International Journal of Hygiene and Environmental Health 240 (2022) 113922

Contents lists available at ScienceDirect

International Journal of Hygiene and Environmental Health

journal homepage: www.elsevier.com/locate/ijheh

Measurement invariance of a household water insecurity metric in Greater Accra, Ghana: Implications for test-retest reliability

Melissa N. Sidote [a], Zachary T. Goodman [b], Christina L. Paraggio [c], Raymond A. Tutu [d], Justin Stoler [a,e,*]

[a] Department of Public Health Sciences, Miller School of Medicine, University of Miami, Miami, FL, USA
[b] Department of Psychology, University of Miami, Coral Gables, FL, USA
[c] Department of Microbiology and Immunology, University of Miami, Coral Gables, FL, USA
[d] Department of Sociology and Criminal Justice, Delaware State University, Dover, DE, USA
[e] Department of Geography and Sustainable Development, University of Miami, Coral Gables, FL, USA

ARTICLE INFO

Keywords:
Global health
Measurement
Reliability
Water insecurity
West Africa

ABSTRACT

The mitigation of household water insecurity is recognized as an important component of global poverty alleviation, but until recently was difficult to measure. Several new metrics of household water insecurity have been proposed and validated, but few have been field-tested for reliability in diverse contexts. We used confirmatory factor analysis to test the psychometric equivalence of one such metric—the Household Water Insecurity Experiences (HWISE) scale—across two survey waves administered 18 months apart in similar climatic conditions among households in a peri-urban community outside of Accra, Ghana. The HWISE metric was not equivalent across survey waves, which may be attributable to the metric itself, sample size, subtle instrumentation changes, or other unobserved factors. Test-retest reliability may also be difficult to achieve given the dynamic nature of household water use, and we discuss the implications of using household water insecurity metrics as longitudinal measures of well-being in global anti-poverty programs.

1. Introduction

Globally, 1.42 billion people face high, or extremely high, water scarcity (UNICEF, 2021), and the United Nations predicts that the number of people living in water-stressed or water-scarce countries will increase to 3.4 billion by 2025 (Alhassan and Kwakwa, 2014). Among the most pressing of these global water challenges is water insecurity: a lack of access to water, including limited access to water that is affordable, safe, and socially acceptable (Jepson et al., 2017). Water insecurity poses a significant risk to global health security, increasing population risks across a range of health issues, such as vector-borne diseases, gastrointestinal conditions, and injuries (Adams et al., 2020), as well as psychological distress (Gaber et al., 2020; Wutich et al., 2020a), and feelings of anger, frustration, embarrassment, and depression (Kangmennaang et al., 2020). Although water insecurity is present in many high-income settings (Meehan et al., 2020), most of its burden is borne by low- and middle-income countries and constitutes a major public health issue in regions such as sub-Saharan Africa (Prüss-Ustün et al., 2019). Water demand in West Africa, specifically, is expected to rise from 8 to 20% of the share of total African demand by 2050 (Burek et al., 2016), underscoring the importance of quantifying insecurity in the region and developing interventions to safely meet growing demand.

Efforts to measure household water insecurity more precisely have increased over the past decade, with the development of both locally-grounded and cross-site scales (Wutich, 2020). Many local water insecurity scales have been implemented in low- and middle-income countries (Octavianti and Staddon, 2021) but were developed for specific regions (e.g., Tsai et al. (2016) in Uganda; Stevenson et al. (2016) in Ethiopia; or Tomaz et al. (2020) in Brazil), without validation across multiple settings. The 12-item Household Water Insecurity Experiences scale (HWISE-12) represented a significant advancement because it was cross-culturally validated across 11 sites in different low- and middle-income nations (Young et al., 2019a, 2019b). A 4-item version (HWISE-4) was subsequently proposed that captures most of the variation in the original scale in a shorter form preferred by the development community (Young et al., 2021). The HWISE scale is a useful tool for researchers to monitor and evaluate water security over time and facilitates the comparison of data from distinct collection sites (Slaymaker

* Corresponding author. Department of Geography and Sustainable Development, University of Miami, Coral Gables, FL, USA.
E-mail address: stoler@miami.edu (J. Stoler).

https://doi.org/10.1016/j.ijheh.2022.113922
Received 2 November 2021; Received in revised form 27 December 2021; Accepted 6 January 2022
Available online 10 January 2022

et al., 2020). However, the HWISE scale does not include regionally specific items, which may provide more valid data within the context of specific communities. Shorter screener-type items have previously been implemented to assess basic issues (due to the lower respondent burden and ease of use in multicultural contexts) (Wutich et al., 2020b), and studies have begun to deconstruct the HWISE scale into short scales that reflect specific subdomains of household water insecurity, such as hygiene or water worry (Jepson et al., 2021). The tradeoff of screeners or short scales is that the questions are either less precise or provide narrow information about the nature of water insecurity and how it is experienced.

These differences aside, water insecurity can demonstrate considerable variation in sub-Saharan Africa not only seasonally, but even day-to-day given changes to drinking water availability associated with intermittency, limited vendor operating hours, or mechanical breakdown of water access points (Price et al., 2021). The sensitivity of these new household water insecurity metrics to these factors remains unknown, as these metrics are commonly tested and validated in cross-sectional studies. Given the dynamic nature of water insecurity in many contexts, it is important to establish test-retest reliability of these new household water insecurity metrics before using them for water resources policy and management.

The HWISE scale was originally created using measurement invariance techniques, which can be used to assess the psychometric equivalence of an instrument across two or more groups or points in time (Meredith, 1993). Invariance is typically tested by using confirmatory factor analysis to sequentially impose an increasing number of equality constraints across groups on measurement parameters (Meredith, 1993). These constraints force estimated parameters to take the same value across groups, testing the assumption that properties of the test are equivalent. First, configural invariance tests whether the number of factors, and the items which belong to those factors, are the same across time points. Metric invariance imposes constraints that require factor loadings to be equal. Factor loadings represent how strongly each item contributes to that factor, and violations of metric invariance suggest that the factor represents a different construct at different times. Scalar invariance is tested by adding equality constraints on item intercepts. Residual terms represent the amount of variance each item does not share with a factor, also described as measurement error. Residual invariance tests for differences in residual error for each item between models (or, in our application, over time) by adding equality constraints on item residuals (Meredith, 1993). In the context of measurement invariance across time, residual invariance is analogous to measuring test-retest reliability.

Ghana is a fitting setting for assessing the reliability of household water insecurity metrics; access to piped water was only 33% nationwide and 40% in urban areas (WHO/UNICEF, 2019). Ghana Water Company Limited (GWCL), the public water utility company responsible for the management and distribution of urban water supply, rations water to different urban areas to meet growing demand (Adams and Vásquez, 2019; Tutu and Stoler, 2016). Despite numerous urban water sector reforms, GWCL is unable to supply water on a consistent basis to the Greater Accra Region, meeting only about 70% of demand (Adams and Vásquez, 2019; Asante-Wusu and Yeboah, 2020). The insufficient water supply in the region has been attributed to both rapid urban growth and inadequate investment in water infrastructure (Kangmennaang et al., 2020; Tutu and Stoler, 2016). Greater Accra experienced the fastest population growth in the country, growing at 4.4% annually between 1984 and 2010 (Kangmennaang et al., 2020). Even with GWCL's implementation of pro-poor piped water expansion projects aiming to increase access to Accra's piped water network, most peri-urban areas in Greater Accra remain unconnected (Asante-Wusu and Yeboah, 2020).

This study presents household water insecurity data from a peri-urban community of Accra, Ghana. We apply measurement invariance methods to examine the psychometric properties, including temporal stability, of the HWISE scale between two survey waves—meaning that the scale measured the same underlying construct of water insecurity at baseline and upon revisit. Household water insecurity metrics intended to quantify water insecurity should be tested in a wider variety of contexts, and this repeated measure study offers a first glimpse at the HWISE scale's test-retest reliability. We close with discussion of the implications of this case study and general challenges with assessing the stability of household water insecurity metrics.

2. Methods

2.1. Study site

This study began in 2017 in Pokuase, a community located in the Ga West Municipal Assembly, one of the fastest-growing peri-urban Districts in Ghana's Greater Accra Region (Ashiagbor et al., 2019). Pokuase is part of a peri-urban stretch adjacent to Nsawam Road (N6, the Accra–Kumasi motorway, see Fig. 1), approximately 16 km northwest of Central Accra, that is home to nearly a quarter million residents (African Development Bank, 2021). Pokuase is also a growing transportation hub on the fringe of the Accra Metropolitan Area with a lively market and retail shopping area just north of the 4-tier Pokuase Interchange completed in 2021, a key project of the Accra Urban Transport Project (Ayeh-Paye and Amoako-Atta, 2019), and the largest stack interchange in West Africa.

Pokuase's water situation was typical of growing peri-urban communities in West Africa; many residents traditionally drew water from the local rivers, but cholera outbreaks in the early 2010s (Twumasi, 2011) shifted household water use toward well water and vended sources, particularly packaged drinking water. Pokuase was also one of several communities selected for a pro-poor piped water expansion project by GWCL that began in 2017, given its importance as a regional transportation corridor. For these reasons, Pokuase was selected as one of several water-stressed sites around the world surveyed for the parent HWISE study that led to the compilation of the HWISE scale (Young et al., 2019a, 2019b). Although water mains were installed along Pokuase's main roads by early 2019, the project was delayed for years due to bureaucracy and land disputes, and again in 2020 due to the COVID-19 pandemic.

2.2. Household survey

Survey data were collected in Pokuase in June 2017 and December 2018. The intent of these two survey waves was to first establish baseline water insecurity data before the GWCL project's completion, and to then assess test-retest reliability of select HWISE survey modules, particularly those related to the HWISE scale.

The sampling zones created to facilitate random sampling in Pokuase are presented in Fig. 1. We first enumerated the number of housing structures in each zone using aerial imagery, and then determined the proportion of households in each zone that would collectively yield the target total of 250 that was determined *a priori* by the parent HWISE study (Young et al., 2019b). Trained Ghanaian interviewers started at a corner of an assigned sampling zone and selected households using a mobile app that rolled a virtual pair of six-sided dice. The first die determined the number of households to skip, and the second dictated direction with the numbers 1–6 representing angles from 0 to 180°, thus simulating a random walk (with additional protocols for remaining within the sampling zone to avoid overlapping with another interviewer). The participant inclusion criteria were being at least 18 years old, being knowledgeable about water acquisition and use within their household, and consenting to participate (Young et al., 2019b).

The interview team used tablet computers to collect data on socio-demographics and core components of household water insecurity such as experiences with water availability, accessibility, reliability, and use (Jepson et al., 2017). The HWISE survey items asked each

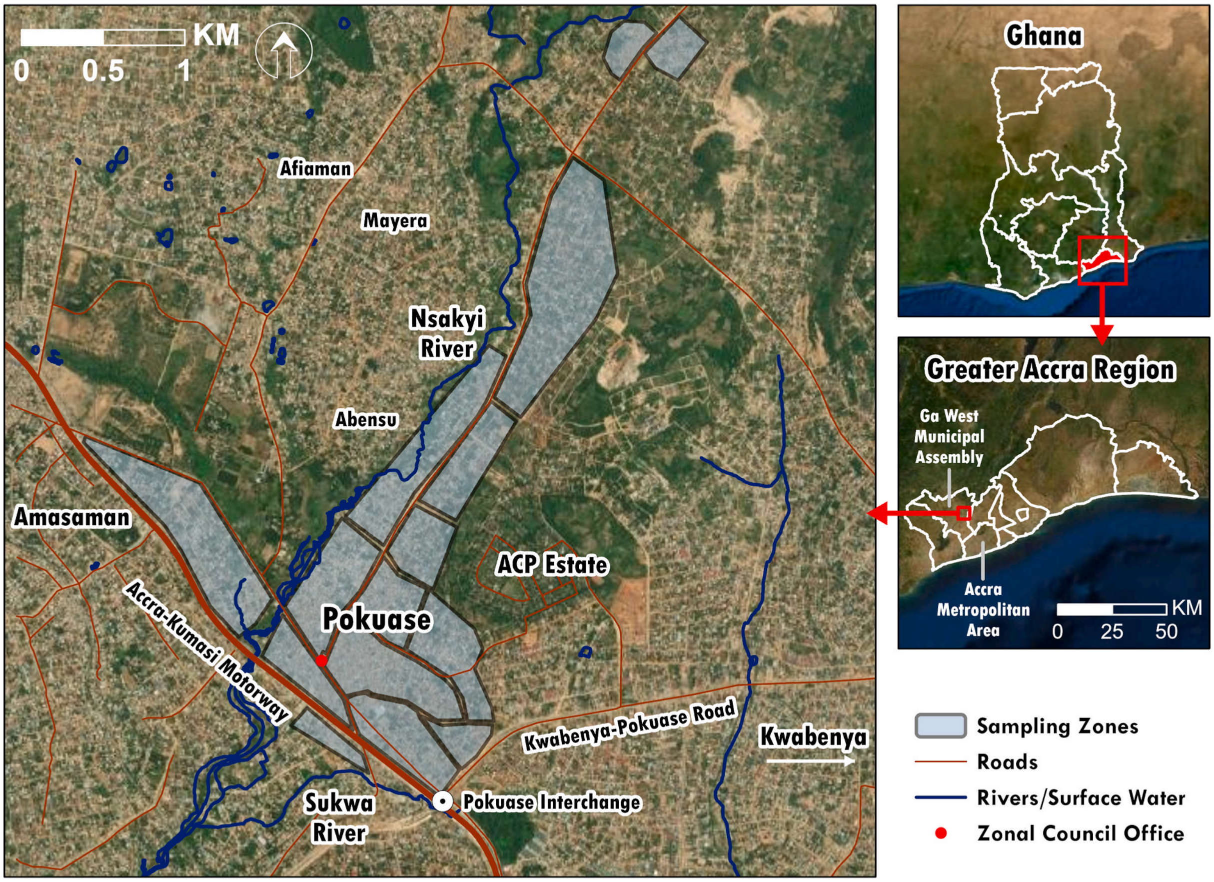

Fig. 1. The study community, Pokuase in Greater Accra Region, with sampling zones used to implement the household water insecurity survey.

participant to report the frequency of various household experiences related to water in the four weeks prior to survey implementation, with responses categorized as follows: never (0 times; scored as 0), rarely (1–2 times; scored as 1), sometimes (3–10 times; scored as 2), and often or always (respectively 11–20 times and more than 20 times; these are combined and scored as 3). For example, the *clothes* item was worded: "In the last four weeks, how frequently has there not been enough water in the household to wash clothes?" (see Table 1 for item phrasing). The HWISE-11 was computed by scoring and summing responses to the 12 original HWISE scale items excluding the *shame* item, which was added to the questionnaire during a survey refinement workshop in August 2017 after the baseline survey wave (Young et al., 2019a). The HWISE-11 has been used elsewhere and accounted for >99% of the variation in HWISE scale scores with minimal additional error (Stoler et al., 2020; Venkataramanan et al., 2020).

The household survey also included the 4-item Perceived Stress Scale (PSS-4), a measure of everyday stress (Cohen et al., 1983) that has been used in sub-Saharan Africa (e.g., Garcia et al., 2013; Lemma et al., 2012), but to our knowledge not validated there (Hjelm et al., 2017). The Household Food Insecurity Access Scale (HFIAS) (Coates et al., 2007), which has been validated in sub-Saharan African settings (e.g., Becquey et al., 2010; Desiere et al., 2015; Gebreyesus et al., 2015), was asked during the baseline survey, but was omitted for expediency during the follow-up wave because food insecurity scores in Pokuase were, fortunately, very low.

Although 229 Pokuase households participated in the baseline survey, we completed just 100 follow-up interviews, of which 98 had complete response data. The lower sample size was not attributable to participant factors typically associated with loss to follow-up, but rather due to budget constraints that prevented the project team from adequately reconnecting with participants who had especially busy work and domestic schedules.

2.3. Statistical analysis

We began by computing the mean and standard deviation for each HWISE-11 item in both survey waves, using Spearman's correlation analysis and paired *t*-tests to compare responses between waves. We also created scatterplots, incorporating a jitter function to account for the ordinal data, to visualize the reliability of each scale item between waves. We computed the same descriptive statistics for HFIAS and PSS-4 as well.

We used confirmatory factor analysis (CFA) to test the fit of a single factor model for the HWISE-11. Multiple-group CFA was then used to test three forms of measurement invariance across survey waves—configural, metric, and scalar invariance—with the ultimate goal of assessing residual invariance as a measure of test-retest reliability. In the case of non-invariance at any stage, partial invariance can be tested by removing constraints on non-invariant parameters (Byrne et al., 1989). To establish a partially invariant model, we sequentially remove constraints one parameter at a time and compare model fit to the previous invariant model. We retained models which fit significantly and meaningfully better after relaxing constraints. We compared model fit by using a chi-squared difference ($\Delta\chi^2$) test between models, subtracting the more constrained model's chi-squared statistic from that of the prior accepted model, using the difference in the models' degrees of freedom. A significant chi-squared difference test indicates that the model fit is significantly worse in the constrained model compared to the previous model. In addition to significant chi-squared difference tests, we interpreted decreases in the comparative fit index (CFI) $\geq$ 0.01 and increases in root mean square error of approximation (RMSEA) $\geq$ 0.01 as a meaningful decrease in model fit.

Table 1
Phrasing of the HWISE-11 survey items in 2017 and 2018.

Item	In the last four weeks …	
	2017	2018
Worry	how frequently did you or anyone in your household worry you would not have enough water for all of your household needs?	No change
Interrupt	how frequently has your household water supply from your main water source been interrupted?	No change
Clothes	how frequently has there not been enough water in the household to wash clothes?	No change
Plans	how frequently has the time spent getting water prevented you or anyone in your household from doing household chores (such as cooking, preparing food, washing clothes, etc.)?	how frequently have you or anyone in your household had to change schedules or plans due to problems with your water situation, such as problems getting or distributing water within the household? Activities that may have been interrupted include caring for others, doing household chores, etc.
Food	how frequently have you or anyone in your household had to change what was being eaten because there wasn't enough water (e.g., for washing foods, cooking, etc.)?	No change
Hands	how frequently have you or anyone in your household had to go without washing hands after dirty activities (e.g., defecating or changing diapers, cleaning animal dung) because you didn't have enough water?	how frequently have you or anyone in your household had to go without washing hands after dirty activities (e.g., defecating or changing diapers, cleaning animal dung) because of problems with water?
Body	how frequently have you or anyone in your household had to go without washing their body because there wasn't enough water?	No change
Drink	how frequently has there not been as much water to drink as you would like for you or anyone in your household?	No change
Angry	how frequently did you or anyone in your household feel upset about your water situation?	how frequently did you or anyone in your household feel angry about your water situation?
Sleep	how frequently have you or anyone in your household gone to sleep thirsty?	No change
None	how frequently has there been no water whatsoever in your household?	No change

We began by testing configural variance by constructing a CFA model with two factors: a factor representing the HWISE in 2017, and a separate factor representing the HWISE in 2018. Items only loaded onto factors from the same wave. To account for item autocorrelation, item residuals were correlated across waves. To confirm that the single-factor model of HWISE appropriately fit the data for each wave, we did not apply equality constraints.

We tested metric invariance by fully constraining the factor loadings to be equal across 2017 and 2018 in a single model. We only explored partial metric invariance if we observed worse model fit relative to the configural model. To establish a partially invariant model, we sequentially unconstrained the factor loadings of each item one at a time and compared each model fit to the configural model. When a model yielded improved fit compared to the fully constrained metric model, but was still significantly worse in fit compared to the configural model, we continued to relax constraints of additional factor loadings one at a time until we found a model with a non-significant change in the chi-squared statistic and minimal degradation of the CFI and RMSEA.

Then, we tested scalar invariance by constraining item intercepts to be equal across each factor, and we compared model fit relative to the partial metric model. We did not constrain the intercepts of items with non-invariant factor loadings, as revealed in testing for metric invariance. If we observed worse model fit relative to the partial metric model, we then fit a partial scalar model, in which we sequentially unconstrained intercepts for one item at a time and compared each model fit to the partial metric model.

Finally, we tested residual invariance by fully constraining item residuals to be equal across factors, and only tested a partial residual model if model fit was worse relative to the partial scalar model. Once acceptable fit was confirmed, the residual model helped us assess test-retest reliability by comparing the model's standardized loadings between 2017 and 2018.

We also tested single-factor model fit of the two additional scales in our survey, HFIAS (in the first survey wave only) and PSS-4 (in 2017 and 2018) to provide additional context around our assessment of HWISE-11. Given the wide use of these indicators, we expected that a good model fit of HFIAS—albeit just for 2017—and any evidence of test-retest reliability for PSS-4 would affirm the strength (or limitations) of these data and aid our interpretation of the results for HWISE-11.

2.4. Ethics statement

This study was approved by the Institutional Review Board of Northwestern University (protocol STU00204884), through an institutional agreement with the University of Miami, as part of the original study, "A novel tool for the assessment of household-level water insecurity: scale refinement, validation, and manual development."

3. Results

Table 2 summarizes select household characteristics of all survey participants in each survey wave. Relative to the full sample of 229 in 2017, the 98 participants we re-interviewed in 2018 were similar in terms of age, gender, ethnic distribution, gender of the household head, the number of children and adults in the household, the type of housing, and primary drinking and non-drinking water sources.

Table 3 presents the scores in each survey wave for each HWISE-11 item and the aggregated HWISE-11, HFIAS, and PSS-4 scores. All HWISE-11 items were significantly correlated between survey waves (Table 3). The means of the two survey waves were significantly different for four items: *interrupt* ($t = 3.29$, $p < .01$), *body* ($t = 2.15$, $p < .05$), *drink* ($t = 3.17$, $p < .01$), and *none* ($t = 2.23$, $p < .05$). The HWISE-11 scores from the two waves were significantly correlated ($\rho = 0.46$, $p < .001$), but the 2018 mean (6.0) was higher than the 2017 mean (4.3) ($t = 2.05$, $p < .05$). Notably, both of these means were slightly lower than the average HWISE-11 score of 6.95 across 27 sites from the original parent study (Stoler et al., 2021).

Fig. 2 presents scatterplots of the HWISE-11 and each individual survey item across survey waves. The intercept of each scatterplot's trend line in Fig. 2 suggests that, on average, households who reported zero or near-zero frequencies of water insecurity experiences in 2017 reported higher scores on most items in 2018. This is consistent with the higher mean HWISE-11 score in 2018. But the slopes, which are less than 45° for every item, suggest that this pattern disappears—or even reverses—at higher levels of water insecurity. This may also simply be an artefact of bounded scoring: if you reported a 0 for any item in 2017, there was nowhere to go but up, and vice versa if reporting a 3.

The configural model, without parameter constraints across groups, demonstrated good model fit (χ^2 (197) = 100.17, $p = 1.00$; CFI = 0.95, RMSEA = 0.02). In the configural model, the *interrupt*, *drink*, and *hands* items displayed the largest differences in loadings between 2017 and

Table 2
Demographic and household characteristics of the original 229 households surveyed in 2017 in Pokuase, Ghana, and the sub-sample of 98 households surveyed in 2018. Categorical measures are expressed as a percentage, and continuous measures as mean (M) and standard deviation (SD).

	N = 229	N = 98
Age	M = 37.39 (SD = 12.85)	M = 39.23 (SD = 11.96)
Female	78%	77%
Ethnicity		
Akan	40%	44%
Ewe	31%	35%
Ga Dangme	21%	16%
Other	7%	4%
Female Household Head	30%	31%
Number of Children in Household	M = 2.45 (SD = 2.78)	M = 2.56 (SD = 3.28)
Number of Adults in Household	M = 3.59 (SD = 3.22)	M = 3.53 (SD = 3.56)
Type of Housing		
Owned	51%	52%
Rented	38%	37%
Informal Settlement	7%	6%
Other	3%	5%
Primary Drinking Water Source		
Standpipe	3%	2%
Borehole	6%	5%
Protected Dug Well	2%	1%
Tanker Truck	<1%	0%
Sachet Water	86%	90%
Surface Water	2%	1%
Primary non-drinking water source		
Standpipe	13%	11%
Borehole	47%	44%
Protected Dug Well	14%	14%
Unprotected Dug Well	1%	1%
Rainwater Collection	3%	6%
Small Water Vendor	8%	14%
Tanker Truck	10%	7%
Surface Water	3%	2%

Table 3
Mean (M) and standard deviation (SD) of item and scale scores by survey wave with paired *t*-tests of the differences between means, and Spearman's correlations (ρ).

Survey Item or Scale	2017 n = 229 (M, SD)	2018 n = 98 (M, SD)	t	ρ
Drink	0.22 (0.60)	0.55 (0.91)	3.17**	0.23*
Hands	0.17 (0.69)	0.31 (0.75)	1.53	0.29**
Plans	0.34 (0.75)	0.33 (0.75)	−0.10	0.23*
Worry	1.01 (1.35)	0.96 (1.19)	−0.33	0.27**
Angry	1.00 (1.27)	1.14 (1.09)	1.04	0.37***
Body	0.28 (0.65)	0.48 (0.91)	2.15*	0.41***
Clothes	0.42 (0.80)	0.54 (0.92)	1.12	0.34***
Food	0.16 (0.47)	0.16 (0.55)	0.00	0.31**
Interrupt	0.44 (0.78)	0.81 (1.01)	3.29**	0.27**
None	0.32 (0.67)	0.56 (0.95)	2.32*	0.23*
Sleep	0.24 (0.70)	0.28 (0.62)	0.38	0.30**
HWISE-11 score (range 0–33)	4.33 (5.65)	6.02 (6.90)	2.05*	0.46***
HFIAS (food insecurity) score (range 0–27)	5.70 (5.77)			
Perceived Stress Scale score (range 0–16)	6.78 (2.97)	6.56 (3.25)	0.64	0.41***

*P < .05, **P < .01, ***P < .001.

2018 (Table 4). In the metric model, the loadings were fully constrained to be equal across 2017 and 2018, and we observed worse model fit relative to the configural model ($\Delta\chi^2$ (10) = 16.53, p = .08; ΔCFI = −0.03, ΔRMSEA = +0.01). The model was deemed non-invariant given the decrease in CFI of 0.01 and increase RMSEA of 0.01, though the change in the chi-squared statistic was acceptable. We therefore proceeded with a partial metric invariance model in which we sequentially unconstrained intercepts for one item at a time and compared each model fit to the configural model. When the *hands* and *drink* items were unconstrained, model fit was not significantly different from the configural model ($\Delta\chi^2$ (8) = 9.69, p = .28; ΔCFI < −0.01, ΔRMSEA < +0.01).

The scalar model, which is critical for justifying the use of a sum score (as opposed to a factor score), was significantly different from the partial metric model ($\Delta\chi^2$ (8) = 16.23, p = .039; ΔCFI = −0.01, ΔRMSEA = +0.01). Because of these differences, we then fit a partial scalar model, which we sequentially unconstrained intercepts for one item at a time and compared each model fit to the partial metric model. One item, *interrupt*, was non-invariant in this partial scalar model; by unconstraining its intercept, the resulting model was not significantly different from the partial metric model ($\Delta\chi^2$ (7) = 9.49, p = .219; ΔCFI < −0.01, ΔRMSEA < +0.01). We also see that the intercept increased for *interrupt* from 2017 to 2018, meaning that households reported higher frequency of interruptions in 2018 compared to 2017 given the same level of water insecurity.

The residual model tests for differences in residual error for each item between models, which is analogous to measuring test-retest reliability. The residual model is also the most sensitive to Type I errors, and thus the most stringent of the measurement invariance procedures used in this analysis. The residual variances were fully constrained to be equal across 2017 and 2018 in a single model—with the exception of *hands* and *drink*, which were non-invariant in factor loadings in the partial metric model, and *interrupt*, which had a non-invariant intercept in the partial scalar model—and we compared model fit relative to the partial scalar model. The residual model was non-invariant given the significantly higher chi-squared value and decrease in CFI of 0.01, though the increase in RMSEA was acceptable ($\Delta\chi^2$ (8) = 15.70, p = .047; ΔCFI = −0.01, ΔRMSEA < +0.01). Because of these differences, we then fit a partial residual model, in which we sequentially unconstrained residuals for one item at a time and compared each model fit to the partial metric model. When the *angry* item was unconstrained, model fit achieved comparable fit relative to the partial-scalar model ($\Delta\chi^2$ (7) = 11.33, p = .125; ΔCFI < −0.01, ΔRMSEA < +0.01). The non-invariant item *angry* had a slight change of wording between 2017 (which focused on the word *upset*) and 2018 (*upset* was replaced with *angry*), yet was more reliable in 2018 as demonstrated by the decrease in residual variance from 2017 to 2018 (standardized residual of 0.72 vs. 0.54). Fit indices for all measurement invariance CFA model tests are summarized in Table 5.

The HFIAS scale demonstrated good single-factor model fit for both the full 2017 sample of 228 households (χ^2 (27) = 30.98, p = .272; CFI = 0.99, RMSEA = 0.03) and 2017 subsample of 98 households corresponding to the 2018 wave (χ^2 (27) = 13.11, p = .989; CFI = 1.00, RMSEA <0.01). The PSS-4 demonstrated very poor single-factor model fit for the full 2017 sample (χ^2 (2) = 65.59, p < .001; CFI = 0.70, RMSEA = 0.37), and a single-factor model would not converge for the 2017 subsample of 98 households corresponding to the 2018 wave, which prevented us from assessing test-retest reliability. It did demonstrate good single-factor model fit for the 2018 wave (χ^2 (2) = 1.15, p = .563; CFI = 1.00, RMSEA <0.01).

4. Discussion

This study evaluated the HWISE-11 scale for test-retest reliability in Pokuase, Ghana. We found that the scale was invariant, but four items were non-invariant to varying degrees. These findings should not be over-interpreted, given that they come from a modest sample in just one study site, whereas the original HWISE-12 was based on thousands of households across 11 sites. But our results do suggest the need for more robust psychometric evaluation—particularly focused on reliability testing—of the HWISE scale using longitudinal data from multiple study sites with larger sample sizes.

The *hands* and *drink* items had to be unconstrained in our partial

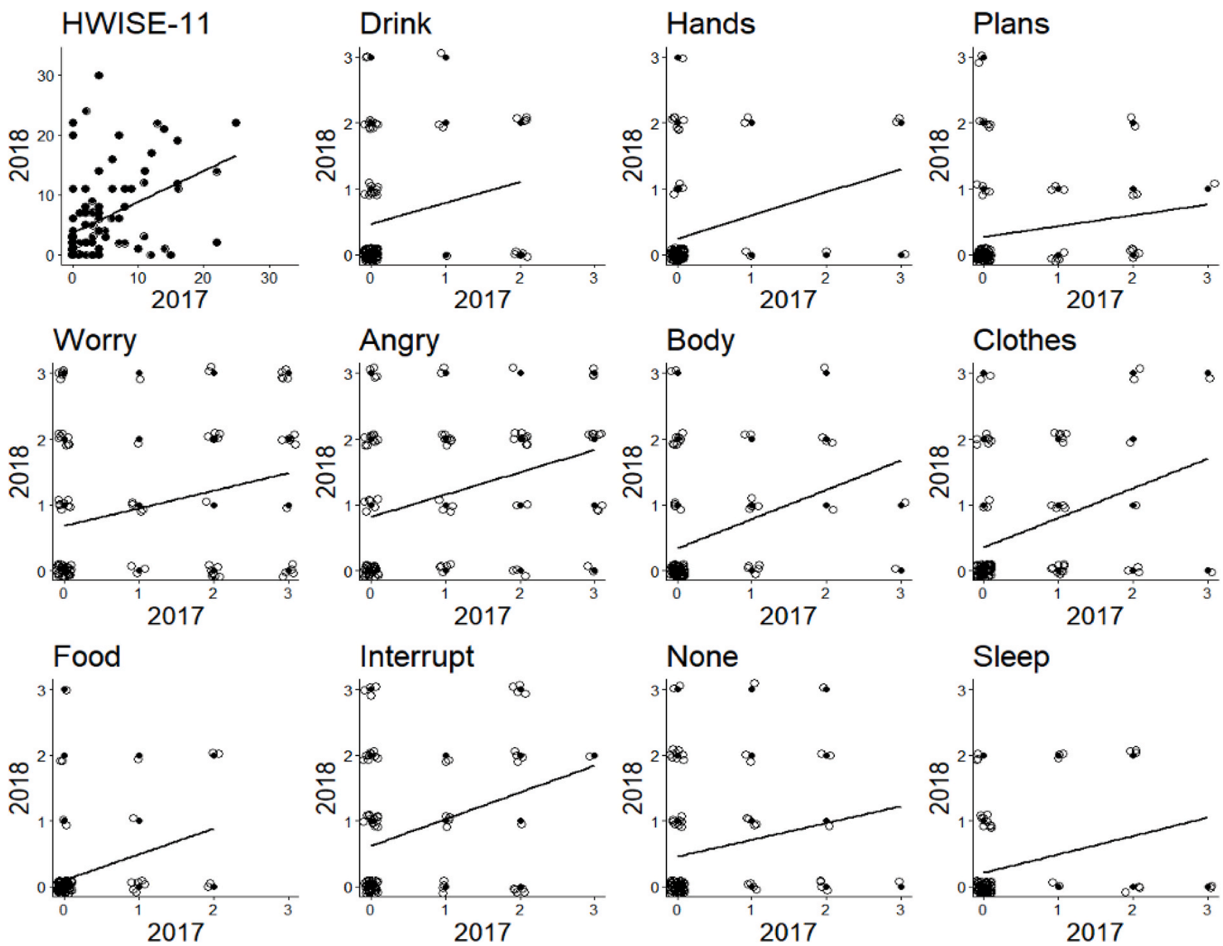

Fig. 2. Scatterplots of the HWISE-11 scores and each individual item between 2017 and 2018 survey waves in Pokuase, Ghana. The scatterplots incorporate a jitter function to visualize points that would otherwise be plotted on top of each other at each discrete value.

Table 4
Standardized factor loadings from the configural model of HWISE-11.

HWISE-11 Item	2017	2018
Drink	0.50	0.74
Hands	0.46	0.80
Plans	0.66	0.68
Worry	0.61	0.73
Angry	0.65	0.61
Body	0.70	0.66
Clothes	0.85	0.80
Food	0.49	0.58
Interrupt	0.52	0.73
None	0.65	0.78
Sleep	0.72	0.57

metric model to increase model fit. This non-invariance reveals the *hands* and *drink* items were stronger indicators of water scarcity in 2018 compared to 2017; however, it is important to note that the ending of the question for *hands* changed slightly between survey waves, going from more specific in 2017 (not having enough water) to less specific in 2018 (problems with water). The scalar model, which is critical for justifying the use of a sum score as opposed to a factor score, was also only partially invariant. The *interrupt* item was found to be non-invariant in the scalar model. This raises the potential for excluding *interrupt* from sum score versions of the HWISE scale if this finding were to persist in future studies.

The residual model, analogous to measuring test-retest reliability, is most sensitive to Type I errors, and thus the most stringent of the

Table 5
Fit indices for all measurement invariance models in the confirmatory factor analysis of the HWISE-11.

Factor	X^2	*df*	*P*	CFI	RMSEA	SRMR
HWISE-11						
Configural	100.17	197	1.00	0.95	0.02	0.08
Metric	153.54	207	0.99	0.92	0.03	0.10
Partial Metric	129.12	205	1.00	0.95	0.02	0.09
Scalar	136.98	213	1.00	0.94	0.02	0.09
Partial Scalar	133.45	212	1.00	0.95	0.02	0.09
Residual	144.20	220	1.00	0.94	0.02	0.10
Partial Residual	140.69	219	1.00	0.95	0.02	0.10

X^2 = chi-squared; *df* = degrees of freedom; *P* = *P*-value; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

measurement invariance procedures used in this analysis. The residual model was deemed non-invariant given the significantly higher χ^2 value and decrease in CFI of 0.01, though the increase in RMSEA was acceptable, and an argument could have been made to use the fully constrained residual model. We proceeded with a partial residual model, in which model fit improved when the *angry* item was unconstrained. The non-invariant item *angry* also had a slight change of wording between 2017 ("How frequently did you or anyone in your household feel upset about your water situation?") and 2018 ("How frequently did you or anyone in your household feel angry about your water situation?"), yet was more reliable in 2018 as demonstrated by the decrease in residual variance from 2017 to 2018 (standardized residual of 0.72 vs.

0.54). This may explain the need to unconstrain this item in order to improve model fit.

The single-factor models of food security (HFIAS) and perceived stress (PSS-4) did not ultimately provide information to help interpret the HWISE-11 results. HFIAS demonstrated good model fit, which we expected given that it has been validated in many international contexts. But because we did not collect these responses in our second survey wave, we could not attempt a configural model and proceed to assess test-retest reliability of HFIAS. Still, the single-factor model fit for 2017 increased our confidence in the data quality of our sample.

PSS-4 demonstrated good single-factor model fit for the 2018 wave. But we observed very poor model fit for the 2017 wave using the full sample, and lack of model convergence for the subset of 98 households that was needed to assess test-retest reliability. Because PSS-4 has not been validated in sub-Saharan Africa, these results were not completely surprising. These results provided no additional information regarding whether the non-invariance of HWISE-11 items is more likely attributable to the items or something about our sample.

Future reliability studies should focus on the four items identified in this analysis to be non-invariant. The slight change in question wording between survey waves may have contributed to our findings and should obviously be avoided (these changes were unfortunate artefacts of the original parent study that developed the HWISE scale), though the *plans* item was invariant in all models despite a similar change in phrasing. Notably, two of the non-invariant items, *hands* and *drink*, are included in the short-form HWISE-4. Because the HWISE-4 is intended to assess household water insecurity in a similar manner as the HWISE-12 (Young et al., 2021) and these scales are being implemented by nearly 100 organizations and researchers around the world (Institute for Policy Research, 2021), we recommend additional analysis of these items in other study sites.

The greater context surrounding any attempt to assess the temporal stability of any water insecurity metric is the dynamic nature of household water insecurity itself. One of the hallmarks of household water insecurity is source intermittency—i.e., temporal variability in water availability—which can have social, political, seasonal, and technical drivers (Galaitsi et al., 2016; Kumpel and Nelson, 2016). Intermittency can be compounded by variability in related factors such as water prices, vendor and domestic storage practices, work and domestic responsibilities, fetching times and risks (Smiley and Stoler, 2020). These factors ultimately interact and can lead to daily variability in water quality (Price et al., 2019, 2021), and fluctuating mental health burdens (Wutich et al., 2020a). Increased climate variability presents yet another driver in regions where water sources are most vulnerable to changing patterns of rainfall (Hall et al., 2014). Finally, novel household water insecurity metrics such as the HWISE scale capture the frequency of water insecurity experiences, but they do not capture household resiliency, i.e., how households adapt to those experiences, and their capacity for sustained adaptation. Two households affirming a particular dimension of water insecurity do not necessarily bear the same impact of that experience if they adapt differently; these adaptations may also vary subjectively vs. objectively, and over time. The integration of adaptation or adaptive capacity is an important consideration for the next generation of household water insecurity metrics.

Given these realities, it becomes even more important to consider test-retest reliability at multiple temporal scales. It is plausible (in lieu of contrary evidence) that a metric such as the HWISE-12 may be, for example, reliable within days, weeks, or even months within the same season, but unreliable between seasons or other social-environmental contexts. By no means would this necessarily invalidate the metric; rather, it helps us understand the extent to which metrics should be used for decision making. The dynamic nature of water insecurity forces us to acknowledge inherent limitations in the current generation of water insecurity metrics, especially in development contexts where absolute scores may be used for important resource allocation decisions. Much was made of the "nonsense statistics" formerly used to guide progress toward the MDGs (Nganyanyuka et al., 2014; Satterthwaite, 2003), and we must avoid reproducing these mistakes in monitoring progress toward the Sustainable Development Goals for water, sanitation, and hygiene.

5. Conclusion

The HWISE scale is an important tool for measuring water insecurity at the household level in low- and middle-income countries, and can contribute to clinical practice, public health, and policy decision making (Slaymaker et al., 2020). As water insecurity remains a significant public health issue globally, the lack of a "silver-bullet" cross-cultural, valid, and reliable tool for measuring water insecurity must not slow our progress in measuring and prioritizing communities in greatest need of water solutions. To avoid unintended consequences in development initiatives, we must recognize the limitations of household water insecurity metrics until we fully understand their respective equivalence and reliability across a wider sample of settings.

Declaration of competing interest

The authors report no declarations of interest.

CRediT authorship contribution statement

Melissa N. Sidote: Conceptualization, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Zachary T. Goodman:** Formal analysis, Resources, Writing – review & editing. **Christina L. Paraggio:** Writing – original draft, Writing – review & editing. **Raymond A. Tutu:** Conceptualization, Investigation, Writing – review & editing. **Justin Stoler:** Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Acknowledgments

We acknowledge support from the Household Water Insecurity Experiences Research Coordination Network (HWISE RCN) (National Science Foundation Award BCS-1759972), and from the Office of the Vice Provost for Research at the University of Miami. Neither funder had any involvement in this study. We thank Sera Young, Shalean Collins, and Joshua Miller for IRB and logistical support at Northwestern University; our interviewer team of Rita Antanah, Maxwell Akosah-Kusi, David Okai Nunoo, and Michael Nyoagbe; Richard Appiah Otoo at Ghana Water Company Ltd. for field support in Greater Accra; and Maria M. Llabre for methodological support.

References

Adams, E.A., Stoler, J., Adams, Y., 2020. Water insecurity and urban poverty in the Global South: implications for health and human biology. Am. J. Hum. Biol. 32, e23368.
Adams, E.A., Vásquez, W.F., 2019. Do the urban poor want household taps? Community preferences and willingness to pay for household taps in Accra, Ghana. J. Environ. Manag. 247, 570–579.
African Development Bank, 2021. Ghana - the Awoshie-Pokuase Road and Community Development Project. 46002-P-GH-DB0-012. Abidjan, Côte d'Ivoire. African Development Bank Group.
Alhassan, H., Kwakwa, P., 2014. When water is scarce: the perception of water quality and effects on the vulnerable. J. Water, Sanit. Hyg. Dev. 4, 43–50.
Asante-Wusu, I., Yeboah, I.E.A., 2020. Geography of water insecurity in three localities in the Accra-Tema City-Region, Ghana. African Geographical Review 39, 308–323.
Ashiagbor, G., Amoako, C., Asabere, S.B., Quaye-Ballard, J.A., 2019. Landscape transformations in rapidly developing peri-urban areas of Accra, Ghana: results of 30 years. Open Geosci. 11, 172–182.
Ayeh-Paye, S., Amoako-Atta, A., 2019. Report on the committee on roads and Transport on the annual budget estimates of the ministry of roads and highways for the 2020 financial year. In: Ministry of Roads and Highways, vol. 13. Ghana Parliament, Accra, Ghana.
Becquey, E., Martin-Prevel, Y., Traissac, P., Dembélé, B., Bambara, A., Delpeuch, F., 2010. The household food insecurity access scale and an index-member dietary

diversity score contribute valid and complementary information on household food insecurity in an urban West-African setting. J. Nutr. 140, 2233–2240.
Burek, P., Satoh, Y., Fischer, G., Kahil, M.T., Scherzer, A., Tramberend, S., et al., 2016. Water Futures and Solutions—Fast Track Initiative, Final Report. International Institute for Applied Systems Analysis, Laxenburg, Austria.
Byrne, B.M., Shavelson, R.J., Muthén, B., 1989. Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. Psychol. Bull. 105, 456.
Coates, J., Swindale, A., Bilinsky, P., 2007. Household Food Insecurity Access Scale (HFIAS) for Measurement of Food Access: Indicator Guide: Version 3. Food and Nutrition Technical Assistance Project Academy for Educational Development, Washington, DC.
Cohen, S., Kamarck, T., Mermelstein, R., 1983. A global measure of perceived stress. J. Health Soc. Behav. 385–396.
Desiere, S., D'Haese, M., Niragira, S., 2015. Assessing the cross-sectional and inter-temporal validity of the household food insecurity access scale (HFIAS) in Burundi. Publ. Health Nutr. 18, 2775–2785.
Gaber, N., Silva, A., Lewis-Patrick, M., Kutil, E., Taylor, D., Bouier, R., 2020. Water insecurity and psychosocial distress: case study of the Detroit water shutoffs. J. Public Health fdaa157.
Galaitsi, S.E., Russell, R., Bishara, A., Durant, J.L., Bogle, J., Huber-Lee, A., 2016. Intermittent domestic water supply: a critical review and analysis of causal-consequential pathways. Water 8, 274.
Garcia, J., Hromi-Fiedler, A., Mazur, R.E., Marquis, G., Sellen, D., Lartey, A., et al., 2013. Persistent household food insecurity, HIV, and maternal stress in Peri-Urban Ghana. BMC Publ. Health 13, 215.
Gebreyesus, S.H., Lunde, T., Mariam, D.H., Woldehanna, T., Lindtjørn, B., 2015. Is the adapted Household Food Insecurity Access Scale (HFIAS) developed internationally to measure food insecurity valid in urban and rural households of Ethiopia? BMC Nutrition 1, 2.
Hall, J.W., Grey, D., Garrick, D., Fung, F., Brown, C., Dadson, S.J., et al., 2014. Coping with the curse of freshwater variability. Science 346, 429–430.
Hjelm, L., Handa, S., de Hoop, J., Palermo, T., 2017. Poverty and perceived stress: evidence from two unconditional cash transfer programs in Zambia. Soc. Sci. Med. 177, 110–117.
Institute for Policy Research, 2021. Who Is Already Using the WISE Scales? Northwestern University.
Jepson, W.E., Stoler, J., Baek, J., Morán Martínez, J., Uribe Salas, F.J., Carrillo, G., 2021. Cross-sectional study to measure household water insecurity and its health outcomes in urban Mexico. BMJ Open 11, e040825.
Jepson, W.E., Wutich, A., Collins, S.M., Boateng, G.O., Young, S.L., 2017. Progress in household water insecurity metrics: a cross-disciplinary approach. WIREs Water 4, e1214.
Kangmennaang, J., Bisung, E., Elliott, S.J., 2020. 'We are drinking diseases': perception of water insecurity and emotional distress in urban slums in Accra, Ghana. Int. J. Environ. Res. Publ. Health 17, 890.
Kumpel, E., Nelson, K.L., 2016. Intermittent water supply: prevalence, practice, and microbial water quality. Environ. Sci. Technol. 50, 542–553.
Lemma, S., Gelaye, B., Berhane, Y., Worku, A., Williams, M.A., 2012. Sleep quality and its psychological correlates among university students in Ethiopia: a cross-sectional study. BMC Psychiatr. 12, 237.
Meehan, K., Jepson, W., Harris, L.M., Wutich, A., Beresford, M., Fencl, A., et al., 2020. Exposing the myths of household water insecurity in the global North: a critical review. WIREs Water 7, e1486.
Meredith, W., 1993. Measurement invariance, factor analysis and factorial invariance. Psychometrika 58, 525–543.
Nganyanyuka, K., Martinez, J., Wesselink, A., Lungo, J.H., Georgiadou, Y., 2014. Accessing water services in Dar es Salaam: are we counting what counts? Habitat Int. 44, 358–366.
Octavianti, T., Staddon, C., 2021. A review of 80 assessment tools measuring water security. WIREs Water 8, e1516.
Price, H.D., Adams, E., Quilliam, R.S., 2019. The difference a day can make: the temporal dynamics of drinking water access and quality in urban slums. Sci. Total Environ. 671, 818–826.
Price, H.D., Adams, E.A., Nkwanda, P.D., Mkandawire, T.W., Quilliam, R.S., 2021. Daily changes in household water access and quality in urban slums undermine global safe water monitoring programmes. Int. J. Hyg Environ. Health 231, 113632.
Prüss-Ustün, A., Wolf, J., Bartram, J., Clasen, T., Cumming, O., Freeman, M.C., et al., 2019. Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: an updated analysis with a focus on low- and middle-income countries. Int. J. Hyg Environ. Health 222, 765–777.
Satterthwaite, D., 2003. The Millennium Development Goals and urban poverty reduction: great expectations and nonsense statistics. Environ. Urbanization 15, 179–190.
Slaymaker, T., Johnston, R., Young, S., Miller, J., Staddon, C., 2020. WaSH Policy Research Digest Issue #15: Measuring Water Insecurity. WaSH Policy Research Digest. University of North Carolina, Chapel Hill, Chapel Hill, NC.
Smiley, S.L., Stoler, J., 2020. Socio-environmental confounders of safe water interventions. WIREs Water 7, e1438.
Stevenson, E.G.J., Ambelu, A., Caruso, B.A., Tesfaye, Y., Freeman, M.C., 2016. Community water improvement, household water insecurity, and women's psychological distress: an intervention and control study in Ethiopia. PLoS One 11, e0153432.
Stoler, J., Miller, J.D., Adams, E.A., Ahmed, F., Alexander, M., Asiki, G., et al., 2021. The household water insecurity experiences (HWISE) scale: comparison scores from 27 sites in 22 countries. J. Water, Sanit. Hyg. Dev. 11, 1102–1110.
Stoler, J., Pearson, A.L., Staddon, C., Wutich, A., Mack, E., Brewis, A., et al., 2020. Cash water expenditures are associated with household water insecurity, food insecurity, and perceived stress in study sites across 20 low- and middle-income countries. Sci. Total Environ. 716, 135881.
Tomaz, P., Jepson, W., Santos, J.d.O., 2020. Urban household water insecurity from the margins: perspectives from Northeast Brazil. Prof. Geogr. 72, 481–498.
Tsai, A.C., Kakuhikire, B., Mushavi, R., Vořechovská, D., Perkins, J.M., McDonough, A. Q., et al., 2016. Population-based study of intra-household gender differences in water insecurity: reliability and validity of a survey instrument for use in rural Uganda. J. Water Health 14, 280–292.
Tutu, R.A., Stoler, J., 2016. Urban but off the grid: the struggle for water in two urban slums in greater Accra, Ghana. African Geographical Review 35, 212–226.
Twumasi, P., 2011. Pokuase Sits on Time Bomb (Modern Ghana).
UNICEF, 2021. Reimagining WASH: Water Security for All. UNICEF, New York, NY.
Venkataramanan, V., Geere, J.A., Thomae, B., Stoler, J., Hunter, P.R., Young, S.L., et al., 2020. In pursuit of "safe" water: the burden of personal injury from water-fetching in 21 low- and middle-income countries. BMJ Global Health 5, e003328.
WHO/UNICEF, 2019. Progress on Household Drinking Water, Sanitation and Hygiene 2000-2017: Special Focus on Inequalities. World Health Organization.
Wutich, A., 2020. Water insecurity: an agenda for research and call to action for human biology. Am. J. Hum. Biol. 32, e23345.
Wutich, A., Brewis, A., Tsai, A.C., 2020a. Water and mental health. WIREs Water 7, e1461.
Wutich, A., Rosinger, A.Y., Stoler, J., Jepson, W., Brewis, A., 2020b. Measuring human water needs. Am. J. Hum. Biol. 32, e23350.
Young, S.L., Boateng, G.O., Jamaluddine, Z., Miller, J.D., Frongillo, E.A., Neilands, T.B., et al., 2019a. 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, e001750.
Young, S.L., Collins, S.M., Boateng, G.O., Neilands, T.B., Jamaluddine, Z., Miller, J.D., et al., 2019b. Development and validation protocol for an instrument to measure household water insecurity across cultures and ecologies: the Household Water InSecurity Experiences (HWISE) Scale. BMJ Open 9, e023558.
Young, S.L., Miller, J.D., Frongillo, E.A., Boateng, G.O., Jamaluddine, Z., Neilands, T.B., 2021. Validity of a four-item Household Water Insecurity Experiences Scale for assessing water issues related to health and well-being. Am. J. Trop. Med. Hyg. 104, 391–394.