

RESEARCH ARTICLE

Investigating impacts of gender-transformative interventions in water, sanitation, and hygiene: Structural validity, internal reliability and measurement invariance of the water, sanitation, and hygiene–Gender equality measure (WASH-GEM)

Jess MacArthur^{1*}, Rachel P. Chase², Diana Gonzalez¹, Tyler Kozole³, Chris Nicoletti⁴, Veasna Toeur³, Juliet Willetts¹, SNV Beyond the Finish Line Team[¶]

1 Institute for Sustainable Futures, University of Technology Sydney, Sydney, Australia, **2** Independent Researcher, Columbus, Ohio, United States of America, **3** iDE Cambodia, Phnom Penh, Cambodia, **4** iDE Global, Denver, Colorado, United States of America

[¶] Membership of the SNV Beyond the Finish Line Team is provided in the Acknowledgments.

* Jessica.MacArthur@uts.edu.au



OPEN ACCESS

Citation: MacArthur J, Chase RP, Gonzalez D, Kozole T, Nicoletti C, Toeur V, et al. (2024) Investigating impacts of gender-transformative interventions in water, sanitation, and hygiene: Structural validity, internal reliability and measurement invariance of the water, sanitation, and hygiene–Gender equality measure (WASH-GEM). *PLOS Water* 3(10): e0000233. <https://doi.org/10.1371/journal.pwat.0000233>

Editor: Shafi Mohammad Tareq, Jahangirnagar University, BANGLADESH

Received: January 27, 2024

Accepted: September 21, 2024

Published: October 17, 2024

Copyright: © 2024 MacArthur et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: De-identified data has been provided in [supplementary materials](#).

Funding: This work was supported by the Australian Government's Water for Women Fund - Grant number WRA-034 (JW). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Links between gender equality and water, sanitation, and hygiene (WASH) are well documented in qualitative studies, and emerging literature is recognizing the transformative potential of WASH interventions towards improvements in gender equality, of which empowerment is an important aspect. However, quantitative approaches to measure changing gender equality within WASH programs remain nascent. This study builds on previous work conducted in collaboration with the multi-country Water for Women Fund, establishing the theoretical development of the multi-dimensional WASH-gender equality measure (WASH-GEM) with five domains: resources, agency, critical consciousness, structures, and wellbeing. Starting from a strong conceptual foundation—this article demonstrates the WASH-GEM's empirical rigor in balance with practical considerations. We present the measure's staged development; examine its structural validity, internal consistency, and measurement invariance from an empirical basis; providing analysis from concurrent validation studies in Cambodia and Nepal (n = 3056). Twelve of the WASH-GEM's 15 testable themes demonstrated high internal reliability and nine structural validity. Further refinement is recommended for the remaining four themes. Validation results from Cambodia and Nepal provide evidence of women and men's differentiated experiences in relation to their WASH access, decision-making and roles, and in their lives more broadly. Results also reinforce the inter-sectional experiences of inequality and the importance of measurement beyond women's empowerment. Overall, we provide evaluators and researchers with twelve curated and validated measures to examine changing gender dynamics within WASH-programs. These

Competing interests: The authors have declared that no competing interests exist.

measures can be used as diagnostic or evaluation tools to support design and implementation of gender-transformative WASH interventions, programs, and policies.

1. Introduction

Embedded in the Sustainable Development Goals (SDGs) is a recognition that improvements in health, education, food security, and poverty reduction are only strengthened through improvements in gender and societal equalities (Goals 5 and 10). This same desire is also illustrated in Goal 6's emphasis on water and sanitation for all access with "specific attention to the needs of women, girls and those in vulnerable situations." Referring to the Millennium Development Goals, but with continued validity for the SDGs, Willetts et al. argue that "*water, sanitation and hygiene issues and gender equality can be constructively addressed together, with the former three providing an entry point for the latter*" [1]. Such an approach to water, sanitation, and hygiene (WASH) programming has been more recently described as a gender-transformative approach to WASH [2–4] aiming to "*synergistically both address gender and social inequalities and improve WASH outcomes*" [4].

Building on the wealth of qualitative explorations of gender equality in WASH [5–11], programs increasingly require comparable quantitative data sets to evaluate the salience and scale of transformative change resulting from interventions. Such measures hope to strengthen evidence for gender-focused WASH interventions, offer monitoring and evaluation opportunities, and inform better gender-WASH practice.

Over the last decade there has been a surge in focus on this type of quantitative evidence within the international development sector, including in the WASH sub-sector. For example, recent quantitative measurement of empowerment, an important component of gender equality investigations, has utilized the empowerment modules in the demographic and health surveys (DHS) to define and evaluate measures for the African context [12–15]. Development of the Women's Empowerment in Agricultural Index (WEAI) to measure differences in empowerment between women and men related to agriculture has also catalyzed an increase in quantitative approaches to assessing intervention changes [16]. The WEAI uses a threshold of empowerment to identify the relative frequency of empowered and disempowered individuals. Subsequent iterations of the WEAI have aimed to shorten it [17] and use it within programs [18]. Adaptations of the WEAI have been used within development sub-sectors such as livestock [19], fisheries [20], nutrition [21], solar energy [22] and most recently sanitation [23]. Other recent quantitative measures within the WASH sector focus on measuring women's distress in association with water insecurity [24], sanitation insecurity [25], and empowerment in urban sanitation [26, 27].

All of these tools are useful in the measurement and evaluation of change within programs; however, none explicitly seek to assess gender-transformative approaches to WASH. The WASH-GEM addresses this gap, with an aim to create a practical and conceptually grounded tool for WASH programs that can both: 1) inform where and how they should intervene by better understanding gender dynamics within and beyond WASH, and 2) provide evidence of program influence on gender equality in relation to WASH access, decisions, and roles as well as more broadly.

This paper complements other publications on the WASH-GEM. Carrard et al. provides detailed justification of the theoretical foundation of the WASH-GEM and its domains, including a critical discussion on the tool's intentional focus on gender-equality rather than

women's empowerment [4]. Gonzalez et al. discusses the collaborative partnership approach to measure design, involving academics, practitioners, and civil society in Cambodia and Nepal, highlighting how practitioners shaped its design and how these partners were able to use the information it provides in their WASH program implementation [28].

This paper presents the most recent empirical development stage of the WASH-GEM with a focus on its validity and reliability. The article begins with a brief discussion of gender-transformative WASH and provides an overview of the structure and format of the WASH-GEM. Next, the methodology section describes the process of developing and validating the WASH-GEM in Cambodia and Nepal, and discusses limitations. The results follow. We present descriptive, reliability and validity statistics from the concurrent validation pilots ($n = 3056$) including coefficients of internal consistency and measurement invariance. We then review the measure's internal construct validity using the same dataset. Finally, we present the thematic scores from the validation study. We conclude with discussion of the opportunities and challenges to continue developing, refining, and using the measure within WASH programs.

2. Background: Gender-transformative WASH

As explored during the conceptual design of the WASH-GEM, previous scholarship has identified important connections between gender dynamics in households and communities and the traditional burden often placed on women and girls related to WASH [4]. This includes aspects such as cooking, cleaning, water collection, and childcare [29, 30]. Additionally, menstruation, pregnancy and menopause give rise to specific WASH needs for women and girls [31]. The many gender dynamics embedded within WASH-related needs and activities, at both community and household levels, provide a unique entry point to address wider gender inequalities. This approach is described as gender-transformative, in which WASH interventions can influence wider gender dynamics, norms and social structures [4].

Historically, WASH development programs have adopted a range of welfare, efficiency, or equity approaches to addressing gender dynamics often through the provision and management of water or sanitation infrastructure [32]. Typically, these programs focus on women rather than gendered social structures [4, 33]. However, over the last thirty years, a sub-set of programs have thought more strategically about how to better engage with gender dynamics in programs. This has been done by linking practical gender needs such as access to water and sanitation, with strategic gender interests such as empowerment and changing gender norms [1, 34, 35]. By focusing on wider societal transformations, gender-transformative approaches also aim to lift the burden of change from women, a challenge inherent in many women-focused interventions [4, 36].

3. Overview of the WASH-GEM

Building on this gender-transformative approach to WASH and articulated in parallel publications [4, 37], the WASH-GEM is ultimately rooted in the Capability Approach of human development, which identifies aspects of individual empowerment as pathways towards transformative equality [38–40]. As such, the WASH-GEM centers on personal empowerment as defined by Kabeer [32], but situates empowerment within the broader social and structural environment—to reflect a societal system of gender equality. At its core the WASH-GEM includes the personal empowerment domains of **Resources**, **Agency**, and **Wellbeing** (Kabeer's achievements or outcomes). This empowerment triad is then blended with two additional domains that articulate the influence of gendered societal norms on personal empowerment as evidenced in beliefs and community structures. Focused on beliefs, the WASH-GEM includes a **Critical Consciousness** domain which highlights one's ability to recognize societal

inequalities. This is distinct to other conceptualizations of Critical Consciousness as agency (WEAI) or resources (ARISE scales). As such, the WASH-GEM emphasizes that the Critical Consciousness of family and community members can significantly influence one’s personal empowerment. Next, addressing social dynamics, the WASH-GEM includes the domain of **Structures** by measuring community gender norms. The Structures domain also includes the theme of mobility, which while often connected with agency (as in the WEAI and ARISE scales), was conceptualized within the WASH-GEM as being more limited by societal norms and practices, than intrinsic agency.

As a result, the WASH-GEM comprises five domains: Resources, Agency, Critical Consciousness, Structures, and Wellbeing, and a total of 16 themes (Fig 1). These themes are designed to help diagnose gender dynamics or assess changes associated with WASH programs in both WASH-related and beyond-WASH activities. The multi-thematic approach has also recently been taken in the development of other quantitative measures, such as the ARISE scales which measure women’s empowerment in urban sanitation and are a complementary set of measures to the WASH-GEM [27]. As a gender-transformative tool, the WASH-GEM curates a set of measures related to both WASH and beyond-WASH, with an aim of demystifying what types of gender equality aspects can be influenced by changes in WASH [37]. It should be noted that this conceptual framework has evolved since previous WASH-GEM publications based on the validation study results.

At the start of the project, the WASH-GEM was conceptualized as a single multi-dimensional index or a series of domain-wise indices; however through the collaborative conceptual design, and rapid and exploratory piloting this approach was deemed less judicious for practical and empirical reasons. A single summary index has been adopted by the developers of other quantitative measures, such as the WEAI [17, 41] and the EWI [23], and include thresholds or levels of empowerment. However, with the WASH-GEM’s broader focus on equality (of which empowerment is a key component) the team decided during the exploratory piloting phase, that it was not appropriate to distill individual experiences into a single number with a threshold. As such the WASH-GEM, became set of measures—such as the ARISE scales [27]

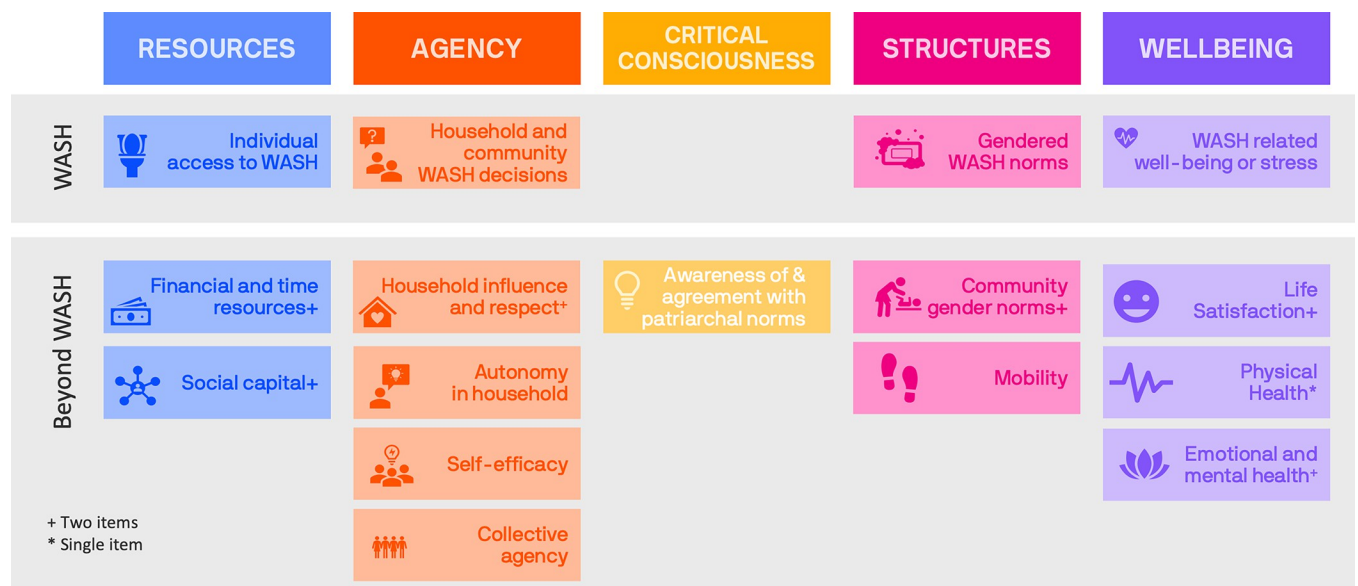


Fig 1. Themes within the WASH-GEM indicating the WASH and beyond WASH construction.

<https://doi.org/10.1371/journal.pwat.0000233.g001>

and thematic interpretation of the DHS empowerment module from Nigeria [15]. While a set of measures is less concise than a single index or domain-indices, it is arguably more prudent and empirically grounded.

Additionally, during the initial conceptual framework development, and piloting phases of the measure design, the team explored and tested a wide variety of existing and novel items with a goal of creating a short, practical and simple measurement tool focused on highlighting gender differences with a strong do-no-harm approach. As such the team explored measures such as the *Self-Reporting Questionnaire (SRQ-20)* [42] for mental health, the *Psychological Wellbeing measure (PWB-18)* [43], the *Perceived Physical Health Survey (RAND-36)* [44], and the *Critical Consciousness Scale (CCS-22)* [45]; yet these measures were deemed to be too long for a multi-dimensional measurement system. Other measures such as the *Psychological Wellbeing and Self-esteem Measure* [46] and critical consciousness measures [47, 48] had items that would require significantly more robust do-no-harm approaches than would be appropriate for the WASH-GEM (through items about self-harm, suicide ideation or violence). The developers of the WASH-GEM believe that much more care must be taken when exploring such sensitive topics than is usual in program-based studies. Additional useful scales such as the *Individual Water Insecurity Experiences* [49] only explored aspects to water and did not incorporate aspects of sanitation and hygiene—a core part of the WASH-GEM’s conceptualization. The ARISE sanitation scales [27] were not yet developed during this piloting phase, but also did not include water and hygiene. Lastly, the team also piloted existing agency focused tools such as the *General Self-efficacy Scale (GSE-6)* [50], the *Relative Autonomy Index* [41, 51], the *Household Decision-making Index* [52], the *Power Ladder* [53], and the *Sense of Agency Scale* [54]. However, these measures all performed poorly during cognitive interviews and piloting, with confused and frustrated respondents. For example, concepts such as ‘goals’ and ‘achievements’ proved complicated and non-contextualized for respondents, sometimes even causing distress. Often the nuance in the English wording was lost in translation leading to repetition especially in local dialects. As such, only two existing measures were piloted in the validation study: the *Relative Autonomy Index* [51] and the *General Self-Efficacy Scale* [50] used to explore criterion validity.

4. Methodology: Measure validation

Since the beginning, the WASH-GEM was designed to balance first **conceptual**, then **practical**, and finally **empirical** considerations. As such, the measure was not initially designed using traditional scale development techniques. Hence the empirical validity of the measure, the focus of this article, was done at a later stage once the key aspects of the measure had been determined using practical and conceptual considerations. In this section we outline the process of testing the measure’s empirical validity based on data collected through a third iteration of the WASH-GEM in 2020.

4.1 Measure development process

The WASH-GEM’s multi-phase development process began in 2019 as outlined in Fig 2 and included multiple rounds of data collection and measure refinement. This paper specifically presents results from Phase 4: Validation Pilot. Details of the other phases can be found in Table A in S1 Text.

Each phase of the process focused on a different aspect of validity, however also engaged with all aspects of conceptual, content, face, and construct validity to some extent. The measure was refined, and its utility interrogated across all stages of its development in collaboration with WASH practitioner partners in Cambodia and Nepal: iDE and SNV respectively [4]. The validation pilot took place in 2020, with continued analysis through 2023.

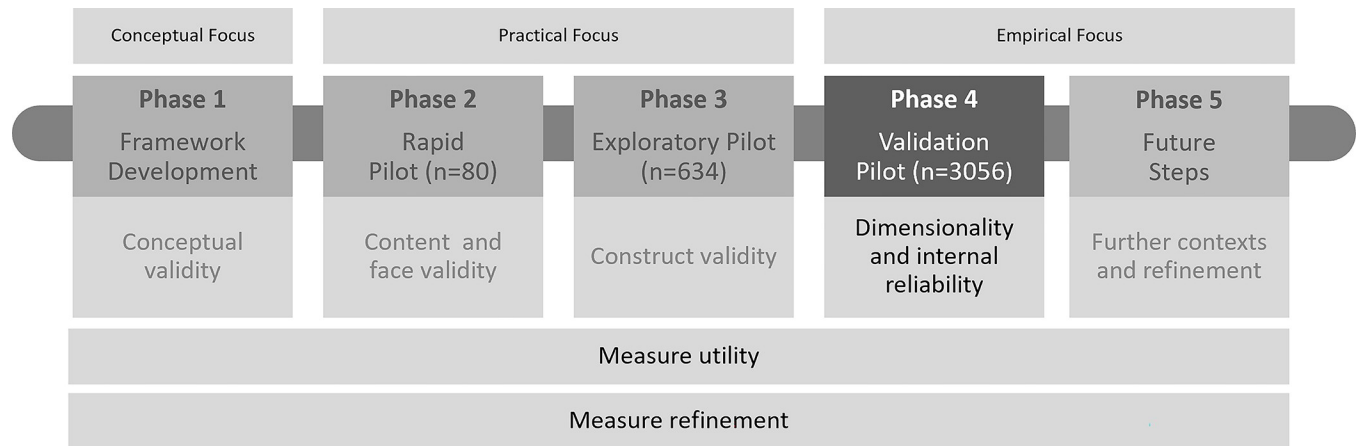


Fig 2. Phased measure development process and phase focus.

<https://doi.org/10.1371/journal.pwat.0000233.g002>

The validation pilot utilized a pre-shortened version of the WASH-GEM, relying on results from the rapid and exploratory pilots which reviewed conceptual validity, content validity, construct validity and initial refinement. While not the focus of this article, Phases 1–3 included expert consultation (n = 30), global expert review (n = 12), local expert review (n = 10), expert translation, rapid pilot observations (n = 40), cognitive interviews (n = 20), sensemaking workshops (n = 4), item difficulty and discrimination analysis, exploratory factor analysis, inter-item and intra-measure correlations, and known-groups validity.

There is further work required to explore the external validity of the WASH-GEM and measurement invariance in new contexts as indicated in Phase 5. Nonetheless, this work offers a starting point for gender-transformative measurement.

4.2 Programmatic context

The WASH-GEM was developed in collaboration with two rural WASH programs within the Australian Government's Water for Women Fund in South and Southeast Asia. Each program adopted a systems-focused approach, one through sanitation markets in Cambodia and the other through water governance in Nepal.

- *Sanitation Marketing Scale-Up Program 3 (SMSU3)* implemented by iDE Cambodia operated in six rural provinces, and supported the promotion and sale of latrines, fecal sludge management and handwashing systems through private enterprises.
- *Beyond the Finish Line* leveraged recent decentralization in governance structures in Nepal to support inclusive, sustainable, and resilient water supply services and hygiene behavior change communications in households, schools, and health care facilities in two rural districts.

4.3 Data collection

Data was collected through structured enumerated interviews between October and December 2020, between waves in the COVID-19 pandemic. COVID protocols were followed in both as required. The WASH-GEM items (questions) used in the Phase 4 study can be found in Table B in [S1 Text](#).

A stratified multistage random sample design was used in both countries, with a balanced sample of women and men respondents. The sampling consisted of four stages: 1) selection of

programmatic sub-national locations, 2) purposeful selection of districts/municipalities to represent the breadth of variance in program catchment areas (using DHS data to identify areas with variations in WASH status, electricity, land ownership, poverty levels and female headed households), 3) random selection of communes/wards from lists of the program working areas, 4) random selection of villages. Two villages were selected in each commune/ward; one primary and one alternate if the goal of just over 80 respondents in a village was not feasible. At the village level, enumerators sought socio-economic variation where relevant in their selection of households. At the sub-national level, in Nepal both districts were selected: Sarlahi and Dailekh, and in Cambodia, a subset of three provinces (out of six) were selected to represent the breadth of variance in the program operational areas: Kampong Thom, Prey Veng, and Kandal. In Cambodia, the WASH-GEM tool was also used with male and female latrine business owners ($n = 21$), to investigate the impacts on WASH service providers, who are the primary touchpoint for iDE's programming.

The survey tool aimed to interview, wherever possible, a man and a woman from the same household—a dyad of the male primary decision-maker (amongst men) and the female primary decision-maker (amongst women). Challenges related to migration and individuals working outside the home led to the inclusion of 32% of participants who were interviewed as “singletons” in the exploratory and validation phases.

A two-day enumerator training was conducted and included detailed exploration of the WASH-GEM items and ethical considerations including distress protocols. Lessons learned from prior phases informed enumerator recruitment and training, as well as the process of data collection. Recruited enumerators were an equal number of women and men; and women enumerators interviewed women respondents and men interviewed men respondents. Data was collected through tablets or smartphones using Taroworks (Cambodia) and Akvo-Flow (Nepal) and stored in secure cloud-based data servers.

4.4 Data analysis

The empirical focus of the validation analysis sought to interrogate the validity and reliability of the measure using the available data from the validation data collection. In hindsight, additional data could have been collected to further strengthen this process of analysis. Nonetheless, the analysis procedures aimed to analyze the structural validity, construct validity, internal reliability and measurement invariance for the WASH-GEM.

Data analysis for the validation pilot was conducted in RStudio by the first author, supported by the second author. Sensemaking workshops with each in-country program team utilized item and theme insights to create action plans and leveraged contextual knowledge to triangulate and interpret the analysis.

Data was pooled from Cambodia and Nepal. Missing data was filled using imputed means calculated by gender and region, and a subset of data was used which removed two responses who self-identified as ‘other gender’, as with such a small sample meaningful conclusions could not be drawn. The data was also randomly split in half for analysis indicated by ‘Split a’ and ‘b’ in the subsequent documentation. Split a was used for the exploratory aspects of analysis and Split b for the confirmatory aspects. Tests of structural validity and invariance were only possible for themes with three items or more (nine WASH-GEM themes) and tests of internal reliability were only possible for themes with two or more items (15 WASH-GEM themes). As such, one theme which only had one-item, *Physical Health* was only tested for known-group construct validity.

The specific analysis procedures in the validation pilot are summarized in [Table 1](#) and described in the subsequent text.

Table 1. Summary of Phase 4 measure validation procedures.

Focus	Activity	Purpose	Procedure	Dataset	References
Structural Validity	Exploratory factor analysis (EFA) for Item Reduction	Explore the factor structure of the <i>nine</i> WASH-GEM themes with more than two items and explore structural validity.	Scree plots, parallel analysis, and exploratory factor analysis were used to determine the optimal number of factors for each theme. Exploratory factor analysis was also used for item reduction based on cross-loadings greater than 0.3 and pattern coefficients of less than 0.3 . Fit was assessed using the Tucker Lewis Index (TLI), and Diagonal Values Fit (DVF), Root Mean Square Error of Approximation (RMSEA) and Root Mean Square Residual (RMSR). Fit was deemed good with TLI >0.90 or DVF >0.90 and RMSR < 0.08.	Split a	[55]
Structural Validity	Confirmatory factor analysis (CFA)	Confirm the factor structure of the <i>nine</i> WASH-GEM themes with more than two items and clarify structural validity.	Confirmatory factor analysis was conducted for each theme to confirm structural validity and model fit. Fit was assessed using the Comparative Fit Index (CFI), TLI, RMSEA and (Standardized) Root Mean Square Residual (SRMR). Fit was deemed good with TLI >0.90 and CFI > 0.90 and RMSR < 0.08 or RMSEA < 0.08.	Split b	[55, 56]
Measurement invariance	Multiple-Group Confirmatory factor analysis (MGCFA)	Assess whether the <i>nine</i> WASH-GEM themes with more than two items hold across gender and country.	Conducted multiple-group confirmatory factor analysis (MGCFA) testing for configural, metric and scalar invariance across relevant themes. For configural model fit, the same thresholds were used from the CFA analysis. Invariance thresholds of $\Delta 0.01$ for CFI and $\Delta 0.03$ for RMSEA, with priority given to the CFI threshold.	Pooled	[27, 55, 56]
Internal Reliability	Internal consistency	Assess the internal consistency the of 15 WASH-GEM themes with more than one item.	Coefficients of internal consistency were calculated for each theme, to confirm reliability. Standardized Cronbach's Alpha, Spearman-Brown Statistic and Omega Total were used to assess consistency of unidimensional, two-item, and multidimensional themes respectively. A threshold of 0.70 for both coefficients was used, with 0.80 preferred.	Split b	[55, 57–59]
Construct validity and item reduction	Inter-item and intra-measure correlations	Evaluate if items have been allocated to the correct themes.	Calculated correlation coefficients (Pearson's r) between items and themes.	Pooled	[55]
Construct validity	Known-groups validity	Bivariate descriptive statistics for items, and all themes.	Quantified the associations between themes and other variables such as education, age, and wealth quintile. Calculated statistical significance (unpaired t-tests) and effect size (Cohen's d) for relevant co-factors by item, and theme.	Split b	[55]
Construct validity	Dyad analysis	Assess agreement between dyads with regard to household asset ownership and WASH access.	Assessed for difference in trends in dyad and non-dyad responses using correlation coefficients (Pearson's r).	Split b—Cambodia	[55]
Criterion validity	Criterion correlations	Assess agreement between the new Self-efficacy scale and General Self-Efficacy Scale	Assessed for broad alignment between the specific self-efficacy scale and the existing General Self-Efficacy scale (Pearson's r).	Pooled	[55]

<https://doi.org/10.1371/journal.pwat.0000233.t001>

4.4.1 Structural validity

Structural validity and reliability were interrogated in Phase 4 for the WASH-GEM through exploratory factor analysis, confirmatory factor analysis, measurement invariance, and tests of internal consistency for nine of the testable WASH-GEM themes. Structural validity was a critical aspect of the validation of the WASH-GEM due to its theoretical combination of indices (“*deliberately multidimensional aggregate measures*” [60]) and scales. It should be noted that the importance of such testing was not incorporated into the initial design of the measure and as such, seven of the WASH-GEM's themes did not include more than two items, a requirement for such testing. As such, the analysis was run on the nine testable themes.

Exploratory factor analysis (EFA) was conducted on the exploratory random split half of the validation dataset (Split a) for each of the nine testable WASH-GEM themes (with more than two items) to explore underlying factors and to refine the measure. EFA sought to identify low loading items and items with potential cross loading. EFA was conducted using the most appropriate factoring method or rotation for the theoretical makeup of each theme ('varimax' and principal factors for deliberately multidimensional aggregates and 'quartimax' with weighted least squares for scales). Additionally, detailed EFA and IRT was done in Phase 3 of this study (outside the scope of the article).

Confirmatory factor analysis (CFA) was conducted on the confirmatory random split half of the validation dataset (Split b) to test the factor structures identified through EFA, for each of the nine testable WASH-GEM themes. Two additional shortened themes were also tested for fit as these included items in which men and non-menstruating women were automatically scored fully (*Access to WASH* and *WASH Related Wellbeing*). All themes were tested using both the maximum likelihood and weighted least squares estimator.

Tests for measurement invariance were conducted on the nine testable WASH-GEM themes using multiple-group confirmatory factor analysis (MGCFA) between respondents by gender and country as configural (factor structure), metric (factor loadings) and scalar (mean values) invariance [56] using the confirmatory dataset. For the two themes which included menstrual hygiene management questions, it was not possible to test for gendered measurement invariance as all men and non-menstruating women were automatically given full scores. As such, invariance has been tested on a shortened version of these two scales. This indicates that comparison between women and men would rely on the shortened version rather than the full version (only comparable for women). It should be noted that the significant contextual differences between Cambodia and Nepal gives an opportunity to further test invariance in other contexts.

4.4.2 Internal consistency

Internal consistency of the WASH-GEM themes was calculated for each of the 15 testable themes using Cronbach's alpha, Spearman-Brown (for two-item themes) and McDonald's Omega Total [55, 57–59]. McDonald's Omega Total was particularly important for themes that were conceptualized as indices rather than scales, for example *Access to WASH*, where the unidimensional pre-requisite would not be met. When the internal consistency scores were greater than 0.70, the measure was considered reliable [61].

4.4.3 Internal construct validity

Internal construct validity was tested using the Split a dataset using: 1) inter-item and inter-measure correlations, 2) known-group validity, and 3) dyad-agreement analysis. Inter-item and intra-measure correlations (Person's r) were calculated to identify items for removal and ensure that items were correctly allocated to themes. Known-group validity was explored using t-tests, ANOVA, and the calculation of effect sizes (Cohen's d) for relevant themes against key groups including age, highest level of education, and wealth quintile. Notably, gender was not included in known-group analysis based on experiences in Phases 1–3 which highlighted that women may be more empowered than men, in particular amongst women-income earning homes and caste minorities. Lastly, agreement between paired dyads was also analyzed using Person's r for asset ownership, wealth quintile and WASH access items. This helped to examine the construct validity of access and ownership items. Cross-dyad identification numbers were only available for the Cambodia dataset and as such dyad agreement was not reviewed for Nepal.

4.4.4 Criterion validity

Within the validation dataset, one of the WASH-GEM themes—*Self Efficacy*, was also testable for criterion validity. Based on the experiences of cognitive testing and exploratory piloting, the team had decided to create a specific self-efficacy measure which was anticipated to have broad alignment with the General Self-efficacy Scale (GSE-6) [50]. Alignment was assessed through correlation (Pearson's r) between the existing and novel scales.

4.4.5 Scoring

Since its inception, the WASH-GEM's scoring methodology has been a critical area of collaborative debate and deliberation with an objective to identify the simplest, yet most robust, scoring approach for each theme.

To score the WASH-GEM themes for the empirical validation study, the team drew on the confirmed structural validity of each theme. For two-item themes, simple scoring was utilized. For multidimensional themes, sensitivity analysis was conducted against gender, age, education, and wealth quintile to determine if a simple sum instead of a weighted sum (based on factors) would produce broadly the same statistical results. Full details of scoring can be found in Table B in [S1 Text](#). This is in alignment with recent recommendations by Widaman & Revelle who argue based on empirical evidence that “*if sum scores are computed in exactly the same fashion in different studies, the results of the studies will be more easily compared than if some other, sample-specific method is used in each study to composite item scores*” [62]. Additionally, some of the themes were designed as measures of structural experiences rather than psychosocial measures, however they were still analyzed and scored through the same procedures. Scoring sensitivity analysis results can be found in Table G in [S1 Text](#).

It should be noted that previous publications related to the WASH-GEM [28, 37], were written prior to the decision to shift from domain-wise to thematic aggregation.

4.5 Ethical statement

The study was approved by University of Technology Sydney (UTS HREC REF NO. ETH18-2599-Project 17232), prior to the start of the research, covering all phases of the research. Formal informed consent was obtained by each participant prior to the start of the interview. A checklist ensured that enumerators discussed each key aspect of the study and ensured comprehension. Consent was given verbally and recorded digitally through a checkbox ticked by the enumerator.

As research on gender topics is sensitive and respondents can experience backlash, care was taken to mitigate risks for both male and female participants. For all rounds of piloting, this included 1) focused training for the enumeration teams, 2) informed consent from participants conducted privately to reduce risk that a respondent's decision to participate would be influenced by others, 3) private and separate interviews (gender-paired with the enumerator based on the respondent's self-identified gender), and 4) do-no-harm and distress protocols to respond to potential emotional responses during and after the interviews. The survey tool was purposefully designed to commence with less sensitive questions before moving into more challenging topics.

The study also piloted the use of a distress measure to identify and diagnose potential experiences of distress by respondents. This measure included four questions related to interview ease, stress, privacy, and comprehension to be completed by the enumerator at the end of the survey interview. The distress scale was also validated during Phases 4 and 5 of the study and results are anticipated in a forthcoming publication.

4.6 Limitations

This study must be interpreted through its limitations. Namely, constraints on appropriate validation approaches, complicated logistics due to COVID-19, and restraints on the extent to which the tool can explore inequalities related to sexual and gender minorities and individuals with disability.

The analysis presented in this paper is intended to explore the validity of a measure designed firstly with conceptual and practical considerations. As such, there were not opportunities to explore other validation procedures such as criterion validity and test-retest validity. Attempts to conduct criterion validity for some aspects of the measure (such as self-efficacy and relative autonomy) were dropped due to low performance of the externally validated components in cognitive testing (Phase 1). Additionally, test-retest was purposefully not conducted as part of a do-no-harm strategy. This was related to the sensitive nature of the survey tools in exploring potentially distressing topics as observed during Phases 1–3. Lastly, as the measure had already been refined by Phase 4, seven of the themes were already reduced to single or two-items—limiting opportunities for validation procedures. Within these confines, the team has, to the best of our abilities, aimed to examine the validity and reliability of the measure.

The third and fourth phases of the development of the WASH-GEM coincided with the COVID-19 pandemic and subsequent country-wide lockdowns in Cambodia and Nepal. All data collection was conducted during safe travel periods. Nonetheless, the pandemic did create several logistical challenges, such as conducting enumerator training and sensemaking workshops remotely through video conferencing. Despite these constraints, the team was able to collect and verify a full sample to ensure that the measure could be validated as required.

The WASH-GEM has not been tailored to address specific disability issues; it also currently focuses on cis-gender individuals in heterosexual relationships. These decisions were made in collaboration with organizations focusing on individuals with disability and sexual and gender minorities (SGMs) along with the pilot programs. The WASH-GEM's use of a binary definition of gender reflects the previous research on which the measure builds—which addresses roles, relationships, and distribution of power between cis-gender women and men—and acknowledges the practical challenges and protection risks associated with collecting and managing data involving SGMs. Two respondents in the validation study from Nepal self-identified as other gender and their responses have not been displayed in the validation study gender-segregated analysis. With only two individuals, responses are not assumed to be indicative of the wider experiences of SGMs and drawing attention to them is not aligned with the study's do-no-harm approach. Tailored tools for the inclusion of SGMs and of people with a disability would be a helpful evolution for the next stage of development of the WASH-GEM.

In forthcoming work, not the focus on this article, the WASH-GEM will be further tested and refined. In particular, the measure will be piloted as a diagnostic tool in four further countries (Ghana, Laos, Bhutan, and Bangladesh) and used to evaluate changing gender dynamics in Cambodia and Nepal through a baseline-endline analysis. Future work could also include test-retest analysis as well as tests of criterion or external validity as appropriate.

5. Results

Building on the phased development of the WASH-GEM, we now present the results of the Phase 4 validation pilot. We begin with socio-demographic characteristics of the full validation pilot sample before presenting results of the measure's validity and reliability. We conclude the results section by presenting thematic WASH-GEM scores. Deidentified data and analysis code can be found in [S1](#) and [S2](#) Data respectively.

5.1 Descriptive statistics

Descriptive characteristics of the respondents for the validation pilot are included in [Table 2](#) and explored by gender and country. The results indicate that a large proportion of respondents had relatively low levels of education. In particular, 77% of female respondents in Nepal had either never attended school or only attended through preschool, compared to 51% of men in Nepal. Most households (53%) comprised between four and six members who regularly lived within the home. This number excluded family members who had migrated. 93% of respondents were married or living together with their partner. The Cambodian respondents were all ethnically Khmer and almost all Buddhist, which is consistent with the population generally. The ethnicities of respondents in Nepal were also consistent with the general population and almost all respondents were Hindu. Approximately a quarter of the households from Cambodia and Nepal included a member with a disability based on the Washington Group Short Set on Functioning questions. Two respondents in Nepal self-identified as other gender and have been excluded from demographic characteristics. Approximately 70% of respondents in Nepal (73% for women, 69% for men) had experienced a change related to their WASH situation within the last year reflecting the recent Government-led open-defecation-free campaigns which had taken place in two provinces. Within both countries, some 16% of respondents were engaged in some way as change agents within programs (such as part of a local sanitation business, WASH committee members or community mobilizers).

5.2 Reliability and validity

Structural validity and reliability were interrogated in Phase 4 for the WASH-GEM through exploratory factor analysis, confirmatory factor analysis, measurement invariance, and tests of internal consistency for each testable theme of the WASH-GEM.

5.2.1 Exploratory factor analysis and inter-item correlations

Through the EFA and inter-item/theme correlations, a total of 19 items were removed from the *Critical Consciousness*, *Gender Norms in WASH*, and *Mobility* themes because of low pattern coefficients, low correlations and feedback from the project teams related to the understandability of the items. After refinement, all exploratory factor models showed overall good fit (TLI 0.934–1.000 with SRMR 0.01–0.02 and pattern coefficients ranging from 0.29–0.99). The models all indicated between one and three factors. The *Access to WASH* theme did not have any degrees of freedom and as such TLI was not estimated instead relying on DVF of 0.998. Full details of the final EFA model fit are included in Table D in [S1 Text](#). The final list of items can be found in Table B in [S1 Text](#).

5.2.2 Confirmatory factor analysis (CFA)

All of the nine testable themes demonstrated overall good fit through CFA (CFI 0.988–1.000, TLI 0.936–1.000, RMSEA 0.000–0.063, and SRMR 0.009–0.041 through weighted least squares estimation). Detailed CFA results for each relevant theme are included in Table E in [S1 Text](#). For the two themes which included menstrual hygiene items - *Access to WASH* and *WASH Related Wellbeing*- the shorted models which excluded the menstrual hygiene items (for gender comparison), also demonstrated good fit.

5.2.3 Internal consistency

Internal consistency was calculated for each of the 15 testable themes using Cronbach's alpha, Spearman-Brown and Omega total as shown in [Table 3](#). Three of the two-item themes did not

Table 2. Sociodemographic characteristics of respondents (n = 3054). Counts are of individuals unless indicated as a count of households with “h =”.

Sociodemographic characteristics	Overall	Cambodia		Nepal	
		Women	Men	Women	Men
Sample					
Individuals	3054	748	748	801	757
Households	h = 2011	h = 606	h = 606	h = 413	h = 387
Dyad Response					
Individuals	2086 (68.3%)	285 (38.1%)	284 (38.0%)	776 (96.9%)	741 (97.9%)
Households	h = 1043	h = 142	h = 142	h = 388	h = 370
Region					
Kampong Thom	502 (16.4%)	251 (33.6%)	251 (33.6%)	-	-
Kandal	498 (16.3%)	249 (33.3%)	249 (33.3%)	-	-
Prey Veng	496 (16.2%)	248 (33.2%)	248 (33.2%)	-	-
Dailekh	626 (20.5%)	-	-	321 (40.1%)	305 (40.3%)
Sarlahi	932 (30.5%)	-	-	480 (59.9%)	452 (59.7%)
Wealth Quintile					
Lowest	681 (22.3%)	168 (22.5%)	139 (18.6%)	197 (24.6%)	177 (23.4%)
Second	571 (18.7%)	151 (20.2%)	143 (19.1%)	129 (16.1%)	148 (19.6%)
Middle	712 (23.3%)	156 (20.9%)	147 (19.7%)	215 (26.8%)	194 (25.6%)
Fourth	489 (16.0%)	152 (20.3%)	141 (18.9%)	103 (12.9%)	93 (12.3%)
Highest	601 (19.7%)	121 (16.2%)	178 (23.8%)	157 (19.6%)	145 (19.2%)
Highest Level of Education					
Preschool or below	1264 (41.4%)	143 (19.1%)	120 (16.0%)	617 (77.0%)	384 (50.7%)
Primary	1083 (35.5%)	464 (62.0%)	346 (46.3%)	94 (11.7%)	179 (23.6%)
Secondary or higher	707 (23.2%)	141 (18.9%)	282 (37.7%)	90 (11.2%)	194 (25.6%)
Number of Household Members					
1–3	498 (16.3%)	168 (22.5%)	149 (19.9%)	92 (11.5%)	89 (11.8%)
4–6	1629 (53.3%)	438 (58.6%)	453 (60.6%)	373 (46.6%)	365 (48.2%)
7–9	712 (23.3%)	118 (15.8%)	129 (17.2%)	248 (31.0%)	217 (28.7%)
10+	206 (6.7%)	24 (3.2%)	16 (2.1%)	85 (10.6%)	81 (10.7%)
NA	9 (0.3%)	0 (0%)	1 (0.1%)	3 (0.4%)	5 (0.7%)
Age					
18–30	431 (14.1%)	116 (15.5%)	49 (6.6%)	171 (21.3%)	95 (12.5%)
31–49	1448 (47.4%)	349 (46.7%)	336 (44.9%)	423 (52.8%)	340 (44.9%)
50+	1175 (38.5%)	283 (37.8%)	363 (48.5%)	207 (25.8%)	322 (42.5%)
Marital Status					
Divorced / separated	34 (1.1%)	28 (3.7%)	5 (0.7%)	0 (0%)	1 (0.1%)
Married or living together	2841 (93.0%)	609 (81.4%)	701 (93.7%)	789 (98.5%)	742 (98.0%)
Migrated spouse	27 (0.9%)	25 (3.3%)	2 (0.3%)	0 (0%)	0 (0%)
Unmarried	62 (2.0%)	25 (3.3%)	14 (1.9%)	10 (1.2%)	13 (1.7%)
Widowed	90 (2.9%)	61 (8.2%)	26 (3.5%)	2 (0.2%)	1 (0.1%)
Ethnicity/Caste					
Ethnic Khmer	1496 (49.0%)	748 (100%)	748 (100%)	-	-
Brahmin / Chhetri	401 (13.1%)	-	-	204 (25.5%)	197 (26.0%)
Dalits	497 (16.3%)	-	-	271 (33.8%)	226 (29.9%)
Janjati	257 (8.4%)	-	-	129 (16.1%)	128 (16.9%)
Madhesi	262 (8.6%)	-	-	146 (18.2%)	116 (15.3%)
Muslim	99 (3.2%)	-	-	48 (6.0%)	51 (6.7%)
Other	42 (1.4%)	-	-	3 (0.4%)	39 (5.2%)

(Continued)

Table 2. (Continued)

Sociodemographic characteristics	Overall	Cambodia		Nepal	
		Women	Men	Women	Men
Religion					
Buddhist	1496 (49.0%)	746 (99.7%)	747 (99.9%)	1 (0.1%)	2 (0.3%)
Christian	13 (0.4%)	2 (0.3%)	1 (0.1%)	8 (1.0%)	2 (0.3%)
Hindu	1446 (47.3%)	0 (0%)	0 (0%)	744 (92.9%)	702 (92.7%)
Muslim/Islamic	99 (3.2%)	0 (0%)	0 (0%)	48 (6.0%)	51 (6.7%)
Disability in household	680 (22.3%)	121 (16.2%)	182 (24.3%)	211 (26.3%)	166 (21.9%)
WASH-specific Engagements					
Program involvement	491 (16.1%)	11 (1.5%)	10 (1.3%)	279 (34.8%)	191 (25.2%)
Changes in WASH situation in last 12 months	1540 (50.4%)	177 (23.7%)	252 (33.7%)	586 (73.2%)	525 (69.4%)
WASH skill development	788 (25.8%)	367 (49.1%)	97 (13.0%)	202 (25.2%)	122 (16.1%)
Income related to WASH	312 (10.2%)	65 (8.7%)	70 (9.4%)	114 (14.2%)	63 (8.3%)

<https://doi.org/10.1371/journal.pwat.0000233.t002>

demonstrate internal reliability (*Control over Resources, Gender Norms and Mental Health*). Four themes demonstrated *high* internal reliability with scores over 0.70 including: the two-item *Household Respect* and *Life Satisfaction* themes; and the multidimensional *Relative Autonomy* and *Collective Agency* themes. Eight themes demonstrated *very high* internal reliability with scores over 0.80 including: *Access to WASH, Decision Making in WASH, Awareness of Patriarchal Norms (Critical Consciousness), Gender Norms in WASH, Self-efficacy* and *WASH Related Wellbeing* themes and the two-item *Social Capital* theme.

5.2.4 Measurement invariance

Table 4 includes a summary of the Comparative Fit Index (CFI) scores for each of the nine measures tested for gender and country invariance. Full results are provided in Table F in [S1 Text](#). Configural models had good fit (CFI>0.95) for all nine tested measures across both gender and country. Full invariance was found for the *Self-efficacy* and *Mobility* measures across both gender and country. Gendered invariance was found in all themes except the *Decision Making in WASH, Collective Agency* and *Gender Norms in WASH*. However, gendered metric and scalar invariance was found for these three themes when using a dataset of only community members (excluding the 21 business owners in Cambodia indicating how unique these business owners are from the general population). Metric country-based invariance was not found in the *Relative Autonomy, Collective Agency* measure, and *Perceptions of WASH Wellbeing*. Additionally, the *Critical Consciousness (Awareness of Patriarchal Norms)* theme did not indicate country-based scalar invariance indicating just how strong these norms differ across contexts.

These results indicate that the measures are valuable for comparing results between gender and a sub-set are currently effective for cross-country comparison.

5.3 Construct and criterion validity

Construct validity was tested using the Split b dataset using both known-group validity and dyad-agreement analysis.

5.3.1 Known group and criterion validity

Known-group validity was explored using t-tests, ANOVA, and the calculation of effect sizes (Cohen's-d) for relevant themes against key known groups including region/province, age,

Table 3. Internal validation results (Phase 4b) for WASH-GEM themes with more than one item.

Domain	Theme	Factors / Items	Standardized alpha (α)	Spearman-Brown	Omega Total	Interpreted Reliability
Resources	Access to WASH+	3 factors, 6 items	0.53	-	0.80	Very High
	Control over Resources*	1 factor, 2 items	0.15	0.15	-	Very Low
	Social Capital	1 factor, 2 items	0.85	0.85	-	Very High
Agency	Decision Making in WASH	2 factors, 5 items	0.68	-	0.80	Very High
	Household Respect	1 factor, 2 items	0.81	0.78	-	High
	Adapted Relative Autonomy	2 factors, 5 items	0.67	-	0.79	High
	Self-efficacy	2 factors, 5 items	0.85	-	0.89	Very High
	Collective Agency	2 factors, 5 items	0.62	-	0.79	High
Critical Consciousness	Awareness of Patriarchal Norms	2 factors, 7 items	0.86	-	0.90	Very High
Structures	Gender Norms in WASH*	3 factors, 6 items	0.69	-	0.86	Very High
	Gender Norms^	1 factor, 2 items	0.30	0.43	-	Very Low
	Mobility	2 factors, 5 items	0.72	-	0.83	Very High
Wellbeing	WASH related Wellbeing+	3 factors, 8 items	0.79	-	0.88	Very High
	Mental Health*	1 factor, 2 items	0.32	0.33	-	Very Low
	Life Satisfaction	1 factor, 2 items	0.74	0.73	-	High

- Alpha was calculated for all 9 themes, however recognizing that unidimensionality is frequently noted as a requirement.
- Bolded values are > 0.70
- Plus (+) indicates themes which included questions focused on menstruation.
- Asterixis (*) indicates that work is required to strengthen these themes.
- Dash (-) indicates that the test was not appropriate.
- Carrot (^) indicates that the theme has been strengthened in Phase 5

<https://doi.org/10.1371/journal.pwat.0000233.t003>

Table 4. Measurement invariance summary as CFI scores of internally validated measures of the WASH-GEM by country and gender.

Theme	Configural	Metric	Scalar
Access to WASH	Country 0.988 Gender+ 0.975	Country 0.986 (yes) Gender+ 0.968 (yes)	Country 0.746 (no) Gender+ 0.733 (no)
Decision Making in WASH	Country 0.981 Gender 0.968	Country 0.947 (no) Gender 0.949 (no)*	- _*
Relative Autonomy	Country 0.971 Gender 0.971	Country 0.924 (no) Gender 0.985 (yes)	- Gender 0.982 (yes)
Self-efficacy	Country 0.998 Gender 0.998	Country 0.998 (yes) Gender 0.996 (yes)	Country 0.994 (yes) Gender 0.995 (yes)
Collective Agency	Country 0.980 Gender 0.989	Country 0.954 (no) Gender 0.969 (no)*	- _*
Awareness of Patriarchal Norms	Country 0.996 Gender 0.990	Country 0.983 (yes) Gender 0.983 (yes)	Country 0.923 (no) Gender 0.978 (yes)
Mobility	Country 0.996 Gender 0.982	Country 0.990 (yes) Gender 0.991 (yes)	Country 0.990 (yes) Gender 0.986 (yes)
Gender Norms in WASH	Country 1.000 Gender 0.975	Country 1.000 (yes) Gender 0.998 (no)*	Country 0.962 (no) _*
Perceptions of WASH related Wellbeing	Country 0.991 Gender+ 0.984	Country 0.961 (no) Gender 0.978+ (yes)	- Gender 0.975+ (yes)

- Invariance is indicated through the bracketed yes and no responses.
- Dashes indicate that tests are irrelevant.
- Asterisks (*) indicates that the test was valid using a subset of the data which did not include Latrine Business Owners (LBOs) in Cambodia
- Plus sign (+) indicated that a shortened version of the theme was tested

<https://doi.org/10.1371/journal.pwat.0000233.t004>

highest level of education, and wealth quintile. Themes related to wealth including *Access to WASH*, *Control of Resources*, *Collective Agency*, *WASH Wellbeing* and *Mental Health* all performed as hypothesized. Themes related to education including *WASH Related Decision Making*, *Self-efficacy*, *Collective Agency*, *Critical Consciousness*, *Mobility*, *Mental Health*, and *Life Satisfaction* performed as hypothesized. Themes related to age including *Household Influence*, *Self-efficacy*, and *Physical Health* all also performed as hypothesized. Additionally, regions which were more remote or harder to reach were also associated with lower scores for nearly all themes in Nepal. Such analysis was more complex for Cambodia due to more regional variations. Notably, aspects such as ethnicity, education, wealth, and age have more impact than gender in several of the themes, indicating that a women-focus on empowerment and equality is insufficient in understanding differences. Detailed results are included in Figs A through E in [S1 Text](#).

For the *Self-efficacy* scale, criterion validity was assessed against the GSE-6 with Pearson's r values ranging from 0.54 for Cambodian women to 0.70 for Nepali women ($p < 0.001$). While this does not indicate perfect agreement, it is in alignment with the hypothesized direction and extent of statistically significant correlation.

5.3.2 Dyad pair agreement

Agreement between paired dyads related to household asset ownership and access to WASH was analyzed using Pearson's r for the Cambodian dataset. The data was not analyzed for Nepal as dyad pair identification numbers were recorded as phone numbers in Nepal, which led to incorrect coding of households without access to a phone number. This process has been refined for the fifth phase of data collection. Household asset ownership as indicated through wealth quintile showed high agreement between male and female dyad pairs ($r = 0.72$, $p < 0.001$). Notably, the shortened *Access to WASH* theme was not aligned between women and men—highlighting the importance of intrahousehold and individual measures of WASH access ($r = 0.3$, $p < 0.001$). At the item level, toilet access was the most correlated ($r = 0.42$, $p < 0.001$), however women identified that they have higher toilet access, likely due to the lack of toilets in workplaces and fields for men working outside the home. Women who had less mobility were more likely to have access to a household toilet when needed. Water conflict was the least statistically correlated within dyads; women and men within dyad pairs did not agree whether there was conflict in accessing drinking water ($r = 0.11$, $p < 0.01$). Dyad agreement on access to soap was also poorly correlated with women indicating higher levels of soap access.

5.4 Scoring

In an effort to keep the WASH-GEM simple for future implementation, sample-based scoring processes were not adopted, rather themes were either scored using a simple sum or weighted sum approach. Thematic scores are displayed as scores between zero and one by gender and country, with distributions represented in the boxplots of [Fig 3](#). Comparing responses within countries by gender, there were statistically significant differences in means in most of the themes. The largest valid gender differences across both countries were seen in the *Access to WASH*, *Mobility*, *Collective Agency* and *Life Satisfaction* themes, with women having significantly lower scores than men.

6. Discussion

6.1 Summary and interpretation

Within the constraints of the validation study dataset, our analysis explored the structural and conceptual validity, internal consistency, and measurement invariance of the testable

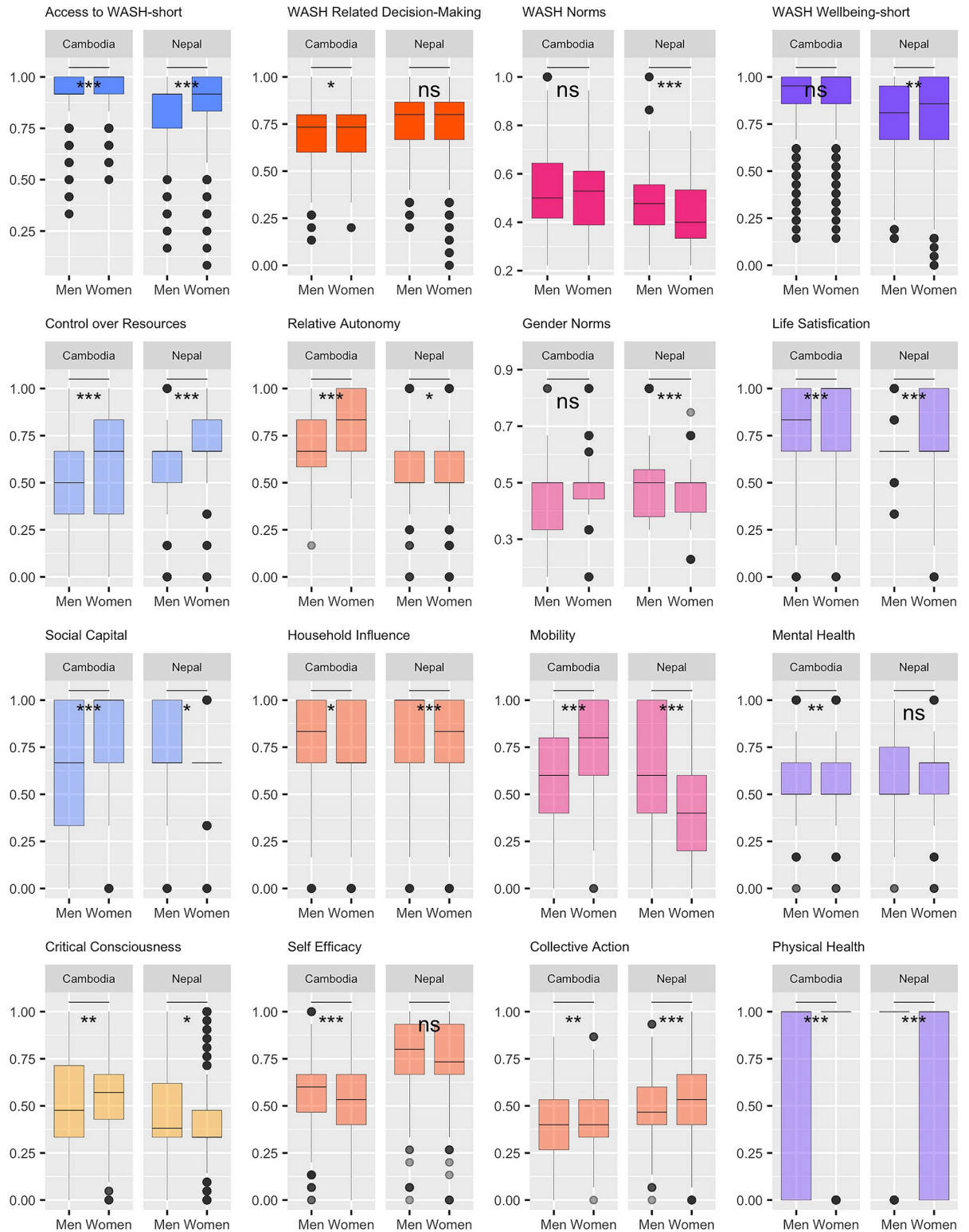


Fig 3. Theme scores by country and gender with statistical significance (Significance levels: * $p < 0.05$, ** $p < 0.01$, * $p < 0.001$).**

<https://doi.org/10.1371/journal.pwat.0000233.g003>

Table 5. Summary interpretation for each measure, ordered by reliability and validity.

Measure	Known-Groups Validity	Internal Reliability	Structural Validity	Country Invariance	Gender Invariance
1. Self-efficacy	Yes	Very High	Yes	Scalar	Scalar
2. Mobility	Yes	Very High	Yes	Scalar	Scalar
3. Critical Consciousness	Yes	Very High	Yes	Metric	Scalar
4. Relative Autonomy	Yes	High	Yes	Configural	Scalar
5. Access to WASH	Yes	Very High	Yes	Metric	Metric (short)
6. WASH related Wellbeing	Yes	Very High	Yes	Configural	Metric (short)
7. Gender Norms in WASH	Yes	Very High	Yes	Metric	Configural
8. Decision Making in WASH	Yes	Very High	Yes	Configural	Configural
9. Collective Agency	Yes	High	Yes	Configural	Configural
10. Social Capital	Yes	Very High	-	-	-
11. Life Satisfaction	Yes	High	-	-	-
12. Household Respect	Yes	High	-	-	-
13. Gender Norms	Yes	Very Low	-	-	-
14. Control over Resources	Yes	Very Low	-	-	-
15. Mental Health	Yes	Very Low	-	-	-
16. Physical Health	Yes	n/a	-	-	-

<https://doi.org/10.1371/journal.pwat.0000233.t005>

WASH-GEM themes. A summary interpretation is found in [Table 5](#). All sixteen themes exhibited construct validity as assessed through known-group analyses and the *Self-efficacy* theme (only theme tested) showed good criterion validity. Nine themes (all which were testable) exhibited structural validity as assessed through EFA and CFA and demonstrated configural invariance. Two of these themes demonstrated scalar invariance by country and four also demonstrated scalar invariance by gender. In sum, two of our measures perform very well (*Mobility* and *Self-efficacy*), seven of our measures perform well (*Access to WASH*, *Decision Making in WASH*, *Relative Autonomy*, *Collective Agency*, *Critical Consciousness*, *Gender Norms in WASH* and *WASH Related Wellbeing*), three perform well as two-item measures and require further criterion validation (*Social Capital* and *Life Satisfaction and Household Respect*), and four measures require further strengthening (*Control over Resources*, *Gender Norms*, *Mental Health* and *Physical Health*).

Based on our findings, only some of the measures indicate scalar invariance across both gender and country—a result anticipated due to the strong contextual understanding of gender equality and empowerment not only across geographies, but also between women and men. While in our results we have presented gender comparisons for each of the themes, the measurement invariance assessment indicates that the most valid gender comparisons are the *Relative Autonomy*, *Self-efficacy*, *Critical Consciousness* and *Mobility* themes. Country comparisons are also most valid for *Self-efficacy* and *Mobility*. Nonetheless, we still recommend that teams conduct gender and geographic comparisons for all the themes, noting that there is still value in exploring differences in conceptualization even if statistically valid inferences are not possible. Lastly, for the measures initially conceptualized as indices (notably the four WASH related themes), their measurement invariance is less of a concern as the tool was not necessarily designed to achieve invariance.

6.2 Implications for research and practice

The WASH-GEM has been designed as a multi-purpose diagnostic and an evaluation tool to be used across the WASH program cycle. In the WASH-GEM validation pilot, the tool was

used in a diagnostic, one-point-in time application, to explore gender dynamics in the different themes, and the extent to which gender equality more broadly was correlated with equality within the WASH sector in selected provinces in Cambodia and Nepal.

The 12 validated themes can be used individually or included in a wider summary of equality within the WASH-GEM. For the four non-validated themes (*Gender Norms*, *Control over Resources*, *Mental Health* and *Physical Health*), future work can help to refine and strengthen them, nonetheless they remain useful components in visualizing a gender equality context for programs and interventions. Importantly, seven of the validated themes from the Beyond-WASH part of the tool may be useful for other programmatic measurement outside of the WASH sector. These include *Social Capital*, *Self-efficacy*, *Collective Agency*, *Critical Consciousness*, *Mobility*, and *Life Satisfaction*.

While other tools do exist for some of our validated themes, some were purposely not included in the validation phase of this research due to practical (length or poor performance during cognitive interviews) or ethical reasons, others were not identified during our initial scoping. It is important to note that the WASH-GEM was initially conceptualized as an index and as such the integration of existing tools was less of a priority during the scoping stage. Only once the validation pilot was completed, was it clear that the best approach would be a collection of thematic measures rather than a single index.

In the next iteration of the project, we anticipate comparing our validated WASH-GEM themes with further existing tools (such as the *Community Social Capital (CSC-6)* [63] measure, *Satisfaction with Life Scale (SWLS-5)* [64] and *Freedom of Movement Index* [65]) that were not excluded in the initial rapid piloting and cognitive interviews. Additionally, the areas of the WASH-GEM which were not validated in this study require further refinement including *Control over Resources*, *Gender Norms*, *Mental Health* and *Physical Health*. A subsequent iteration of the *Gender Norms* measure in Phase 5 (forthcoming) has already showed strong internal consistency and structural validity. To re-design the other three themes, the team will need to go back to the piloting stage and review other existing tools and items that show promise from other work. Lastly, we will continue to explore opportunities to add additional validated tools to the WASH-GEM toolbox. Ultimately the WASH-GEM aims to be a set of measurement tools curated for WASH practitioners to explore the impact of improved WASH on gender equality at home and in the community.

Now to several cautions. Gender equality is a complex and dynamic phenomenon and only some aspects will be captured in any quantitative tool. The WASH-GEM has also highlighted that inequalities are not limited to gender, but are intersectional. Our data suggests that men also experience disempowerment related to social standing, ethnicity, religion, caste, and financial standing. As such we cannot recommend comparisons based on gender with a starting assumption that men have higher levels of equality and empowerment.

These complexities can lead to misapplication and misuse, which need to be managed. However risks are a companion to the benefits of having new quantitative insights, rather than a problem that can be eliminated. Given the limitations of what can be understood through a quantitative approach, we anticipate the tool's application as part of a suite of methodological approaches, including qualitative approaches to understand gender equality dynamics. We discourage its stand-alone use. In addition, we have developed procedures to be used in conjunction with the tool. These include training modules on the use of the WASH-GEM to support planning, data collection, analysis, and sensemaking processes to bring the measure's use from a potentially extractive process to a transformative one. Our pilots in Cambodia and Nepal have also highlighted transformative outcomes for the teams in applying the WASH-GEM in their programs. This has been articulated by one team member in Cambodia as 'the learning

value' for the teams deploying the WASH-GEM, due to the type of thinking these questions and concepts provoke.

The tool presented in this paper is one iteration of an evolving measurement approach, and as such there are a variety of opportunities to interrogate and further strengthen the WASH-GEM. As a program evaluation tool, there is scope to test the measure with comparative base- and end-line responses from a program catchment. From a measure validity perspective, test-retest reliability, enumerator 'inter-rater' reliability, and further context explorations of measurement invariance remain important to further refine and strengthen the tool. Lastly, additional data has been collected as part of Phase 5 in three more study countries (Bhutan, Laos, Ghana) and a second round of data collection has been conducted in both Cambodia and Nepal. This paper focuses on the validation data set from Phase 4, but additional analysis and piloting would help to explore the WASH-GEM's relevance and empirical validity in other geographic, programmatic, and social contexts.

7. Conclusions

Building on conceptual and practical foundations of the WASH-GEM's development, this article evaluated the empirical validity of the WASH-GEM's themes with a validation dataset from Cambodia and Nepal. Twelve of the themes have been validated and are recommended for use as individual measures in WASH programs, of which seven are appropriate for non-WASH contexts. For the four non-validated themes, their inclusion within the WASH-GEM is still valuable to provide insight into a wider conceptual framework of (gender) equality building on the capability approach. This novel approach to investigate the gender-transformative potential of WASH interventions is only an iteration and future studies will aim to strengthen the conceptual, evaluative, and practical considerations of the tool.

Notably, the team began from a conceptual foundation followed by analysis of the measure's empirical validity—in other words, we navigated and sought an appropriate balance between conceptual rigor and empirical validity. As such, items were not selected to ensure the most robust empirical model, but rather crafted to provide the most value for program teams based on a conceptually defensible model of gender equality as relevant for WASH. Most importantly, this work does not seek to supplant other gender and WASH measurement initiatives, rather to expand the sector's toolbox in the pursuit of tools and practices to explore, diagnose and evaluate change.

Supporting information

S1 Text. Further information on the development and validation of the WASH-GEM. This text file includes background information on the phased development of the WASH-GEM (Table A) and the Phase 4 tested items (Table B). Additionally, the file includes details of the factor analyses (Tables C-E), measurement invariance (Table F), scoring sensitivity analysis (Table G), and known-group analysis (Figs A-E).
(DOCX)

S1 Data.
(CSV)

S2 Data. Analysis code.
(RMD)

Acknowledgments

Special thanks to Jo Crawford and Melissa Meinhart from the International Women's Development Agency for their insightful edits and comments, helping to strengthen the gender

equality perspective of this paper. Acknowledgement is also made to the many participants in Cambodia and Nepal who shared their time, insights, and experiences. The SNV Beyond the Finish Line Team includes: Min Prasad Basnet, Ratan Budhathoki, Aastha Chhetri, Harishova Gurung, Gabrielle Halcrow, and Ambika Yadav.

Author Contributions

Conceptualization: Chris Nicoletti, Juliet Willetts.

Data curation: Jess MacArthur.

Formal analysis: Jess MacArthur, Tyler Kozole.

Funding acquisition: Rachel P. Chase, Juliet Willetts.

Investigation: Diana Gonzalez, Tyler Kozole, Veasna Toeur.

Methodology: Rachel P. Chase, Chris Nicoletti, Veasna Toeur, Juliet Willetts.

Project administration: Diana Gonzalez, Tyler Kozole, Veasna Toeur, Juliet Willetts.

Supervision: Veasna Toeur, Juliet Willetts.

Validation: Jess MacArthur, Rachel P. Chase, Chris Nicoletti.

Visualization: Jess MacArthur.

Writing – original draft: Jess MacArthur.

Writing – review & editing: Diana Gonzalez, Tyler Kozole, Veasna Toeur, Juliet Willetts.

References

1. Willetts J, Halcrow G, Carrard N, Rowland C, Crawford J. Addressing two critical MDGs together: Gender in water, sanitation and hygiene initiatives. *Pac Econ Bull.* 2010; 25: 211–221. Available from: <https://openresearch-repository.anu.edu.au/handle/1885/30589>
2. Cavill S, Francis N, Grant M, Huggett C, Leahy C, Leong L, et al. A call to action: Organizational, professional, and personal change for gender transformative WASH programming. *Waterlines.* 2020; 39: 219–237. <https://doi.org/10.3362/1756-3488.20-00004>
3. MacArthur J, Carrard N, Willetts J. WASH and Gender: A critical review of the literature and implications for gender-transformative WASH research. *J Water Sanit Hyg Dev.* 2020; 10: 818–827. <https://doi.org/10.2166/washdev.2020.232>
4. MacArthur J, Carrard N, Mott J, Raetz S, Siscawati M, Willetts J. Gender equality approaches in water, sanitation, and hygiene programs: Towards gender-transformative practice. *Front. Water.* 2023; 5: 1090002. <https://doi.org/10.3389/frwa.2023.1090002>
5. Willetts J, Carrard N, Crawford J, Rowland C, Halcrow G. Working from strengths to assess changes in gender equality. *Dev Pract.* 2013; 23: 991–1006. <https://doi.org/10.1080/09614524.2013.840564>
6. Leahy C, Winterford K, Nghiem T, Kelleher J, Leong L, Willetts J. Transforming gender relations through water, sanitation, and hygiene programming and monitoring in Vietnam. *Gend Dev.* 2017; 25: 283–301. <https://doi.org/10.1080/13552074.2017.1331530>
7. Assefa GM, Sherif S, Sluijs J, Kuijpers M, Chaka T, Solomon A, et al. Gender equality and social inclusion in relation to water, sanitation and hygiene in the Oromia region of Ethiopia. *Int J Environ Res Public Health.* 2021; 18. <https://doi.org/10.3390/ijerph18084281> PMID: 33920729
8. Cairns MR, Workman CL, Tandon I. Gender mainstreaming and water development projects: analyzing unexpected enviro-social impacts in Bolivia, India, and Lesotho. *Gend Place Cult.* 2017; 24: 325–342. <https://doi.org/10.1080/0966369X.2017.1314945>
9. Soeters S, Siscawati M, Ratnasari, Anggriani SNailah, Willetts J. Gender equality in the government water, sanitation, and hygiene workforce in Indonesia: An analysis through the Gender at Work framework. *Development Studies Research.* 2021; 8: 280–293. <https://doi.org/10.1080/21665095.2021.1978300>
10. Grant M, Soeters S, Bunthoeun IV V., Willetts J. Rural piped-water enterprises in Cambodia: A pathway to women's empowerment? *Water (Switzerland).* 2019; 11: 1–18. <https://doi.org/10.3390/w11122541>

11. Indarti N, Rostiani R, Megaw T, Willetts J. Women's involvement in economic opportunities in water, sanitation and hygiene (WASH) in Indonesia: Examining personal experiences and potential for empowerment. *Development Studies Research*. 2019; 6: 76–91. <https://doi.org/10.1080/21665095.2019.1604149>
12. Asaolu IO, Alaofè H, Gunn JKL, Adu AK, Monroy AJ, Ehiri JE, et al. Measuring Women's Empowerment in Sub-Saharan Africa: Exploratory and Confirmatory Factor Analyses of the Demographic and Health Surveys. *Front Psychol*. 2018;9. <https://doi.org/10.3389/fpsyg.2018.00009> PMID: 29410642
13. Ewerling F, Lynch JW, Victora CG, van Eerdewijk A, Tyszler M, Barros AJD. The SWPER index for women's empowerment in Africa: Development and validation of an index based on survey data. *Lancet Glob Health*. 2017; 5: e916–e923. [https://doi.org/10.1016/S2214-109X\(17\)30292-9](https://doi.org/10.1016/S2214-109X(17)30292-9) PMID: 28755895
14. Miedema SS, Haardörfer R, Girard AW, Yount KM. Women's empowerment in East Africa: Development of a cross-country comparable measure. *World Dev*. 2018; 110: 453–464. <https://doi.org/10.1016/j.worlddev.2018.05.031>
15. Pratley P, Sandberg JF. Refining the conceptualization and measurement of women's empowerment in sub-Saharan Africa using data from the 2013 Nigerian Demographic and Health Survey. *Soc Indic Res*. 2018; 140: 777–793. <https://doi.org/10.1007/s11205-017-1811-1>
16. Alkire S, Meinzen-Dick R, Peterman A, Quisumbing AR, Seymour G, Vaz A. The Women's Empowerment in Agriculture Index. *World Dev*. 2013. <https://doi.org/10.1016/j.worlddev.2013.02.005> PMID: 25125791
17. Malapit H, Pinkstaff C, Sproule K, Kovarik C, Quisumbing A, Meinzen-Dick R. The Abbreviated Women's Empowerment in Agriculture Index (A-WEAI). Washington, D.C. IFPRI; 2017. Discussion Paper 01647. Available from: <http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/131231>
18. Yount KM, Cheong YF, Maxwell L, Heckert J, Martinez EM, Seymour G. Measurement properties of the project-level Women's Empowerment in Agriculture Index. *World Dev*. 2019; 124: 104639. <https://doi.org/10.1016/j.worlddev.2019.104639> PMID: 31798204
19. Galiè A, Teufel N, Korir L, Baltenweck I, Webb Girard A, Dominguez-Salas P, et al. The Women's Empowerment in Livestock Index. *Soc Indic Res*. 2019; 142: 799–825. <https://doi.org/10.1007/s11205-018-1934-z>
20. Cole SM, McDougall C, Kaminski AM, Kefi AS, Chilala A, Chisule G. Postharvest fish losses and unequal gender relations: Drivers of the social-ecological trap in the Barotse Floodplain fishery, Zambia. *Ecol Soc*. 2018;23. <https://doi.org/10.5751/ES-09950-230218>
21. Narayanan S, Lentz E, Fontana M, De A, Kulkarni B. Developing the Women's Empowerment in Nutrition Index in two states of India. *Food Policy*. 2019; 89: 101780. <https://doi.org/10.1016/j.foodpol.2019.101780>
22. Burney J, Alaofè H, Naylor R, Taren D. Impact of a rural solar electrification project on the level and structure of women's empowerment. *Environmental Research Letters*. 2017;12. <https://doi.org/10.1088/1748-9326/aa7f38>
23. Dickin S, Bisung E, Nansi J, Charles K. Empowerment in water, sanitation and hygiene index. *World Dev*. 2021; 137: 105158. <https://doi.org/10.1016/j.worlddev.2020.105158>
24. Stevenson EGJ, Ambelu A, Caruso BA, Tesfaye Y, Freeman MC. Community water improvement, household water insecurity, and women's psychological distress: An intervention and control study in Ethiopia. *PLoS One*. 2016;11. <https://doi.org/10.1371/journal.pone.0153432> PMID: 27124391
25. Caruso BA, Clasen T, Yount KM, Cooper HLF, Hadley C, Haardörfer R. Assessing women's negative sanitation experiences and concerns: The development of a novel sanitation insecurity measure. *Int J Environ Res Public Health*. 2017;14. <https://doi.org/10.3390/ijerph15010014> PMID: 29295470
26. Sinharoy SS, Conrad A, Patrick M, Mcmanus S, Caruso BA. Development and validation protocol for an instrument to measure women's empowerment in urban sanitation across countries: The Agency, Resources, and Institutional Structures for Sanitation-related Empowerment (ARISE) Scales. *BMJ Open* 2022; 12:e053104. <https://doi.org/10.1136/bmjopen-2021-053104> PMID: 35177447
27. Sinharoy S, McManus S, Conrad A, Patrick M, Caruso BA. The Agency, Resources, and Institutional Structures for Sanitation-Related Empowerment (ARISE) Scales: Development and validation of measures of women's empowerment in urban sanitation for low- and middle-income countries. *World Dev*. 2023; 164: 106183. <https://doi.org/10.1016/j.worlddev.2023.106183> PMID: 37013085
28. Gonzalez D, Abdel Sattar R, Budhathoki R, Carrard N, Chase RP, Crawford J, et al. A partnership approach to the design and use of a quantitative measure: Co-producing and piloting the WASH gender equality measure in Cambodia and Nepal. *Development Studies Research*. 2022; 9: 142–158. <https://doi.org/10.1080/21665095.2022.2073248>
29. White AU, Bradley DJ, White GF. *Drawers of water: Domestic water use in East Africa*. Chicago: University of Chicago Press; 1972.

30. van Wijk-Sijbesma C. Participation of women in water supply and sanitation: Roles and realities. Hague, the Netherlands: IRC; UNDP; 1985. Available from: <https://www.ircwash.org/sites/default/files/202.1-85PA-2977.pdf>
31. Fisher J, Cavill S, Reed B. Mainstreaming gender in the WASH sector: Dilution or distillation? *Gend Dev*. 2017; 25: 185–204. <https://doi.org/10.1080/13552074.2017.1331541>
32. Kabeer N., Resources agency, achievements: Reflections on the measurement of women's empowerment. *Dev Change*. 1999; 30: 435–464. <https://doi.org/10.1111/1467-7660.00125>
33. Caruso BA, Conrad A, Patrick M, Owens A, Kviton K, Zarella O, et al. Water, sanitation, and women's empowerment: A systematic review and qualitative metasynthesis. *PLOS Water* 1(6): e0000026. <https://doi.org/10.1371/journal.pwat.0000026>
34. Moser CON. Gender planning and development: Theory, practice and training. London; New York: Routledge; 1993.
35. Carrard N, Crawford J, Halcrow G, Rowland C, Willetts J. A framework for exploring gender equality outcomes from WASH programmes. *Waterlines*. 2013; 32: 315–333. <https://doi.org/10.3362/1756-3488.2013.033>
36. Halcrow G, Rowland C, Willetts J, Crawford J, Carrard N. Working effectively with women and men in water, sanitation and hygiene programs. International Women's Development Agency and Institute for Sustainable Futures, University of Technology Sydney, Australia; 2010. Available from: www.genderinpacificwash.info
37. Carrard N, MacArthur J, Leahy C, Soeters S, Willetts J. The water, sanitation and hygiene gender equality measure (WASH-GEM): Conceptual foundations and domains of change. *Womens Stud Int Forum*. 2022; 91: 102563. <https://doi.org/10.1016/j.wsif.2022.102563>
38. Sen A. Development as capability expansion. In: Griffin K, Knight J, editors. Human development and the international development strategy for the 1990s. London: Macmillan; 1990.
39. Sen A. Development as Freedom. New York: Knopf; 1999.
40. Nussbaum MC. Women and human development: The capabilities approach. Cambridge: Cambridge University Press; 2000. <https://doi.org/10.1017/CBO9780511841286>
41. Malapit H, Quisumbing A, Meinzen-Dick R, Seymour G, Martinez EM, Heckert J, et al. Development of the project-level Women's Empowerment in Agriculture Index (pro-WEAI). *World Dev*. 2019; 122: 675–692. <https://doi.org/10.1016/j.worlddev.2019.06.018> PMID: 31582871
42. Beusenbergh M, Orley JH, World Health Organization. User's guide to the self reporting questionnaire (SRQ). Geneva: World Health Organization; 1994. Available from: <https://iris.who.int/handle/10665/61113>
43. Ryff CD, Keyes CLM. The structure of psychological well-being revisited. *J Pers Soc Psychol*. 1995 Oct; 69(4):719–27. <https://doi.org/10.1037//0022-3514.69.4.719> PMID: 7473027
44. Hays RD, Sherbourne CD, Mazel RM. The rand 36-item health survey 1.0. *Health Econ*. 1993; 2: 217–227. <https://doi.org/10.1002/hec.4730020305> PMID: 8275167
45. Diemer MA, Rapa LJ, Park CJ, Perry JC. Development and validation of the Critical Consciousness Scale. *Youth Soc*. 2017; 49: 461–483. <https://doi.org/10.1177/0044118X14538289>
46. Cilliers J, Dube O, Siddiqi B. Reconciling after civil conflict increases social capital but decreases individual well-being. *Science (1979)*. 2016; 352: 787–794. <https://doi.org/10.1126/science.aad9682> PMID: 27174981
47. O'Hara C, Clement F. Power as agency: A critical reflection on the measurement of women's empowerment in the development sector. *World Dev*. 2018; 106: 111–123. <https://doi.org/10.1016/j.worlddev.2018.02.002>
48. Kishor S, Subaiya L. Understanding women's empowerment: A comparative analysis of DHS Data. Calverton, Maryland, USA: Macro International; 2008 DHS Comparative Reports No. 20. Available from: <https://dhsprogram.com/publications/publication-cr20-comparative-reports.cfm>
49. Young SL, Bethancourt HJ, Ritter ZR, Frongillo EA. The Individual Water Insecurity Experiences (IWISE) Scale: Reliability, equivalence and validity of an individual-level measure of water security. *BMJ Glob Health*. 2021; 6: 1–14. <https://doi.org/10.1136/bmjgh-2021-006460> PMID: 34615660
50. Romppel M, Herrmann-Lingen C, Wachter R, Edelmann F, Düngen H-D, Pieske B, et al. A short form of the General Self-Efficacy Scale (GSE-6): Development, psychometric properties and validity in an inter-cultural non-clinical sample and a sample of patients at risk for heart failure. *Psychosoc Med*. 2013; 10: Doc01. <https://doi.org/10.3205/psm000091> PMID: 23429426
51. Vaz A, Pratley P, Alkire S. Measuring women's autonomy in Chad using the Relative Autonomy Index. *Fem Econ*. 2016; 22: 264–294. <https://doi.org/10.1080/13545701.2015.1108991>

52. Lombardini S, Bowman K, Garwood R. A 'how to' guide to measuring women's empowerment. Oxfam GB; 2017. <https://doi.org/10.21201/2017.9750>
53. Petesch P, Bullock R. Ladder of power and freedom: A qualitative data collection tool to understand local perceptions of agency and decision-making. CDMX, Mexico: CIMMYT; 2018. Available from: <https://hdl.handle.net/10568/92128>
54. Tapal A, Oren E, Dar R, Eitam B. The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Front Psychol*. 2017; 8: 1–11. <https://doi.org/10.3389/fpsyg.2017.00001> PMID: 28197108
55. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Front Public Health*. 2018; 6: 1–18. <https://doi.org/10.3389/fpubh.2018.00001> PMID: 29404319
56. van de Schoot R, Lugtig P, Hox J. A checklist for testing measurement invariance. *Eur J Dev Psychol*. 2012; 9: 486–492. <https://doi.org/10.1080/17405629.2012.686740>
57. Rodríguez A, Reise SP, Haviland MG. Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychol Methods*. 2016; 21: 137–150. <https://doi.org/10.1037/met0000045> PMID: 26523435
58. Flora DB. Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Adv Methods Pract Psychol Sci*. 2020; 3: 484–501. <https://doi.org/10.1177/2515245920951747>
59. Eisinga R, Te Grotenhuis M, Pelzer B. The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *Int J Public Health*. 2013; 58: 637–642. <https://doi.org/10.1007/s00038-012-0416-3> PMID: 23089674
60. Peters G-JY. The alpha and the omega of scale reliability and validity. *The European Health Psychologist*. 2014; 16: 56–69. Available from: <https://www.ehps.net/ehp/index.php/contents/article/view/ehp.v16.i2.p56>
61. Rodríguez A, Reise SP, Haviland MG. Applying bifactor statistical indices in the evaluation of psychological measures. *J Pers Assess*. 2016; 98: 223–237. <https://doi.org/10.1080/00223891.2015.1089249> PMID: 26514921
62. Widaman KF, Revelle W. Thinking thrice about sum scores, and then some more about measurement and analysis. *Behav Res Methods*. 2023; 55: 788–806. <https://doi.org/10.3758/s13428-022-01849-w> PMID: 35469086
63. Paek H-J, Lee B, Salmon CT, Witte K. The contextual effects of gender norms, communication, and social capital on family planning behaviors in Uganda: A multilevel approach. *Health Educ Behav*. 2006; 35: 461–477. <https://doi.org/10.1177/1090198106296769> PMID: 17513691
64. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. *J Pers Assess*. 1985; 49: 71–75. https://doi.org/10.1207/s15327752jpa4901_13 PMID: 16367493
65. Bloom SS, Wypij D, Das Gupta M. Dimensions of women's autonomy and the influence on maternal health care utilization in a North Indian city. *Demography*. 2001; 38: 67–78. <https://doi.org/10.1353/dem.2001.0001> PMID: 11227846