Evaluating changes in gender equality related to water, sanitation and hygiene interventions in rural Nepal: findings from a quasi-experimental evaluation

Jess MacArthur, Min Prasad Basnet, Ratan Budhathoki, Aastha Chhetri, Ambika Yadav & Juliet Willetts

View supplementary material

Published online: 04 Dec 2025.

Submit your article to this journal

Article views: 294

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Evaluating changes in gender equality related to water, sanitation and hygiene interventions in rural Nepal: findings from a quasi-experimental evaluation

Jess MacArthur[a], Min Prasad Basnet[b], Ratan Budhathoki[b], Aastha Chhetri[b], Ambika Yadav[b] and Juliet Willetts[a]

[a]Institute for Sustainable Futures, University of Technology Sydney, Ultimo, NSW, Australia; [b]SNV Nepal, Kathmandu, Nepal

ABSTRACT

Growing evidence highlights the gender-transformative potential of water, sanitation, and hygiene (WASH) interventions, which can shape gender equality and social inclusion alongside and through improvements in WASH practices, service delivery, and governance. This paper presents results from a quasi-experimental evaluation of the gender-transformative impacts of WASH interventions in the districts of Dailekh and Sarlahi in Nepal between 2020 and 2022. This timeframe also overlapped with the COVID-19 pandemic. The evaluation utilized propensity-score-matching and difference-in-difference techniques with the 16-theme water, sanitation, and hygiene – gender equality measure (WASH-GEM) to differentiate results between women and men of varying levels of program involvement. Average scores increased in 10 of the 16 themes for men, while decreasing in 10 themes for women; suggesting that men experienced greater improvements in gendered outcomes. Nonetheless, active participation in the program led to higher benefits or mitigated negative impacts in 15 themes, and women had better outcomes than men in 13 of these. These results suggest that active participation in the program had a positive net benefit for women participants, mitigating the negative gender impacts of the pandemic. This study illustrates the value of gender-transformative WASH programming as a pathway not only to improved WASH but also advancing equality.

1 Introduction

Gender equality and social inclusion have been shown to be both a pathway to and an outcome of water, sanitation, and hygiene (WASH) programs in international development (Willetts et al. 2010; MacArthur et al. 2023). However, many WASH interventions focus on the welfare, efficiency and equity of service provision, rather than opportunities for empowerment and social transformation (Caruso et al. 2022; MacArthur et al. 2023). The sector has seen increased momentum towards WASH programs which explicitly aim to transform gender norms, relations, and structures towards equality and inclusion which are known as 'gender-transformative' (MacArthur et al. 2023). Such programs aim to address inequalities in behaviors, practices, service delivery, and governance of water and sanitation. However, tools to evaluate gender-transformative changes alongside WASH interventions remain nascent (Caruso et al. 2021).

To address this absence of measurement tools, several research teams have begun building a toolbox of relevant measurement approaches for understanding the state of gender equality in the WASH sector. Such tools build from an evidence base of gender outcomes found to be associated with WASH interventions (Carrard et al. 2013). Tools developed so far are complementary to one another and each focus on different applications or subtopics. For example 1) the Water InSecurity Experiences (WISE) scales help to explore household or individual experiences of water security (Young et al. 2019, 2021), 2) the Agency, Resources

---

and Institutional Structures for Sanitation Empowerment (ARISE) scales are designed to explore empowerment related to urban sanitation (Sinharoy et al. 2021), and 3) the Empowerment in WASH index (EWI) to investigate agency, participation and empowerment in the water and sanitation sector (Dickin et al. 2021).

One tool, the water, sanitation and hygiene – gender equality measure (the WASH-GEM), is specifically designed to explore the intersection of gendered WASH outcomes and wider gender equality dynamics (Carrard et al. 2022). The five-domain WASH-GEM was codeveloped with WASH researchers and practitioners (Gonzalez et al. 2022) and to date has been piloted in Cambodia, Nepal, Laos, Bhutan and Ghana through cross-sectional assessments (MacArthur et al. 2024). These cross-sectional assessments explored gender dynamics in WASH programs at a single point in time, as a diagnostic tool to strengthen program interventions. Yet, until this present study, the WASH-GEM had not yet been used to evaluate changing gender equality over time.

Building on existing baseline data, this paper presents results from a quasi-experimental evaluation of transforming gender equality through WASH in two rural districts in Nepal. The study builds on previous articles on the WASH-GEM's development and validation (MacArthur et al. 2024), the theoretical foundations of the WASH-GEM (Carrard et al. 2022), and its development in partnership with implementation partners (Gonzalez et al. 2022). Notably, this study offers one of the few quantitative examinations of the COVID-19 pandemic's impact on gender equality outcomes, highlighting how active participation in a gender-transformative WASH program shaped these outcomes differently than passive receipt of improved WASH services.

The article contains four further sections. We begin with an overview of the WASH-GEM measure, an introduction to gender equality and WASH in Nepal, and a brief description of the impacts of COVID-19 on development interventions. This is followed by an overview of the quasi-experimental evaluation approach and its limitations. Next, we present the results from the study in three parts: 1) the sociodemographic background of the respondents; 2) a comparison of gendered theme scores between baseline and midline; and 3) an estimation of the effects of active program participation on theme scores. We then discuss key insights from the study related to the five WASH-GEM domains, the strengths and challenges of using the WASH-GEM for impact evaluation, and the concurrent impacts of COVID-19 on the efficacy of WASH interventions and gender equality. The article concludes by reflecting on key findings.

2 Background and theoretical foundations

2.1 Overview of the WASH-GEM

The WASH-GEM is a quantitative measure designed to assist practitioners and researchers in the exploration of gender outcomes associated with water, sanitation and hygiene (WASH) programs. The WASH-GEM seeks to expand beyond women's empowerment in WASH by including a focus on the gender dynamics between women and men, both within and beyond the sphere of WASH activities. The WASH-GEM builds on existing evidence that WASH interventions can influence changes in wider gender equality dynamics (Carrard et al. 2013; Willetts et al. 2013; Leahy et al. 2017). Through structured interviews with an equal number of male and female community members, the WASH-GEM measures aspects of gender equality both directly related to WASH and broader gender equality dynamics beyond WASH.

The conceptual model for the WASH-GEM was developed through a collaborative and iterative process informed by critical review of relevant literature and deepened through engagement with practitioners and specialists in the fields of gender, WASH, and international development (Carrard et al. 2022). The first iteration of the measure was codeveloped during 2019–2021 in collaboration with iDE Cambodia and SNV Nepal within the DFAT Water for Women Fund (Gonzalez et al. 2022) and further refined in 2022 through piloting in Ghana, Cambodia, Bhutan and Laos. This theoretical framework centers empowerment through a human development approach [resources → agency → wellbeing outcomes], while situating this process in external social and cultural structures (rules and norms) (Kabeer 1999; Keleher 2014). Internal personal beliefs (critical consciousness), typically understood as part of agency was broken out as a separate domain in the WASH-GEM to provide more targeted attention and interrogation (Carrard et al. 2022). This extends an empowerment framework to more holistically reflect equality. While the WASH-GEM was designed to

understand gendered outcomes for women and men, it has also been used to explore other dimensions of equality, taking an intersectional approach.

Illustrated in Figure 1, the WASH-GEM currently comprises five domains made up of 16 themes. The WASH-GEM domains and their themes are **Resources** [*WASH Access, Control over Resources,* and *Social Capital*], **Agency** [*WASH Decisions, Household Influence, Household Autonomy, Self-Efficacy,* and *Collective Agency*], **Critical Consciousness**, **Structures** [*WASH Norms, Gender Norms,* and *Mobility*], and **Wellbeing** [*WASH Wellbeing, Life Satisfaction, Physical Health* and *Mental Health*]. *Critical Consciousness* is both a theme and a domain. For the WASH-GEM, desired gendered outcomes are indicated by higher scores across all 16 themes and reduced disparity in the scores of women and men. Better outcomes include, higher levels of access and control over resources (Resources), higher levels of personal agency (Agency), stronger awareness of gender inequalities (Critical Consciousness), more equitable perceived norms (Structures), and better wellbeing (Wellbeing).

The WASH-GEM has been designed to be use both descriptively (single point in time) and evaluatively (multiple points in time). Existing uses of the WASH-GEM have been descriptive, as formative research upon which to build interventions that address contextual gender inequalities. However, this study used the WASH-GEM as an evaluation tool to compare the impact of the intervention between women and men through varying forms of program involvement.

Previous work has investigated the reliability and validity of the 16 themes within the WASH-GEM (MacArthur et al. 2024). As shown in Figure 1, four of the themes exhibited construct validity, internal reliability, structural validity, and scalar gender invariance, as indicated with three plus symbols (+++very strong). Five of the themes exhibited construct validity, internal reliability, structural validity, and configural or metric gender invariance, as indicated with two plus symbols (++strong). An additional three of the themes only exhibited construct validity and internal reliability, indicated through a single plus symbol (+moderate), and four of the themes only exhibited construct validity, indicated through a minus symbol (−fair). These final four themes were designed using robust item development processes – but still have room for further refinement. As such, the less reliable and valid themes provide less stable ground on which to draw conclusions; however, at the lack of other established and validated options, offer a proxy for this evaluation.

2.2 Gender equality and WASH in Nepal

Changes in WASH are interconnected with gender equality in Nepal. Nepal is ranked 110th on UNDP's Gender Inequality Index (GII),[1] with a score of 0.452 (UNDP, 2022). The global GII average is 0.465, with Nepal scoring better than Pakistan and Bangladesh (0.534 and 0.430, respectively), yet worse than India

Figure 1. WASH-GEM domains and themes.

(0.490). Nepali women are responsible for household WASH and are more vulnerable than men to WASH insecurity and distress (Wali et al. 2020). Related to WASH, in 2020, only 16% of the rural Nepali population had access to safely managed drinking water, and 50% to safely managed sanitation services (JMP, 2021). Notably, Nepal has achieved significant improvements related to access to basic sanitation in the last few years (JMP, 2021). At the same time, this process has been highly gendered in Nepal, as with many development practices more broadly (Regmi 2021).

2.3 Impacts of COVID-19 on development interventions

This evaluation took place during COVID-19 and the ramifications of this are important to consider. Between 2020 and 2021, the COVID-19 pandemic significantly impacted communities and development programs across the globe. This included both impacts within families, and communities due to state-led responses and the drastic halting of program activities and support. These impacts spanned decreases in resources, individual agency, and wellbeing (De et al. 2021; Leach et al. 2021). Within families, the pandemic led to significant gendered impacts. This included an increase in gender-based violence and women's care role responsibilities alongside a decrease of women's work outside the home (Paz et al. 2020; De et al. 2021). Within communities, lockdowns and other state-led responses replicated '*biases within development, reinforcing alienation, marginalization and stigmatization*' (Leach et al. 2021, p. 5). Social connections and capital were strained especially in low resource environments and volunteerism dropped globally (OEDC 2021). Programmatically within the WASH sector, interventions related to governance, service delivery and behavior change communication were halted and program support was paused.

In Nepal, these gender, social and economic impacts were significant during the COVID-19 pandemic due to prolonged lockdowns and drastic economic decline (World Bank 2020). Notably, in Nepal many internal and international migrant workers were required to return home which led to a drastic reduction in remittance funds. This led to an increased reliance on subsistence agriculture (Bista et al. 2022), increases in confinement, conflict, and violence against women (Dahal et al. 2020).

3 Evaluation approach

As a quasi-experimental evaluation, this study had three key objectives:

1. To investigate gender differentiated changes in equality articulated in the 16 WASH-GEM themes. This objective hypothesizes that women will have greater improvements than men.
2. To assess the effects of direct WASH program involvement on gender equality through the 16 WASH-GEM themes. This objective hypothesizes that individuals with higher levels of involvement will have a greater change in equality.
3. To assess the heterogeneity of direct WASH program involvement on gender equality separately for women and men.

3.1 Program setting: Beyond the Finish Line

The study evaluated the program impacts of a gender-transformative WASH program operating in two rural districts in Nepal. Implemented by SNV, the program '*Beyond the Finish Line*'[2] supported inclusive, sustainable, and resilient water supply services and hygiene behavior change communications in households, schools, and health care facilities in primarily rural or peri-urban areas Dailekh and Sarlahi. The program adopted a systems approach. Dailekh is a heavily forested hilly district in midwestern Nepal. Boarding India, Sarlahi is in Nepal's terai plains and features flat farming land.

Within the districts, the program had two different types of participants living within targeted geographic areas with improved services (Box 1). The first type of participants, termed here 'passive', were participants living within geographic areas which received improvements in water supply and access to improved sanitation products and services. This included improvements in water access through gravity-fed

water supply systems, primarily for those living in hilly areas (Dailekh) and groundwater-based systems (boreholes), primarily for those living in the terai plains (Sarlahi). Additionally, related to both water and sanitation the project sought to professionalize and strengthen WASH service delivery by working with both local government and private operators. As the strategy was highly contextual – it is not feasible to detail exactly what occurred in each area – however, the overarching premise is that in all the program working areas, steps were taking to strengthen WASH service provision, benefiting the wider population. This group is treated as the comparison group.

The second type of participants, termed here 'active', were agents of change within the program at the community level. This meant that they opted in to volunteer or work with the program actively supporting positive change towards improved services. These individuals received targeted face-to-face engagement from the program with illustrative activities included in **Box 1**. Each of these face-to-face touch points were also specifically designed through a gender-transformative lens with a goal of not only improving WASH services, but also to transform harmful gender-norms. This was often seen in the content, context and modality of the activities. Once again, as these strategies were highly contextual, it is not possible to detail exactly what occurred. We also were not able to cross-reference active participant self-reports with project process documentation. Future analysis could also explore the effect of active participation not only as a binary variable (passive–active), but looking at the number of regular touch points as a scale of involvement. One important challenge with the passive–active approach is that active participation was an opt-in strategy, and therefore active participants may have started the program with higher levels overall. To mitigate this, we have utilized propensity score matching to minimize differences between the 'treatment and comparison' groups.

Box 1. Types of program involvement

- **Passive** (comparison) = Accessing improved WASH services from public and private sector
- **Active** (treatment) = (Participated in training OR Participated in group discussions OR Participated in planning meetings OR Participated in monitoring OR Participated in day celebrations or campaigns OR Received WASH-related skill development OR Earning income from WASH related activities) AND (Accessing improved WASH services from public and private sector)

3.2 Sampling strategy

The cross-sectional study ($n = 3172$) included 1549 individual respondents in 2020 (women $n = 796$; men $n = 753$) and 1623 in 2022 (women $n = 811$; men $n = 812$). Study respondents were from male-female dyad pairs from the same household. Dyads were made up of two adult members of each household and as much as possible, represented dyads of the key decision maker among women in the household, and the key decision maker among men in the household. This was often married couples.

Respondents were randomly selected to include both community members who benefited from improvements in WASH services (comparison group) and individuals actively involved within the program (treatment group). In 2020, this included 46% of women respondents and 34% of men respondents who were actively involved; and in 2022, this included 78% of women respondents and 67% of men who were actively involved. This indicates a strong increase in active involvement over time as the pandemic waned.

The sample was divided 40% in Dailekh and 60% in Sarlahi based on population density. Eight project wards in each Sarlahi and Dailekh were randomly chosen for the baseline, and the same wards were surveyed in the midline; however, the same households were not targeted. Households were randomly selected within these wards to be representative of different poverty levels, ethnicities/castes, and religions; however, they were oversampled with respect to disability.[3]

An adapted version of the WASH-GEM was used to compare the baseline and midline datasets as minor changes were made to the tool during the program. The baseline study was conducted in November and December 2020, during a break in lockdowns, but during the COVID-19 pandemic. The midline study was conducted in September and October 2022. The sample wards and number of target interviews remained

the same in the baseline and midline; however, the households were not the same – and as such this is not a panel study. Different enumerators were locally contracted in each district for the two phases of the evaluation and were trained on the survey instrument and distress protocol (three-day training). The survey was conducted in Nepali, with explanations in Maithili (local language in Sarlahi) as required. Daily debriefs were conducted by the second author to clarify questions that arose during the data collection.

3.3 Data analysis and sensemaking

The quasi-experimental study aimed to understand the impact of active participation within a WASH program, on 16 aspects of gender equality as articulated in the WASH-GEM themes (dependent outcome variables). The 16 WASH-GEM themes were developed as continuous scale scores between 0 and 1, with higher scores indicating better gendered outcomes such as strengthened agency, better wellbeing or greater awareness of inequalities (MacArthur et al. 2024). A comparison of scores between women and men also was used to illustrate differences in their experiences and any disparities. This program also coincided with the COVID-19 pandemic, further complicating analysis. As such, we have aimed to untangle the impacts of active participation on mitigating the substantial consequences of COVID-19 on community members.

Data analysis was conducted in four parts. First, descriptive analysis was conducted at the item and theme levels to best understand the situational context at baseline and midline. Second, focused on objective-1, at baseline and midline, differences in gender equality more broadly (across each of the 16 themes) were calculated separately for women and men. T-tests were run to show statistical differences between the scores of women and men, and Cohen's *d* was calculated and reported for effect sizes greater than |0.5|. Third, focused on objective-2, the impact of active program involvement was explored by comparing the results over time of active and passive respondents. As we did not have panel data, rather cross-sectional data at baseline and midline, we adopted the approach outlined by Binci et al. (2018) blending Propensity Score Matching (PSM) and Difference-in-Differences (DiD). The approach has also been adopted and recommended by other researchers tasked with untangling cross-sectional data (Blundell and Costa Dias 2000; Aerts and Schmidt 2008). To start, a pseudo treatment panel was made by matching active participants across baseline and midline. Next the pseudo panel was split into baseline and midline and further matching done between active and passive participants at the two points in time. In essence, this approach requires three rounds of PSM – across the treatment group (1) and between the control and treatment at baseline and midline (2 and 3). A comparison of the final selected approach against direct DiD calculations and PSM at baseline and midline with DiD can be found in supplementary materials. Using this matched data, effects were determined using multivariate regression models to calculate DiD scores (White and Sabarwal 2014; Gertler et al. 2016; White and Raitzer 2017; Lutz 2018). DiD regressions controlled for involvement, education, wealth and ethnicity. Forth, focused on objective-3, differentiated (heterogenous) effects for program involvement were compared by gender by estimating effects using triple difference (DDD) estimators (White and Sabarwal 2014; Gertler et al. 2016; White and Raitzer 2017). Multivariate regressions compared each theme score between baseline and midline by gender, district and project involvement

Sensemaking of the results was conducted collaboratively with the *Beyond the Finish Line* project leadership in an online workshop. In particular, the workshop highlighted the significance of the COVID-19 pandemic scores related to *Resources*, *Agency,* and *Wellbeing*. This was triangulated with a qualitative study that the project had conducted during the same time frame, which found similar results.

3.4 Ethical considerations

Ethical considerations were an important aspect of both the baseline and midline data collection. Formal ethical approval was obtained prior to the start of data collection from the Institute for Sustainable Futures at the University of Technology Sydney (UTS HREC REF NO. ETH18–2599 Project–17232 and Project–21051). As a sensitive research topic, care was taken to mitigate risks for all participants. This included: 1) focused

training for the enumeration teams; 2) informed consent from participants individually, before beginning the interviews; 3) an emphasis on ensuring privacy in interviews; 4) gender-pairing between the enumerator and interviewee based on the respondent's self-identified gender; and 5) distress protocols developed in tandem with the enumerators and program team. Distress was also tracked by enumerators for each respondent at the end of the interview through questions related to comprehension, privacy, stress and ease.

3.5 Limitations and future research

This study must be interpreted through its limitations, in particular due to the complexities in the data analysis and due to COVID-19. The analysis adopted a flawed assumption that all individuals were equally impacted by COVID-19 and therefore that the control and treatment groups were still valid. However, as other literature has evidenced (Barbier and Burgess 2020; Paz et al. 2020), impacts are intersectional and therefore the results may represent artificially inflated benefits or negative impacts.

While significant efforts were taken to reduce differences between the treatment and control groups using PSM, the approach is not perfect. As summarized in Binci et al. (2018), there are several important limitations to the PSM-DiD approach. First, we are only able to control for observable characteristics such as gender, age and education – rather than psychosocial aspects which may influence participation. Second, active participation is a time-variant variable, and different levels of participation were compiled into a single binary variable. Binci et al. (2018) highlight that while using PSM, the calculation of standard errors of estimated treatment is more complex and there is not consistency in literature about how to address these errors. As the data represents only two points in time, we were also not able to check for the parallel assumption required for DiD; however, the PSM approach helps to control for these differences.

Finally, the evaluation relied on the WASH-GEM, which includes four less reliable themes and four themes which did not exhibit full (scalar) gender invariance using multiple-group confirmatory factor analysis – which assesses the validity of comparing scores across gender. While these tests are important in the development of robust measures, in the absence of other fully reliable options, we have chosen to utilize these less robust themes as a proxy for the constructs they represent. Nonetheless, future work can continue to expand the toolbox of validated scales to explore gender equality constructs in the Global South. Using the 2022 dataset, we were able to validate a new version of the *'Gender Norms'* scale, the results of which are included in Supplementary Materials – however, this version of the scale is not comparable with 2020 data. As such, a comparable version using the same items has been used in the evaluation. In alignment with Figure 1 – we have used the same nomenclature (+++,++,+,−) throughout the article to indicate which measures are the most valid and which ones should be read with more caution. Despite these challenges, the approach taken here offers a proxy to track improvements and overcome challenges for different types of program participants.

Future work could 1) explore the impact of different levels of active participation on outcomes, 2) utilize the WASH-GEM with a panel study rather than cross-sectionally – which would provide more robust results, and 3) continue to refine the validity of the WASH-GEM's themes – with particular emphasis on the three themes which only show construct and content validity.

4 Results

4.1 Descriptive characteristics

Sociodemographic characteristics from both the baseline and midline datasets are included in Table 1 by gender. Respondents were approximately 40% from Dailekh and 60% from Sarlahi with slight variations due to strategic extra sampling. Ninety-three percent of the respondents were the key decision makers of their gender in the home and 56% of respondents were the primary income earners in their home. Respondents were on average 44 years old at baseline and 47 years old at midline. Additionally at both baseline and midline, men were slightly older than women, and individuals in Sarlahi were older than in Dailekh. The results indicate that women had significantly lower levels of education than men in both districts; however,

Table 1. Descriptive characteristics of respondents between baseline and midline.

	2020 (n = 1549)		2022 (n = 1623)	
	Women (N = 796)	Men (N = 753)	Women (N = 811)	Men (N = 812)
District				
Dailekh	321 (40.3%)	305 (40.5%)	315 (38.8%)	317 (39.0%)
Sarlahi	475 (59.7%)	448 (59.5%)	496 (61.2%)	495 (61.0%)
Dyad	775 (97.4%)	740 (98.3%)	786 (96.9%)	788 (97.0%)
Decision Maker[4]	788 (99.0%)	747 (99.2%)	634 (78.2%)	766 (94.3%)
Income earner[5]	305 (38.3%)	576 (76.5%)	225 (27.7%)	657 (80.9%)
Age (grouped)				
18−30	166 (20.9%)	91 (12.1%)	114 (14.1%)	60 (7.4%)
31−49	423 (53.1%)	340 (45.2%)	443 (54.6%)	375 (46.2%)
50+	207 (26.0%)	322 (42.8%)	254 (31.3%)	377 (46.4%)
Education status (grouped)				
Preschool-	612 (76.9%)	381 (50.6%)	656 (80.9%)	393 (48.4%)
Primary	94 (11.8%)	179 (23.8%)	90 (11.1%)	225 (27.7%)
Secondary+	90 (11.3%)	193 (25.6%)	65 (8.0%)	194 (23.9%)
Married or partnered	788 (99.0%)	742 (98.5%)	805 (99.3%)	800 (98.5%)
Wealth quintile[6]				
Poorest	197 (24.7%)	179 (23.8%)	229 (28.2%)	191 (23.5%)
Poor	117 (14.7%)	126 (16.7%)	126 (15.5%)	122 (15.0%)
Middle	226 (28.4%)	214 (28.4%)	176 (21.7%)	168 (20.7%)
Rich	104 (13.1%)	93 (12.4%)	132 (16.3%)	163 (20.1%)
Richest	152 (19.1%)	141 (18.7%)	148 (18.2%)	168 (20.7%)
Religion				
Buddhist	1 (0.1%)	2 (0.3%)	26 (3.2%)	15 (1.8%)
Christian	8 (1.0%)	2 (0.3%)	3 (0.4%)	5 (0.6%)
Hindu	739 (92.8%)	700 (93.0%)	745 (91.9%)	760 (93.6%)
Muslim	48 (6.0%)	49 (6.5%)	37 (4.6%)	32 (3.9%)
Ethnicity				
Brahmin/Chhetri	204 (25.6%)	197 (26.2%)	151 (18.6%)	157 (19.3%)
Dalits	268 (33.7%)	225 (29.9%)	188 (23.2%)	185 (22.8%)
Janjati	129 (16.2%)	127 (16.9%)	194 (23.9%)	201 (24.8%)
Madhesi	144 (18.1%)	116 (15.4%)	229 (28.2%)	226 (27.8%)
Muslim	48 (6.0%)	49 (6.5%)	35 (4.3%)	32 (3.9%)
Newar	0 (0%)	0 (0%)	1 (0.1%)	0 (0%)
Other	3 (0.4%)	39 (5.2%)	13 (1.6%)	11 (1.4%)
Involvement				
Active	368 (46.2%)	257 (34.1%)	636 (78.4%)	542 (66.7%)
Passive	428 (53.8%)	496 (65.9%)	175 (21.6%)	270 (33.3%)

Table 2. Levels of active program involvement between baseline and midline by gender.

	2020		2022		Overall	
	Women (N = 796)	Men (N = 753)	Women (N = 811)	Men (N = 812)	Women (N = 1607)	Men (N = 1565)
Active	368 (46.2%)	257 (34.1%)	636 (78.4%)	542 (66.7%)	1004 (62.5%)	799 (51.1%)
Passive	428 (53.8%)	496 (65.9%)	175 (21.6%)	270 (33.3%)	603 (37.5%)	766 (48.9%)

overall education levels were lower in Sarlahi than Dailekh; which aligns with the slightly higher literacy rate found in Dailekh (National Statistics Office, 2025). The vast majority of participants were married and Hindu. Slightly higher numbers of the Brahmin/Chhetri ethnicity were included in the baseline as compared to the midline, with more Madhesi individuals in the midline than the baseline. Across both years, 27% of the respondents were Dalit (historically marginalized caste), 22% were Madhesi (historically marginalized community) and 20% Janjati (indigenous community); ethnicities which experience broad levels of social, political and economic marginalization (Minority Rights Group, 2024). The study population of marginalized community members is slightly higher than the general population indicating an oversampling of marginalized groups which could lead to lower levels of equality than can be generalized to the full population.

At both baseline and midline women were significantly ($p < 0.001$) more likely to actively participate than men in both districts. As shown in Table 2, active program engagement increased by 32% between 2020 and 2022, in alignment with post-COVID-19 opportunities. At baseline, participation was higher in Sarlahi than Dailekh, but nearly equal at midline (at approximately 50% of respondents). The difference in participation over time was primarily due to the COVID-19 timing, with program activities restarting after the main impacts of the pandemic in 2021. Additionally, the program purposefully targeted more women

than men within its activities and had different forms of activities between the two districts. For example, with community meetings, leadership trainings and support of marginalized voices in government planning activities.

4.2 Descriptive analysis: differentiated gendered outcomes

Focused on objective-1, we now estimate the scores for the 16 WASH-GEM themes separately for baseline and midline results by gender as shown in the boxplots in Figure 2. The plots also highlight the statistical
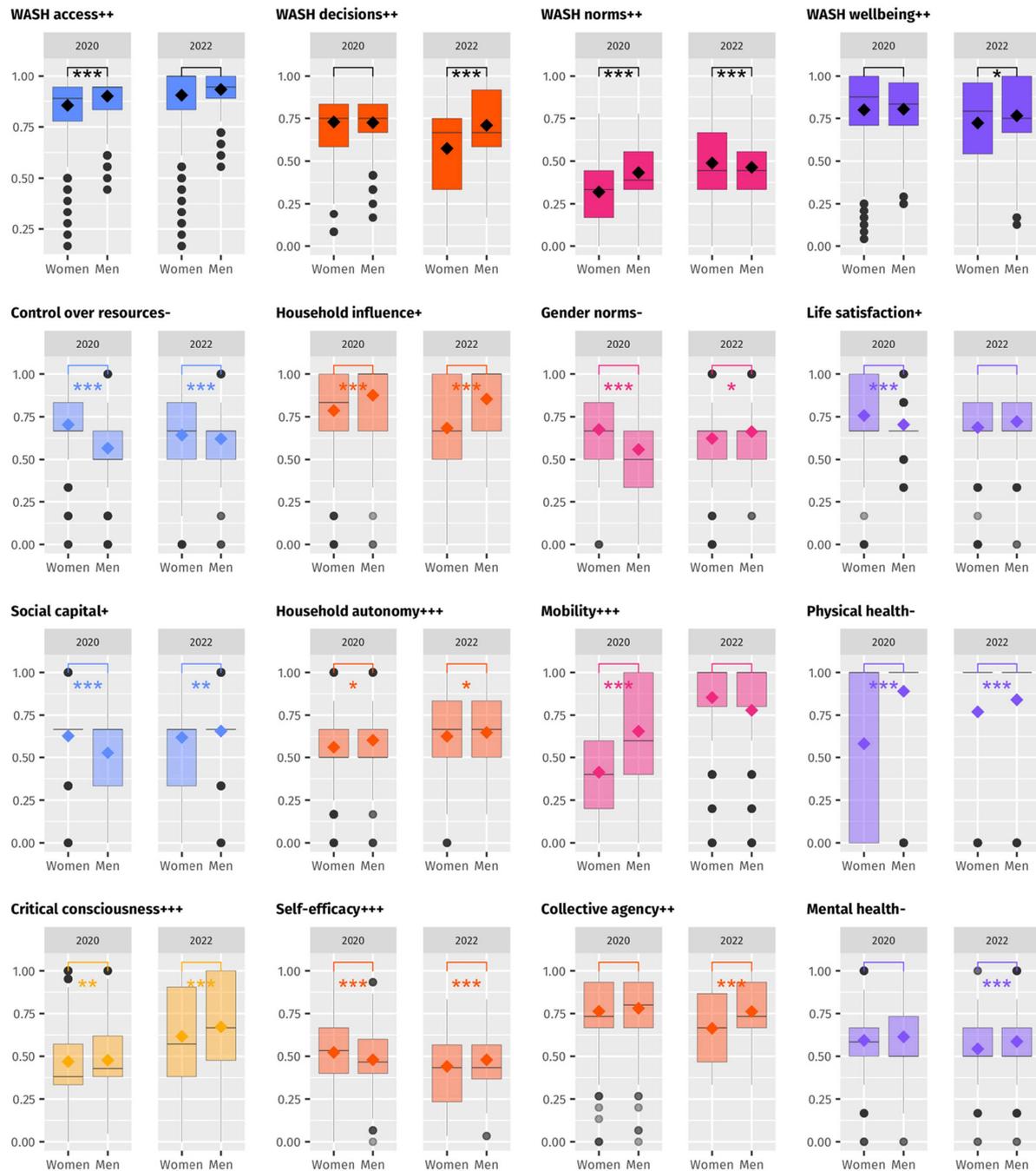

Figure 2. Boxplot illustrations of changes in theme scores between baseline and midline for women and men.[7,8]

significance of the differences in mean scores between women and men and visualize the extent of the differences.

At baseline, men had higher scores than women in 11 of the 16 WASH-GEM, indicating that men generally experienced better overall outcomes than women (e.g. greater access to and control over resources, more agency, improved well-being). When investigating the magnitude of these differences, men had much higher scores (Cohen's d > |0.5|) than women for three themes *'WASH Norms', 'Mobility'* and *'Physical Health'* consistent with entrenched gender roles granting men greater freedom of movement and less exposure to household-level health burdens. Nonetheless, women had much higher scores than men (Cohen's d > |0.5|) for two themes: *'Control over Resources'* and *'Gender Norms'*, suggesting that in the broader catchment population, women may have exercised greater influence in household financial management and held more progressive attitudes than men about household roles. This pattern could reflect pre-existing community characteristics or the cumulative impact of prior social and WASH initiatives in the area that emphasized women's participation.

At midline, the differences between men's and women's scores were less stark, yet men still had higher scores than women in 13 of the 16 WASH-GEM themes, indicating that gender gaps persisted across most areas. Women's scores did not surpass men's by a meaningful effect size in any theme at midline. However, men had much higher scores than women with medium effect sizes for *'Social Capital'* and *'WASH Decision-making'*. These differences may relate to post-lockdown recovery dynamics, as men were more likely to re-engage in community and public spaces, while women's social networks and decision-making opportunities may have remained more in the home.

When examining changes over time, men's scores increased in ten of the themes, whereas women's scores increased in only six. This suggests that broadly, men experienced greater improvements in gendered outcomes compared to women. Overall, this indicates that the WASH-GEM themes were able to highlight differences between the experiences of women and men throughout the project, yet women were not always disempowered relative to their male counterparts. A summary of these average theme scores by gender can be found in Supplementary Materials Table A.

4.3 Impact evaluation: gendered effects of active program involvement

Moving to the second and third objectives, we now present the estimated effects of active program involvement on scores. Figure 3 visualizes the differences between baseline and midline theme scores for active and passive involvement. The plots indicate the statistically significant differences between baseline and midline. Additionally, the impacts of the program were estimated through DiD multivariate regressions and DDD multivariate regressions with heterogeneous (gender) effects. These results are presented in Table 3 by year and involvement level alongside average scores for passive and active participants.

When reviewing the mean scores across the two-year period for active-passive involvement (not gender disaggregated), there was a mix of increasing and decreasing scores over time across the 16 WASH-GEM domains. Notably, just five of the themes showed increasing mean scores for both passive and active groups (*'WASH Access', 'Social Capital', 'Critical Consciousness, 'WASH Norms'* and *'Physical Health'*). Many scores in fact went down towards worse gender outcomes. This result highlights just how significant the negative impacts of COVID-19 on gender equality over time were for both women and men in Nepal, but also that there is some societal movement towards gender equality that has countered the impacts of the pandemic in some themes.

When reviewing the DID results (changes over time by program involvement, not yet gender disaggregated), active program involvement appeared to buffer participants from the negative impacts of COVID-19 on gendered outcomes in all but one theme (*'Life Satisfaction'*), as indicated by the broadly positive DiD scores. Notably, no significant difference between active and passive participants was observed for *'WASH Access'*, which was expected since physical infrastructure improvements benefitted all community members regardless of participation level. Additionally, there were benefits greater than 10% and statistically significant at $p > 0.001$ for active participants in four themes (*'WASH Decisions', 'Household Autonomy', 'Critical Consciousness'* and *'WASH Wellbeing'*). In this analysis, benefits are understood as either an increase in scores over time or a smaller decline compared to passive participants. These results
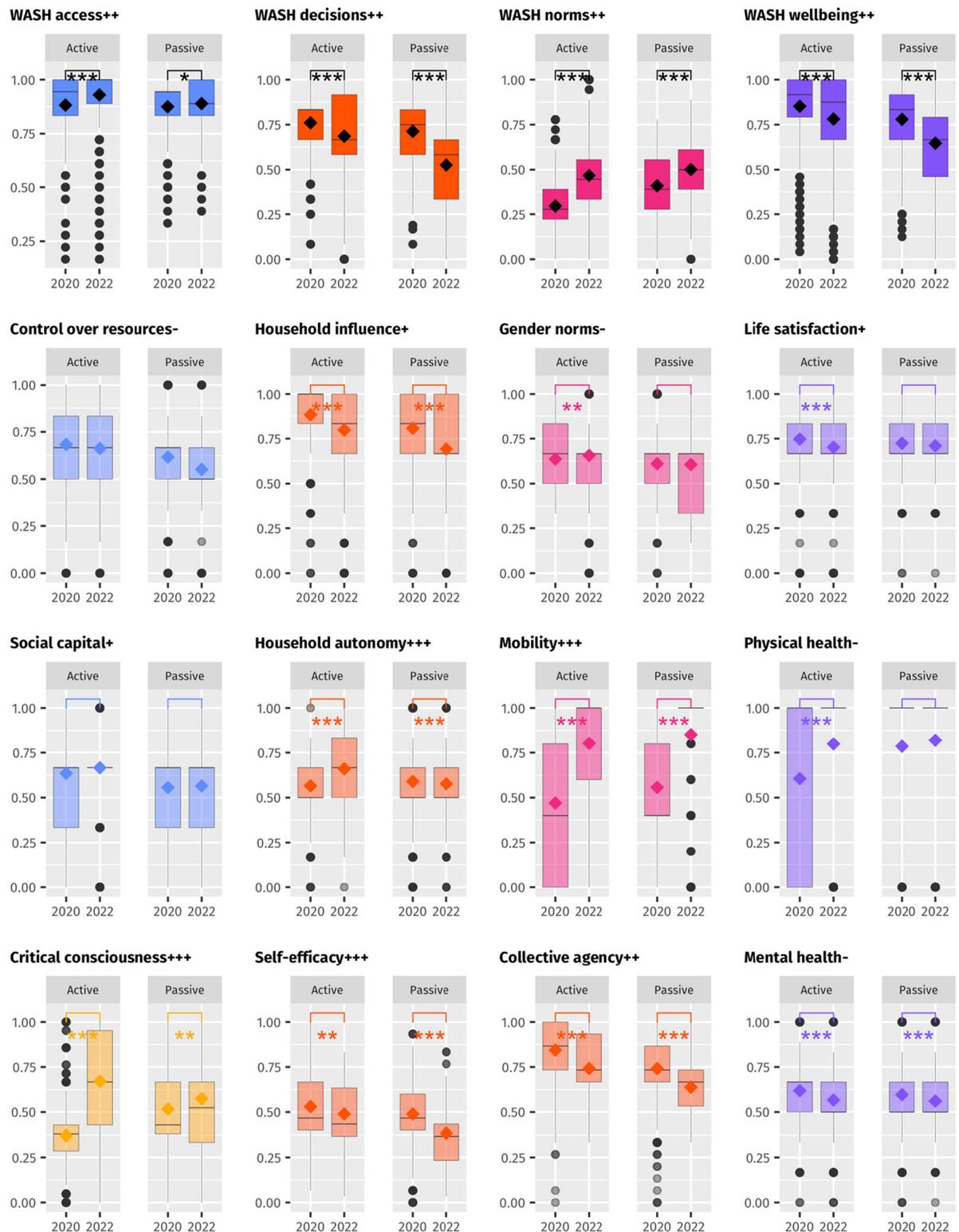
Figure 3. Boxplot illustrations of changes in thematic WASH-GEM scores between baseline and midline for active participants and passive beneficiaries (plot does not visualize the PSM).[9,10]

align with the program's participatory design, which emphasized reflection on norms, shared learning, and community planning. The results suggest that through these processes, active participants were able to build confidence in decision-making, strengthen household autonomy, and link improved access to WASH with broader wellbeing benefits.

Table 3. Mean theme scores by year and involvement and multivariate DiD and DDD scores between baseline and midline for program involvement (DiD) and gendered involvement (DDD).

	2020	2020	2022	2022	Involvement Multivariate DiD[11]	Gendered Involvement Multivariate DDD[12]
	Active	Passive	Active	Passive	Mean (SD) significance level	Mean (SD) significance level
WASH access++	0.885	0.871	0.930↑	0.889↑	0.011 (0.009) ns	0.007 (0.019) ns
Control over resources−	0.685	0.604	0.662↓	0.551↓	0.080 (0.014)***	**0.111 (0.029)*****
Social capital+	0.634	0.543	0.666↑	0.566↑	0.044 (0.018)*	0.083 (0.035)*
WASH decisions++	0.774	0.695	0.686↓	0.523↓	**0.136 (0.016)*****	0.029 (0.03) ns
Household influence+	0.883	0.794	0.798↓	0.691↓	0.085 (0.018)***	**0.259 (0.034)*****
Household autonomy+++	0.565	0.594	0.660↑	0.577↓	**0.114 (0.013)*****	0.055 (0.026)*
Self-efficacy+++	0.532	0.481	0.489↓	0.383↓	0.050 (0.014)***	−**0.114 (0.024)*****
Collective agency+++	0.846	0.722	0.742↓	0.636↓	0.024 (0.014)	**0.163 (0.028)*****
Critical consciousness++	0.382	0.535	0.671↑	0.573↑	**0.260 (0.018)*****	**0.158 (0.035)*****
WASH norms++	0.298	0.426	0.467↑	0.500↑	0.033 (0.015)*	−0.041 (0.029) ns
Gender norms−	0.678	0.578	0.658↓	0.604↑	0.041 (0.016)*	**0.126 (0.032)*****
Mobility+++	0.466	0.575	0.803↑	0.850↑	0.055 (0.025)*	**0.153 (0.048)****
WASH wellbeing++	0.848	0.770	0.781↓	0.647↓	**0.111 (0.016)*****	**0.304 (0.031)*****
Life satisfaction+	0.752	0.717	0.702↓	0.710↓	−0.028 (0.013)*	0.048 (0.025)
Physical health−	0.637	0.797	0.800↑	0.818↑	0.068 (0.033)*	−0.033 (0.062) ns
Mental health−	0.723	0.669	0.702↓	0.664↓	0.032 (0.012)**	0.073 (0.023)**

↑ Improvement in average score over time, ↓ Decline in average score over time, Bold text indicates differences over 0.100, Significance levels: $p < 0.001$ '***'; $p < 0.01$ '**', $p < 0.05$ '*', $p < 0.1$ '.', $p > 0.1$ 'ns'.

When reviewing the DDD results (changes over time by program involvement, gender disaggregated), actively involved women had more benefits than actively involved men over the two-year period in 11 of the 16 themes, as indicated by positive DDD scores. These benefits could be either an increase in scores or less of a decrease than men. This was most notable for '*Household Influence*' (+0.259 for active women), and '*WASH Wellbeing*' (+0.304). Modest improvements in '*Collective Agency*', '*Critical Consciousness*', '*WASH Norms*', '*Gender Norms*' and '*Mobility*' also suggest a gradual, but not transformative, shift in social expectations and women's public participation between baseline and midline. These findings suggest that women may have leveraged their participation to expand household roles and gain greater confidence and visibility in WASH-related discussions. This could also reflect how program spaces, particularly group activities and training, served as social platforms for women to rebuild social capital disrupted during lockdowns. Additionally, actively involved women had increased awareness on how to best support clean, private and safe sanitation and bathing facilities, important for *WASH Wellbeing* and a role predominately delegated to women. In contrast, men's active involvement was only associated with lower '*Self-efficacy*' scores over time. This could reflect the economic and social pressures men faced during and after the pandemic, as traditional provider roles were disrupted, possibly reducing perceptions of personal control and confidence; which could have been reinforced by program activities.

5 Discussion

5.1 Summary and interpretation

This study aimed to evaluate the impact of a gender-transformative WASH intervention in Nepal between baseline (2020) and midline (2022) using the WASH-GEM. The WASH-GEM's 16 themes were compared between baseline and midline with regards to gender and level of program involvement (passive and active) using DiD and DDD regressions. As the program was conducted during a period of global uncertainty due to COVID-19, the results provide an interesting spotlight on the gendered impacts of program involvement during economic and social shocks. While a standard control group was not feasible in this context, the use of passive and active levels of involvement provides insights into the program's impact.

At first glance, men appeared more likely than women to experience improvements across the WASH-GEM themes over the two-year period. However, accounting for active program participation reveals a different pattern: actively involved women experienced significantly greater benefits than men in 11 of the WASH-GEM themes. These benefits reflect both higher overall scores and the mitigation of COVID-19–related declines. In contrast, for passive participants who only received improved WASH services,

men generally benefited more across most themes. Women's advantages became most pronounced when they actively engaged in participatory activities and capacity-building opportunities, highlighting the value of gender-transformative approaches (MacArthur et al. 2023). These findings suggest that WASH interventions can achieve more equitable and meaningful impacts by combining infrastructure improvements with strategies that intentionally strengthen agency, leadership, and influence within households and communities.

5.2 Changing gendered outcomes

As highlighted throughout the results, while the program had significant benefits in many of the themes, these impacts were affected by COVID-19, when explored through the lens of the five WASH-GEM domains. The results suggest that while there were negative impacts of the COVID-19 pandemic on key gendered outcomes, active program participants experienced less of an impact than passive beneficiaries. We now revisit each of the five WASH-GEM domains while reflecting on the results.

5.2.1 Resources

Within the results, all three *Resources* themes were strongly gendered, and two themes were strongly influenced by the pandemic. The *Resources* domain includes *'WASH Access', 'Control over Resources'* and *'Social Capital'*. For the three *Resources* themes, there were disparities between women and men, but women had higher levels of control and support than men at baseline. This could be due to a long history of programs working with women in these communities (Gonzalez et al. 2022) and supporting community development and WASH services. Disparities reduced over the two-year period. The results also suggest that *'Social Capital'* and *'Control over Resources'* were both impacted by the lockdowns and strained social connections, a finding that mirrors other studies focused on South Asia (World Bank, 2020) and globally (OEDC, 2021). Importantly, the results suggest that program involvement had a small positive impact on increasing *'Social Capital'* as communities came together post lockdown to support one another within the program activities. Additionally, active program participants were less likely to experience the decrease in *'Control over Resources'*. This could be due to the program providing participants with greater access to information, financial tools, and collective opportunities, which helped buffer the negative effects of the lockdowns on individual resource control.

5.2.2 Agency

From the results, men broadly had higher levels of agency than women, and there were decreasing scores across all five *Agency* themes over the two-year period, illustrating how much COVID-19 impacted personal agency. Within the WASH-GEM, *Agency* includes *'WASH Decisions', 'Household Influence', 'Household Autonomy', 'Self-efficacy'* and *'Collective Agency'*. With the exception of *'Self-efficacy'*, men had higher agency than women across the study timeframe, aligning more clearly with existing literature which suggests higher empowerment of men than women and focus more on agency-related topics such as decision-making (Adhikari and Sharma 2022). *Agency* was also strongly and negatively impacted by COVID-19 and in particular for individuals who were not involved in the program. This could be directly related to the coming home of many migrant workers negatively impacting aspects such as gender-based violence and volunteerism (Dahal et al. 2020; Leach et al. 2021). Importantly, the program seems to have mitigated these impacts potentially by modeling gender-transformative interactions, championing the voices of women and encouraging more equal household roles and responsibilities.

5.2.3 Structures and critical consciousness

In contrast to the results from *Agency*, the *Structures* and *Critical Consciousness* domains saw improvements across all themes for both passive and active participants over the two-year period, indicating a broader social transformation towards structural equality. Structures include: *'WASH Norms', 'Gender Norms'* and *'Mobility',* while *'Critical Consciousness'* is both a theme and a domain. *Structures* are related

to one's perceived beliefs of society, while *Critical Consciousness* refers to internal personal beliefs, both which shape *Resources, Agency* and *Wellbeing*. The data suggest that women who were actively involved in the program experienced a positive shift toward more equal gender norms and improved mobility than men who were actively involved. This social change is aligned with other scholarship on the shifting dynamics around the 'normal' roles and responsibilities for women and men in Nepali society as well as freedom of movement and awareness of inequalities (Adhikari and Sharma 2022). The results suggest that while personal agency was significantly impacted by COVID-19, the undercurrent of social change was strong enough to continue on an upward trend through the pandemic. Perhaps people had additional time to reflect, discuss, read and even watch television advocating for more equal gender dynamics. Further research could explore to what extent the pandemic actually accelerated structural social change.

5.2.4 Wellbeing

The impacts by gender and time were less meaningful for *Wellbeing* than for the other domains. The *Wellbeing* domain includes *'WASH Wellbeing', 'Life Satisfaction', 'Physical Health'* and *'Mental Health'*. *'WASH Wellbeing'* broadly declined over time, potentially because of an increased awareness of safety, cleanliness, and privacy in bathing and sanitation facilities, yet women who were actively involved reported higher *'WASH Wellbeing'* than women who were not involved. Importantly, apart from the *'WASH Wellbeing'* theme, the other *Wellbeing* themes are less reliable scales, and they require further refinement. As such, their results should be taken more lightly. Additionally, while *Wellbeing* differences between women's and men's scores in both 2020 and 2022 were statistically significant, when normalized the effects were small or negligible. The one exception was the higher *'Physical Health'* scores for men than women at the baseline. These results indicate that although there were some improvements in other domains, it appears that wellbeing outcomes beyond WASH may take longer to influence than the two-year time period, or that COVID-19 impacted this area.

Overall, these findings align with previous research highlighting that levels of gendered access, agency, perceptions and wellbeing can vary significantly across different domains (Adhikari and Sharma 2022; Parajuli and Neupane 2025), even for the same people experiencing similar program interventions and pandemic stresses. Importantly, the results suggest that improvements in *Resources*, *Structures,* and *Critical Consciousness* do not always equate with improvements in *Agency* or *Wellbeing.* This key finding reinforces WASH-GEM's conceptual framework away from a linear pathway from *Resources → Agency → Wellbeing,* towards a series of overlapping and interconnected domains (Carrard et al. 2022).

The findings also highlight the important and complex intersections of ethnicity/caste, age and education with gendered outcomes. Notably, our results contradict the typical perspective that women are less empowered than men in Nepal. We illustrate that in marginalized groups and for older populations men can be less empowered than their female counterparts. For example, with the *Resources* domain, as the male study population skewed to older men and marginalized communities, these men may have had less opportunity to engage in community groups or felt less in control of their own resources. This pattern is noteworthy, as studies directly comparing empowerment across both age and gender groups are relatively rare; most existing research tends to focus on a single demographic dimension, such as either gender or age (Bakshi and Pathak 2012; Malapit et al. 2015; van Dongen, Obrizan, and Shymanskyi 2024).

5.3 Reflections on using the WASH-GEM for impact evaluation

The WASH-GEM remains a robust and conceptually grounded tool connecting WASH and gender (Carrard et al. 2022). It has been collaboratively designed (Gonzalez et al. 2022), statistically validated[13] (MacArthur et al. 2024) and demonstrated to be valuable in exploring gender dynamics in diverse WASH contexts. This study extends its application to the evaluation of gendered change over time, illustrating how the WASH-GEM can generate evidence on the intersections between WASH and gender, and support policy, program, and project-level monitoring.

5.3.1 Strengths

The WASH-GEM comprises 16 curated themes rather than a single index or score. Together, these themes provide a comprehensive framework for identifying changes in gendered outcomes within WASH programs or other relevant development programs aiming to transform gender norms, structures or dynamics. While this study focused on gender, the tool can also be adapted to analyze differences by other covariates, such as ethnicity, disability status, or age, depending on the sampling strategy. The use of Difference-in-Differences (DiD) and Difference-in-Difference-in-Differences (DDD) estimations enabled comparison between active and passive participants, offering a pragmatic quasi-control group approach that helps assess how deeper engagement in program activities influences outcomes, in line with many programs' theories of change. Furthermore, the WASH-GEM's visual design, using color and pattern to represent domains, supports accessibility and encourages uptake of findings among nontechnical audiences, facilitating data use by implementers and policymakers. Finally, the results of this study have wider and further implications to other sectors and geographic contexts. In particular, the WASH-GEM shows strong promise as a multidimensional gender measure foregrounding individual measurement practices.

5.3.2 Challenges

Despite these strengths, several challenges emerged in applying the WASH-GEM for impact evaluation. Thematic-level analysis, while useful for summarizing results, can sometimes obscure important insights at the item (question) level. For this reason, the WASH-GEM team recommends complementing quantitative results with participatory sensemaking workshops (Gonzalez et al. 2022), where item-level data can be explored collaboratively to generate richer interpretation. In addition, the statistical techniques, such as DiD and DDD, can be complex and difficult for program teams to interpret, especially when reviewing results across 16 themes. Presenting findings in clear, plain-language summaries and highlighting key areas of change before dissemination can greatly support understanding and uptake. Finally, as seen in this study, external shocks such as the COVID-19 pandemic significantly influenced gender equality outcomes. This underscores that empowerment and equality are shaped by broader contextual factors, making claims of causation in quasi-experimental settings inherently uncertain.

6 Conclusions

This study has explored the effects of a WASH intervention on gendered outcomes for rural Nepalis living in Sarlahi and Dailekh between 2020 and 2022. The study utilized the multidimensional WASH-GEM previously developed and validated for use in South and Southeast Asia. The WASH-GEM contains 16 themes across five domains, which were curated to capture important aspects of equality and empowerment for both women and men to identify disparities and trace changes. The evaluation overlapped in time with the COVID-19 pandemic, creating interesting and complex compounded impacts on rural populations.

The results suggest that changes in gendered outcomes over time were influenced by program involvement, the pandemic, and a broader undercurrent of social change towards equality in Nepal; all intersecting with individual characteristics such as gender, age, education, ethnicity and wealth. The results remained complex and nuanced. Additionally, the impacts of COVID-19 were varied across different respondents and across domains. Overall, both women and men experienced improvements in the *Critical Consciousness* and *Structures* domains over time, these domains are related to beliefs and norms. These types of internal or attitudinal changes may have continued to progress despite the pandemic, partly due to ongoing community engagement and information exposure even during lockdowns. However, both women and men also experienced decreasing scores in the *Resources*, *Agency* and *Wellbeing* domains. These domains were likely more sensitive to the pandemic's effects on livelihoods, social interactions, and self-confidence, which disproportionately limited opportunities for decision-making and control over resources. Most notably, women who were actively involved in the program experienced statistically significant benefits in all five WASH-GEM domains (across eleven themes), when compared to those who were passively involved and only had improved WASH services. This suggests that WASH interventions can achieve more equitable and meaningful impacts by combining infrastructure

improvements with strategies that intentionally build women's agency, leadership, and influence within households and communities.

As part of ongoing development, validation and use of the WASH-GEM, this work has illustrated the tool's value in comparing results over time and untangling differences between gender and program involvement. Further testing in other contexts, and outside the constraints of COVID-19 is recommended to continue the refinement of the WASH-GEM. Given the important linkages between gender and WASH, the sector requires a full toolbox with measures such as the WASH-GEM to generate, interrogate and utilize evidence to strengthen ongoing and future policies, projects and programs.

Endnotes

1. The GII is a composite measure of gender inequality in three dimensions: reproductive health, empowerment and the labor market. A low GII score indicates more equality between women and men.
2. Beyond the Finish Line has been a five-year multi-country program (2018–2022), which is being implemented by SNV in Nepal, Bhutan and Lao PDR. Supported by the Water for Women Fund of the Australian government, the project is a collaborative initiative between SNV, the Institute for Sustainable Futures of University of Technology Sydney (ISF-UTS), and the Ministry of Water Supply, together with the National Federation of Disabled in Nepal and CBM Australia.
3. Over sampling was followed to identify at least 5% of households in which a household member had a disability as identified through the Washington Group Criteria.
4. Primary decision maker of the men, or primary decision maker of the women in the household.
5. Primary income earner for the household.
6. Wealth quintiles were calculated separately for the baseline and midline datasets based on asset data.
7. Significance levels: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 'ns' 1. Mean theme scores are indicated in the colored diamond point. Measure robustness: +++ very strong, ++ strong, + moderate, – fair
8. Measure robustness: +++ very strong, ++ strong, + moderate, – fair
9. Significance levels: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 'ns' 1. Mean theme scores are indicated in the colored diamond point. Measure robustness: +++ very strong, ++ strong, + moderate, – fair
10. Significance levels: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1. Mean domain scores are indicated in the colored diamond point. Measure robustness: +++ very strong, ++ strong, + moderate, – fair
11. Passive involvement taken as standard for DiD calculations (positive estimates indicate that active participants had a greater difference between baseline and midline, while negative DiD estimates suggest that passive participants had a greater difference over time)
12. Women are taken as standard for DDD calculations (positive estimates indicate that women had a greater increase between baseline and midline, while negative DDD estimates suggest that men had a greater benefit over time)
13. 12 of the 16 WASH-GEM themes passed structural validity, internal reliability, and measurement invariance testing. The final four themes were designed using content, conceptual and Item Response Theory – but still have room for further refinement.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

References

Adhikari, R., and J. R. Sharma. 2022. "Gendered Consequences of Social Changes in Nepal: Rich Possibilities." *European Bulletin of Himalayan Research* (58): 105–124. https://doi.org/10.4000/ebhr.549.

Aerts, K., and T. Schmidt. 2008. "Two for the Price of One? Additionality Effects of R&D Subsidies: A Comparison between Flanders and Germany." *Research Policy* 37 (5): 806–822. https://doi.org/10.1016/j.respol.2008.01.011.

Bakshi, S., and P. Pathak. 2012. "An Analysis of the State of Empowerment of females Vis-à-vis Males at Old Ages in India." *Gender Issues and Empowerment of Women* 104667: 83–108.

Barbier, E. B., and J. C. Burgess. 2020. "Sustainability and Development After COVID-19." *World Development* 135: 105082. https://doi.org/10.1016/j.worlddev.2020.105082.

Binci, M., M. Hebbar, P. Jasper, and G. Rawle. 2018. Matching, differencing on repeat. www.opml.co.uk.

Bista, R., R. Parajuli, K. Giri, R. Karki, and C. Song. 2022. "Impacts of COVID-19 Pandemic on the Livelihoods of Rural Households in the Community Forestry Landscape in the Middle Hills of Nepal." *Trees, Forests and People* 9 (August): 100312. https://doi.org/10.1016/j.tfp.2022.100312.

Blundell, R., and M. Costa Dias. 2000. "Evaluation Methods for Non-experimental Data." *Fiscal Studies* 21 (4): 427–468. https://doi.org/10.1111/j.1475-5890.2000.tb00031.x.

Carrard, N., J. Crawford, G. Halcrow, C. Rowland, and J. Willetts. 2013. "A Framework for Exploring Gender Equality Outcomes from Wash Programmes." *Waterlines* 32 (4): 315–333. https://doi.org/10.3362/1756-3488.2013.033.

Carrard, N., J. MacArthur, C. Leahy, S. Soeters, and J. Willetts. 2022. "The Water, Sanitation and Hygiene Gender Equality Measure (WASH-GEM): Conceptual Foundations and Domains of Change." *Women's Studies International Forum* 91 (2022): 102563. https://doi.org/10.1016/j.wsif.2022.102563.

Caruso, B. A., A. Salinger, M. Patrick, A. Conrad, and S. Sinharoy. 2021. A Review of Measures and Indicators for Gender in WASH (Issue June). https://washdata.org/sites/default/files/2021-10/jmp-2021-gender-review-final-report.pdf.

Caruso, B. A., A. Conrad, M. Patrick, A. Owens, K. Kviten, O. Zarella, H. Rogers, and S. S. Sinharoy. 2022. "Water, Sanitation, and Women's Empowerment: A Systematic Review and Qualitative Metasynthesis." *PLOS Water* 1: e0000026. https://doi.org/10.1371/journal.pwat.0000026.

Dahal, M., P. Khanal, S. Maharjan, B. Panthi, and S. Nepal. 2020. "Mitigating Violence Against Women and Young Girls During COVID-19 Induced Lockdown in Nepal: A WAKE-UP CAll." *Globalization and Health* 16 (1): 1–3. https://doi.org/10.1186/s12992-020-00616-w.

De, C., Nieves, P., Gaddis, I., and Muller, M. 2021) Gender and COVID-19: What have we Learnt, One Year Later? (9709; Policy Research Working Paper, Issue June). http://www.worldbank.org/prwp

Dickin, S., E. Bisung, J. Nansi, and K. Charles. 2021. "Empowerment in Water, Sanitation and Hygiene Index." *World Development* 137: 105158. https://doi.org/10.1016/j.worlddev.2020.105158.

Gertler, P. J., S. Martinez, P. Premand, L. B. Rawlings, and C. M. J. Vermeersch. 2016. *Impact Evaluation in Practice (Second Edi)*. Washington, DC: World Bank Group. https://doi.org/10.1109/WI-IATW.2006.145

Gonzalez, D., R. Abdel Sattar, R. Budhathoki, N. Carrard, R. P. Chase, J. Crawford, G. Halcrow, et al. 2022. "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* 9 (1): 142–158. https://doi.org/10.1080/21665095.2022.2073248.

JMP. 2021. Progress on Household Drinking Water, Sanitation and Hygiene. https://washdata.org/report/jmp-2021-wash-households-highlights.

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

Keleher, L. 2014. "Sen and Nussbaum: Agency and Capability-Expansion." *Éthique et économique/Ethics and Economics* 11 (2): 54–70. http://hdl.handle.net/1866/10936.

Leach, M., H. MacGregor, I. Scoones, and A. Wilkinson. 2021. "Post-Pandemic Transformations: How and Why COVID-19 Requires us to Rethink Development." *World Development* 138: 105233. https://doi.org/10.1016/j.worlddev.2020.105233.

Leahy, C., K. Winterford, T. Nghiem, J. Kelleher, L. Leong, and J. Willetts. 2017. "Transforming Gender Relations through Water, Sanitation, and Hygiene Programming and Monitoring in Vietnam." *Gender and Development* 25 (2): 283–301. https://doi.org/10.1080/13552074.2017.1331530.

Lutz, C. 2018. Difference-in-Difference-in-Differences CFPLs Case Study. https://chandlerlutz.github.io/difference-in-difference-in-differences-CFPLs/

MacArthur, J., N. Carrard, J. Mott, S. Raetz, M. Siscawati, and J. Willetts. 2023. "Gender Equality Approaches in Water, Sanitation, and Hygiene Programs: Towards Gender-transformative Practice." *Frontiers in Water* 5: 1090002. https://doi.org/10.3389/frwa.2023.1090002.

MacArthur, J., R. P. Chase, D. Gonzalez, T. Kozole, C. Nicoletti, V. Toeur, and J. Willetts. 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 October): 1–24. https://doi.org/10.1371/journal.pwat.0000233.

Malapit, H. J. L., S. Kadiyala, A. R. Quisumbing, K. Cunningham, and P. Tyagi. 2015. "Women's Empowerment Mitigates the Negative Effects of Low Production Diversity on Maternal and Child Nutrition in Nepal." *Journal of Development Studies* 51 (8): 1097–1123. https://doi.org/10.1080/00220388.2015.1018904.

Minority Rights Group. 2024. Nepal - Minority Rights Group. https://minorityrights.org/country/nepal/.

National Statistics Office. 2025. National Population and Housing Census 2021.

OEDC. 2021. COVID-19 and Well-being. OECD. https://doi.org/10.1787/1e1ecb53-en.

Parajuli, C., and B. K. Neupane. 2025. "Assessing Gender Equality in Nepal: A Comparative Study of Progress and Disparities." *NPRC Journal of Multidisciplinary Research* 2 (8): 200–219. https://doi.org/10.3126/nprcjmr.v2i8.83845.

Paz, C. de, M. Muller, A. M. M. Boudet, and I. Gaddis. 2020. *Gender Dimensions of the COVID-19*. World Bank Group, (Policy Note, Vol. 18, Issue April). http://pubdocs.worldbank.org/en/232551485539744935/WDR17-BP-Gender-based-violence-and-the-law.pdf%0Awww.adelaide.edu.au/writingcentre/%0Awww.iucn.org

Regmi, S. C. 2021. *Engendering the Development: Real or Rhetoric!: (with Specific Reference to Nepal's Water Sector 2000-2020)*. New Delhi: Adroit Publishers.

Sinharoy, S. S., A. Conrad, M. Patrick, S. Mcmanus, and B. A. Caruso. 2022. "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* 12 (2): e053104. https://bmjopen.bmj.com/content/12/2/e053104.

UNDP. 2022. Human Development Report 2021/2022: Uncertain times, unsettled lives Shaping our future in a transforming world. https://hdr.undp.org/content/human-development-report-2021-22.

van Dongen, D. M., M. Obrizan, and V. Shymanskyi. 2024. "Determinants of Women's Empowerment in Nepal." *PLoS One* 19 (9): 1–16. https://doi.org/10.1371/journal.pone.0310266.

Wali, N., N. Georgeou, O. Simmons, M. S. Gautam, and S. Gurung. 2020. "Women and WASH in Nepal: A Scoping Review of Existing Literature." *Water International* 45 (3): 222–245. https://doi.org/10.1080/02508060.2020.1754564.

White, H., and S. Sabarwal. 2014. Quasi-Experimental Design and Methods. www.unicef-irc.org.

White, H., and Raitzer, D. A. 2017) Impact Evaluation of Development Interventions: A Practical Guide. https://www.adb.org/sites/default/files/publication/392376/impact-evaluation-development-interventions-guide.pdf

Willetts, J., G. Halcrow, N. Carrard, C. Rowland, and J. Crawford. 2010. "Addressing Two Critical MDGs Together: Gender in Water, Sanitation and Hygiene Initiatives." *Pacific Economic Bulletin* 25 (1): 211–221. https://openresearch-repository.anu.edu.au/handle/1885/30589.

Willetts, J., N. Carrard, J. Crawford, C. Rowland, and G. Halcrow. 2013. "Working from Strengths to assess Changes in Gender Equality." *Development in Practice* 23 (8): 991–1006. https://doi.org/10.1080/09614524.2013.840564.

World Bank. 2020. Beaten or Broken? Informality and COVID-19. https://doi.org/10.1596/978-1-4648-1640-6.

Young, S. L., H. J. Bethancourt, Z. R. Ritter, and E. A. Frongillo. 2021. "The Individual Water Insecurity Experiences (IWISE) Scale: Reliability, Equivalence and Validity of an Individual-level Measure of Water Security." *BMJ Global Health* 6 (10): 1–14. https://doi.org/10.1136/bmjgh-2021-006460.

Young, S. L., G. O. Boateng, Z. Jamaluddine, J. D. Miller, E. A. Frongillo, T. B. Neilands, S. M. Collins, A. Wutich, W. E. Jepson, and J. Stoler. 2019. "The Household Water Insecurity Experiences (HWISE) Scale: Development and Validation of a Household Water Insecurity Measure for Low-income and Middle-income Countries." *BMJ Global Health* 4 (5): e001750. https://doi.org/10.1136/bmjgh-2019-001750.