Journal of Undergraduate Chemistry Research, **2024**, 23 (4), 54

ASSESSMENT OF THE PRECISION AND ACCURACY OF AT-HOME MULTI-TEST WATER QUALITY TEST STRIPS

Gabriel G. Swinney*, Sarah A. Mullins*, Jackson Martin*, Jakeb B. Kinsey*, and Daniel A. King†

Department of Chemistry and Biochemistry, Taylor University, Upland, IN 46989

Abstract

One-fourth of the global population and 77 million Americans are supplied drinking water from contaminated, unsafe water supplies which prompts need for a reliable and efficient method for household water testing. Water test strips fulfill the need for a consumer-friendly form of water testing, though their reliability remains unestablished. In efforts to evaluate test strip reliability, the precision and accuracy of commercially available test strips were assessed. The test strips appeared consistent for individuals, but large variability was observed between users. Results also suggest that much of the variability between users relates to their differing perception of the color pads. The test strips also revealed modest inaccuracies and generally had smaller dynamic ranges than the packaging would suggest. However, the consistency observed for individual users suggests that the routine use of test strips for monitoring household drinking water supplies may be sufficient for detecting significant changes in contamination levels.

†Corresponding author: dnking@taylor.edu *Undergraduate researchers and co-authors

Keywords: drinking water, water testing, test strips

Received: August 5, 2024 Accepted: October 26. 2024 Published: November 4, 2024

Introduction

In 2024, one in four individuals across the globe remain without access to a safe water supply, and one in three individuals remain without access to safe sanitation.[1,2] This inequality disproportionally affects impoverished urban neighborhoods and low-income populations across the world. Contaminated drinking water also remains to be a primary determinant in decreased life expectancies within these regions.[3,4] One UNICEF study discovered that waterborne illnesses are more than twice as likely to kill children relative to conflict-related deaths in low-income countries.[5] While this inequity is most pervasive in undeveloped countries, the United States' population also experiences adverse health effects due to contaminated water supplies. To alleviate this crisis, the United States' Environmental Protection Agency (EPA) first enacted the Safe Water Drinking Act in 1974 to help prevent Americans from unsafe drinking water. Nearly fifty-years later, there are still 77 million Americans who receive water from highly contaminated systems that violate of the Safe Water Drinking Act's water requirements.[6,7]

Acute or chronic exposure to contaminated drinking water can lead to several adverse effects. Acute exposure to polluted drinking water may cause metal poisoning, respiratory infection, and gastrointestinal illness,[8,9,10,11] while chronic exposure to polluted drinking water can be associated with cancers, neurological disorders, reproductive issues like infertility and birth defects, and fluorosis in bones and teeth.[8,12,13,14] The 2014 Water Crisis in Flint, Michigan is one example of a government-regulated water supply that delivered households elevated concentrations of lead and other hazardous metals. The error in water regulation caused over half of the city's households to report one or more family members suffering from unprecedented mental and physical illnesses.[15,16,17] American Consumer Confidence Reports regarding drinking water are at all-time lows in response to the frequent regulatory violations of drinking supplies that are causing or perceived to be causing adverse health effects for consumers. The unfortunate reality of gaps within governmental regulation of water leads many Americans to perform at-home water testing to ensure their water source is safe.[18,19]

The current methods available for at-home drinking water testing include shipping water samples to specialized laboratories, liquid at-home test kits, and at-home test strips. Of these testing approaches, the most reliable option for a high-quality analysis of contaminants is to use laboratory testing.[20,21] The disadvantages of this method are that it is costly and incurs a slow turnaround time relative to the alternative testing options. A water supply's concentration of metals and other contaminants can change daily as seen in Flint, Michigan, which demands a quicker method that permits more frequent water testing. Liquid at-home test kits provide consumers with nearly immediate results while also delivering a high-quality analysis of household water supplies. However, the liquid test kits are more complicated and may cause novice consumers to struggle to perform and accurately interpret results. Additionally, liquid test-kits remain costly and have limited shelf lives.[18] At-home multi-test test strips are inexpensive and easy to read, which gives consumers access to an approach that can be conveniently and frequently used.[22] Test strips may be ideal for consumers between its cost-effectiveness and speed of use, as long as the quality of results are sufficiently useful. Previous studies determined that consumers report low levels of confidence regarding the reliability their water testing strips provide.[18,23] This study's intent is to evaluate the reliability of at-home multi-test test strips by evaluating the accuracy and precision of five commercially available water multi-test test strips.

Experimental Methods

Materials and Data Normalization.

16-in-one multi-test test strips were purchased from Varify (A), Bestprod (B), Pamasana (C), JNW Direct (D), and Tespert (E). In an effort to evaluate the behavior of the test strips generally rather than each test individually and because each parameter may have very different ranges and even non-linear scales it became essential that the results be normalized. To this end, each result (either a concentration or pH) was converted to the color pad number the value represents on the color charts provided by the manufacturer. Normalizing the data to integer Pad Numbers provided two significant benefits – all parameter results would be on a similar scale to allow for general comparisons and standard deviations would be of a useful scale where anything less than +/- 0.5 would suggest the ability to distinguish between adjacent concentration levels, particularly between safe and unsafe level thresholds.

Application of Test Strips.

A lab experiment was conducted in an introductory chemistry course to evaluate whether test strip instructions are sufficient for consumers to generate consistent results. 80 novice students were asked to determine the concentrations of a prepared standard solution first by following the manufacturer's instructions and then a second time following more detailed instructions developed by our research assistants. The primary difference between the two was that the expanded instructions directed students to take a photo of their test strip immediately after testing to alleviate the concern of assigning all of the results before colors began to change or fade. Students did not receive any additional guidance from lab instructors beyond the written instructions included below.

The standard evaluation solution contained 200ppm Ca, 50ppm Zn, 1ppm Fe, and 1ppm Cu in deionized water and students used test kit A.

The manufacturer's instructions consisted of three steps:

1. Dip test strip into water for 2 seconds and remove.
2. Gently shake off any excess water from the test strip.
3. Compare test pads to the color chart immediately (discard strip after 60 seconds).

The research assistants' expanded instructions were as follows:

1. Obtain a water sample with a depth greater than the length of the strip, have your camera app opened on the phone, have a paper towel prepared, and try to be near a source of natural light. Have the chart ready to use.
2. Completely Dip the strip into the water sample with tweezers held on the head of the strip for 2 seconds and then remove the strip from the sample (avoid touching any of the pads with exposed hands).
3. Remove excess water from the strip by tapping its side against the towel, before laying it down on a flat surface (do not lay the strip on the towel).
4. Immediately after dipping, take a photo of the strip in natural light to compare to the color chart.
5. Compare the photo of the strip to the color chart.

Differences between test 1 (manufacturer's instructions) and test 2 (expanded instructions) results were analyzed via comparison of means and standard deviations.

Survey.

A color identification survey was completed by the same 80 novice students to simulate the consumer experience of matching the colors on a test strip to the colors on the color chart of commercial bottle. The survey contained a color identification question for each of the 16 parameters. Each question included a picture of the color chart and a picture of one of the color pads taken from the chart, and students were asked to match the color to the chart (Figure 1). The survey's results were evaluated by comparing means and standard deviations.

Test Strip Accuracy.

The accuracy of five commercially available multi-test test trips was evaluated for pH, copper, iron, and lead. For pH analysis, standard pH buffers (pH 4.0 and 10.0) (Fisher Scientific) were used to calibrate a digital pH probe (Pasco Scientific) and the test solution was observed to have a pH of 7.1. For metal analysis, a set of calibration standards were prepared from purchased standard stock solutions (Fisher Scientific) with the following concentrations: copper (0.3125ppm, 0.625ppm, 1.25ppm, 5ppm, and 10ppm), lead (1ppb, 20ppb, 100ppb, 200ppb, 1000ppb) and iron (0.3125ppm, 0.625ppm, 1.25ppm, 5ppm, and 10ppm). The concentrations of the standards were designed to cover the range values on the color charts provided by the manufacturers. The concentration of each metal standard was validated by an iCE3000 atomic absorption spectrometer (Thermo Fisher Scientific).

Results and Discussion

Application of Test Strips.

80 novice students tested a common water sample with multi-test test strips once using the brief manufacturer's instructions and once using expanded instructions. Means and standard deviations were calculated to evaluate interpersonal precision of the test strips and any potential effect of the quality of the instructions.

As seen in Figure 2, the standard deviations showed small

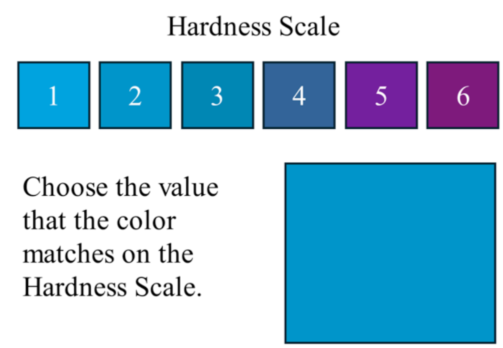

Figure 1. A sample survey question is shown.

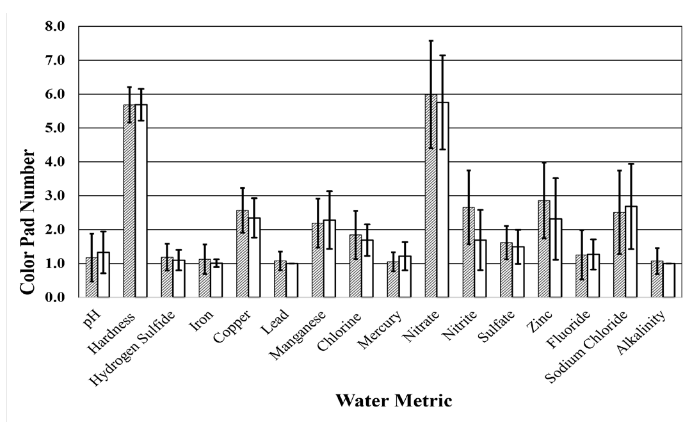

Figure 2. Mean of outcomes from Test 1 (black) using bottle instructions and Test 2 (white) using expanded set of instructions of novice users conducting test strip analysis. Plus and minus one standard deviation for each data set is represented by error bars.

improvements for 11 of the 16 metrics when using expanded instructions (test 2) instead of the commercial bottle's instructions (test 1). A paired t-test comparing each metric's standard deviations from test 1 and test 2 indicated that there is a statistically significant difference in the standard deviations between test 1 and test 2 (p-value < 0.05) despite their not being a significant difference in mean outcomes (p-value > 0.05). It should be noted that a t-test is run in this case to compare these two paired sets of standard deviations rather than an F-test, which would typically only be used to compare two standard deviations of two sets of repeated measures. The average standard deviation for test 2 was 0.6, only slightly smaller than the 0.7 average standard deviation of test 1. Therefore, the precision of interpreting test strip results improved when using an expanded set of instructions, though the improvement was modest.

Survey.

80 novice students were asked to complete a color matching survey mimicking the step of assigning test strip results to color charts. The data was compiled and the resulting standard deviations, in terms of color pad numbers, are shown in Figure 3.

The highest standard deviation from the color identification survey was 0.71 (hardness), and 12 of the 16 measures reported standard deviations less than 0.5. The standard deviations from the color identification survey were generally lower than the standard deviations resulting from students using the test strips and then also assigning values during Tests 1 and 2. However, the average observed standard deviation from the survey represents 70-75% of that observed during the full tests suggesting that most of the variance may be attributed to different people inconsistently matching color pads to the charts. Therefore, the high precision observed among replicates when individual researchers used the test strips repeatedly (data presented in the next section) suggests that the primary source of variance is purely interpersonal. Essentially, test strips yield consistent colors, and each individual consistently matches the color pads to the charts, but different people tend to interpret those colors differently.

Test Strip Accuracy.

For five commercially available test strips, their accuracy was assessed by evaluating the difference between calibrated solution concentrations and concentrations measured by the test strip for each of the following four important water metrics: pH, Copper, Lead, Iron (II).

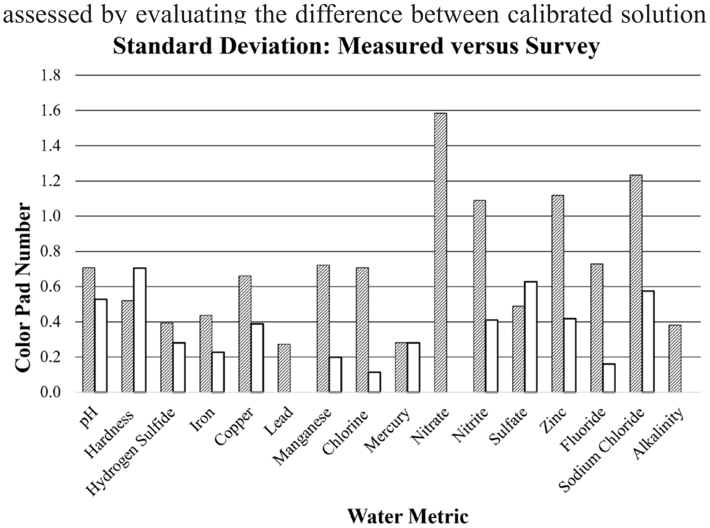

Figure 3. Standard Deviations of Test 1 (black) and the color identification survey (white) are shown.

In contrast to the higher variability observed between many users, Table 1 shows that for an individual researcher, pH measurements remained precise through triplicate tests. One of the five strips (E) reported a pH within 0.1 pH of the pH determined by a digital pH meter and four of the five strips reported a pH within 1.0 pH of the expected value. This indicates commercial test strips are precise in pH testing, though there is modest error in the test strips' accuracy.

A set of copper standards were tested in triplicate for all five commercial test strips (Figure 4). All observed results were consistent between triplicate measurements. One of the five commercial test strips detected copper at concentrations of 1.25ppm or less. Each manufacturer test strip bottle reported that copper concentrations above 1ppm are unsafe for human consumption, although three of the five test strips report a copper concentration of less than 1ppm for the 2.5ppm copper solution, and two of the five test strips report a copper concentration of less than 1ppm when testing a solution containing 5ppm copper. Therefore, if a consumer tests their water supply with test strips that report safe copper concentrations, it is possible consumers are still subject to unsafe levels of copper due to the inaccuracy of their test strips.

Table 1. Triplicate measurements of pH testing from the five selected test strip companies are shown

Test Strip pH Measurement

Test Strip Brand	Trial 1	Trial 2	Trial 3
Varify	8	8	8
Bestprod	8.2	8.2	8.2
Pamasana	7.6	7.6	7.6
JNWD	8	8	8
Tespert	7.2	7.2	7.2

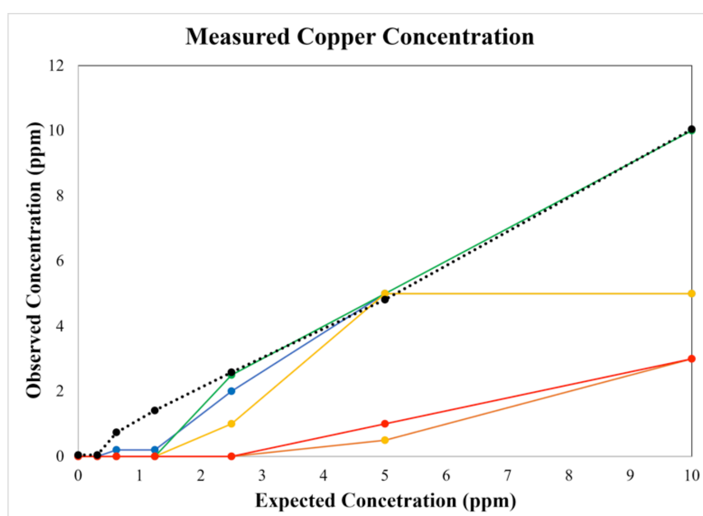

Figure 4. Copper concentrations were tested at 0.3125ppm, 0.625ppm, 1.25ppm, 2.5ppm, 5ppm, and 10ppm. The line graph shows the solution concentrations reported by the Atomic Absorption Spectroscopy (black), Varify (blue), Bestprod (orange), Pamasana (green), JNWD (yellow), and Tespert (red).

The results from the lead analysis are shown in Figure 5. Consistency was observed between replicates and consistency between manufacturers at higher concentrations of lead (200ppb and 1000ppb). At concentrations between 1ppb and 100ppb, the five commercial test strip brands lacked agreement. The test strip bottles indicate any trace of lead is unsafe for consumption, though only one of the five commercial test strips detected lead in solution when 1ppb was present.

Test strips showed precision among replicates and consistency between manufacturers for Iron (II) analysis (Figure 6). Accurate measurements were only observed for the highest concentration, 10ppm, but for solutions with concentrations between 0ppm and 5ppm, Iron (II) remained undetected by each of the test strips. According to the safe concentration ranges listed on each of the bottles, any iron amount between 0-5 ppm is safe for consumption. While the test strips were flawed in detecting exact iron concentrations, they were sufficient for indicating the unsafe iron level and need of treatment.

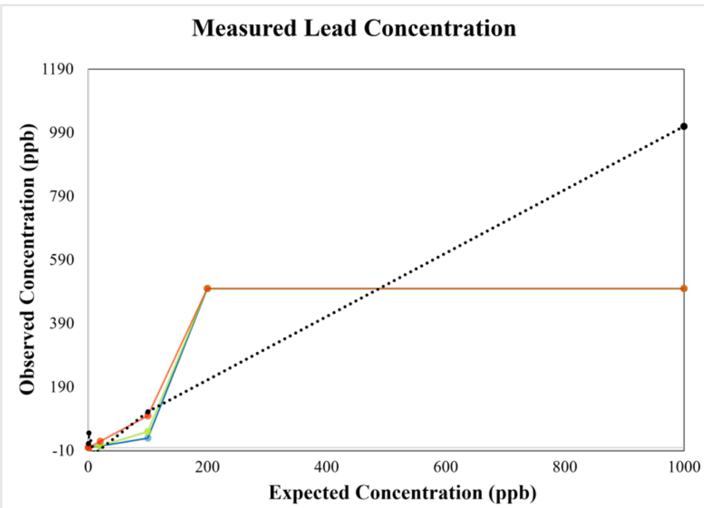

Figure 5. Lead concentrations were tested at 1ppm, 20ppm, 100ppm, 200ppm, and 1000ppm. The line graph shows the solution concentrations reported by the Atomic Absorption Spectroscopy (black), Varify (blue), Bestprod (orange), Pamasana (green), JNWD (yellow), and Tespert (red).

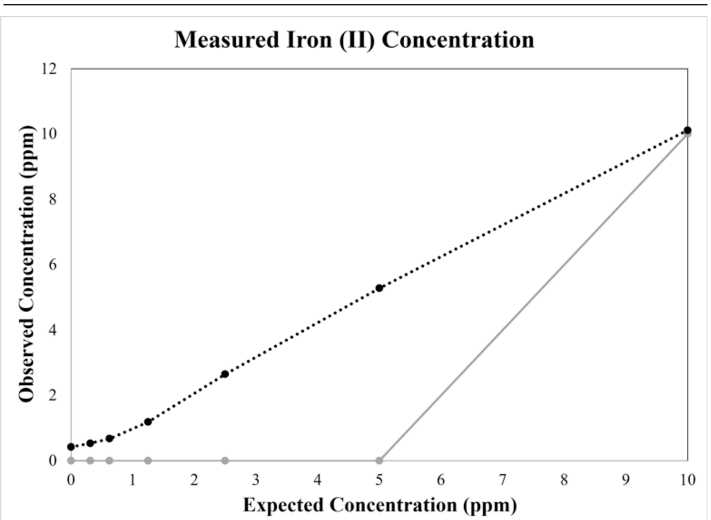

Figure 6. Iron (II) concentrations were tested at 0.3125ppm, 0.625ppm, 1.25ppm, 2.5ppm, 5ppm, and 10ppm. The dashed line shows the solution concentrations reported by the Atomic Absorption Spectroscopy (black). Each sample reported identical outcomes (gray).

In conclusion, an individual homeowner using the same multi-test strip kit should expect to see consistent behavior of the test strips. Test strip measurements may not be very accurate at low levels, but seem to be more accurate at higher levels, particularly levels of concern. Consequently, during routine testing, any significant change in water chemistry would likely result in an observable change in test strip behavior. A set of instructions that was more expansive than commercial bottles' instructions modestly improved precision between different users, but only marginally impacted mean measured values. The color identification survey confirmed that much of the error associated with imprecise test strip readings among the 80 subjects is a result of different color interpretations that occur with interpersonal analysis. Consequently, it is expected that a homeowner would observe consistent response when using test strips for routine water monitoring. Additionally, the color charts provided with the kits significantly overestimate the dynamic range the kits are capable of measuring. Although most of the lower levels on the charts are below the detection limits of the test strips, the test strips seem to be generally sensitive at the levels of concern for the parameters explored in this work.

References

1. Thomas, E. Turning Global Water Security Research into Policy and Action. *PLOS water* **2024**, *3*.
2. WHO; UNICEF. *State of the World's Drinking Water: An Urgent Call to Action to Accelerate Progress on Ensuring Safe Drinking Water for All*; Geneva; World Health Organization, **2022**.
3. Mújica, O. J.; Haeberer, M.; Teague, J.; Santos-Burgoa, C.; Galvao, L.A.C. Health Inequalities by Gradients of Access to Water and Sanitation between Countries in the Americas, 1990 and 2010. *PubMed,* **2015**, *38*, 347–354.
4. Swanston, S.F. *Race, Gender, Age, and Disproportionate Impact: What Can We Do about the Failure to Protect the Most Vulnerable?*; Fordham Urban Law Journal, **1994,** *21*, 577–604.
5. UN News. *More children killed by unsafe water, than bullets, says UNICEF chief*. UN News. **2019**, https://news.un.org/en/story/2019/03/1035171.
6. US EPA. *Understanding the Safe Drinking Water Act*; **2024**.
7. Fedinik, K. P.; Taylor, S.; Roberts, M. *Natural Resource Defense Council Report,* **2019**, 12–46.
8. Vaishaly, A.G.; Mathew, B.B.; Krishnamurthy, N.B. Health Effects Caused by Metal Contaminated Ground Water. *International Journal of Advances in Scientific Research,* **2015**, *1*, 60-64.
9. Jadhav, S. H.; Sarkar, S. N.; Patil, R. D.; Tripathi, H. C. Effects of Subchronic Exposure via Drinking Water to a Mixture of Eight Water-Contaminating Metals: A Biochemical and Histopathological Study in Male Rats. *Archives of Environmental Contamination and Toxicology,* **2007**, *53*, 667–677.
10. Strauss, B.; King, W.; Ley, A.; Hoey, J. R. A Prospective Study of Rural Drinking Water Quality and Acute Gastrointestinal Illness. *BMC Public Health,* **2001**, *1*, 1-6.
11. Pizarro, F.; Olivares, M.; Gidi, V.; Araya, M. The Gastrointestinal Tract and Acute Effects of Copper in Drinking Water and Beverages. *Reviews on Environmental Health*, **1999**, *14*, 231-238.

12. Bove, F. J.; Greek, A.; Gatiba, R.; Boehm, R. C.; Mohnsen, M. M. Evaluation of Mortality among Marines, Navy Personnel, and Civilian Workers Exposed to Contaminated Drinking Water at USMC Base Camp Lejeune: A Cohort Study. *Environmental Health,* **2024**, *23*, 61.
13. Veeramachaneni, D. N. R.; Palmer, J. S.; Amann, R. P. Long-Term Effects on Male Reproduction of Early Exposure to Common Chemical Contaminants in Drinking Water. *Human Reproduction*, **2001**, *16*, 979–987.
14. Rodriquez, A.G.P.; Sandoval, G.A.; Alfaro, J.L.G.; Burgos, H.S. *Exposure to Contaminants in Karst Groundwater*, Springer International Publishing **2024**, 1–20.
15. CDC. *Story: Flint Water Crisis*. CASPER. **2024**, https://www.cdc.gov/casper/php/publications-links/flint-water.html
16. Cuthbertson, C. A.; Newkirk, C.; Ilardo, J.; Loveridge, S.; Skidmore, M. Angry, Scared, and Unsure: Mental Health Consequences of Contaminated Water in Flint, Michigan. *Journal of Urban Health* **2016**, *93*, 899–908.
17. Fortenberry, G. Z.; Reynolds, P.; Burrer, S. L.; Johnson-Lawrence, V.; Wang, A.; Schnall, A.; Pullins, P.; Kieszak, S.; Bayleyegn, T.; Wolkin, A. Assessment of Behavioral Health Concerns in the Community Affected by the Flint Water Crisis — Michigan (USA) 2016. *Prehospital and Disaster Medicine* **2018**, *33*, 256–265.
18. Kriss, R.; Pieper, K. J.; Parks, J.; Edwards, M. A. Challenges of Detecting Lead in Drinking Water Using At-Home Test Kits. *Environmental Science & Technology* **2021**, *55*, 1964–1972.
19. Iuliano, K. R.; Fox, M.; Munroe, J.; Nachman, K. E.; Spannhake, B.; Burke, T. A. "Is My Tap Water Safe to Drink?": Assessing a National Sample of Consumer Confidence Reports for Public Health Communication. *EarthArXiv (California Digital Library)* **2024**.
2). Isbell, F.I.; Mercer, D.R.; Iqbal, M. Reliability of Water Quality Test Strips for Field Measurement of Nutrients Essential for Developing Mosquito Larvae. *American Journal of Undergraduate Research*. **2006**, *5*.
21. Bischoff, M.; Hiar, A. M.; Turco, R. F. Evaluation of Nitrate Analysis Using Test Strips: Comparison with Two Analytical Laboratory Methods. *Communications in Soil Science and Plant Analysis* **1996**, *27*, 2765–27.
22. Brouwer, S.; van der Wielen, P.; Schriks, M.; Claassen, M.; Frijns, J. Public Participation in Science: The Future and Value of Citizen Science in the Drinking Water Research. *Water,* **2018**, *10*, 284.
23. Jones, M. Determining the Ability of an At-Home Test Kit to Detect Lead in Drinking Fountains at East Carolina University. *The ScholarShip* – East Carolina University, **2022**.