Check for updates

AUTHORS:
Riyaadh Mayet[1]
Craig M. Sheridan[1]
Simone Dahms-Verster[1]

AFFILIATION:
[1]Wits Water, School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Craig Sheridan

EMAIL:
craig.sheridan@wits.ac.za

DATES:
Received: 28 July 2025
Revised: 17 Mar. 2026
Revised: 24 June 2026
Accepted: 25 June 2026
Published: 29 Sep. 2026

HOW TO CITE:
Mayet R, Sheridan CM, Dahms-Verster S. How accurate are water quality test strips in South Africa? S Afr J Sci. 2026;122(9/10), Art. #23287. https://doi.org/10.17159/sajs.2026/23287

ARTICLE INCLUDES:
☒ Peer review
☒ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Philani Mashazi

KEYWORDS:
water quality monitoring, linear regression analysis, citizen science, smartphone colourimetry

FUNDING:
South African National Research Foundation (MND210620613627), Ninety One, The Claude Leon Foundation, University of the Witwatersrand

© 2026. The Author(s). Published under a Creative Commons Attribution Licence.

How accurate are water quality test strips in South Africa?

Water quality monitoring is essential for public health and sustainable development, yet laboratory-based analysis remains expensive in resource-limited contexts. Four water quality test strip products available in South Africa were evaluated by comparing test strip measurements of pH, nitrate, nitrite and hardness with reference measurements, with strip readings conducted under indoor (1000 lx) and outdoor (5000 lx) lighting conditions. The results indicate that while pH and nitrate measurements are reasonably accurate, water hardness exhibits the highest error rates (normalised root mean square error up to 43%), highlighting a need for improved assay chemistry. Outdoor lighting improved accuracy by an average of 12% compared with indoor lighting ($p = 0.0354$), although this improvement was not consistent across all parameters, with hardness remaining a clear exception. Smartphone-based interpretation did not outperform visual reading ($p = 0.0382$), as digital sensors proved more sensitive to white-balance fluctuations and noise. Furthermore, price was not a reliable predictor of accuracy, with cheaper strips often matching or exceeding the performance of expensive ones. An important limitation of the study is that measurements were conducted by a single user, and insights into smartphone-based analysis were based on a single product (the only one available at the time) and two smartphone devices. While the findings are limited to the products and conditions tested, the results suggest that test strips are not yet suitable for regulatory enforcement but are effective for trend detection when used with standardised protocols. These findings provide a vital framework for making low-cost monitoring technologies robust enough to support Sustainable Development Goal reporting and environmental advocacy.

Significance:

- Affordable testing tools empower underserved communities to participate in water management, bridging the gap between professional monitoring and local advocacy.
- Identifying the limitations of specific tests ensures that community-led projects focus on reliable data that can actually influence policy.
- Understanding how lighting and hardware affect digital analysis allows developers to create more resilient, field-ready environmental monitoring software for smartphones.
- These findings provide a framework for developing universal protocols for the use of test strips in citizen science water quality monitoring.

Introduction

Water quality monitoring is the systematic process of monitoring and assessing various physical, chemical and biological parameters of water.[1] Water quality is monitored to evaluate its condition for specific purposes such as drinking, sanitation, agriculture and recreation. To determine whether water is suitable for a specific purpose, these parameters are compared to a set of established numerical standards[2] defined by governmental bodies and international organisations such as the World Health Organization (WHO).[3] Water quality is a complex issue that is affected by both anthropogenic and environmental influences.[4] It is important to collect water quality data to ensure compliance with national and (potentially) international standards[5] and to monitor indicators for sustainable development[6]. By conducting regular monitoring, authorities and researchers can establish trends, address contamination issues promptly and ensure water resources remain safe and sustainable for future generations.

Water quality is monitored through a range of physical, chemical and biological parameters.[3] The process involves the collection of water samples from various sources and the subsequent analysis of relevant indicators such as pH, nutrient concentrations and the presence of microbial contamination. In a laboratory, various methods are used to assess these parameters, including spectroscopy, chromatography, microscopy, titration and gravimetric analysis (amongst others).[7] These techniques are limited by the use of expensive equipment and high operational costs, making them impractical for field-testing and citizen science initiatives.

Among various testing methods, test strips offer an affordable, convenient and, most importantly, rapid means of assessing key chemical parameters. Test strips enable broader participation in water quality monitoring, allowing citizens to contribute to data collection through citizen science initiatives.[1,8] Water quality test strips are less accurate than laboratory methods due to subjective interpretation.[9] The manufacturing and quality control processes are also less rigorous, leading to reduced accuracy and precision. Test strips consist of a plastic or paper substrate with multiple chemically treated reagent pads that undergo analyte-induced chromogenic reactions, producing characteristic colour changes proportional to analyte concentration.[10]

The concentration of each analyte is estimated by comparing the final colour to a calibrated reference colour chart with known concentrations. The results can also be read using a digital reader[11], such as a smartphone camera. Ambient light is reflected from the test strip surface into the camera lens, where it is filtered by the Bayer filter array into red, green, and blue wavelength components. These photons are detected by the complementary

metal oxide semiconductor (CMOS) photodiode array, where photon–electron interactions generate electrical charges proportional to local light intensity. The charge carriers are then digitised, converting analogue signals into digital red, green, and blue pixel values, and the raw images are formatted through image-processing pipelines that may include white balance and gamma corrections, noise-filtering and colour space conversions.[12] These quantitative metrics are transformed into concentrations either through analytical calibration models or via machine learning algorithms, which enable automated prediction while accounting for non-linear responses, illumination variability and device-specific sensor effects. Smartphone applications also enable data logging, trend tracking, diagnostic feedback and testing reminders.[13]

Several studies have examined the performance of water quality test strips across different parameters and contexts.[9,14,15] Collectively, their findings indicate that, while test strips can be reasonably precise for screening or trend detection, their analytical accuracy varies with the product, measured parameter, water source, and lighting conditions, limiting their suitability for standardised applications. However, these studies did not perform inferential statistical analyses to assess significance, and the lighting conditions during the experiments were not quantified. Studies on test strip analysis using smartphones have focused on developing new applications and algorithms rather than evaluating existing products.[11,16] Some studies have shown that smartphones tend to overestimate high-concentration samples because of scattering and are prone to errors caused by sample interferences and variations in ambient lighting.[17,18] Despite their widespread availability, the analytical performance of test strip products sold in South Africa remains largely uncharacterised.

In this study, four water quality test strip products available in South Africa were evaluated under indoor and outdoor lighting conditions. Their accuracy and precision were assessed by comparing values obtained from the test strips with reference measurements. The evaluation focused on pH, nitrate, nitrite and general hardness. By offering a comprehensive comparison of water quality test strips in South Africa, this research seeks to guide manufacturers, consumers, environmental professionals, farmers and policymakers in selecting the most appropriate tools and practices for water quality assessment using test strip technology.

Methods

Test strip selection

Four commonly used and commercially available water quality test strip products were selected for comparison in this research (Table 1). The strips were purchased from Takealot, South Africa. The products were selected according to their price, availability, number of parameters and the method used to obtain data and therefore provide good representation.

Standard solution preparation

Standard solutions of pH, nitrate, nitrite and calcium carbonate were prepared in deionised water. This was done to minimise chemical matrix interferences and assess test strip performance under best-case conditions. In natural water samples, additional constituents may introduce interferences that could further reduce analytical accuracy. This is a recognised limitation of this study.

pH

A series of standard solutions with a pH of 6.5–8.5 ± 0.05 were prepared. The pH range was selected to span the testing range specified by each manufacturer. The selected range also falls within the drinking water limits defined by the South African National Standard.[2] Solutions with a pH of 6.5–7 were prepared using 10 mM 2-morpholin-4-ylethanesulfonic acid (MES) hydrate ($C_6H_{13}NO_4S.xH_2O$), while those with a pH of 7–8.5 were prepared using 10 mM 2-amino-2-(hydroxymethyl) propane-1,3-diol (Tris-HCl) ($C_4H_{11}NO_3.HCl$). These chemicals were selected specifically for pH adjustment and were not intended to assess other parameters. The pH of the standard solutions was adjusted to the final pH, where necessary, with 1 M HCl or 1 M NaOH. The final pH of the solutions was verified, in triplicate, using a Crison Model 20 pH meter (Merck, Darmstadt, Germany) (Supplementary figure 1). The solutions were prepared immediately before the experiment, and all experiments were conducted at 20 °C.

Nitrate, nitrite and total hardness

A series of standard solutions for nitrate (0–500 mg/L ± 5 mg/L NO_3^-), nitrite (0–10 mg/L ± 0.5 mg/L NO_2^-) and total hardness (0–1000 mg/L ± 10 mg/L $CaCO_3$) were prepared by diluting their respective 1000 mg/L Certipur® Standards (Merck) to the desired concentrations with deionised water. The concentration ranges were selected to span the testing range specified by each manufacturer. These ranges do not correspond to the drinking water limits defined by the South African National Standard.[2] The concentrations were verified (in triplicate) using Spectroquant® photometric tests (Merck) for nitrate (0.4–110.7 mg/L NO_3^-) (109713), nitrite (0.007–3.28 mg/L NO_2^-) (114776) and total hardness (12–537 mg/L $CaCO_3$) (100961) (Supplementary figures 2–4). The spectra were obtained with a Spectroquant® Prove 600 spectrophotometer (Merck) and a quartz cuvette (Merck) with a path length of 10 mm. The solutions were prepared immediately before the experiment, and all experiments were conducted at 20 °C.

Testing procedure of the test strips

The products were used according to the manufacturer's instructions. In general, each test strip was immersed into 500 mL of the respective standard for 2–3 s. The strip was then removed and read after the specified waiting period (Table 1). Products A and B refer to the same product (test strip) but were analysed using different smartphone devices. Product A was analysed using a Huawei Nova 5T (Huawei, Shenzhen, China), while Product B was analysed using an Honor 200 (Honor Device Co, Ltd., Shenzhen, China), both loaded with the JBL ProScan application (JBL GmbH & Co. KG, Neuhofen, Germany).[13] Products C, D and E were interpreted visually by a single user through direct comparison with the manufacturer-provided charts. The entire testing procedure performed in the laboratory was repeated outdoors in a shaded area on a sunny day to assess whether environmental lighting influenced test strip performance. Illuminance levels were quantified in lux (lx, lm/m²), with indoor experiments conducted at approximately 1000 lx and outdoor measurements conducted at approximately 5000 lx. To assess the accuracy of each parameter, standard solutions were prepared and assessed separately for each analyte, product and lighting condition. This was done to minimise potential interference effects

Table 1: Description of the water quality test strip products used in this study

Product	Price (ZAR) per test	Reading time (s)	No. of parameters	Measurement method	Country
JBL ProScan (A/B)	25	60	7	Smartphone	Germany
JBL ProAquaTest (C)	8.10	60	7	Visual	Germany
DNX Medical 16 in 1 (D)	4.50	30	16	Visual	South Africa
The Chemistry Solutions Company 21 in 1 (E)	17.50	30	21	Visual	South Africa

that could arise from testing multiple variables simultaneously, which might affect colour development or cause cross-reactivity between test pads. The tests were repeated five times at each concentration, and the arithmetic mean and standard deviation were then calculated.

Data analysis

Precision was evaluated using the standard deviation of replicate measurements. Linear regression analysis was performed to evaluate the relationship between test strip values and the corresponding concentrations determined by spectroscopy and pH measurements. For nitrate, nitrite and hardness, the regression was constrained through the origin because the strips always produce a zero response at zero concentration. The slope was used to assess proportional agreement between strip measurements and reference concentrations. Analytical accuracy was quantified using the root mean square error (RMSE) between the observed test strip values and the corresponding concentrations from spectroscopy and pH measurement. To enable comparison across analytes with different concentration ranges, RMSE values were normalised to the tested concentration range and expressed as a percentage (nRMSE). Differences among analytes and among products were assessed using one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test. Differences between measurement approaches and between lighting conditions were evaluated using unpaired two-tailed *t*-tests.

Results and discussion

pH Analysis

Under indoor conditions (Figure 1A), all products exhibited strong linear relationships with reference measurements (R^2 = 0.92–0.96), although regression slopes (m = 0.94–0.98) indicate consistent proportional underestimation of pH. Visually interpreted Products C (m = 0.98, R^2 = 0.95) and E (m = 0.97, R^2 = 0.96) had slopes closest to unity, while smartphone-based Products A (m = 0.95, R^2 = 0.92) and B (m = 0.95, R^2 = 0.94) deviated slightly more but remained strongly correlated. Replicate measurements were generally precise across all products, as indicated by low standard deviations. Descriptive statistics for all data sets are given in Supplementary tables 1 and 2, while full linear regression results are provided in Supplementary table 3.

In outdoor conditions (Figure 1B), linear relationships remained strong (R^2 = 0.91–0.99), and slopes moved closer to unity for several products (m = 0.87–1.03), suggesting improved proportional agreement under natural lighting. Visually interpreted Products C, D and E showed slopes near the ideal (m = 0.99–1.03), while smartphone-based Product A (m = 0.87, R^2 = 0.91) showed greater underestimation outdoors. Although proportional agreement improved, some visually interpreted strips showed slightly higher variability between replicates. The improvement in visual strip accuracy outdoors likely reflects enhanced colour contrast under brighter, full-spectrum light, while increased variability for some visually interpreted strips indicates the challenge of consistent human reading under high-intensity lighting.

Nitrate analysis

As observed for pH indoors, nitrate measurements under indoor conditions (Figure 2A) showed strong linear responses with the reference measurements (R^2 = 0.85–0.91) but consistently underestimated concentrations (m = 0.63–0.95). This underestimation was more pronounced than that for pH. Smartphone-based products showed the largest proportional bias (m = 0.63–0.71), while visually interpreted strips were closer to the ideal line, particularly Product E (m = 0.95). The R^2 values indicate that, despite this bias, all products tracked increasing nitrate concentrations consistently across the tested range. Replicate measurements were more precise for smartphone products, whereas visual strips showed greater variability at higher concentrations. This likely reflects the difficulty of visually distinguishing the darker nitrate colour intensities that develop at elevated concentrations under indoor lighting.

Outdoors (Figure 2B), proportional agreement improved for all products, with slopes approaching unity (m = 0.96–1.02). Linear relationships remained strong (R^2 = 0.91–0.97), slightly lower than those for pH (R^2 = 0.91–0.99), indicating that nitrate concentrations were still reliably tracked but with marginally greater scatter. However, the improved accuracy observed for visual strips C and E was accompanied by reduced precision, as replicate variability increased under bright light. In contrast, smartphone products maintained both slopes near unity and consistent replicate agreement, suggesting that digital colour analysis is less sensitive to high-intensity outdoor lighting.

Nitrite analysis

The assessment of nitrite test strips (Figure 3) revealed substantial differences between measurement approaches. Under indoor conditions (Figure 3A, 3C), all products underestimated nitrite concentrations, with slopes well below unity (m = 0.59–0.94), indicating consistent proportional bias. Smartphone-based products showed the greatest underestimation (m = 0.59–0.61), whereas visually interpreted strips were closer to the ideal line, particularly Product E (m = 0.94). Linear responses were moderate to strong (R^2 = 0.59–0.91), indicating that most products still tracked increasing nitrite concentrations, though less consistently than observed for pH or nitrate. Replicate variability was also higher than for the other analytes, likely because the nitrite assay produces subtle pink colour changes that are more difficult to distinguish consistently, especially at low concentrations.

Outdoor measurements (Figure 3B, 3D) produced clear differences in performance. Visual products showed improved proportional agreement, with slopes close to unity (m = 0.94–1.00) and strong linear responses

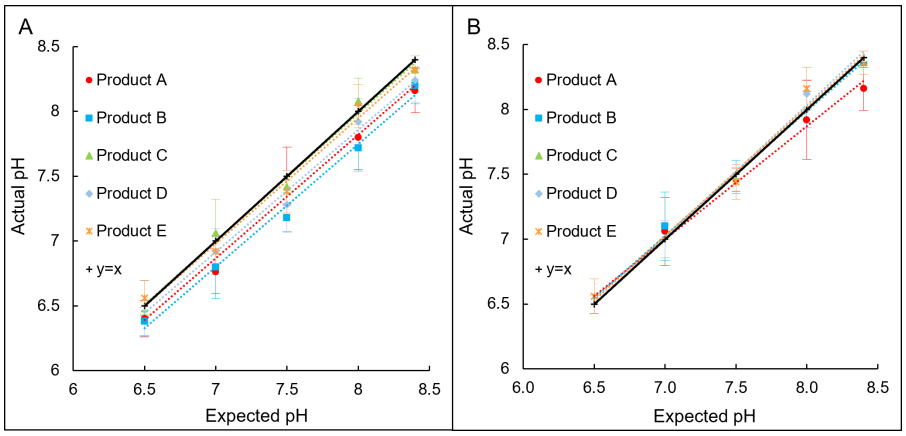

Figure 1: Comparison of test strip pH measurements with lab analysis under indoor (A) and outdoor (B) conditions. Each point represents the mean observed value, with error bars indicating one standard deviation. The solid black line represents a perfectly accurate test strip. Linear regression lines were fitted to each data set. Indoors: A: $y = 0.95x + 0.22$, $R^2 = 0.92$; B: $y = 0.95x + 0.17$, $R^2 = 0.94$; C: $y = 0.98x + 0.13$, $R^2 = 0.95$; D: $y = 0.94x + 0.33$, $R^2 = 0.93$; E: $y = 0.97 + 0.21$, $R^2 = 0.96$. Outdoors: A: $y = 0.87x + 0.89$, $R^2 = 0.91$; B: $y = 0.96x + 0.28$, $R^2 = 0.95$; C: $y = 0.99x$, $R^2 = 0.99$; D: $y = 1.03x - 0.18$, $R^2 = 0.99$; E: $y = 0.99x + 0.07$, $R^2 = 0.96$.

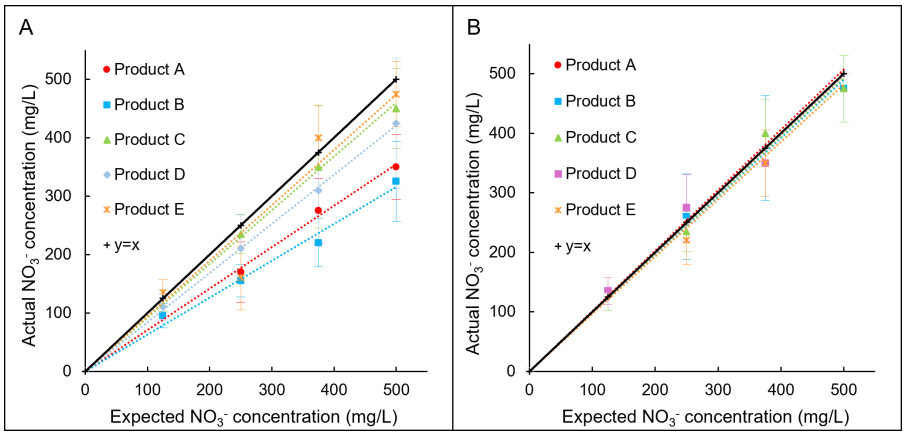

Figure 2: Comparison of test strip nitrate measurements with spectroscopic analysis across indoor (A) and outdoor (B) conditions. Each point represents the mean observed value, with error bars indicating one standard deviation. The solid black line represents a perfectly accurate test strip. Linear regression lines, constrained through the origin, were fitted to each data set. Indoors: A: $y = 0.71x$, $R^2 = 0.91$; B: $y = 0.63x$, $R^2 = 0.90$; C: $y = 0.92x$, $R^2 = 0.90$; D: $y = 0.84x$, $R^2 = 0.85$; E: $y = 0.95x$, $R^2 = 0.90$. Outdoors: A: $y = 1.02x$, $R^2 = 0.94$; B: $y = 0.98x$, $R^2 = 0.91$; C: $y = 0.99x$, $R^2 = 0.95$; D: $y = x$, $R^2 = 0.96$; E: $y = 0.96x$, $R^2 = 0.97$.

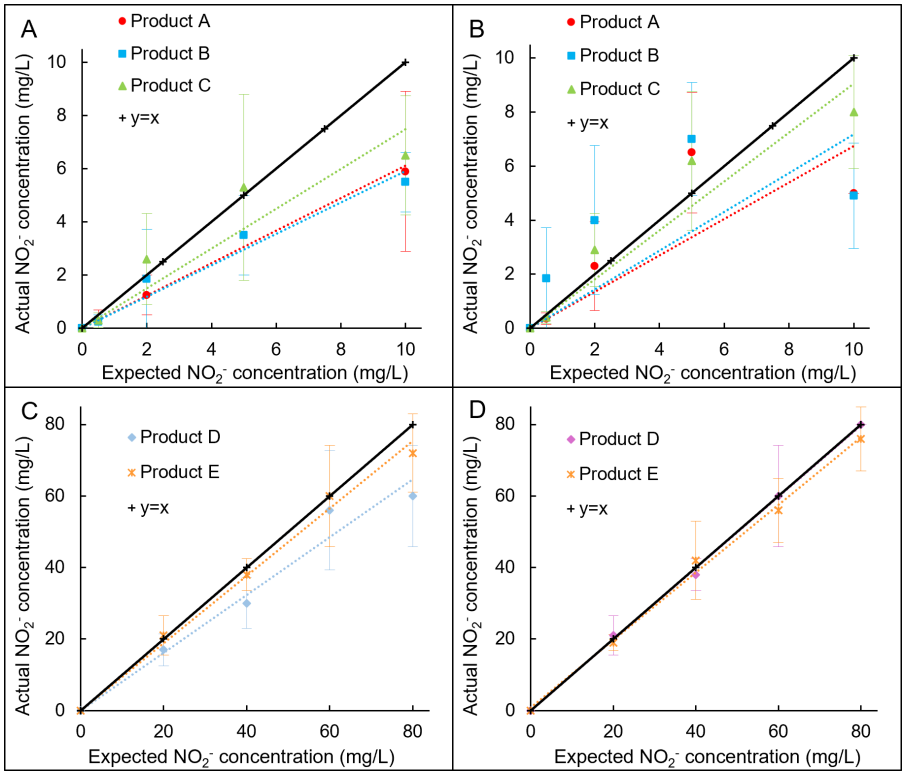

Figure 3: Comparison of test strip nitrite measurements with spectroscopic analysis across indoor (A, C) and outdoor (B, D) conditions. Each point represents the mean observed value, with error bars indicating one standard deviation. The solid black line represents a perfectly accurate test strip. Linear regression lines, constrained through the origin, were fitted to each data set. Indoors: A: $y = 0.61x$, $R^2 = 0.71$; B: $y = 0.59x$, $R^2 = 0.76$; C: $y = 0.75x$, $R^2 = 0.59$, D: $y = 0.81x$, $R^2 = 0.82$; E: $y = 0.94x$, $R^2 = 0.91$. Outdoors: A: $y = 0.68x$, $R^2 = 0.48$; B: $y = 0.72x$, $R^2 = 0.08$; C: $y = 0.94x$, $R^2 = 0.74$; D: $y = x$, $R^2 = 0.95$; E: $y = 0.96x$, $R^2 = 0.94$.

(R^2 = 0.74–0.95). In contrast, smartphone-based systems showed poor linearity (R^2 = 0.08–0.48) despite similar slopes (m = 0.68–0.72), indicating unreliable quantification under natural light. Replicate variability also increased outdoors for several products, particularly Product B. Unlike the pH and nitrate data sets, nitrite results suggest that digital colour analysis was strongly affected by illumination variability and glare, whereas visual interpretation remained comparatively robust. This finding highlights that smartphone-based colourimetry may not universally outperform visual interpretation, particularly for analytes that produce subtle or light-sensitive colour reactions.

Hardness analysis

The analysis of calcium carbonate (hardness) test strips (Figure 4) revealed the greatest divergence in performance among products. Under indoor conditions (Figure 4A, 4C), all products underestimated hardness concentrations, with slopes below unity (m = 0.62–0.87), indicating consistent proportional bias. Smartphone-based Product B and visual Product C showed the best proportional agreement (m = 0.87), whereas Product D exhibited poor linearity (R^2 = 0.15), indicating measurements did not consistently track increasing concentrations.

Linear responses for the remaining products were moderate (R^2 = 0.70–0.82), suggesting most systems followed the concentration trend despite systematic underestimation. Replicate variability was also greater than for pH and nitrate, particularly at higher concentrations, likely because hardness strips rely on broad colour transitions rather than discrete changes, making consistent interpretation more difficult.

Outdoor measurements (Figure 4B, 4D) produced a pronounced divergence in behaviour. Smartphone products and visual Product C showed improved proportional agreement (m = 0.92–0.95) with moderate to strong linear responses (R^2 = 0.72–0.87), indicating that natural light improved discrimination for these systems. In contrast, visual Products D and E exhibited substantial proportional overestimation (m = 2.08–2.40) and increased variability, suggesting that strong sunlight intensified the perceived response of the hardness indicator. Unlike the pH and nitrate data sets, where outdoor measurements generally improved agreement without major bias, hardness strips showed analyte-specific sensitivity to illumination, particularly for visually interpreted products.

Overall analysis of test strip accuracy

To evaluate the accuracy of the four test strip products, RMSE was used as a single error metric. RMSE was calculated as the deviation between test strip measurements and the corresponding reference measurements:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \text{Equation 1}$$

This metric captures both systematic bias and random variance among replicates. Larger deviations contribute disproportionately to RMSE, meaning occasional large errors are penalised more strongly than consistent moderate errors. RMSE values were normalised to the full measurement range of each analyte and expressed as percentages:

$$nRMSE = \left(\frac{RMSE}{y_{max} - y_{min}}\right) \times 100 \qquad \text{Equation 2}$$

This allowed for comparison both across analytes with different scales and within the same analyte when products use different measurement ranges.

An initial survey of the data (Figure 5) reveals a broad spectrum of accuracy across the experimental conditions. The nRMSE values ranged from 3.49% (Product C, pH, Outdoor) to 43.07% (Product B, Hardness, Indoor). Generally, the test strips exhibited greater variability under indoor lighting, where the majority of error values clustered between 10% and 25% (Figure 5A). In contrast, outdoor conditions resulted in a visible reduction in error for most parameters; notably, nitrate and pH errors in the outdoor data set dropped almost universally below 12%, regardless of the product used (Figure 5B).

The nRMSE values differed substantially among the four analytes (Figure 5). pH consistently showed the lowest nRMSE across all products and environments, remaining below 15% and reaching a minimum of 4% for Product C outdoors. In contrast, hardness showed the largest errors and greatest variability, with indoor nRMSE values ranging from 18% (Product C) to 43% (Product B), and similarly high values outdoors (up to 42%). Nitrate and nitrite displayed intermediate accuracy, with nitrate ranging from 7% to 24% and nitrite from 4% to 26% depending on product and environment. An ordinary one-way ANOVA confirmed that the differences among analytes were statistically significant (F(3,16) = 8.952, p = 0.0001) (Supplementary table 4). Post-hoc analysis with Tukey's test further showed that hardness performed significantly

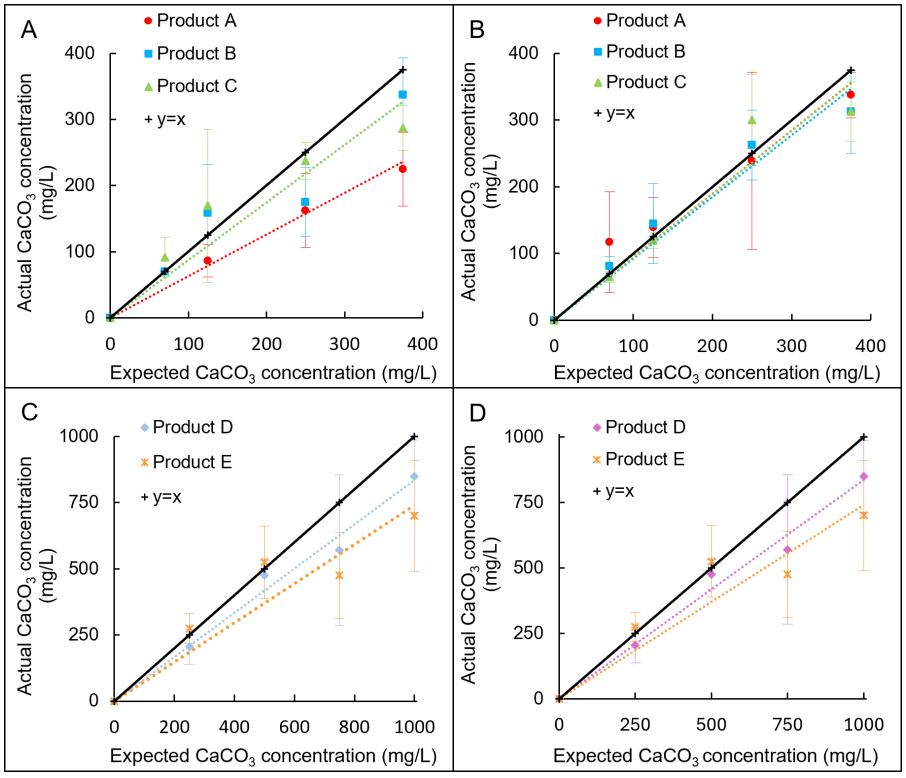

Figure 4: Comparison of test strip calcium carbonate measurements with expected values from spectroscopic analysis across indoor (A, C) and outdoor (B, D) conditions. Each point represents the mean observed value, with error bars indicating one standard deviation. The solid black line represents a perfectly accurate test strip. Linear regression lines, constrained through the origin, were fitted to each data set. Indoors: A: $y = 0.63x$, R^2 = 0.82; B: $y = 0.87x$, R^2 = 0.82; C: $y = 0.87x$, R^2 = 0.70, D: $y = 0.62x$, R^2 = 0.15; E: $y = 0.75x$, R^2 = 0.76. Outdoors: A: $y = 0.95x$, R^2 = 0.72; B: $y = 0.92x$, R^2 = 0.84; C: $y = 0.95x$, R^2 = 0.87; D: $y = 2.4x$, R^2 = 0.77; E: $y = 2.08x$, R^2 = 0.53.

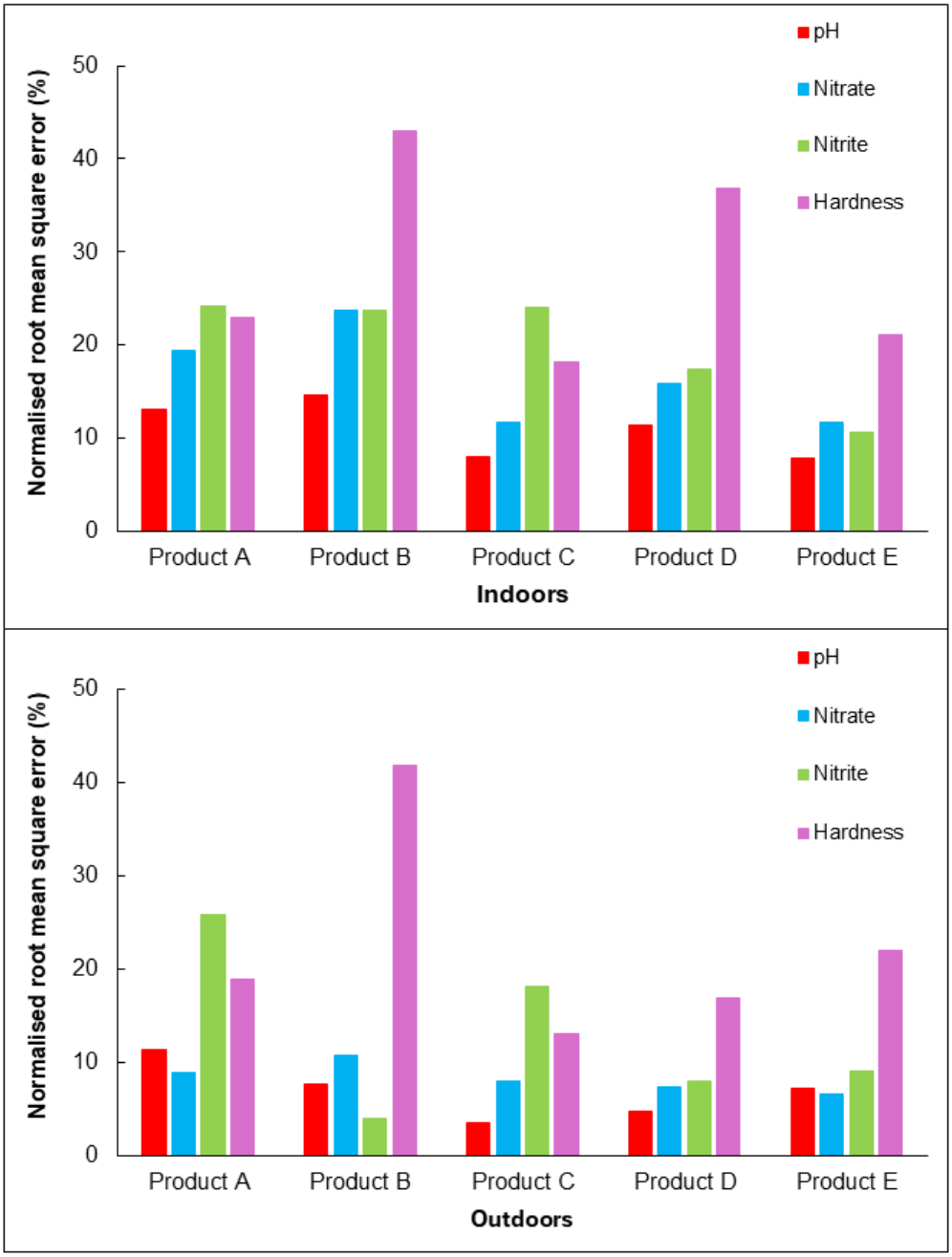

Figure 5: Normalised root mean square error values between test strip measurements and lab measurements under indoor (A) and outdoor (B) conditions. Root mean square error values were normalised to the concentration range specified for each product and parameter and expressed as a percentage. Indoor and outdoor measurements were conducted at approximately 1000 lux and 5000 lux, respectively.

worse than both pH ($p = 0.0001$) and nitrate ($p = 0.0024$), indicating that hardness measurements were the primary contributor to overall analytical error.

These differences likely reflect the underlying colourimetric chemistries of the test strips. pH strips rely on acid–base indicator dyes that produce clear and continuous transitions[19], enabling both visual interpretation and digital analysis to resolve small differences reliably. In contrast, hardness detection involves more complex reactions with metallochromic indicators to detect calcium and magnesium, producing broader colour changes that are more sensitive to matrix effects[20]. In some products (A and B), hardness is determined using multiple reaction pads, further increasing interpretation complexity. Consequently, hardness measurements show greater variability and may require confirmatory laboratory analysis. Previous findings are consistent with our results, showing that pH measurements are generally accurate and easily resolved, whereas nitrate and hardness are more difficult to quantify reliably.[11] pH strips have previously been shown to yield highly reproducible and consistent results when used by a single operator.[9]

Across both indoor and outdoor conditions, visually interpreted products (Products C, D and E) consistently outperformed smartphone-based products (Products A and B), showing lower and more stable measurement error profiles (Figure 5). Indoors, Product B showed the largest errors, ranging from 14% to 44% across analytes, with particularly pronounced deviations for hardness and nitrite, whereas Product E maintained moderate and consistent accuracy, with errors mostly between 8% and 21%. Outdoors, Product B again produced elevated errors, especially for hardness and pH (8–42%), while Product E remained relatively stable, with errors generally confined to 6–22% across all analytes. An unpaired *t*-test confirmed that the visually interpreted products were significantly more accurate

than the smartphone-based products ($t(38) = 2.147$, $p = 0.0382$) (Supplementary table 5). A statistical limitation is that the pooled nRMSE values were not strictly independent, because the same products and analytes recurred across multiple combinations. Furthermore, because the smartphone data sets used the same product while the visual data sets used three different products, the observed difference may reflect product chemistry, reading method or both. This result should therefore be interpreted as specific to the products and reading approaches tested. The findings are in line with previous work showing that, for nitrate, smartphone analysis was not necessarily more accurate than visual reading, even though that study used the Hach© nitrate test strip (Hach Company, Loveland, CO, USA).[21]

These results suggest that smartphone-based analyses may be limited by challenges such as white-balance calibration or edge detection on specific strip geometries, introducing digital noise that the human eye can naturally filter. This is consistent with previous studies indicating that image quality suffers from non-uniform and non-reproducible lighting conditions.[18,22] Various strategies have been developed to ensure more consistent image capture, such as using imaging boxes[18] and phone stands[23]. Products A and B, which use identical strip chemistry, were read on different smartphones (Huawei Nova 5T for A and Honor 200 for B), and the lack of a statistically significant difference suggests the strips may be consistent between devices. However, the study was limited to a single operator, only two devices and one type of smartphone-based strip. Future work should include multiple users and additional smartphone-based products to better evaluate inter-device, inter-person and inter-product variability, clarifying whether precision deficits arise from software, hardware or operator factors.

Across all products, outdoor measurements consistently improved test strip accuracy relative to indoor conditions, with reductions in nRMSE observed for most analytes (Figure 5). Visually interpreted Products (C, D and E) remained highly accurate, with the greatest improvements seen in hardness. For example, the hardness error for Product D decreased from 37% indoors to 17% outdoors, while Product E exhibited a more modest reduction from 21% to 18%. Among smartphone-based products, Product A displayed the largest improvement, with the nitrate error decreasing from 20% indoors to 9% outdoors, whereas Product B showed minimal change, with hardness remaining largely unchanged. Across all products and analytes, outdoor lighting reduced errors by an average of 12%. An unpaired *t*-test confirmed that outdoor readings were significantly more accurate than indoor measurements ($t(38) = 2.181$, $p = 0.0354$) (Supplementary table 6). This interpretation is subject to the same statistical limitations regarding the pooled nRMSE values.

These results show that even shaded outdoor light on a sunny day enhances the contrast and visibility of the colourimetric bands, improving the accuracy of both visual and smartphone-based readings. Consistent with previous findings, smartphone analyses remain more sensitive to factors such as white-balance calibration, edge detection and digital noise[24–26], which the human eye can naturally filter. Future work should explore the development of algorithms that better compensate for lighting variability, test a wider range of smartphone-based products and include multiple operators and devices to assess inter-person and inter-device variability. Evaluating additional lighting conditions, including partial sun and artificial illumination, would also clarify how environmental factors influence accuracy and help optimise both manual and digital strip-reading protocols.

Two notable analytical trends emerged from the data set. First, analyte concentrations were systematically underestimated, particularly under indoor conditions, where gradients for nitrite ($m = 0.59–0.81$) and hardness ($m = 0.62–0.87$) remained consistently below 1. Previous studies have shown that test strips systematically underestimated hardness[11] and nitrate[27], respectively. Underestimation may allow harmful substances to exceed regulatory thresholds, so users should always confirm measurements when values are near regulatory limits. Second, cost was not a reliable predictor of accuracy. The most expensive product (Product A/B) had moderate to high nRMSE values across several analytes, whereas lower-cost alternatives (including Products C and D) produced lower nRMSE values and more accurate readings. Mid-priced products (such as Product E) showed variable performance across analytes. The relationship between product cost and analytical accuracy of water quality test strips remains largely unexplored in the literature.

The results highlight that, while test strips show variability in accuracy across analytes, products and lighting conditions, they remain valuable tools for citizen science if appropriate protocols are followed. Given the systematic underestimation observed for some analytes (such as nitrite and hardness) and the influence of environmental lighting on accuracy, standardised operating procedures are essential to ensure data comparability.[28] Citizen science projects should therefore implement fixed reading times as specified by the manufacturer; standardised lighting conditions where possible[11]; and clear training materials to guide correct sampling, timing and interpretation. To account for the variability observed between individual readings, at least three replicate measurements per sample should be performed to improve precision.[29] Additionally, the performance of any strip used in monitoring programmes should be quantified through parallel laboratory validation using calibrated analytical techniques, with blind testing employed to confirm reliability.[28] Where possible, project organisers should also quantitatively document the accuracy of each product–analyte combination so that the limitations of the data are understood. These safeguards are particularly important if citizen-generated data sets are intended to inform environmental management, support Sustainable Development Goal reporting or contribute to holding polluters and regulators accountable. For affordable monitoring technologies to influence policy and regulatory decisions, the data they generate must be sufficiently accurate, transparent and reproducible.

Limitations and future work

This study has several limitations that also highlight directions for future research. Visual interpretations were performed by a single operator, and smartphone-based analyses used only one application and operating system (Android). In addition, only one smartphone-based test strip product (JBL ProScan) was evaluated, and the study was limited to four products available in South Africa. These constraints may not fully capture variability associated with different users, smartphone hardware, operating systems or a broader range of commercial strips. Future work should therefore include multiple operators, additional devices (including iOS platforms) and a wider selection of products. Further research is also needed to develop reliable rapid assays for microbiological indicators such as *Escherichia coli*, which remain largely absent from current strip-based water quality monitoring tools.

Conclusion

This study evaluated four water quality test strip products in South Africa under indoor and outdoor conditions to assess their suitability for low-cost environmental monitoring. Accuracy varied by analyte, product and lighting: pH and nitrate were generally reliable, particularly in outdoor shaded conditions, while hardness showed the greatest error. Outdoor readings were more accurate than indoor readings on average, although hardness remained a clear exception, and smartphone analysis did not consistently outperform visual interpretation. Regression gradients indicated that test strips frequently underestimated concentrations, potentially allowing harmful levels to go undetected. Product price did not predict accuracy; lower-cost visual strips sometimes matched or outperformed more expensive smartphone-based systems.

Despite these limitations, test strips are useful for rapid screening, citizen science and trend detection when standardised protocols are applied, including fixed reaction times, consistent lighting, user training and replicate measurements. In parallel, independent laboratory validation of specific product–analyte combinations is recommended to quantify accuracy. While current strips are not precise enough for regulatory enforcement, SDG reporting or holding polluters accountable, implementing these measures is a crucial first step toward making low-cost monitoring technologies robust and reliable for such formal applications in the future.

Funding

This research was funded by the National Research Foundation of South Africa (MND210620613627), Ninety One, The Claude Leon Foundation and the University of the Witwatersrand (Wits Water).

Data availability

The data supporting the results of this study are available upon request to the corresponding author.

Declarations

We have no competing interests to declare. To support clarity and accessibility, we used ChatGPT (OpenAI, GPT-5.5 Thinking) to provide editorial suggestions aimed at improving the readability and coherence of the text. The prompt used was: "Provide editorial suggestions aimed at improving the readability and coherence of the text." Preliminary findings from this study were presented as a conference abstract (Mayet R, Sheridan CM, Dahms-Verster S. How accurate are water quality test strips? 9th Annual IIES Science & Policy Workshop, South Africa: Improving Water Security, 2025). This article presents an expanded and more detailed analysis of the research presented.

Authors' contributions

R.M.: Conceptualisation, data curation, formal analysis, investigation, methodology, project administration, visualisation, writing – original draft. C.M.S.: Conceptualisation, funding acquisition, resources, supervision, visualisation, writing – review and editing. S.D-V.: Conceptualisation, resources, supervision, visualisation, writing – review and editing. All authors read and approved the final manuscript.

References

1. Chapman DV, Sullivan T. The role of water quality monitoring in the sustainable use of ambient waters. One Earth. 2022;5(2):132–137. https://doi.org/10.1016/j.oneear.2022.01.008
2. South African Bureau of Standards (SABS). SANS 241:2025 — Drinking water. 7th ed. Pretoria: SABS Standards Division; 2025.
3. World Health Organization. Guidelines for drinking-water quality: Fourth edition incorporating the first addendum [document on the Internet]. c2017 [cited 2024 Sep 23]. Available from: https://www.who.int/publications/i/item/9789241549950
4. Khatri N, Tyagi S. Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. Front Life Sci. 2014;8(1):23–39. https://doi.org/10.1080/21553769.2014.933716
5. McDowell RW, Noble A, Kittridge M, Ausseil O, Doscher C, Hamilton DP. Monitoring to detect changes in water quality to meet policy objectives. Sci Rep. 2024;14, Art. #1914. https://doi.org/10.1038/s41598-024-52512-7
6. Sustainable Development Solutions Network. Indicators and a monitoring framework for the Sustainable Development Goals [document on the Internet]. c2015 [cited 2024 Jun 11]. Available from: https://sdgs.un.org/publications/indicators-and-monitoring-framework-sustainable-development-goals-17958
7. Ting JON, Yee SK. Review on water quality monitoring technologies. Indones J Electr Eng Comput Sci. 2020;18(3):1416–1423. https://doi.org/10.11591/ijeecs.v18.i3.pp1416-1423
8. San Llorente Capdevila A, Kokimova A, Sinha Ray S, Avellán T, Kim J, Kirschke S. Success factors for citizen science projects in water quality monitoring. Sci Total Environ. 2020;728, Art. #137843. https://doi.org/10.1016/j.scitotenv.2020.137843
9. Swinney GG, Mullins SA, Martin J, Kinsey JB, King DA. Assessment of the precision and accuracy of at-home multi-test water quality test strips. J Undergrad Chem Res. 2024;23(4), Art. #54.
10. Hillscher LM, Liebich VJ, Avrutina O, Biesalski M, Kolmar H. Functional paper-based materials for diagnostics. ChemTexts. 2021;7(2), Art. #14. https://doi.org/10.1007/s40828-021-00139-w
11. McCarty DA, Kim MA, Jo H, Yim E, Yun H, Sims S, et al. Accessible water quality monitoring through hybrid human–machine colorimetric methods. Environ Monit Assess. 2025;197, Art. #555. https://doi.org/10.1007/s10661-025-13983-x
12. McCracken KE, Yoon J-Y. Recent approaches for optical smartphone sensing in resource-limited settings: A brief review. Anal Methods. 2016;8(36):6591–6601. https://doi.org/10.1039/C6AY01575A
13. JBL PROSCAN. Water test with analysis via smartphone for freshwater aquariums, ponds, and tap/well water [webpage on the Internet]. 2014 [cited 2023 Jan 16]. Available from: https://www.jbl.de/en/products/detail/6774/jbl-proscan?country=us
14. D'Andrea L, Kumpel E. Evaluation of drinking water quality test kits for home use in the United States. J Water Health. 2025;23(3):350–363. https://doi.org/10.2166/wh.2025.289
15. Naigaga S, Boyd CE, Gaillard P, Abdelrahman HS, Molnar JJ. Assessing the reliability of water-test kits for use in pond aquaculture. J World Aquac Soc. 2017;48(4):555–562. https://doi.org/10.1111/jwas.12377
16. Doğan V, Isık T, Kılıç V, Horzum N. A field-deployable water quality monitoring with machine learning-based smartphone colorimetry. Anal Methods. 2022;14:3458–3466. https://doi.org/10.1039/d2ay00785a
17. Yazawa T, Hasid AS, Carré – Poussin VM. Validating smartphone-based water quality monitoring for rapid environmental assessment. Water Pract Technol. 2025;20(3):653–660. https://doi.org/10.2166/wpt.2025.031
18. Chen G, Fang C, Chai HH, Zhou Y, Li WY, Yu L. Improved analytical performance of smartphone-based colorimetric analysis by using a power-free imaging box. Sens Actuators B Chem. 2019;281:253–261. https://doi.org/10.1016/j.snb.2018.09.019
19. Alberti G, Zanoni C, Magnaghi LR, Biesuz R. Disposable and low-cost colorimetric sensors for environmental analysis. Int J Environ Res Public Health. 2020;17(22), Art. #8331. https://doi.org/10.3390/ijerph17228331
20. Phuangsaijai N, Jakmunee J, Kittiwachana S. Investigation into the predictive performance of colorimetric sensor strips using RGB, CMYK, HSV, and CIELAB coupled with various data preprocessing methods: A case study on an analysis of water quality parameters. J Anal Sci Technol. 2021;12, Art. #19. https://doi.org/10.1186/s40543-021-00271-9
21. Topping M, Kolok A. Assessing the accuracy of nitrate concentration data for water quality monitoring using visual and cell phone quantification methods. Citiz Sci Theory Pract. 2021;6(1), Art. #5. https://doi.org/10.5334/cstp.346
22. Mirhosseini S, Nasiri AF, Khatami F, Mirzaei A, Aghamir SMK, Kolahdouz M. A digital image colorimetry system based on smart devices for immediate and simultaneous determination of enzyme-linked immunosorbent assays. Sci Rep. 2024;14, Art. #2587. https://doi.org/10.1038/s41598-024-52931-6
23. Cho S, Park T, Reynolds K, Yoon J-Y. Multi-normalization and interpolation protocol to improve norovirus immunoagglutination assay from paper microfluidics with smartphone detection. SLAS Technol. 2017;22(6):609–615. https://doi.org/10.1177/2472630317724769
24. Meng R, Yu Z, Fu Q, Fan Y, Fu L, Ding Z, et al. Smartphone-based colorimetric detection platform using color correction algorithms to reduce external interference. Spectrochim Acta A Mol Biomol Spectrosc. 2024;316, Art. #124350. https://doi.org/10.1016/j.saa.2024.124350
25. Zhang G, Song S, Panescu J, Shapiro N, Dannemiller KC, Qin R. A novel systems solution for accurate colorimetric measurement through smartphone-based augmented reality. PLoS One. 2023;18, e0287099. https://doi.org/10.1371/journal.pone.0287099
26. Geng Z, Miao Y, Zhang G, Liang X. Colorimetric biosensor based on smartphone: State-of-art. Sens Actuators A Phys. 2023;349, Art. #114056. https://doi.org/10.1016/j.sna.2022.114056
27. Searles Nielsen S, Mueller BA, Kuehn CM. An evaluation of semi-quantitative test strips for the measurement of nitrate in drinking water in epidemiologic studies. J Expo Sci Environ Epidemiol. 2008;18:142–148. https://doi.org/10.1038/sj.jes.7500563
28. Ramírez SB, van Meerveld I, Seibert J. Citizen science approaches for water quality measurements. Sci Total Environ. 2023;897, Art. #165436. https://doi.org/10.1016/j.scitotenv.2023.165436
29. D'Alessio M, Rushing G, Gray TL. Monitoring water quality through citizen science while teaching STEM undergraduate courses during a global pandemic. Sci Total Environ. 2021;779, Art. #146547. https://doi.org/10.1016/j.scitotenv.2021.146547