Research Paper

Assessing peri-urban sanitation quality using a theoretically derived composite measure in Lusaka, Zambia

James B. Tidwell, Jenala Chipungu, Roma Chilengi and Robert Aunger

ABSTRACT

Despite ongoing debates about what constitutes adequate sanitation, there is a lack of sanitation quality measures that are theoretically grounded in ways that allow empirical comparisons of quality across different types of sanitation. The Healthy Sanitation Framework (HSF) was developed to capture universal aspects of sanitation quality from a public health perspective. From this, the Peri-Urban Healthy Toilet Index (PUHTI) was created for measuring on-site, peri-urban sanitation quality. This PUHTI score was used to assess sanitation quality in a peri-urban area in Lusaka, Zambia. The HSF identified five categories for capturing sanitation quality: *hygiene, use*, *sustainability, desirability,* and *accessibility*. A composite index derived from these categories had high reliability and plausible validity, despite barriers to rigorously evaluating validity. Applying the PUHTI tool showed that while 87% of toilets were classified as 'improved, but shared,' there were frequent concerns about doors that could not be locked, dirty user interfaces, unhygienic containment, limited emptyability, and lack of handwashing facilities. The HSF allows granular measures of sanitation quality to be developed in any setting using a reproducible and theoretically grounded process. However, lack of a unified basis on which to compare different types of sanitation overall or evidence to compare within narrower categories currently limits comparisons across types of sanitation.

Key words | measure, peri-urban, quality, sanitation, shared, toilet

James B. Tidwell (corresponding author)
Robert Aunger
London School of Hygiene & Tropical Medicine,
Keppel St, London WC1E 7HT,
UK
E-mail: *ben.tidwell@lshtm.ac.uk*

Jenala Chipungu
Roma Chilengi
Center for Infectious Disease Research in Zambia,
Plot 34620, Alick Nkhata Road, Lusaka,
Zambia

INTRODUCTION

Poor sanitation is a major public health problem, but it is difficult to establish priorities to address it without a detailed measure of sanitation quality to understand what is adequate for different types of toilets in different settings. The Sustainable Development Goal (SDG) for Sanitation (6.2) aims to eliminate open defecation (OD) by 2030, and the global indicator is the percentage of the population using safely managed sanitation services, defined as 'an improved sanitation facility that is not shared with other households, and where excreta are disposed of *in situ* or transported and treated off-site' (WHO and UNICEF 2017). As of 2015, about 4.5 billion people globally lacked access to safely managed sanitation, according to this standard, with 892 million still practicing OD (WHO and UNICEF 2017). An additional 600 million use a limited sanitation service, which is shared between households, but otherwise improved, and although rates of OD are decreasing globally, many are transitioning to limited sanitation, especially in sub-Saharan Africa

doi: 10.2166/washdev.2018.029

(UNICEF and WHO 2015). There is ongoing debate about whether some limited sanitation should be considered 'adequate' (Evans *et al.* 2017), as is stated in the target for SDG 6.2, and evidence that it may be no dirtier than (Exley *et al.* 2015) and structurally better than household sanitation (Jenkins *et al.* 2014a).

To understand when limited sanitation might be considered adequate, there is a need for more appropriate tools for measuring different types of sanitation, including household, shared, and public options, and for planning sanitation investments across different settings. Recent efforts to expand measurement beyond lists of 'improved technologies' have focused on functional definitions that capture broader ecological aspects of sanitation (Kvarnström *et al.* 2011), measuring variations in intra-household sanitation use (Jenkins *et al.* 2014b), gathering more detail about toilet quality (Jenkins *et al.* 2014a), and providing tools to understand community- and city-level sanitation status (Mehta & Mehta 2013). However, the major limitations of these tools are that they lack the foundation of a theoretical framework for sanitation quality needed for any composite measure (OECD 2008) or provide only a high-level overview (A detailed assessment of several existing tools is available in the supplementary material (Appendix A).) The gap between high-level definitions and comprehensive measurement tools also leads to the exclusion of many important components of sanitation, particularly those beyond the most straightforward implications of excreta management in the environment, such as privacy, safety, and sustainability.

To develop a detailed, comprehensive measure of peri-urban sanitation quality and examine its potential to be compared to measures of other types of sanitation, we created a general theoretical framework for sanitation abstracted from any particular context and used this framework to create a concrete composite measure of peri-urban sanitation quality via a systematic and reproducible process adaptable to other settings.

METHODS

First, a general framework for measuring sanitation quality from a public health perspective was derived. Using this framework as a guide, we then conducted a literature review by forward reference searching from documents defining Millennium Development Goal (MDG) and SDG monitoring as well as consulting ten sanitation measurement experts (Zambian and foreign academics as well as consultants) to identify other important measurement protocols. We identified key aspects of sanitation quality in our context, and chose appropriate measures from the literature where possible.

Based on this review, we created a composite measure of on-site, peri-urban sanitation quality and assessed its validity and reliability. Each field assistant independently observed the same set of toilets to assess measurement accuracy and reliability. The researchers agreed on 'correct' measures in cases where objective distinctions that did not change over time were possible. However, the subjectivity of some measures (e.g., floor and pan cleanliness), as well as variation over time in fly presence, meant that the team did not assess the accuracy of these variables. Krippendorf's alpha (Hayes & Krippendorff 2007) was used to capture inter-rater reliability for each item because of its robustness to missing data, and a standard adjustment for prevalence was used to identify the effect of sample prevalence on reliability results (Byrt *et al.* 1993). Several other kinds of reliability were not evaluated. Test-retest reliability will be assessed using baseline and endline data from the trial, as temporal variation in measures such as cleanliness are important considerations. As each item included in the final index was designed to capture a unique aspect of sanitation quality, internal consistency reliability was evaluated on a per-item basis when several potential measures existed (e.g., child feces disposal and handwashing with soap), but not across index items. The small number of potential measures also made assessing parallel-forms reliability infeasible.

Finally, we incorporated the PUHTI scoring tool into a baseline data collection for the 'Creating Demand for Peri-Urban Sanitation' (SanDem) trial to assess the existing quality of sanitation in the Bauleni community. A team of 24 research assistants were trained for 1 week and collected data from landlords and tenants on the same plots from 9 June to 6 July 2017. Plots were selected by a random walk from the center of demarcated zones, and plots with a landlord and at least one tenant household living on them were eligible for enrollment. More detail about the data collection process is available in the full study protocol (Tidwell *et al.* 2018). The trial was reviewed and approved by the University

of Zambia Biomedical Research Ethics Committee (ref: 023-06-16) as well as the London School of Hygiene and Tropical Medicine Ethics Committee (ref: 11714).

RESULTS

Generic framework

We propose the following working definition to motivate a 'Healthy Sanitation Framework' (HSF) with an explicit theoretical foundation from a public health perspective:

> *To have a significant impact on population-level health, any sanitation solution must be effective at reducing exposure to pathogens, desirable and accessible to its users so that it is used, and is usable for a reasonably long time.*

Breaking this statement down into elements suggests five major areas of public health concern: hygiene, desirability, accessibility, sustainability, and use. We define and describe each of these constructs as follows:

- *Hygiene* refers to the sanitation system separating human excreta from human contact. First, bodily wastes deposited into the environment must be safely contained. This covers aspects such as flushing toilets connected to sewer lines, lined latrine pits, and additional components of the sanitation value chain such as the transport, treatment, and safe reuse or disposal of waste. Second, residual excreta should be safely removed, which includes handwashing and anal cleansing as well as having cleanable interface surfaces like a ceramic pan or a concrete slab. Both the cleanability of human and toilet surfaces and the effectiveness of cleaning practices impact the hygiene of sanitation.
- *Desirability* encompasses all psychological factors associated with using a sanitation system. It must not be discomforting to use due to foul odor or nuisances (e.g., insects, lack of roof for shelter from bad weather). It should be private, so that users are not directly visible to others and private practices cannot be inferred. It should be reachable and usable safely without fear of predators (human or animal) or being injured during use, such as if a toilet collapses under the weight of a user because of poor construction. It should also be convenient, so that there are no major delays in use due to a high number of users or distance from daily activities of the user, and so that use fits into the user's daily routine. All of these aspects, while not related to the technological separation of excreta from the local environment, ensure that the sanitation solution is used, valued by its users, and improves public health.
- *Accessibility* means that the sanitation facility *can be* used by the largest possible number of 'in-group' members. In-group members may be members of a family, those living in an area with communal sanitation, or all children regardless of age or sex at a school. But, inappropriate users should be excluded since it may be undesirable for sanitation systems to be accessible to all without limitation, such as patrons at a local bar accessing a nearby private household's toilet (Mara 2016). Physical barriers should not prevent use, such as a full pit or a washed-out bridge preventing a person from walking to access a facility. Biological barriers such as disability due to age or illness should be overcome. Socio-cultural rules related to access, such as caste, gender, or kinship norms, must not prevent access. A solution must also be economically accessible, whether for direct costs of use (e.g., fees charged for public toilets) or indirect costs (e.g., availability of an acceptable sanitation solution at an affordable rental price).
- *Sustainability* includes both the ability to maintain good condition and functionality of the sanitation system and the ability to recover good condition when failures occur. Maintainability includes durability of facilities, a functioning system for waste treatment, and a functioning system for maintaining hygienic condition (e.g., a responsible family member, shared cleaning rota, or a paid individual). Recoverability spans a continuum of repairs, ranging from simply fixing a door that has come unattached to the feasibility of replacement of the entire system using locally available and affordable parts, especially if the system has been provided or subsidized by the government or a non-governmental organization.
- *Use* means that the appropriate population uses the sanitation system for the disposal of bodily waste. The disposal of bodily wastes can be either direct or indirect, such as depositing the excreta of children who use

'potties' or other processes developed for people unable to access the sanitation system directly.

In this framework, hygiene, use, and sustainability are the main measures of the quality of a sanitation system per se. If a system is not desirable, it is not used; if it is not used, it is irrelevant; if it is not hygienic, using it provides no health benefit; if it is not accessible, significant portions of the population will go unserved and population-level health indicators will not improve; if it is not sustainable, any health benefit will not continue into the future. The specific procedures used to aggregate a measure based on the HSF should weigh all of these aspects, as we describe in the PUHTI score development below.

Development of the Peri-Urban Healthy Toilet Index

Setting

The goal of developing the PUHTI score was to create a valid and reliable measure of individual on-site sanitation system quality impacting health, broadly defined to include both physical and psychosocial aspects. Operationalization of the HSF for a particular setting took place in Bauleni, a peri-urban area in Lusaka, Zambia, with a population of approximately 64,000. The compound is divided into plots originally intended to be occupied by one family, but owners have become de facto landlords with two to five tenant households on a typical plot. Most toilets have concrete slabs and dry pits, and almost all are shared between multiple families. These toilets were the primary focus, rather than household toilets, used only by one family, or communal/public toilets, where access is generally unrestricted and maintenance is rarely shared among all users. OD by those other than children was rare in the area. There is no sewerage present in the compound, although public and private pit emptying services were available. Few toilets could be connected to planned sewerage lines in their present form (Tidwell *et al.* 2017).

Adaptation of the framework into a usable index began with field-based qualitative formative research along with local expert consultation. The research was conducted to understand the state of sanitation in a high-density peri-urban area and the context within which it is practiced.

Measurement

An initial list of proposed indicators derived from the HSF for use in the PUHTI score were identified during our formative research, and existing quantitative measures for each element were selected. New measures were created when no suitable measure was found. (An in-depth review of available measures, including those not selected for this setting, is provided in the supplementary materials (Appendix B) for consideration in other contexts.)

Hygienic

As we aimed to assess on-site sanitation at the plot level, we measured the hygiene of waste containment and the safe disposal of solid waste. We assessed containment through landlord self-reports, as is common practice (Program 2014). Special consideration was given to menstrual hygiene material disposal as solid waste. Waste bins were uncommon inside toilets and reported menstrual waste disposal behavior is likely to be problematic, and so a specific indicator for menstrual hygiene management was not included. A variety of measures of interface cleanliness were included in piloting, including observation of the cleanliness of multiple components and the cleanability of materials, self-reported cleaning behaviors, and presence of relevant props (Alam *et al.* 2017). The presence of a place for handwashing with soap and water is recommended as a cost-effective measure for evaluations (Ram 2013). We included the 'hygiene ladder' version of this measurement, where a place for handwashing with just one of soap or water present receives partial credit. Formative research revealed that resources purchased individually were rarely shared, so materials for handwashing will necessarily be kept in individual homes. Thus, assessing the validity of this measure in the peri-urban context is needed. No anal cleansing behaviors were targeted for the intervention, thus no anal cleansing item was included.

Desirable

Desirability is assumed to be related to the ability of the sanitation facility to provide a motivating experience. Motives

mentioned in the formative research included comfort, privacy, and safety. Aspects of comfort included exposure to rain, presence of flies, and foul odor, with many overlapping measures available. To avoid rain and reduce internal contamination, presence of a roof without significant holes was observed. Nuisance insects and bad smell have both been found to affect use in peri-urban, shared latrines (Tumwebaze *et al.* 2013; Kwiringira *et al.* 2014). Insect traps and smell measurement devices were too costly for the study, and subjective measurement seemed unreliable. Measures of the intensity of fly presence and of solid covers, water-seals, or ventilation pipes to reduce smell were included.

Privacy is important, not just as a mediator of use in some contexts, but as it relates to stress caused by using a sanitation system (Sahoo *et al.* 2015). No quantitative measures associated with stress across the broad range of potential plot residents were found in the literature, so we measured the presence of a solid, attached door with an internal lock.

Safety from sexual abuse or violence would also be improved if such a door were present, and the other major fear of a toilet collapsing during use was alleviated by a lined pit (even if holes reduced its effectiveness for hygienic containment). No major issues of convenience were identified in formative investigations, and so no measure of perceived inconvenience was included in the index for this setting. Subsequent data collection revealed a fear of using the toilet at night, which did not come up in our semi-structured formative research interviews, but this oversight was corrected by including the presence of a light near the toilet in the score at endline (see Appendix B).

Accessible

Proper physical access requires that the 'in-group' are able to access the toilet, while outsiders cannot. All toilets were located on plots such that no respondent had to walk more than 20 m to access the toilet. But, to measure the impact of physical disabilities, landlords were asked about any residents currently unable to access the latrine on the plot due to disability. Excluding outsiders can be done through a variety of physical structures or social mechanisms, such as plot residents confronting outsiders. Simple observations of doors made of a solid material and functioning adequately along with the presence of a lock at the time of the observation were therefore included in the PUHTI score. We did not find any widespread exclusion for socio-cultural reasons and thus excluded it from the score. Economic access can be limited due to user fees for public sanitation, high costs of materials, and poor access to financial services such as loans, microcredit, or subsidies for sanitation. In this setting, communal facilities are rare, no formal financial services are available in the compound from public or private sources, and acquisition costs were unlikely to vary by plot, so no economic access measure was included in the PUHTI score.

Sustainable

Maintenance of good functionality in this context means that facilities are physically durable, waste can be treated or removed, and that there is an effective system for regular cleaning. A lined pit could be durable without being hygienic, so the type of lining was captured in the same question and scored separately for the two aspects of the PUHTI score. Since there is no sewerage system, waste is commonly dealt with by construction of a new pit latrine once the existing one fills or by emptying existing toilets. Self-reported mechanical emptyability was selected for the PUHTI score as the most appropriate measure, as landlords generally seemed aware of the kinds of trucks used for emptying in the compound, whether these trucks could access the plot, and whether there was sufficient room for access through the pan or otherwise; although depositing solid waste into pits certainly hinders emptyability, and was reported in our formative research. However, no evaluation of an existing measure was found and the high likelihood of inaccurate self-reporting by tenants led us to exclude this from consideration, with the understanding that this remains an important area for further investigation.

The most common system for maintaining cleanliness was a rota, where households took turns cleaning. Self-reported presence of a cleaning rota was therefore selected for initial consideration in the PUHTI score, but additional items capturing the duration of each household's turns and how many times a day the toilet was usually cleaned were included with observed cleanliness to assess the validity of the measure.

System recoverability is largely affected by the availability of construction materials locally. No latrine components distributed or installed by sanitation projects were observed and no other sanitation promotion or infrastructure development programs were reported by residents or government officials, so no indicator of recoverability was included in the PUHTI score.

Used

Proper measurement of latrine use is essential, as increases in the availability of facilities have been shown in some cases to be poorly correlated with increases in use (Barnard *et al.* 2013). However, OD is uncommon in this context. Thus, we assessed whether toilets were full and child feces disposal practices, which are a major public health concern (O'Connell 2015). Since the PUHTI score was derived from questions asked only of the landlord residing on the plot, the recommended practice of asking caregivers was precluded. Landlords were asked about child feces disposal on their plot, and we validated this measure against tenant responses. See Table 1 for a summary of the final list of included measures.

Table 1 | Final PUHTI score measures and weighting

Sub-scale	Points assigned	Weighting
Hygienic		
Cleanliness of interface	1: Water seal (+ previous) 0.8: Cleaning system in place(+ previous) 0.4: Interface made of cleanable materials (+ previous) 0.2: No visible feces	33%
Excreta hygienically contained	1: Concrete blocks and lining or septic tank 0.5: Concrete blocks and no lining	33%
Place for handwashing	1: Place for handwashing with soap and water 0.5: Place for handwashing with soap or water	33%
Sustainable		
Durable pit lining	1: Concrete blocks used	33%
Containment mechanically emptyable	1: Containment mechanically emptyable	33%
Cleaning system in place	1: System in place	33%
Used		
Toilet usable	1: Pit not full	80%
Child feces disposed into containment	1: All child feces goes into latrine or no children living on plot 0.67: Most child feces goes into latrine 0.33: Some child feces goes into latrine	20%
Desirable		
Few flies	1: 3 or fewer flies observed in toilet	14%
Solid door	1: Solid door present	14%
Inside lock	1: Solid door with internal lock present	14%
Solid walls	1: Concrete or wood	14%
Solid roof	1: Roof present	14%
Strong substructure	1: Concrete blocks used	14%
Smell reduction	1: Flushing toilet, ventilation pipe, or simple cover	14%
Accessible		
Disabled accessibility	1: Anyone living on plot unable to access toilet due to disability	50%
Outside lock to exclude outsiders	1: Solid door with outside lock present	50%

Aggregation

The weighting and aggregation of variables into composite measures like an index or scale can have a significant impact on the relative rankings given to the things being measured. An evidence-based 'common currency' for health impact was generally lacking, as the diarrheal disease impact of specific components is not well established and a QALY-like system for combining categories such as physical and psychological health into a single measure is not available. We therefore decided to construct one measure for each theoretically distinct aspect of the sub-scales and to use equal weights for measures within categories, unless there was a clear justification for doing otherwise.

A simple arithmetic mean was used to combine the five categories of the PUHTI score. This meant that a toilet that received low marks in one category could still receive a high overall score, as opposed to being strongly penalized like the geometric mean used in other community-level sanitation measures (Hawkins & Muximpua 2015). This allows an otherwise high-quality toilet that happened to be full (and thus receive a score of zero for Use) to avoid a drastically reduced score, because only a small change would be required to achieve a high-quality toilet.

Reliability and validity

The only item dropped from the scale due to reliability issues was the presence of holes in toilet walls, as issues around what size of opening at what height and with what level of intentionality (e.g., a window/hole for ventilation) were insurmountable. The final value of Krippendorf's alpha for the scale was 0.885, considered highly reliable (Klaus 1980). (Additional details on this process can be found in the supplementary material.)

While the criterion and content validity of the PUHTI score were ensured by the theory-based derivation of the measures, there is no concrete benchmark for evaluating the construct validity of the PUHTI score, as there is no simple measure of public health, broadly defined, and there is even limited evidence for the associations of individual components with diarrheal disease outcomes. (Additional details on this process can be found in the supplementary material.) Tenant satisfaction with aspects of sanitation on the plot (e.g., privacy) was therefore correlated with observed PUHTI measures (e.g., presence of a solid door) to test measure validity. All associations were found to be statistically significant other than questions about if anyone on the plot was excluded due to disability for socio-cultural reasons, which were not specifically about only the respondent. Adjusted R^2 values were generally between 0.2 and 0.4 when looking at the relationship between specific components and general feelings, which suggested reasonable validity.

A single summary measure for child feces disposal on the plot was collected from the landlord, along with items capturing OD rates and disposal practices from a tenant on each plot for comparison. Landlords ($n = 1{,}096$) reported that approximately 19.5% of child feces remained on the plot, while tenants ($n = 1{,}085$) reported that only 5.5% of their own child's feces did so, for a difference of 14.0 percentage points (95% CI: 11.3–16.8, $p < 0.001$). It is likely that landlords are more likely to notice feces remaining on the plot without a strong understanding of the underlying denominator of the total amount produced on the plot, but parents may also face a social desirability pressure to report less OD on the plot.

Final scale definition and characteristics

The final PUHTI score was assembled from the items whose piloting is described above with only a few minor modifications. The multi-item toilet cleanliness measure combines a lowest rung of having no visible feces, a second rung of having cleanable materials (as those with no visible feces are still likely to be highly contaminated if made of uncleanable materials), and a highest rung of a water seal, implying no contact with excreta spread by flies from the containment. For the desirability sub-scale, seven items were combined with equal weights, regardless of the underlying sub-scale component (e.g., privacy or safety). Desirability consisted entirely of observed measures other than one related to the sub-structure and had a reliability of 0.810 (using a ratio PABAK).

Bauleni situational analysis

We tabulated scores for each PUHTI measure for 918 toilets in Bauleni. The median PUHTI score was 0.663

[IQR: 0.541–0.770]. A detailed analysis of PUHTI measures revealed substantial differences in the presence of individual components and deficits in important aspects of each of hygiene, accessibility, desirability, and sustainability (Figure 1). Few problems with use were reported, and most plots had a toilet with solid walls (87%), a solid door (73%), and a lined pit (80%). About half did not have a solid roof (52%) or a functioning cleaning system (43%), and few were mechanically emptyable (36%) or used a smell-reduction technology (29%). Almost none had a place for handwashing with soap and/or water present (11%). Accessibility to in-group users was not a problem for most, but many toilets (59%) were unprotected from use by outsiders and offered no privacy or safety of an inside lock to users (65%).

DISCUSSION

Developing sanitation measures

Both routine monitoring and research measures should be derived from a sound theoretical framework via a rigorous and transparent development process. The Health Sanitation Framework (HSF) is proposed as the basis for such a process, but there may be arguments for including or excluding constructs or combining them in alternative ways. We hope that the HSF will serve as a starting point and catalyst for a wider discussion and contribute to an eventual consensus on a general framework for sanitation quality.

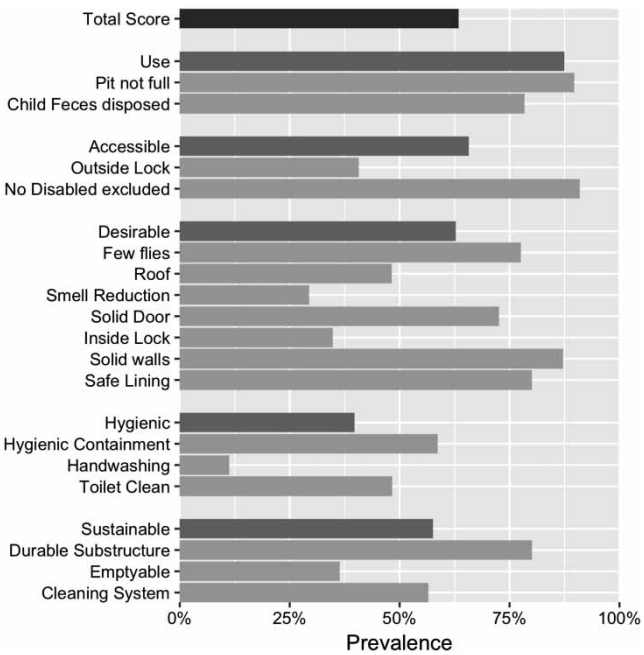

Figure 1 | Bauleni Toilet Quality Summary by overall PUHTI score, five sub-score, and measures.

One useful aspect of coming to consensus on a framework like the HSF would be to develop measures for different types of toilets (e.g., shared, household, public). However, overall comparison would require an objective standard (e.g., diarrheal disease), which we see as using a subset of its abilities, or a development of a QALY-like common currency. Perhaps better would be to look at specific measures that could be directly translated across types and settings (e.g., cleanliness) or that might be measured in different ways, but could still be compared functionally (e.g., a variety of technologies geared towards creating privacy, with subjective psychometric assessment).

The Peri-urban Healthy Toilet Index (PUHTI) score was developed as an outcome for a specific trial, rather than a reusable tool in other settings. However, its development process can be replicated in other settings and for other composite measures, and the resulting PUHTI score should be easily adapted to other peri-urban contexts. The general steps of (1) deriving context-specific adaptations of constructs from a general theoretical framework, (2) selecting and validating measures of those constructs, and (3) creating and justifying a composite measure are uncontroversial, but rare in practice. Many other comprehensive measures of sanitation are not amenable to criticism due to the opaque underlying theoretical links.

Item measures are commonly used across contexts, despite the strong effect that underlying populations have on measure validity and reliability. Even seemingly straightforward measures, such as the presence of a roof, showed significant unreliability in initial pre-testing during this process, demonstrating the need for the explicit analysis of validity and reliability in almost any sanitation research project.

Composite measures are often aggregated by simple addition, and while detailed empirical data for justifying alternative weightings may be lacking, theoretical

overlaps between measures can inform weighting of subscores or the number of measures included. As discussing what makes shared sanitation 'healthy' becomes a more acceptable position in the public health discourse, theoretically sound and well-constructed measures will be necessary to make sound, data-driven decisions (Evans *et al.* 2017).

The PUHTI approach and item selection

Several difficult choices had to be made related to individual measures, the selection of respondents, and the approach to data collection taken in attempting to develop a comprehensive, plot-level measure of sanitation quality. Some constructs simply have no commonly accepted standard of measurement. Handwashing measurement has been discussed extensively in the literature, but the applicability of approaches such as observing a place for handwashing near the toilet is unclear in a shared-sanitation context. The presence of flies exhibits such inter-temporal variability that it seems unlikely any point observation will be reliable. Toilet cleanliness has been judged using a wide range of techniques, from observation of feces or materials used in the toilet to microbiological testing. However, standardized valid and reliable measures for such variables are needed.

The public health importance of other measures is also unclear. However, each item included could plausibly be linked to at least one item of broad health importance, and granular measurement allows for both a variety of analytical approaches depending on what an investigator favors as well as re-calculation as the evidence base expands. Hygienic containment is relatively easy to define, but the actual disease risk to a population depends on characteristics of the soil, water table, and water sources used by residents.

Other measures are complicated or compromised by directing the PUHTI score tool only at landlords. The standard practice is to ask caregivers about child feces disposal for example, and other measures like emptyability may be better judged by local technicians than by landlords who may be unaware of available technologies. However, the decision was made to allow the most efficient collection of data for the largest number of indicators, and special attention was paid to the triangulation of items for which simple observation or landlord reports were not fully trusted.

Some constructs may simply be impossible to measure accurately at the plot level in peri-urban settings, such as measuring access to toilets by those who have a disability. Although we asked landlords if any current residents were prevented from accessing the toilet, inaccessible toilets may have prevented people from living on a plot in the first place, so a straightforward measure of those currently living on a plot unable to access the toilet obscures the scope of the problem. But, few good alternatives exist for measuring this construct at the plot level. Measuring the accessibility of every toilet based on standardized construction parameters would be infeasible and excessive. Additional questions could be asked of the landlord to understand if people with disabilities have ever been prevented from living on the plot, have ever been accommodated by individual equipment, or have ever left due to damage to such equipment. However, these may be poor proxies to capture the size of the problem. Such issues may best be measured using alternative data collection procedures.

Despite the limitations and challenges identified above, the PUHTI score and a majority of individual items are reliable, valid, and easy to collect, and provide more detail than is currently collected on a range of important aspects of on-site, peri-urban sanitation quality.

Policy implications for Bauleni

Almost all toilets observed in Bauleni were shared, but would be considered otherwise improved – however, a wide variety of quality levels was found in the situational analysis. Although most toilets were not full and could be used, the lack of effective cleaning systems or handwashing facilities meant that hygiene in many of these shared facilities is a major concern. Further, serious issues of poor desirability (in particular, a lack of roofs, inside locks, or smell reduction technologies) make using these toilets unpleasant. As the government target for the area is sewered connections for all, these detailed data demonstrate how far the quality of these toilets needs to be improved even if sewerage is provided by the government and the local utility company to allow connections as well as to truly offer high-quality sanitation to peri-urban residents.

CONCLUSION

The theoretically grounded, general Health Sanitation Framework developed here provides a strong foundation for assessing sanitation quality. It also allows the rigorous development of population- and situation-specific measures of sanitation quality. The framework was used here to create the Peri-urban Health Toilet Index (PUHTI) score for investigating peri-urban on-site sanitation services, which was also used as an outcome measure for the SanDem trial (Tidwell *et al.* 2018)).

These developments have highlighted several important measurement challenges remaining in peri-urban contexts, but the transparent process used here openly acknowledges the tradeoffs made along the way and provides a roadmap for both future measurement research and a reproducible set of steps for creating similarly high-quality measures in other settings. Measuring sanitation quality using the PUHTI score has allowed for a detailed understanding of barriers to improving peri-urban sanitation in Lusaka in particular, and can enable policymakers to better understand their own contexts and select the most effective approaches to improve sanitation in diverse settings globally. More work is needed to provide a basis to compare different types of sanitation in the aggregate and stronger granular evidence to compare within narrower categories.

ACKNOWLEDGEMENTS

The SHARE consortium funded this research and approved the overall study design, but played no role in collecting, analyzing, or interpreting the data; writing the report; or deciding to submit this article for publication.

REFERENCES

Alam, M. U., Winch, P. J., Saxton, R. E., Nizame, F. A., Yeasmin, F., Norman, G., Masud, A. A., Begum, F., Rahman, M. & Hossain, K. 2017 Behaviour change intervention to improve shared toilet maintenance and cleanliness in urban slums of Dhaka: a cluster-randomized controlled trial. *Tropical Medicine & International Health* **22** (8), 1000–1011.

Barnard, S., Routray, P., Majorin, F., Peletz, R., Boisson, S., Sinha, A. & Clasen, T. 2013 Impact of Indian Total Sanitation Campaign on latrine coverage and use: a cross-sectional study in Orissa three years following programme implementation. *PLoS One* **8** (8), e71438. DOI: 10.1371/journal.pone.0071438.

Byrt, T., Bishop, J. & Carlin, J. B. 1993 Bias, prevalence and kappa. *Journal of Clinical Epidemiology* **46** (5), 423–429.

Evans, B., Hueso, A., Johnston, R., Norman, G., Pérez, E., Slaymaker, T. & Trémolet, S. 2017 Limited services? The role of shared sanitation in the 2030 Agenda for Sustainable Development. *Journal of Water Sanitation and Hygiene for Development* **7** (3), 349–351. DOI: 10.2166/washdev.2017.023.

Exley, J. L. R., Liseka, B., Cumming, O. & Ensink, J. H. J. 2015 The sanitation ladder, what constitutes an improved form of sanitation? *Environmental Science and Technology* **49** (2), 1086–1094. DOI: 10.1021/es503945x.

Hawkins, P. & Muximpua, O. 2015 *Developing Business Models for Fecal Sludge Management in Maputo, Water and Sanitation Program*. World Bank Group, Washington DC.

Hayes, A. F. & Krippendorff, K. 2007 Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures* **1** (1), 77–89.

Jenkins, M. W., Cumming, O., Scott, B. & Cairncross, S. 2014a Beyond 'improved' towards 'safe and sustainable' urban sanitation: assessing the design, management and functionality of sanitation in poor communities of Dar es Salaam, Tanzania. *Journal of Water, Sanitation and Hygiene for Development* **4** (1), 131–141. DOI: 10.2166/washdev.2013.180.

Jenkins, M. W., Freeman, M. C. & Routray, P. 2014b Measuring the safety of excreta disposal behavior in India with the new Safe San Index: reliability, validity and utility. *International Journal of Environmental Research and Public Health* **11** (8), 8319–8346. DOI: 10.3390/ijerph110808319.

Klaus, K. 1980 *Content Analysis: An Introduction to its Methodology*. Sage Publications, Thousand Oaks, CA, USA.

Kvarnström, E., McConville, J., Bracken, P., Johansson, M. & Fogde, M. 2011 The sanitation ladder – a need for a revamp? *Journal of Water, Sanitation and Hygiene for Development* **1** (1), 3–12. DOI: 10.2166/washdev.2011.014.

Kwiringira, J., Atekyereza, P., Niwagaba, C. & Gunther, I. 2014 Descending the sanitation ladder in urban Uganda: evidence from Kampala Slums. *BMC Public Health* **14**, 624. DOI: 10.1186/1471-2458-14-624.

Mara, D. 2016) Shared sanitation: to include or to exclude? *Transactions of the Royal Society of Tropical Medicine and Hygiene* **110** (5), 265–267. DOI: 10.1093/trstmh/trw029.

Mehta, M. & Mehta, D. 2013 City sanitation ladder: moving from household to citywide sanitation assessment. *Journal of Water, Sanitation and Hygiene for Development* **3** (4), 481–488. DOI: 10.2166/washdev.2013.134.

O'Connell, K. 2015 *Management of Child Feces: Study Design and Measurement Tips*. WSP Sanitation Marketing Toolkit. World Bank Group, Washington DC.

OECD 2008 *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD, Paris, France.

Program, W. B. W. a. S. 2014 *Study Design and Questionnaire Tips for Sanitation Research*. Sanitation Marketing Toolkit. Retrieved from http://www.wsp.org/sites/wsp.org/files/publications/Study-Design-and-Questionnaire-Tips-for-Sanitation-Research.pdf.

Ram, P. K. 2013 *Practical Guidance: Measuring Handwashing Behavior: 2013 Update*. World and Sanitation Program, Washington DC.

Sahoo, K. C., Hulland, K. R., Caruso, B. A., Swain, R., Freeman, M. C., Panigrahi, P. & Dreibelbis, R. 2015 Sanitation-related psychosocial stress: a grounded theory study of women across the life-course in Odisha, India. *Social Science and Medicine* **139**, 80–89. DOI: 10.1016/j.socscimed.2015.06.031.

Tidwell, J., Chipungu, J., Chilengi, R. & Aunger, R. 2017 Theory-driven formative research on sanitation quality and improvement preferences among tenants and landlords in peri-urban Lusaka, Zambia. (in preparation).

Tidwell, J., Bosomprah, S., Chipungu, J., Chilengi, R., Curtis, V. & Aunger, R. 2018 *Creating Demand for Peri-Urban Sanitation Project Website*. Retrieved February 13, 2018, from http://bentidwell.com/sandem/.

Tumwebaze, I. K., Orach, C. G., Niwagaba, C., Luthi, C. & Mosler, H. J. 2013 Sanitation facilities in Kampala slums, Uganda: users' satisfaction and determinant factors. *International Journal of Environment Health Research* **23** (3), 191–204. DOI: 10.1080/09603123.2012.713095.

UNICEF and WHO 2015 *Progress on Sanitation and Drinking Water – 2015 Update and MDG Assessment*.

WHO and UNICEF 2017 *Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and SDG Baselines*. Geneva, Switzerland.

First received 20 February 2018; accepted in revised form 3 July 2018. Available online 27 July 2018