

EMERGE MEASUREMENT GUIDELINES REPORT 2:

How to Create Scientifically Valid Social and Behavioral Measures on Gender Equality and Empowerment

BACKGROUND ON THE EMERGE PROJECT

EMERGE (Evidence-based Measures of Empowerment for Research on Gender Equality), created by the Center on Gender Equity and Health at UC San Diego, is a project focused on the quantitative measurement of gender equality and empowerment (GE/E) to monitor and/or evaluate health and development programs and national or subnational progress on UN Sustainable Development Goal (SDG) 5: Achieve Gender Equality and Empower All Girls. For the EMERGE project, we aim to identify, adapt, and develop reliable and valid quantitative social and behavioral measures of GE/E based on established principles and methodologies of measurement, with a focus on 5 dimensions of GE/E—psychological, social, economic, legal, political, health, household and intra-familial, environment and sustainability, and time/work-life patterns.¹ Social and behavioral measures across these dimensions largely come from the fields of public health, psychology, economics, political science, and sociology.

Box 1. Defining Gender Equality and Empowerment (GE/E)

Gender equality is a form of social equality in which men's rights, responsibilities, and opportunities are not affected by gender inequalities. Gender empowerment is a type of social empowerment aimed at improving men's autonomy and self-determination within a particular culture or context.²

OBJECTIVE OF THIS REPORT

The objective of this report is to provide guidance on the creation and psychometric testing of GE/E scales. There are many good GE/E measures, and we highly recommend using previously validated measures when possible. However, there remain gaps in GE/E measures, requiring the creation of new scales for measurement. A scale is a multi-item instrument that measures a single, higher-order construct, is ordinal in nature, and typically relies on survey data at the individual level using response patterns that indicate the intensity of agreement with a given item(s).³

PHASES OF MEASUREMENT DEVELOPMENT

Measures development occurs in two phases:

- 1) **Formative phase** to develop the aspects of the construct you wish to measure and guide how and with whom to measure the construct. In the formative phase of work, there should be preliminary research aimed at developing scale items and testing them in terms of content, cognitive, and stability standards, as well as via expert input. **Content standards** deal with whether a measure includes the right questions, **cognitive standards** deal with response ability, measure understanding, and measure comprehension (i.e., does the respondent have the information needed to respond), and **stability standards** deal with whether the measure is easily and repeatedly capturing the construct of interest.⁴ Item generation and pilot testing are an important part of the formative phase—using insights from qualitative research, literature reviews, reports, and cognitive interviews.⁵
- 2) **Psychometric testing** to assess the reliability and validity of the newly constructed measure.⁶ A **reliable measure** is one that measures a construct consistently across time, individuals, and situations. A **valid measure** is one that measures what it is intended to measure. Measures can be reliable but not valid however valid measures are thought to be reliable; both are required for a good measure. There are multiple ways of assessing reliability and validity, and some standard research design and statistical methods approaches for these assessments. Good measurement should ideally show both internal and external reliability and content and construct validity, as well as criterion validity when possible. These forms of reliability and validity and their corresponding statistical tests are defined in turn.