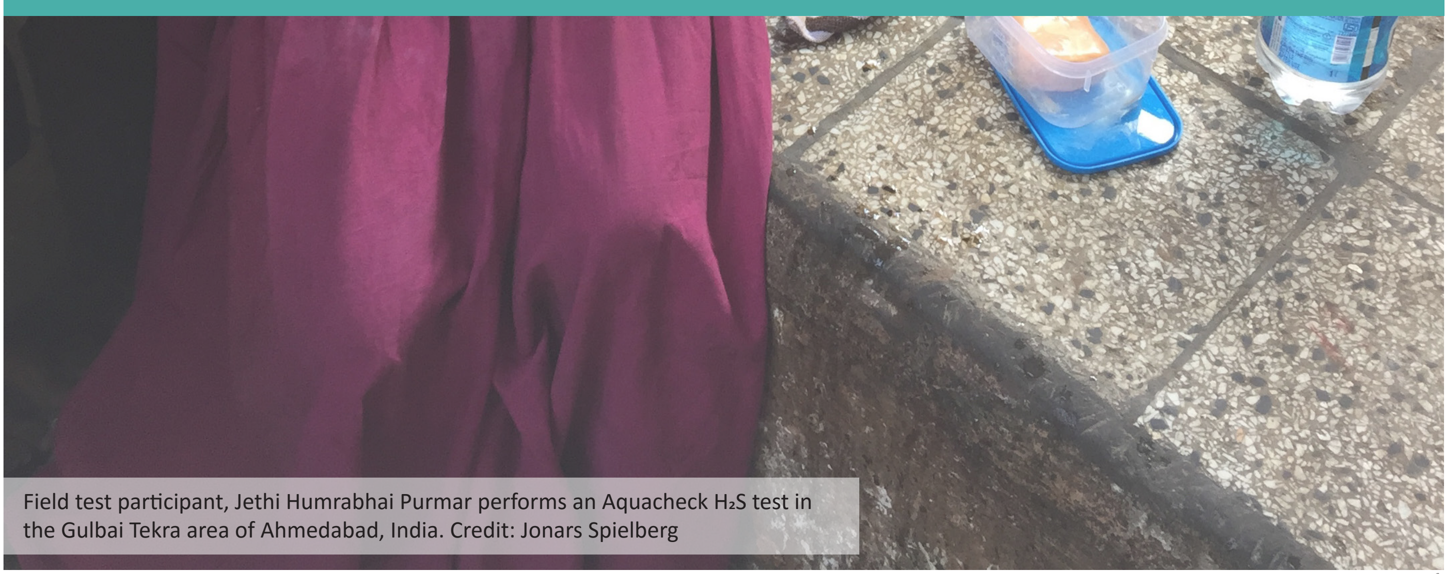
Comprehensive Initiative on Technology Evaluation

Streamlining a Methodology for Product Evaluation

Water Test Kits in Ahmedabad, India



Field test participant, Jethi Humrabhai Purmar performs an Aquacheck H_2S test in the Gulbai Tekra area of Ahmedabad, India. Credit: Jonars Spielberg

This report was produced by the Comprehensive Initiative on Technology Evaluation at the Massachusetts Institute of Technology and made possible through support of the United States Agency for International Development. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the United States Agency for International Development or the US Government.

Read more about our work and other product evaluations at cite.mit.edu.

Table of Contents

INTRODUCTION 3

WHY WATER TEST KITS? 3

HOW TO USE THIS REPORT 4

PRODUCT EVALUATION OVERVIEW 4

METHODOLOGY AT A GLANCE 5

FINDINGS AT A GLANCE 6

SINGLE-PARAMETER WATER TEST KITS 7

MULTI-PARAMETER WATER TEST KITS 16

INCORPORATING LEAN RESEARCH PRINCIPLES 24

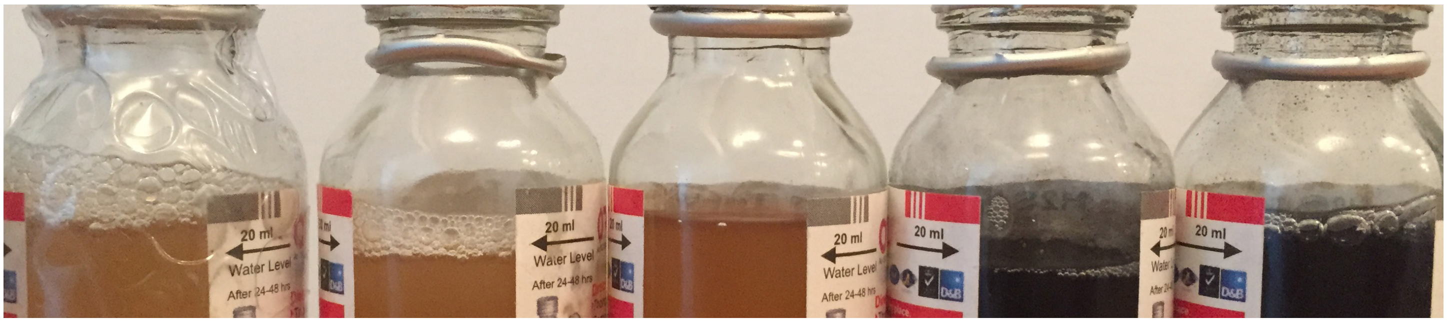
CONCLUSIONS AND RECOMMENDATIONS 24

AUTHORS & ACKNOWLEDGEMENTS 27

Introduction

The Comprehensive Initiative on Technology Evaluation (CITE) at MIT is the first-ever program dedicated to developing methods for product evaluation in global development. CITE is led by an interdisciplinary team at MIT, and draws upon diverse expertise to evaluate products and develop a deep understanding of what makes different products successful in emerging markets. Our evaluations provide evidence for data-driven decision-making by development workers, donors, manufacturers, suppliers, and consumers themselves.

From 2014 to 2015, CITE researchers evaluated single- and multi-parameter water test kits on the market in India's Gujarat State in efforts to streamline CITE's existing evaluation approach, which combines analysis of the suitability, scalability and sustainability of products designed for global development interventions. The team worked closely with students at local universities and NGOs.



Results from single parameter, hydrogen-sulfide (H_2S) water tests taken from samples throughout Ahmedabad, India. Credit: Jonars Spielberg

Why Water Test Kits?

According to the United Nations, 94 percent of India's population has access to an improved water supply.[1] However, using an improved water source does not guarantee that water is safe, and while 91 million in India may be "water poor," many more are likely "*safe* water poor."[2,3] The adverse health effects of unsafe water consumption are several: gastrointestinal illness, reproductive problems, neurological disorders, and developmental issues. How, then, can households regularly discern whether their water is safe to drink? In locations without access to reliable, low cost water testing labs, development organizations and national governments have looked to portable water test kits to answer this question.

There were three main goals for the water test kit evaluation:

1) Advance and streamline CITE's product evaluation approach by applying tailored mixed research methods to an in-process evaluation

2) Evaluate single- and multi-parameter water test kits & develop comparative ratings

3) Expand CITE outreach and partnership activities

---

1 According to WHO/UNICEF, an improved water source is one that, by nature of its construction and proper use, adequately protects the source from outside contamination, particularly fecal matter.

2 Onda, K., LoBuglio, J., and Bartram, J. 2012. Global Access to Safe Water: Accounting for Water Quality and Resulting Impact on MDG Progress. International Journal of Environmental Resources and Public Health, 9, 880-894.

3 A rapid assessment of drinking water quality (RADWQ) in Madhya Pradesh revealed a 40% reduction, equivalent to 3.4 million people in the study region, in safe water coverage relative to official statistics on improved water source coverage. See: Godfrey, S., Labhasetwar, P., Wate, S., and Pimpalkar, S. 2011. How safe are the global water coverage figures? Case study from Madhya Pradesh, India. Environmental Monitoring and Assessment, 176, 561-574.

How to Use This Report

This report contains comparative rating charts and key findings based on a year of rigorous research and analysis undertaken by CITE's multidisciplinary team at MIT.

Key findings are organized by the two categories of water test kits that the CITE team studied:

1) Single-parameter water test kits (SPWTKs), which measure a single aspect or parameter of a water sample

2) Multi-parameter water test kits (MPWTKs), which measure multiple aspects or parameters of a water sample

The findings of this report are especially applicable for the following audiences:

- Government officials and development practitioners seeking to better address water quality issues in Gujarat State

- Consumers in India looking for an inexpensive test to identify the presence or absence of contaminates in their municipal water supply

- Designers, manufacturers, suppliers, and retailers seeking to better understand consumer preferences, use patterns, and needs;

- Development practitioners outside of India who may be unaware of the low cost water test kits made in India and their accuracy compared to other kits

Product Evaluation Overview

In the SPWTK evaluation, the team focused on a simple hydrogen sulfide (H_2S) present-absent (P/A) test which detects microbiological contamination in water and compared two very similar products both technically in a laboratory setting and through surveys and observations. The two products were chosen because they are currently being used in governmental and non-governmental projects in India and are also manufactured in-country. Due to their simplicity, the team expects the evaluation results to be applicable to similar products from other manufacturers.

In the MPWTK evaluation, we sought to greatly expand the number of products under consideration, but in order to do so we had to limit the depth of the research and rely on interviews with manufacturers, distributors, users and other stakeholders for data rather than lab testing or field testing of the actual products. The CITE team also developed a prototype decision support system based on Multi-Criteria Decision Making (MCDM) techniques to allow stakeholders to customize evaluation results based on their personal relative weightings of the criteria.

Methodology at a Glance

The water test kit evaluation included three key components...

	1. Product Evaluation	2. Outreach & Partnership	3. Methodology Development
Single Parameter Test Kits	Lab testing in Ahmedabad, India Survey of 234 low-income users Ahmedabad, India Observation of users performing test in the field	Collaborated with local university students and NGOs Capacity building in rural northern Gujarat State Provision of water quality test results to all study participants	Proposed new standardized framework for CITE evaluations Proposed six new primary criteria, adaptable to all CITE evaluations
Multi Parameter Test Kits	Developed 17 indicators to calculate scores for six criteria based on literature and expert interviews Compiled a list of 56 kits from 25 different manufacturers	Interviewed NGOs working on water test kits	Developed procedures for conducting scoping study Developed Decision Support System Created product database template

And a new streamlined approach to product evaluation:

Technical Performance:
How well does the product perform its function in the lab and in real world settings?

Availability:
Is the product available in local markets? Is the supply chain dependable?

Ease of Use:
How easy or difficult is the product to use by an untrained user in a non-lab setting?

Affordability:
Is the full cost within the ability and willingness to pay for low-income users?

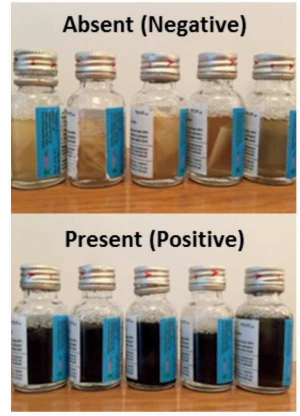
Demand Generation:
How high is the demand, and can the supply chain actors increase demand?

Health & Environment:
Does the product have a negative impact on the environment or the health of its users?

Data sources include: lab testing, field testing, observation, interviews, and surveys.

Findings at a Glance

Single Parameter Water Test Kits



The technical performance of the two water test kits was nearly identical: **94% accurate**

Participants who saw a demo of recommended use were **68% more likely to use it correctly** than those who were just read the instructions

The **cost of the test was acceptable** to low-income survey participants, but these **tests are not available** in retail outlets

Comparative Ratings Chart

	Technical Score	Ease of Use Score	Availability Score	Affordability Score	Demand Score	**Total Score**
Weightings	*20%*	*20%*	*20%*	*20%*	*20%*	*100%*
OrLAB/Bactovial	Excellent	Good	Poor	Excellent	Very Good	Good
Jal-TARA/Aquacheck	Excellent	Fair	Poor	Excellent	Good	Good

Key: Excellent, Very Good, Good, Fair, Poor

Multi-Parameter Water Test Kits

For low-income consumers, **reagent-type kits ranked higher** than incubator-type kits due to the high cost differential.

On average, **potential buyers valued ease of use & affordability** the most, and availability and demand generation least.

Comparative Ratings Chart

Brand & Model	Technical Score	Ease of Use Score	Availability Score	Affordability Score	Demand Score	Environmental & Health Score	**Overall Score**
Incubation Type Test Kits							
Hach	Good	Very Good	Excellent	Fair	Poor	Very Good	64
Wagtech	Excellent	Good	Very Good	Fair	Poor	Very Good	59
Sandberg	Good	Fair	Good	Poor	Poor	Very Good	49
DelAgua	Very Good	Poor	Excellent	Poor	Good	Very Good	49
ELE	Very Good	Poor	Fair	Poor	Poor	Very Good	28
Reagent Type Test Kits							
PSI	Good	Very Good	Very Good	Excellent	Very Good	Excellent	81
Jal-TARA	Good	Very Good	Very Good	Very Good	Good	Excellent	80
CPCB	Very Good	Good	Poor	Excellent	Very Good	Excellent	73
Labsol	Good	Good	Excellent	Very Good	Fair	Excellent	73

Key: Excellent, Very Good, Good, Fair, Poor

Single-Parameter Water Test Kits

A single-parameter test measures one aspect or characteristic of a water sample, such as its hardness, pH, turbidity, chlorine residual, or, microbiological contamination. Many single-parameter water test kits can be packaged together as a multi-parameter water test kits, and often the multi-parameter kits provide enough materials for multiple tests of each single parameter. The single-use, single-parameter kits evaluated in this research measured the presence (or absence) of hydrogen sulfide (H_2S) producing bacteria in a water sample, which indicates that fecal coliform such as *E. coli* may be present.

Results from several single parameter, hydrogen-sulfide (H_2S) water tests taken from samples in Ahmedabad, India. Credit: Jonars Spielberg

If the bacteria are present in detectable concentrations after a requisite 24 to 48 hours of incubation at ambient temperature, the water sample in the vial turns black, otherwise it remains a light yellow, as shown in Figure 1.[4] The test does not indicate the amount of bacteria present. Further investigation using other methods and tests would be required for a quantitative measure of microbiological contaminants.

Figure 1: Research participants in Ahmedabad testing their community and personal drinking water for bacteriological contamination.

4 The shade of yellow can vary quite a bit between tests. Since neither kit provides a color chart, it may be difficult for users to interpret the results with confidence.

Approach & Methodology

This evaluation compared two (2) H_2S tests: the Jal-TARA Aquacheck vial, manufactured and distributed by the nonprofit organization Development Alternatives, and the H_2S bactovials, manufactured and distributed by the private firm, ORLab. The main difference between the two is that the Jal-TARA vial uses a paper strip impregnated with chemicals, while the ORLab vial uses a powder.[5]

Two objectives drove our research:

1) How do the two tests compare in terms of technical performance and ease of use, availability, affordability and demand?

2) Independent of which test is used, how sensitive is ease of use to the way in which individuals are introduced to the test? This question is significant because although there have been numerous studies on the technical performance of H_2S kits in the laboratory conducted by skilled individuals, the technical performance in field tests conducted by real users has not been evaluated to the same extent.[6]

In order to provide a comparative evaluation of these single-parameter water test kits, 11 indicators were identified which mapped to five of the six top-level criteria. The Environmental and Health Impact criterion was not evaluated for the single parameter water test kits since hydrogen sulfide, while potentially toxic at high levels, should not pose any threat to humans or the environment at the low concentrations found in the test kits.

Lakshmi, a resident of the Behrampura neighborhood, completes a H_2S test after viewing a demonstration given by MIT and IIM-A researchers.
Credit: Jonars Spielberg

5 The exact chemical composition of the H_2S test medium used in the Jal-TARA and ORLab vials in not known; however, it is likely to contain some mix of peptone, dipotassium hydrogen phosphate, ferric ammonium citrate, sodium thiosulphate and teepol.

6 A recent meta-analysis included 51 studies of H_2Sperformance, as documented in Wright, J., Yang, H., Walker, K., Pedley, S., Elliott, J., & Gundry, S. (2012, January). The H_2S test versus standard indicator bacteria tests for faecal contamination of water: systematic review and meta-analysis. Tropical Medicine and International Health, 17(I), 94-105.

Technical Performance[7]

In urban Ahmedabad from August 5 – 13, 2014, the ORLab and Jal-TARA H_2S kits and the IDEXX Quanti-tray 2000 were tested side-by-side on 426 water samples.[8] The Quanti-Tray® is regarded as the "gold-standard" in microbiological water testing and is both reliable and accurate. The IDEXX tray has 97 wells that inoculate a 100mL sample and the number of positive wells is converted to a Most Probable Number (MPN) with upper and lower 95% confidence limits. It was used as the positive control to compare the H_2S results. Water samples were collected from homes in pairs of filtered and unfiltered tap water samples using standardized procedures into sterile 500mL bags and processed within 8 hours. The H_2S tests were not diluted.

For the IDEXX Quanti-Tray®, a 100mL sample was inoculated into the 97-well IDEXX tray with its corresponding reagent (Colilert®) and incubated at 35°C for 24 hours. The result of IDEXX Quanti-Tray® was then read to determine the MPN of total coliform and *E. coli*. For the H_2S test, 20 mL of the water from each sample was inoculated to the test and incubated at ambient temperature for 24 and 48 hours. The positive result for H_2S test was indicated by a change in color of the medium from translucent to black, and also by the smell of hydrogen sulfide.

Research participant Sima, a resident of the Khodiyarnagar area of Ahmedabad, completes a H_2S test. Credit: Jonars Spielberg

Because the IDEXX Quanti-Tray® is a test that gives an enumerative result (with detection range of 1~2419.6 Most Probable Number (MPN)/100 mL) and the H_2S tests in this study was conducted in the Presence/Absence (P/A) mode, the two tests could only be compared on the P/A characteristic. For the purpose of this comparison, if the IDEXX Quanti-Tray® detected anything above its lower detection limit of 1 MPN/100 mL, it was considered a "Present" reading. The results of IDEXX Quanti-Tray® and each H_2S test were analyzed using 2x2 contingency tables. General statistical values were calculated according to a method described by Mack and Hewison.[9] Statistical significance between the IDEXX Quanti-Tray® and each H_2S test were calculated using the Phi coefficient value, Chi-squared test and Fisher Exact Probability test.

---

7 This section reports on a subset of the technical results, which have been previously published in Murcott, S., Keegan, M., Hanson, A., Jain, A., Knutson, J., Liu, S., & Wong, T. K. (2015). Evaluation of Microbial Water Quality Tests for Humanitarian Emergency and Development Settings. Procedia Engineering, 107, 237-246.

8 The IDEXX Quanti-Tray® 2000 is an enzyme-substrate coliform test that utilizes the Most Probable Number (MPN) method for enumeration of contamination.

9 Mack K.F. and Hewison K. 1998. The PATH/NEVWRP Tests. Thai-Australian Northeast village water resource project, Report No 47: Evaluation of a Hydrogen Sulphide Screening Test.

As shown in Table 1, three (3) indicators comprise the technical performance criterion: true result (TR), positive predictive value (PPV) and negative predictive value (NPV). For all three indicators, a purposeful sample was used and stratified (by geography) and laboratory testing was used as a method.

Table 1: Technical Indicators for the Single-Parameter Water Test Kit Evaluation

#	Indicator	Indicator Description
1	True Result	TR represents the percentage of samples tested by a new test that yield the same result when compared to the standard (e.g. Absence and Absence)
2	Positive Predictive Value	PPV is the ability of a positive test (by the new test) to predict the presence of a contaminant (in this case, *E. coli,* H_2S, or Total Coliform).
3	Negative Predictive Value	NPV is the ability of a negative test (by the new test) to predict the absence of a contaminant (in this case, *E. coli,* H_2S, or Total Coliform).

Ease of Use, Availability, Affordability and Demand

In this evaluation, the methodological focus was to develop and test an observation-based research protocol designed to add an evidence-based element to the self-reported survey data used to assess the Ease of Use criterion.

At each site, 20 research subjects were chosen based on a purposeful random sample stratified by geography. These participants were divided by H_2S test brand (i.e., ORLab or Jal-TARA) and presentation group, instruction or demonstration. The instruction group was intended to mimic a consumer's experience buying a product at the market: they were read aloud the instructions written on the test vial, and if subjects had clarifying questions, researchers would provide brief answers (similar to what a store employee would provide). In the demonstration presentation, a researcher would read aloud the instructions while also completing the steps to the test himself/herself, a process that took less than five (5) minutes. One site was visited per day (total of 12 days), resulting in the completion of 234 surveys.

Table 2: Indicators for the Single-Parameter Water Test Kit Evaluation

#	Indicator	Indicator Description	Method
4	Perception	Indicates how easy or difficult the survey respondents believe the test will be to perform	Survey
5	Performance	Indicates how easy or difficult it was for the survey respondents to actually perform the test during as observed by the research team	Observation
6	Interpretation	Indicates how easy or difficult the survey respondents believe the test was to interpret	Survey
7	Local Availability	Indicates whether the survey respondents believe they are able to purchase the test locally	Survey
8	Cost	Indicates whether the survey respondents believe they can afford to purchase the test.	Survey
9	Willingness to pay	Indicates whether the survey respondents are willing to buy the test at the current and/or higher prices	Survey

#	Indicator	Indicator Description	Method
10	Relative advantage	Indicates how highly the survey respondents value information on water quality	Survey
11	Recommendability	Indicates whether the survey respondents would recommend the test to a friend	Survey

A six-part survey was developed to gather data to calculate indicator and criterion scores for ease of use, availability, affordability and demand. The six sections were:

1. Demographics, and Knowledge, Attitude and Practices (KAP)
2. Instruction or demonstration protocol
3. Perceptions of the test prior to use
4. Subject testing, which included an observation protocol
5. Immediate post-test follow-up; and
6. 48-hour post-test follow-up, where test results were interpreted by subjects, and researchers collected vials and provided compensation to subjects.

Findings

The following sections present the findings from the single-parameter water test kit evaluation in summary format.[10]

Technical Performance

As shown in Table 3 and Table 4, Jal-TARA and ORLab H_2S products have good statistical results when compared to Quanti-Tray® *E. coli*, with between 88%-95% true result (TR), sensitivity and specificity. This means that these tests yield the same result as the Quanti-Tray® *E. coli* between 88%-95% of the time, having a high ability to determine both true positive and true negative results. It should also be noted that the comparison between the Quanti-Tray® *E. coli* and the H_2S test results was better than a comparison between Quanti-Tray® Total Coliform and H_2S for most statistical values. This suggests that the H_2S tests carried out in India had better correlation with *E. coli*, the WHO-recommended indicator of fecal contamination.

Of particular note is that the Jal-TARA and ORLab kits had identical results across all parameters after 48 hours for the *E. coli* results. These values were triple checked to ensure accuracy, especially since the 48-hour *E. coli* result is the one chosen for inclusion in the top level comparative ratings. Though its false negatives results are high, total coliforms can be of either fecal or environmental origin, especially in tropical areas. As a result, the WHO does not consider total coliforms a useful indicator organism for the presence of pathogens in water supplies.[11]

10 For detailed results, please reference the full version of this report on the CITE website (cite.mit.edu).

11 WHO. 2011. Guidelines for Drinking Water Quality, Fourth Edition. Geneva: World Health Organization. Specifically, "Total coliform bacteria are not acceptable as an indicator of the sanitary quality of water supplies, particularly in tropical areas, where many bacteria of no sanitary significance occur in almost all untreated supplies" (149).

Table 3: Comparison between Quanti-Tray® Total Coliform and H_2S Test Result (shown in %)

n = 426		True Result	False Positive	False Negative	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)
TARA	24 hr	76	0	24	100	64	57	100
	48 hr	80	0	19	99	69	66	99
ORLab	24 hr	73	0	27	100	62	53	100
	48 hr	81	0	19	99	69	66	99

Table 4: Comparison between Quanti-Tray® *E. coli* and H_2S Test Result (shown in %)

n = 426		True Result	False Positive	False Negative	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)
Jal-TARA	24 hr	89	3	8	92	88	78	96
	48 hr	91	4	5	88	92	87	93
ORLab	24 hr	89	1	9	95	87	75	98
	48 hr	91	4	5	88	92	87	93

Ease of Use

Researchers observed 234 participants complete the test, and documented instances when proper procedures were not practiced. Such mistakes included not washing one's hands, touching one's sari or clothing during the test, touching the inside of the cap and bottle, touching the lip of the vial to the water storage container and shaking the bottle inadequately, or forgetting to shake the vial at all.

Table 5: Participants Sorted by Different Groups

Sorted by Brand or Training Group		No.	Sorted by Brand & Training Group	No.
By Brand	ORLab	114	ORLab/Instruction	59
	Jal-TARA	120	ORLab/Demonstration	55
By Training Group	Instruction	119	Jal-TARA/Instruction	60
	Demonstration	115	Jal-TARA/Demonstration	60

Overall, demonstration group participants performed the H_2S test better, as shown in Figure 2. As a group, those who saw a demonstration prior to conducting the test made 62% fewer mistakes and asked for assistance 33% fewer times relative to instruction group participants. When comparing performance by test brand, results are mixed. Those conducting tests with an ORLab test vial made 32% fewer mistakes than Jal-TARA test users. However, those conducting tests with an ORLab test asked for assistance 8% more times than those conducting tests with a Jal-TARA test.

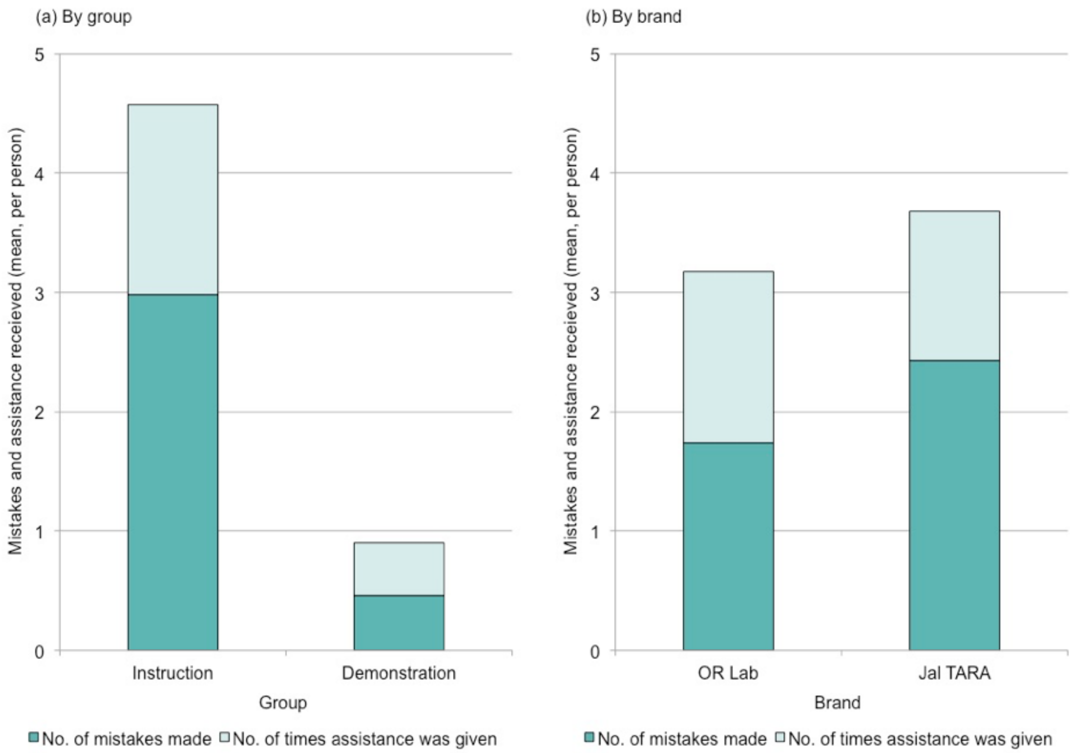


Figure 2: Participants' performance, by group and by test brand

Results indicate that ORLab users were 15% better at interpreting test results correctly than were Jal-TARA users: 87% correct interpretation for ORLab users and 72% for Jal-TARA. This may be indicative of a greater color change or differential in the positive samples for ORLab tests, though more testing would need to be completed to make any definitive statement.

Availability

H_2S tests are currently not available direct to the consumer, at least in Ahmedabad. As a result, both test brands score low in this category. However, this represents an opportunity to expand sales channels, especially given the relatively high willingness to pay for the tests (see Figure 3). Even so, perceived availability proved low: 60% of the 234 survey respondents said that they would not know where to purchase the test, even if it were being sold. This suggests that, should the tests eventually make their way to markets and corner stores, awareness-raising would be required to make sure they would not languish away on shelves.

Affordability

Both tests cost approximately the same, approximately INR 30 ($0.50), and therefore were relatively indistinguishable in terms of affordability. A monadic method of willingness to pay (WTP) revealed that the likelihood of purchase was high, even at triple the current price.[12] Based on these results, we believe both tests are affordable for low-income consumers.

12 See Wright, T. and K. Leith (2013). Scale-Ups Market Research Framework for a discussion on techniques for willingness to pay, including monadic surveys.

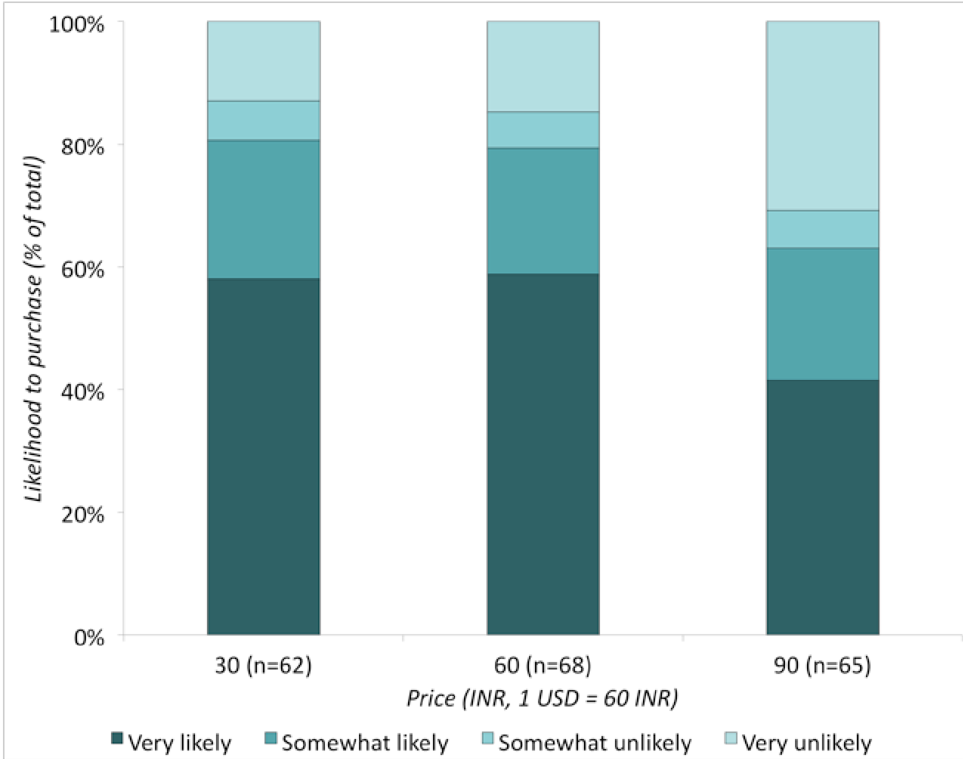


Figure 3: Participant willingness to pay at three different prices

Demand

Demand for the test was high: almost all (95%) respondents said they were likely to recommend this test to neighbors and friends. Knowing the quality of their water was important to the survey participants and when given a choice between receiving information on water quality or one of three other topics that were the subject of local awareness building campaigns (education, credit or nutrition), water quality was preferred between two and eight times more than the alternatives.

Ratings

The decision to report findings by both brand and training group was made because the technology performance of the two kits was nearly identical and the combined results separated by brand alone were very close, except in the ease of use category. This meant that how users are exposed to products and the amount of information and training they receive was perhaps more important than the product itself. This drives at the heart of CITE's comprehensive approach to technology evaluation: technology matters, but so does the context in which it is introduced and used. Overall, our results confirm this: little difference exists between performance of the brands, while a larger difference exists between instruction and demonstration presentation groups.

As shown in Figure 4, with the exception of the ease of use criterion, each indicator is weighted equally since they were seen as equally important. For the ease of use criterion, we weighted the direct empirical observation data (Performance and Interpretation) as twice the weight of the self-reported data (Perception).

Key results include:

- The tests were virtually indistinguishable in terms of technical performance, affordability, and availability
- For ease of use, demonstration groups performed better than instruction groups. Notably, ease of performance was 68% greater among demonstration groups relative to instruction groups.
- Recommendability was high among all tests and presentation groups, while relative advantage varied considerably for no apparent reason.[13]

AVERAGE OF DEMONSTRATION AND INSTRUCTION						
Weightings	20%	20%	20%	20%	20%	Total Score
H_2S Test Manufacturer Name	Technical Score	Ease of Use Score	Availability Score	Affordability Score	Demand Score	
ORLab/Bactovial	90	49	14	82	61	**59**
Jal-TARA/Aquacheck	90	31	16	80	59	**55**

H_2S Test Manufacturer/Group	TECHNICAL PERFORMANCE				EASE OF USE			
	weightings	33%	33%	33%	weightings	20%	40%	40%
	Technical Score	True Result	Positive Predictive Value	Negative Predictive Value	**Ease of Use Score**	Perception	Performance	Interpretation
ORLab/Instruction	**90**	Excellent	Excellent	Excellent	**39**	Fair	Fair	Good
ORLab/Demonstration	**90**	Excellent	Excellent	Excellent	**58**	Good	Good	Very Good
Jal-TARA/Instruction	**90**	Excellent	Excellent	Excellent	**18**	Fair	Poor	Poor
Jal-TARA/Demonstration	**90**	Excellent	Excellent	Excellent	**44**	Good	Good	Fair

H_2S Test Manufacturer/Group	AVAILABILITY		AFFORDABILITY			DEMAND		
	weightings	100%	weightings	50%	50%	weightings	50%	50%
	Availability Score	Local Availability	**Affordability Score**	Cost	Willingness to Pay (60INR)	**Demand Score**	Relative Advantage	Recommend-ability
ORLab/Instruction	**15**	Poor	**83**	Excellent	Very Good	**62**	Good	Very Good
ORLab/Demonstration	**14**	Poor	**80**	Excellent	Very Good	**59**	Fair	Excellent
Jal-TARA/Instruction	**19**	Poor	**79**	Excellent	Very Good	**46**	Poor	Very Good
Jal-TARA/Demonstration	**12**	Poor	**80**	Excellent	Very Good	**71**	Very Good	Excellent

Legend: Excellent, Very Good, Good, Fair, Poor

Figure 4: Single-Parameter Water Test Kit Comparative Ratings Chart (out of 100)

13 Recommendability is relative to others: would you endorse the test (i.e., are you satisfied enough to recommend it?). Relative advantage measures whether you, as an individual, value the information that the test gives you relative to alternative information.

In addition to the evaluation findings, the CITE team also gained valuable experience in streamlining the CITE methodology for future evaluations. Based on this experience, the following activities will be incorporated into the solar water pumping evaluation which is being conducted by CITE in 2016:

- A 3-month desk review/scoping study will be conducted prior to traveling to the field
- An evidence-based observational element will be added to the evaluation (for the solar pumping evaluation this will include the incorporation of remote sensors)

Multi- Parameter Water Test Kits

There were three characteristics that separated the multi-parameter water test kit evaluation approach from the single-parameter water test kit evaluation and other CITE evaluations (e.g., solar lanterns, water filters, etc.):[14]

- The research depended on remote research techniques, including a literature review, telephone interviews with stakeholders, and online surveys, and was significantly shorter and less resource intensive than other evaluations. This was done to increase the number of products included in the water test kit evaluation and to help the CITE team understand how much information could be gathered during a 3-month scoping study for future CITE evaluations
- The primary source of technical performance data was the kit manufacturers and no attempt was made to validate this data independently in the lab.
- This evaluation also included the development of a Group Decision Support System (GDSS) to allow users of this and future evaluations to customize the results to fit their specific context and preferences.

Approach & Methodology

The overall approach for the Multi-Parameter Water Test Kit evaluation included the following steps:

1) Develop a database of potential alternative test kits

2) Identify indicators mapped to the six top level criteria based on a desk review of academic and practitioner literature and expert interviews[15]

3) Develop weightings for the criteria based on a group decision support system

4) Develop a scorecard comparing different brands of water test kits

Database of Alternatives

For the multi-parameter water test kit evaluation, CITE researchers compiled a database of potential alternative test kits. In this study, CITE chose to evaluate water quality test kits that focused on more than one parameter as opposed to the single-parameter tests discussed previously. This was done to ensure that CITE was addressing all of the major types of kits commonly used by development practitioners. Many companies and NGOs have

14 All of CITE’s evaluation reports can be found at http://cite.mit.edu.

15 More than three-dozen articles, reports, and publications were surveyed to understand the most important variables to consider for our evaluation.

developed multi-parameter kits targeted at a wide variety of users, ranging from individual consumers who want to know if their household water is safe to drink to highly trained water engineers working for large international organizations who want to have the same water testing capabilities in the field that they have in the lab.

In order to evaluate kits that measured similar parameters while ensuring that the evaluation was particularly relevant to grassroots level practitioners, CITE only chose to include kits that included bacteriological parameters in addition to physical and chemical parameters. This is because bacteria in water is a major source of disease in the areas where development practitioners work.
The kits that included bacteriological analysis fall into two broad groups: reagent-type and incubator-type, as shown in Figure 5.

Reagent-type kits test for the presence or absence of microbiological contamination as indicated by a chemical reaction between a substance in the kit (e.g., H_2S) and the water sample. The reagent-based kits are relatively inexpensive and can be used by any interested community member. Although they are easy to use, they are not appropriate for all contexts[16] and do not report on the quantity of bacteria in the water, just whether it is present or absent.

Incubator-type kits are portable laboratories that include equipment to accurately measure the amount of microbiological contamination (e.g., total coliforms, fecal coliforms, etc.). In addition to the bacteriological analysis, each of the kits in our evaluation included analysis of basic physical parameters including turbidity, conductivity, total dissolved solids, as well as chemical parameters like chlorine, iron, nitrate, nitrite, and ammonia, as well as others.

The incubator-type kits have very high technical accuracy, but are more complex and require much more training and adherence to specific protocols to ensure uncontaminated results. They are also prohibitively expensive for many rural communities and grassroots NGOs and are used by larger groups and especially in conflict and disaster scenarios where access to quality water supplies are particularly important and might be incredibly stressed.

Figure 5: Incubator-Type (Top) and Reagent-Type (Bottom) Multi-Parameter Water Test Kits

16 Previous research has shown that the H2S kits can report a high rate of false positive when used with groundwater in rural settings.

In addition to the desk review, interviews with project developers and local users provided supplementary data regarding which aspects of the kits were most important during use. A board of decision-makers was formed including representatives from eight organizations who use multi-parameter water test kits in their work including Aga Khan India Rural Support Programme, Arghyam Foundation, Agualimpia Peru, Wello, Gram Vikas, Water Aid India, Water For People, and USAID.

A Group Decision Support System was used to convert expert opinion of the importance of each of the six criteria to weightings used in the multi-parameter water test kit scoring. The focus of the GDSS is to allow stakeholders to choose the weights of the criteria, utilizing expert opinion and their personal preferences.

Ultimately, CITE chose 17 different specific indicators by which each kit would be evaluated, as shown in Table 6. To calculate the scores of each criterion, the lower level indicator scores are averaged. Then, the total score for each kit multiplies the criteria scores by the weighting factor calculated in the Group Decision Support System.

Table 6: Multi-Parameter Water Test Kit Evaluation Criteria and Indicators

#	Indicator	Indicator Description/Rationale	Method
Technical Performance Criteria			
TP1	Technology for bacteriological parameters	Whether the kit has an incubator or is a reagent-based test for bacteriological parameters	Product Brochure
TP2	Technology for physical and chemical tests	Whether the kit has reagent-based chemical tests and visual physical tests or a sensor is used	Product Brochure
TP3	Number of parameters tested	The total number of bacteriological, chemical and physical parameters tested by the kit	Product Brochure
Ease of Use Criteria			
EU1	Number of hours of training required	The amount of training required to ensure accurate results and proper care of the equipment	Survey conducted by the Technology Exchange Lab (TEL)
EU2	Incubator capacity	A higher incubator capacity allows for proper controls for false negatives and false positives	Product Brochure
EU3	Quality of Manual	Simplicity and clarity of manual, including photos and pictorial instructions	CITE Assessment
EU4	Portability (weight)	Weight of the entire kit	Product Brochure
EU5	Portability (battery life)	Incubator battery life (where applicable)	Product Brochure
EU6	Output type (digital/ analog/or color change)	Digital output prevents user error in reading results	Product Brochure
Availability Criteria			
AV1	Time to Delivery	Time, in days, from order to delivery (we used a sample destination of Ahmedabad, India)	Manufacturer Communication
AV2	Consumables Available	Are consumables available locally, in the nearest urban center, or only by import	Survey - TEL
AV3	Expertise/Support Availability	Is technical expertise or support available locally, in the nearest urban center, or repair only offered by shipping back to supplier	Survey - TEL
Affordability Criteria			

#	Indicator	Indicator Description/Rationale	Method
AF1	Price of Kit	Price quoted by manufacturer either in their own literature or by email communication	Manufacturer Communication
AF2	Price per Test	Cost of consumables for each test	Manufacturer Communication
Demand Generation Criteria			
DG1	Outreach Programs	Does the manufacturer organize outreach projects to demonstrate or pilot test their products	Product Website
DG2	Knowledge of Local markets	Does the manufacturer demonstrate a knowledge of grassroots challenges of testing	Product Website
Environmental and Health Impact Criteria			
EHI1	Requires Transport/ Disposal of Hazardous Materials	To what degree does the kit require transport and/or disposal of hazardous materials?	Manufacturer
EHI2	Safety Hazard for Users	To what degree does the kit pose a safety hazard for those using it?	Manufacturer

Alternatives and Indicators

In order to evaluate multi-parameter test kits along the six criteria, we first needed to create an aggregated list of all the multi-parameter test kits on the market. For the purpose of this exercise, we included all kits that measured three or more parameters of water quality as multi-parameter, to ensure that the kits were truly versatile and provided a multidimensional assessment of water quality. These parameters could be physical, chemical, bacteriological, or any combination of those. Through desk research and interaction with the sales teams of water quality test kit manufacturers, we compiled a list of 56 kits from 25 different manufacturers.

In keeping with the methodology of fast, but rigorous evaluation, nine kits were selected from the 56 to allow for efficient data collection that would lead to useful results for the end users of the multi-parameter kits. The nine were chosen because they all measure bacteriological and chemical parameters, come from a variety of manufacturers, and are the most often used tests, as gauged by our interviews with water quality practitioners. Importantly, we included both high-cost incubator-based and/or sensor-based kits as well as the much lower-cost reagent-based kits. Our final list of alternatives for comparison is shown in Table 7.

Group Decision Support System (GDSS)

For multi-parameter water test kits, the Technical and Affordability criteria play a key role in buyer decision-making; however, the product sustainability and social implications also need to be characterized in a robust way. For this evaluation, the CITE team worked with by Dr. Mahdi Zarghami, who was a visiting researcher on the project, to develop a Group Decision Support System (GDSS) based on his previous research.

The GDSS is based on a Multi-Criteria Decision Making (MCDM) technique for ranking alternatives which is similar to the existing weighted criteria method used in other CITE evaluations. Constructing an MCDM model requires both the determination of specific evaluation criteria (shown in Table 6) as well as their relative importance to each other. The relative weights were gauged through interviews with water quality practitioners and experts. Interviews were also a source of stakeholder information to assess risk acceptance and aversion. In addition, a sensitivity analysis on the criteria weights is performed. More information on Dr. Zarghami's MCDM methods can be found in the detailed evaluation reports and in the literature. [17]

Table 7: Characteristics of Chosen Alternatives

Availability	Kit Name	Technology	Alkalinity	Ammonia	Arsenic	Benthos	Calcium	Chloride	Chlorine	Chlorine (Free)	Chlorine (Residual)	Chlorine (Total)	Coliforms (Thermotolerant)	Coliforms (Total)	Color	Conductivity	Dissolved Oxygen	Ecoli	Faecal Streptococci	Fluoride	Hardness	Iron	Magnesium	Nitrate	Odor	PH	Phosphates	Primary productivity	Pseudomonas aeruginosa	Salmonella	Solids (Dissolved TDS)	Solids (Suspended)	Temperature	Transparency	Turbidity	Total Parameter Count	Price
Incubator-Type Water Test Kit																																					
DelAgua	Bacteriological Kit 2	Membrane Filtration, Sensor, Reagent								●		●	●	●		●										●					●		●		●	9	$5,194
ELE	Paqualab 50	Membrane Filtration, Sensor, Reagent														●										●							●		●	4	$3,600
Hach	MEL Potable Water Laboratory	P/A, optional MF and MPN, Sensor, Reagent		●						●		●		●				●						●		●	●				●					9	$2,300
Wagtech	Polalab + Advanced Portable Water Quality Lab	Membrane Filtration, Sensor		●	●					●		●	●	●		●				●				●		●					●				●	12	$2,262
Sandberg	Trawas Complete Kit	Membrane Filtration, Sensor, Reagent							●					●		●										●									●	5	
Reagent-Type Water Test Kit																																					
CPCB India	Water Testing Kit	H_2S, Reagent	●	●		●	●	●			●			●	●		●			●	●	●	●	●	●	●	●	●				●	●		●	21	$31
People's Science Institute	Standard Comprehensive Kit	H_2S, Reagent	●	●		●		●			●			●			●			●	●	●		●		●	●						●	●		14	$70
Development Alternatives	Jal-TARA Standard Water Testing Kit	H_2S, Reagent		●		●		●			●			●			●			●	●	●		●		●	●						●		●	14	$95
Labsol	Water Testing Kit	H_2S, Reagent		●		●		●			●			●			●			●	●	●		●		●	●						●		●	14	$250
	No. of Kits That Measure This Parameter		2	5	1	4	1	4	1	3	4	3	2	8	1	4	4	1	0	5	4	4	1	6	1	9	5	1	0	0	3	1	6	1	7		

17 Zarghami M., Ardakanian R., Memariani A. and Szidarovszky F. 2008. Extended OWA operator for group decision-making on water resources projects. Journal of Water Resources Planning and Management, ASCE, 134(3), 266-275.

Findings

To calculate the group consensus opinion on the weighting of the six criteria, the individual decision maker weightings shown in Table 8 are multiplied by the power of the decision maker. These power weightings reflect the reality that some stakeholder's rankings carry more influence than others. The power weighting is based on the organization's geographic reach and influence and was chosen by CITE researchers.

After describing the criteria to the decision makers, they were asked to assess the importance of each criterion on a 7-point scale (VL, very low; L, low; FL, fairly low; M, medium; FH, fairly high; H, high; and VH, very high). When these linguistic variables are converted to numbers, the group opinion of weightings is calculated and is shown in Figure 7.

Table 8: Criteria Weights from Each Decision Maker

	Weighting Applied by Each Decision Maker					
	Technical Performance	**Ease of Use**	**Availability**	**Affordability**	**Demand Generation**	**Environmental Impact**
Decision Maker 1 (Low Power)	L	VH	H	H	L	L
Decision Maker 2 (Medium Power)	H	VH	VH	M	L	L
Decision Maker 3 (Fairly High Power)	H	H	M	FL	M	H
Decision Maker 4 (Fairly High Power)	VH	VH	L	M	L	VL
Decision Maker 5 (Very High Power)	H	VH	M	VH	L	H
Decision Maker 6 (High Power)	L	H	M	VH	VH	H
Decision Maker 7 (Very Low Power)	H	VH	M	FH	M	L
Decision Maker 8 (Medium Power)	VH	VH	M	VH	L	VH

Researchers (left to right) Jonars Spielberg, Innocent Tumwebaze, Vihar Parikh, Reecha Das and Bianca Shah gather to discuss findings and next steps forward. Credit: Sydney Beasley

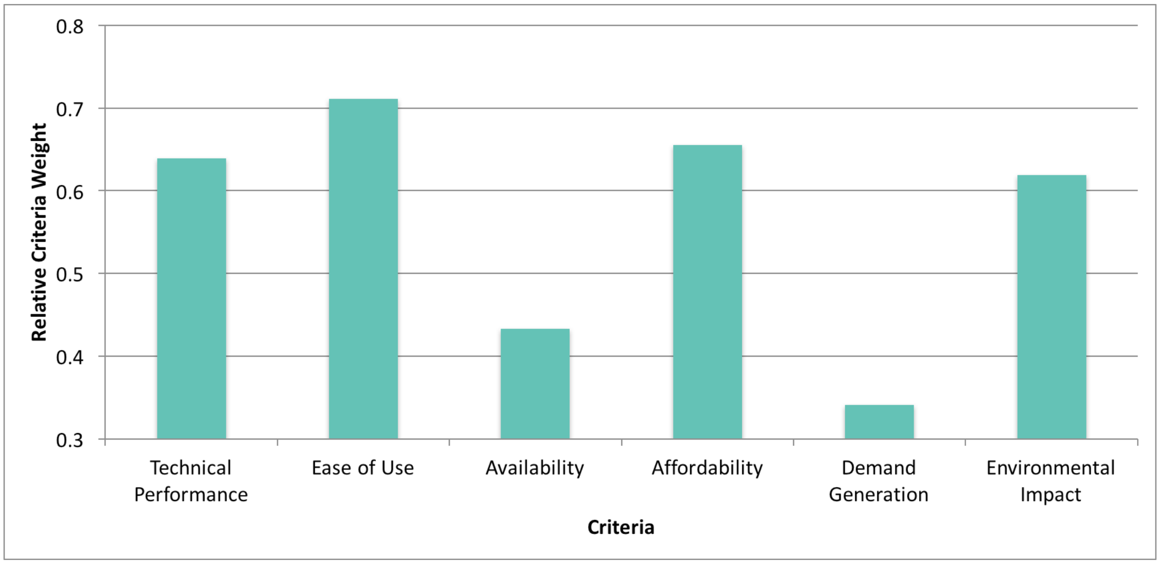


Figure 7: Group weights of the criteria

Several theoretical concepts factor into the calculation of weightings in the GDSS. These include:
Optimism vs. Pessimism, which deals with the number of expert opinions that the decision maker wants to include in their decision (more opinions = more pessimistic; less opinions = higher optimism degree). Consensus, which identifies the number of opinions that weight the criteria in the same manner. The more stakeholders that weigh a criterion the same, the more likely the group will be able to reach consensus on any given decision.

Sensitivity, which relates to how sensitive the weights of the criteria are to changes in the optimism degree.
In general, these factors are used as multipliers on the criteria weights, and details on the decision making theory and formulations can be found in the full water test kit report and in Dr. Zarghami's publications on decision theory, multi-criteria decision making, ordered weighted averaging (OWA), Fuzzy linguistic quantifiers and Stochastic uncertainty.[18]

Ratings

This study evaluated multi-parameter field test kits that fall into two different categories – those kits with incubators and those that are primarily reagent-based. We have included both categories in our evaluation, but the higher scores of the reagent-type over the incubator-type is not an indication that the reagent-based kits are categorically "better" than incubator kits. Because the price differential is so great between the two categories, the reagent types score more highly as a group than the incubation types.

The two categories serve two different purposes – the reagent-type as a very inexpensive and simple to use, widely dispersed field tool to be used in communities at regular intervals, and the incubator-type as a much more expensive but much more accurate measure of water quality. To use this evaluation, practitioners should first decide on a category of multi-parameter test kit, and then consult the ratings matrix below.

18 Zarghami, M., Szidarovszky, F., & Ardakanian, R. (2008). A fuzzy-stochastic OWA model for robust multi-criteria decision making. Fuzzy Optimization and Decision Making, 7(1), 1-15.

Brand & Model	Technical Score	Ease of Use Score	Availability Score	Affordability Score	Demand Score	Environmental & Health Score	**Overall Score**
Incubation Type Test Kits							
Hach	Good	Very Good	Excellent	Fair	Poor	Very Good	64
Wagtech	Excellent	Good	Very Good	Fair	Poor	Very Good	59
Sandberg	Good	Fair	Good	Poor	Poor	Very Good	49
DelAgua	Very Good	Poor	Excellent	Poor	Good	Very Good	49
ELE	Very Good	Poor	Fair	Poor	Poor	Very Good	28
Reagent Type Test Kits							
PSI	Good	Very Good	Very Good	Excellent	Very Good	Excellent	81
JalTARA	Good	Very Good	Very Good	Very Good	Good	Excellent	80
CPCB	Very Good	Good	Poor	Excellent	Very Good	Excellent	73
Labsol	Good	Good	Excellent	Very Good	Fair	Excellent	73

Key: *Excellent* *Very Good* *Good* *Fair* *Poor*

Figure 10: Multi-parameter Field Test Kit Overall and Criterion Scores (out of 100)

While the scores for each criteria are simply the normalized averages of the indicators for each criterion (for example, ease of use combines data on the amount of training required, the quality and flexibility of the accompanying manual, and the portability of the kit), the total rating for each kit uses these criteria scores and multiplies them by the weights determined in the GDSS, which takes into account the relative importance of each criteria from our panel of experts. For example, ease of use is weighted twice as heavily as Demand Generation.

In the incubation category, the Hach MEL Potable Water Lab scored highest, due to high scores in availability, and relatively high scores in ease of use compared to the alternatives in that category. It should be noted that the Hach did not score as highly on the technical performance category, but this does not indicate inaccuracy in their ratings, but rather a smaller number of parameters measured by that particular kit.

In the reagent-based category, People's Science Institute scored the highest, followed closely by Jal-TARA. The deciding characteristics of these kits were low cost, with high ease of use, including manuals and instructions that included photos or pictures, as well as good availability of the kits and replacement consumables.


Left: A H2S test from a water sample taken in Khodiyarnagar shows a positive result. Credit: Vihar Parikh; Right: Makwana, a resident of the Gulbai Tekra neighborhood, shows her H_2S test results. The community water sample came back negative, but the sample from her household matka (an earthen container common throughout South Asia to store water) yielded a positive result. Credit: Jonars Spielberg

Incorporating Lean Research Principles

In addition to the work on the evaluations and capacity building with partners in India, CITE also focused on developing a more standardized process for conducting evaluations, combining experience from CITE's past evaluations with elements of the "Lean Research" approach championed by CITE, the International Development Innovation Network (IDIN) at MIT D-Lab, and other academic and practitioner organizations.[19] In the water test kit evaluations, the CITE team applied Lean Research principles by:

- Ensuring that the information gathered was of high quality and that every survey question was linked directly to a calculated result;
- Involving stakeholders in the evaluation process, including their inclusion in the MPWTK expert panel;
- Providing direct benefit to study subjects and their communities by compensating participants for their time and giving them their own test vial so they could determine the quality of a water source of their own choosing.

Research translator Utsav Patel poses with five women from the Thaltej area of Ahmedabad who completed their water tests after being shown a brief demonstration. Credit: Vihar Parikh

Conclusions and Recommendations

Water Test Kit Evaluations

1) In many ways, the two single parameter H_2S tests performed similarly across all criteria, especially in terms of technical performance and affordability. Differences present themselves more readily when comparing the demonstration group against the instruction group.

This suggests that, even with such a simple technology, a small difference in how consumers are introduced to a product has a sizeable impact on their ability to use the product correctly. Therefore, training should be included as a key component of technology use, whose significance should not be understated.

19 Lean Research is an approach to field research in the context of development work that seeks to maximize benefit and minimize burden and waste for all stakeholders in the research process. For more information, see https://d-lab.mit.edu/lean-research.

Field test participant Asma Patham performs a H_2S test on the steps of her apartment complex in the Khodiyarnagar neighborhood of Ahmedabad. She used to purify her water using a reverse osmosis (RO) filter, but no longer does because she believes that the tap water is clean. Both her test samples yielded positive results. Credit: Jonars Spielberg

2) Overall, the single parameter H_2S test is a good first line defense in monitoring water quality as it is cheap, easy to use, easy to interpret and relatively accurate when compared to standard laboratory methods. The information it provides, however, is limited. Thus, such a test should be ideally paired with other tests and programmatic tools for clean water provision and monitoring, to be carried out by CBOs, NGOs, government or some combination thereof.

3) There are two primary types of multi-parameter water test kits and a wide variety in the price. The average cost of the incubator type of multi-parameter water test kit included in CITE's ratings (e.g., Hach, Wagtech, DelAgua) is over $3,000 and likely is only affordable for larger organizations who can pay the higher premium to improve accuracy. Conversely, the average cost of the reagent type multi-parameter water test kits reviewed by CITE (e.g., Jal-TARA, CPCB India,.) is closer to $100 and therefore more affordable for smaller organizations and individuals that want a simple cost-effective way to test their water quality.

CITE Methodology

1) The process of creating a light, rapid evaluation certainly has value; however, there are several drawbacks to this approach.

One of the challenges that CITE faces is balancing the robustness of the methodology with the required time and cost to complete an evaluation. Since the resources allocated to the multi-parameter water test kit evaluation were limited, we decided to test whether a 3-4 month evaluation without field or lab testing would provide sufficient information to compare products across the same six top level criteria used in other evaluations.

While this method facilitated a relatively fast acquisition of valuable data and analysis to evaluate multi-parameter test kits, its reliance on interviews and publicly available information (including technical performance specifications from manufacturers) does not allow for truly independent and verified comparative ratings of products. Without completing lab or field research, it is impossible to obtain data on the lab or field accuracy of each kit, which is arguably one of the most important components of a technology evaluation. Additionally, the ease of use cannot be measured without working directly with potential users in the field.

That said, the type of information gathered in the multi-parameter water test kit evaluation can help focus the time and resources spent in the lab and field to fill in the gaps not provided by the rapid evaluation approach. Therefore, the CITE team will conduct a similar process as a part of future evaluations.

2) A Group Decision Support System can be used to customize the results of the evaluation based on the decision maker's context and preferences.

A GDDS was developed for this evaluation, which can be applied to other evaluations as well. The focus of the GDSS is to allow stakeholders to choose the weights of the criteria, utilizing expert opinion and their personal preferences. However, since the relative power of each stakeholder is important in calculating the group opinion and those living in poverty often have little power, the technology choice may not reflect their preferences as consumers which can lead to a mismatch between supply and demand.

Left: Thaltej resident Shanthok completes H_2S tests after being shown a demonstration. Credit: Vihar Parikh; Right: Two days after she completed the test, Gulbai Tekra resident Kavita Solanki (left) stands with her mother and displays her negative H_2S test results. Credit: Jonars Spielberg

3) Partnership engagement proved crucial to the evaluation's success.

Working with local academic partners and utilizing their established community partnerships facilitates the research to a great extent. IIM-A faculty and students, in particular, possessed strong relationships with local communities, which made access to them considerably easier.

4) The mix of survey-based data (perceptions) and observation-based tests (actual behavior) proved powerful.

In particular, the observation protocol, which consisted of a checklist of observations researchers noted as subjects complete the test, proved useful and valuable. The CITE team will likely use a similar technique in future evaluations to measure ease of use.

Authors & Acknowledgements

This research would not have been possible without the guidance and support of faculty, staff, and students at the Massachusetts Institute of Technology as well as students and faculty mentors at India Institute of Management, Ahmedabad and TERI University, New Delhi.

Report Authors

Linda Annala
Susan Murcott (Team Lead)
Jonars Spielberg
Sydney Beasley
Vihar Parikh
Innocent K. Tumwebaze, PhD
Jennifer Green (Team Lead)
Sara Lynn Pesek
Mahdi Zarghami, PhD

Additional CITE Team Members

Alison Hanson
Jason Knutson
Teng Ke Wong
Akshay Jain
Shuyue Liu
Megan Keegan
Jenny Tanphanich

Indian Institute of Management - Ahmedabad

Reecha Das
Deepika Saluja
Nandini Singh
Saritha S Vishwanathan
Advaita Rajendra
Prof. Ankur Sarin
Bianca Shah
Ashish Ranjan
Rashi Seth
Vishal Vaghela

TERI University — New Delhi

Sonal Bindal
Nidhi Jha
Chander Singh, PhD

Other Contributors

Vandana A. Pandya
Shrikant Brahmbhatt
Utsav Patel, Government Engineering College – Gandhinagar
Siddharth Kavthekar
Netaji Subhas Institute of Technology (NSIT) – Jetalpur
With unknown last names: Sanjitha, Sulekha, Lalbhai