respondents to interpret the meaning of items and specific terms within the items. If not detected and addressed, such inconsistencies can introduce bias into conclusions drawn

from empirical data. For more details, see De Silva et al. 2006 and Willis 1999.

- 3. Expert review:** Once you have adapted the measurement, it is advised that you circulate it amongst key researchers and practitioners for their expert review. They might be able to point to certain areas that your scale or questionnaire has not covered. For larger studies, it is advisable to consult at least two other researchers and two practitioners, but for smaller studies, one researcher and one practitioner would be adequate. Allow approximately two weeks, or an otherwise reasonable amount of time for the task, for people to provide input.
- 4. Piloting:** It is important to do a 'test run' of the questionnaire before administering the actual survey. Piloting can highlight further problems that were not revealed in back translation or cognitive interviewing. As a rule, you should aim to pilot the questionnaire with at least five members of your target population. However, after the piloting these people should be excluded from your study sample.

DEVELOPING YOUR OWN MEASURE – A LAST RESORT OPTION

In the unlikely event that there are no pre-existing relevant measures, you should develop your own set of questions. This can be done through qualitative one-on-one interviews and/or focus group discussions. These tools should be used to uncover themes that relate to the outcome of interest in the specific population with which you are working. Once a set of questions has been developed, all of the steps listed under 'adapting an existing outcome measure' (see above) should follow.

Additional Considerations

DISTAL AND PROXIMAL OUTCOME MEASURES

Qualitative research for measurement development

One-on-one qualitative interviews are advised for sensitive topics such as sexual risk behaviour. Focus group discussions might be appropriate for non-stigmatized topics.

For more information about how to use one-on-one interviews and focus groups for measurement development, see: Ward & Atkins 2002, Kruegar & Casey 2002, Bauman & Adair 1992.

Proximal outcomes, or intermediate outcomes, are ones that are observed in the short term, such as coping mechanisms or immediate changes in behaviours. By contrast, distal outcomes, or impact outcomes, are outcomes that are observed in the longer term, such as school drop-out, adolescent pregnancy and HIV infection rates. Measuring proximal outcomes rather than distal outcomes often has the advantage of shorter follow-up periods and lower costs of research (Fraser et al. 2009). However, project managers should ensure that there is a strong evidence base to conclude that changes in the proximal outcomes would lead to changes in distal outcomes.

SELF-REPORTED VERSUS HARD OUTCOMES

Ideally, efforts should be made to combine both self-reported and hard outcomes, or verifiable outcomes. While hard outcome data may often be more difficult and costly to collect, self-reported data can be subject to reporting biases. For instance, sexual risk behaviour is likely underestimated when based on

adolescents' self-report due to the effect of social desirability and stigmatisation. Biological data such as HIV/STS testing could be used to complement self-reports, however, these may be problematic from an ethical and practical point of view. If collecting hard outcome data is not possible, methods such as self-interviewing techniques through tablets or mobile devices should be considered to increase disclosure rates of stigmatised outcomes, such as sexual risk behaviours or HIV infection (Gorbach et al. 2013, Malotte et al. 2011).

INCLUDING MEASURES OF POTENTIAL HARMFUL EFFECTS

Was the possibility of harmful effects considered? The only way to assess whether the intervention resulted in negative effects is to also measure potential negative or harmful outcomes. For example, in their review of cash transfer programs, Pettifor et al. (2012) found that although *“concerns were raised that women would suffer negative consequences as a result of receiving the payment, [...] the evidence shows that payments made to female heads of [household] have not resulted in increased intimate partner violence. In fact, women are the more effective beneficiaries with regard to multiple health and education outcomes compared to men.”* Potential harmful effects can be identified before the survey is administered through additional questions asked during the piloting or expert review.

ABOUT THE AUTHORS

Marija Pantelic is a DPhil candidate in the Department of Social Policy and Intervention, University of Oxford. Her research interests lie in methods to minimize bias from observational studies and finding evidence-based

interventions to improve the psychosocial wellbeing of orphaned and vulnerable children.

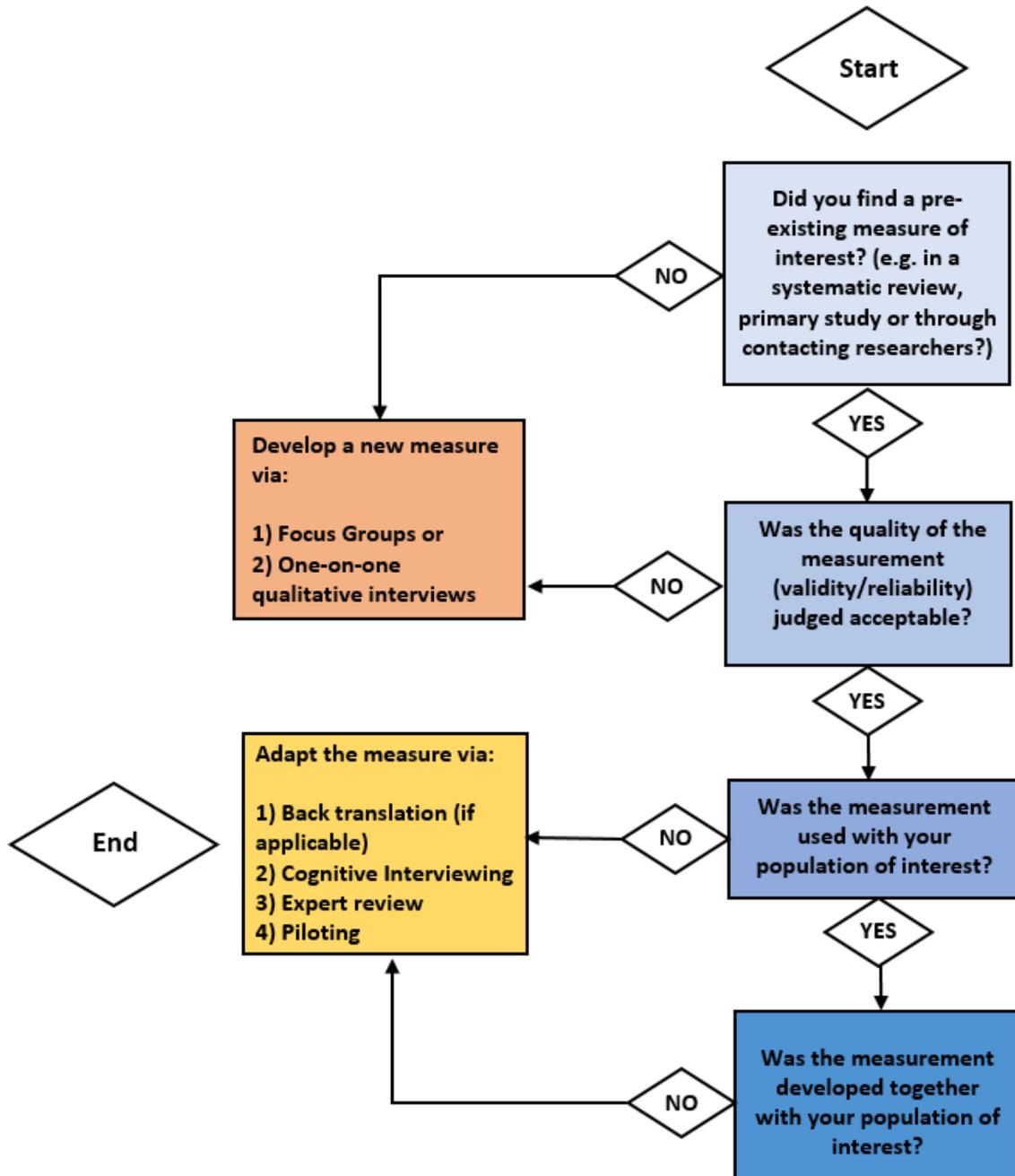
Janina Steinert is a DPhil candidate in the Department of Social Policy and Intervention, University of Oxford. Her research is focused on HES through microsavings and financial literacy and is analyzing economic resilience promotion in a low-income, HIV/AIDS-affected population in rural South Africa.

About the HES Research Dialogues:

In 2014, FHI 360's ASPIRES Project and the SEEP Network recognized that, while HES was a growing area of practice and research, gaps in HES research and evidence remained. To respond to this evidence gap, SEEP facilitated an HES Research Dialogues initiative, bringing together HES researchers and practitioners to define a collaborative learning agenda. Through a series of collaborative activities, the initiative aimed to identify key research questions within HES, as well as draw on existing experience related to appropriate research methods and tools.

This document is complemented by a series of research methods and evidence briefs developed out of the HES Research Dialogues initiative. Access them on FHI 360's ASPIRES Project web page on Microlinks at: <http://bit.ly/1rwRue3>

Figure 1. Decision Tree for the Development and Assessment of Outcome Measures



References

- Bauman, L. J., & Adair, E.G. (1992). "The Use of Ethnographic Interviewing to Inform Questionnaire Construction." *Health Education Quarterly* 19 (1): 9–23.
- Brislin, R.W. (1970). "Back-translation for cross-cultural research." *Journal of Cross-Cultural Psychology* 1(3):185-216. doi: 10.1177/135910457000100301.
- De Silva, M.J., Harpham, T., Tuan, T., Bartolini, R., Penny, M.E., & Huttly, S.R. (2006). "Psychometric and Cognitive Validation of a Social Capital Measurement Tool in Peru and Vietnam." *Social Science & Medicine* 62(4): 941–53. doi:10.1016/j.socscimed.2005.06.050.
- Fraser, M.W. (2009). *Intervention Research: Developing Social Programs*. Oxford University Press, USA.
- Gorbach, P.M., Mensch, B.S., Husnik, M., Coly, A., Mâsse, B., Makanani, B., Nkhoma, C. et al. (2013). "Effect of Computer-Assisted Interviewing on Self-Reported Sexual Behavior Data in a Microbicide Clinical Trial." *AIDS and Behavior* 17 (2): 790–800. doi:10.1007/s10461-012-0302-2.
- Krueger, R.A., & Casey, M.A. (2008). *Focus Groups: A Practical Guide for Applied Research. Fourth Edition*. Los Angeles: SAGE Publications, Inc.
- Malotte, C., Cutting, A., Huettner, S., Matson, P., & Ellen, J. (2011). "P2-S8.12 Feasibility of Using Cell Phones for Daily Data Collection within Adolescent Cohort Studies." *Sexually Transmitted Infections* 87 (1): 260–61. doi:10.1136/sextrans-2011-050108.386.
- Stevenson, S.A.M, Wu, I.C., Voorend, C.G.N, & van Brakel, W.H. (2012). "The Psychometric Assessment of Internalized Stigma Instruments: A Systematic Review." *Stigma Research and Action* 2 (2). doi:10.5463/sra.v1i3.35.
- Streiner, D.L. (2008). *Health Measurement Scales: A Practical Guide to Their Development and Use. 4 edition*. Oxford; New York: Oxford University Press, USA.
- Pettifor, A., MacPhail, C., Nguyen, N., & Rosenberg, M. (2012). "Can Money Prevent the Spread of HIV? A Review of Cash Payments for HIV Prevention." *AIDS and Behavior* 16 (7): 1729–38. doi: 10.1007/s10461-012-0240-z.
- Ward, H.D., Institute for Child and Family Policy (Edmund S. Muskie School of Public Service), and Julie Atkins. (2002). *From Their Lives: A Manual on How to Conduct Focus Groups of Low-Income Parents*. Institute for Child and Family Policy, University of Southern Maine.
- Willis, G.B. (1999). "Reducing Survey Error through Research on the Cognitive and Decision Processes in Surveys". In: *Meeting of the American Statistical Association*. Research Triangle Institute.