

# Validity of Urine Color Scoring Using Different Light Conditions and Scoring Techniques to Assess Urine Concentration

Floris Wardenaar, PhD; Scott Armistead, MS; Kayla Boeckman, BS; Brooke Butterick, BS; Darya Youssefi, BS; Daniel Thompsett, MS; Kaila Vento, MS

College of Health Solutions, Arizona State University, Phoenix

**Context:** Urine color (Uc) is used to assess urine concentration when laboratory techniques are not feasible.

**Objective:** To compare the accuracy of Uc scoring using 4 light conditions and 2 scoring techniques with a 7-color Uc chart. Additionally, to assess the results' generalizability, a subsample was compared with scores obtained from fresh samples.

**Design:** Descriptive laboratory study.

**Samples:** A total of 178 previously frozen urine samples were scored, and 78 samples were compared with their own fresh outcomes.

**Main Outcome Measure(s):** Urine color and accuracy for classifying urine samples were calculated using receiver operating characteristics analysis, allowing us to compare the diagnostic capacity against a 1.020 urine specific gravity cutoff and defining optimal Uc cutoff value.

**Results:** Urine color was different among light conditions ( $P < .01$ ), with the highest accuracy (80.3%) of correct classifica-

tions of low or high urine concentrations occurring at the brightest light condition. Lower light intensity scored 1.5 to 2 shades darker on the 7-color Uc scale than bright conditions ( $P < .001$ ), but no further practical differences in accuracy occurred between scoring techniques. Frozen was 0.5 to 1 shade darker than freshly measured Uc ( $P < .004$ ), but the values were moderately correlated ( $r = 0.64$ ). A Bland-Altman plot showed that reporting bias mainly affected darker Uc without affecting the diagnostic ability of the method.

**Conclusions:** Urine color scoring, accuracy, and Uc cutoff values were affected by lighting condition but not by scoring technique, with greater accuracy and a 1-shade-lower Uc cutoff value at the brightest light (ie, light-emitting diode flashlight).

**Key Words:** hydration status, hydration education, fluid intake monitoring, assessment of hydration status

## Key Points

- Urine color scoring and accuracy were affected by light condition. Assessing urine samples in bright light conditions with intensity  $>1600$  lux resulted in the greatest accuracy.
- In addition to light, other factors, such as container material and volume, should be factored in when defining the optimal urine color cutoff value.
- Light condition will likely influence urine color scoring in many situations, even if athletes assess their urine straight from the toilet or urinal.

High urine concentration has been suggested as a biomarker for detecting underhydration.<sup>1</sup> This can occur without the perception of thirst or a change in plasma osmolality concentration.<sup>1</sup> Evidence is growing for the long-term health benefits from fluid intake that results in a 24-hour urine concentration of  $\leq 500$  mOsm/kg<sup>2</sup> and a urine specific gravity (USG) value of  $\leq 1.012$ .<sup>3</sup> As spot morning urine samples generally have a greater concentration than a full 24-hour urine sample,<sup>4</sup> the cutoff value for well-hydrated status potentially lies somewhat higher. Therefore, concentrations of 700 mOsm/kg<sup>5</sup> or 1.020 USG<sup>4</sup> have often been reported. The most accurate way to assess urine concentration involves laboratory-based techniques. Still, urine color (Uc) has been suggested as an appropriate proxy measurement in an applied setting for assessing the hydration status of athletes.<sup>6</sup> Measuring Uc has several advantages: the method is inexpensive and

noninvasive, does not require technical expertise, and gives immediate results.<sup>7</sup> Despite the fact that urine osmolality is often proposed as the standard for assessing urine concentration,<sup>8</sup> USG displayed similar correlations with Uc ( $r \cong 0.80$ )<sup>9,10</sup> and 5% better accuracy for classifying samples with a high versus low urine concentration compared with urine osmolality.

Urine color charts have been valuable in hydration assessment and education in many settings, including clinical, athletic, and household settings.<sup>11,12</sup> However, the main disadvantage of using the current common Uc charts is their lack of sensitivity ( $\sim 80\%$  sensitivity accuracy).<sup>13</sup> Certain variables, such as vitamins and other bioactive substances, as well as protein, can influence Uc<sup>14</sup> and concentration.<sup>15</sup> On the other hand, the accuracy of correctly predicting urine concentration based on Uc can be improved substantially when using spectrophotometry,

resulting in an analysis with 97.4% sensitivity,<sup>16</sup> showing that assessing light-based urine concentration with high accuracy is possible.

Authors<sup>3,9,13,17</sup> have reported a wide range of average Ucs for different populations, from 3 to 6. Apart from the diameter (volume) and materials used in the urine container, the type of light and light intensity will influence perceived Uc.<sup>10</sup> The Beer-Lambert law states that light absorbance is equal to the concentration of the solution the light is passing through, the solution's volume, and the absorption coefficient. Despite researchers' reports<sup>3,10,18</sup> that urine samples were scored in a well-lit room, not much information is available about the actual light conditions present and how the urine sample was positioned against the light for its color to be scored.

As previous investigators have not clearly defined the light conditions on which the suggested Uc cutoff values have been based, health professionals, such as athletic trainers, coaches, and sport dietitians, would benefit from more information reporting the effect of different light conditions on Uc scoring. Such insight will help further standardize the Uc scoring process, potentially leading to better athlete self-classification of urine samples with low or high urine concentration. Therefore, the objective of our study was to compare the effect of different light conditions on Uc scoring. The aims were as follows: (a) to report Uc scores for 4 light conditions using a 7-color Uc chart and 2 scoring techniques, producing 8 Uc scoring combinations; (b) to evaluate the diagnostic ability and the optimal Uc cutoff value for each of the 8 scoring combinations to assess Uc; and (c) as the samples used were previously frozen, to perform a sensitivity analysis comparing the results of a subsample (n = 78) before and after freezing to determine if the results from aims (a) and (b) can be generalized to freshly collected and measured Uc.

## METHODS

### Design

To evaluate Uc scoring differences among 4 light conditions, we used a 7-color Uc chart and 2 Uc scoring techniques. Three research technicians individually scored each sample. Then the calculated median Uc scores were compared with the measured USG of initially fresh samples to identify the accuracy of correctly classifying urine samples with a low versus high urine concentration for each condition. Urine samples were classified as having a low urine concentration with a USG cutoff value <1.020 or a high urine concentration equal to or above this value.<sup>4</sup> To ensure generalizability of the data, as frozen samples tend to display a slightly darker color, a multilevel comparison was performed in a subset of the samples measured, with Uc and USG assessed before and after freezing. All 3 research technicians were involved in previous studies of Uc scoring and were experienced in comparing color of urine with multiple Uc charts while using both of the scoring techniques described in this Methods section.

### Sample Number

The 178 urine samples were initially collected between 2018 and 2019 during 3 studies approved by the Institutional Review Board at Arizona State University: STUDY00010071,

STUDY00008336, and STUDY00007260. Urine samples were anonymized before the project was started. Apart from the original measured USG value for each sample as described earlier by our laboratory team,<sup>19</sup> no other data were carried over to this study. Samples were collected from a healthy athletic population as a spot urine sample and stored in 30 mL tubes at -20°C for 1 to 18 months.

## Procedures

During each scoring session, the frozen samples were covered with clear film and thawed until they reached room temperature (20°C) for scoring. Four scoring stations were created, 1 per light condition. At each station, the technicians used 2 techniques to score the samples. After a batch of approximately 30 samples was scored, the samples rotated while the technicians remained at their station using the same scoring method and light type before moving over to the next light condition. Figure 1 illustrates all scoring combinations.

**The 7-Color Uc Chart.** The 7-color Uc chart developed by Wardenaar and Bacalzo in 2019<sup>20</sup> was based on a selection of 96 previously collected urine samples.<sup>21</sup> Color prints as described earlier were made on 43.2- × 27.9-cm (17- × 11-in) white paper with a glossy finish.<sup>20</sup>

**Scoring Techniques.** Two scoring techniques were used: the traditional sample-over-chart method, which required the technician to hold the sample while comparing it with the color chart, and the Uc scoring box method that was developed to standardize the distance between the eye and the urine sample, which could potentially affect scoring accuracy. Urine samples were prepared in 30-mL transparent plastic centrifuge tubes (freestanding Evergreen model; Caplugs). Each urine sample was covered using clear Parafilm (Bemis Company, Inc) to seal the sample and prevent color distortion, allowing the scoring of samples without the original green tube cap. Each urine sample was inverted 3 times before being scored.

**The Sample-Over-Chart Method.** During this method, the technicians slid each separate urine sample over the white part of the Uc chart and compared the sample with each color on the 7-color Uc chart.<sup>9</sup> A decision was then made about the color of the sample, and the score was noted.

**The Uc Box Scoring Method.** The second method involved a color scoring box constructed in our laboratory.<sup>21</sup> This box was created to standardize the distance from the observer to the sample and the way the urine sample was positioned in relation to the light for each light condition (36 cm [14 in]) as previously reported.<sup>20</sup> The technicians looked through the box while comparing the color they perceived with the colors on the 7-color Uc chart positioned directly aside the box. A decision was made about the color of the sample, and the score was noted.

**Light Conditions.** Light conditions were selected based on practical relevance. The light intensity was different for each light condition but similar for scoring techniques within each light condition. Each light condition intensity was measured using a foot-candle lux meter (model 407026; Exttech Instruments) at the tungsten/daylight setting. The light-emitting diode (LED) light conditions represent a combination of lights (as the background lights were not turned off).

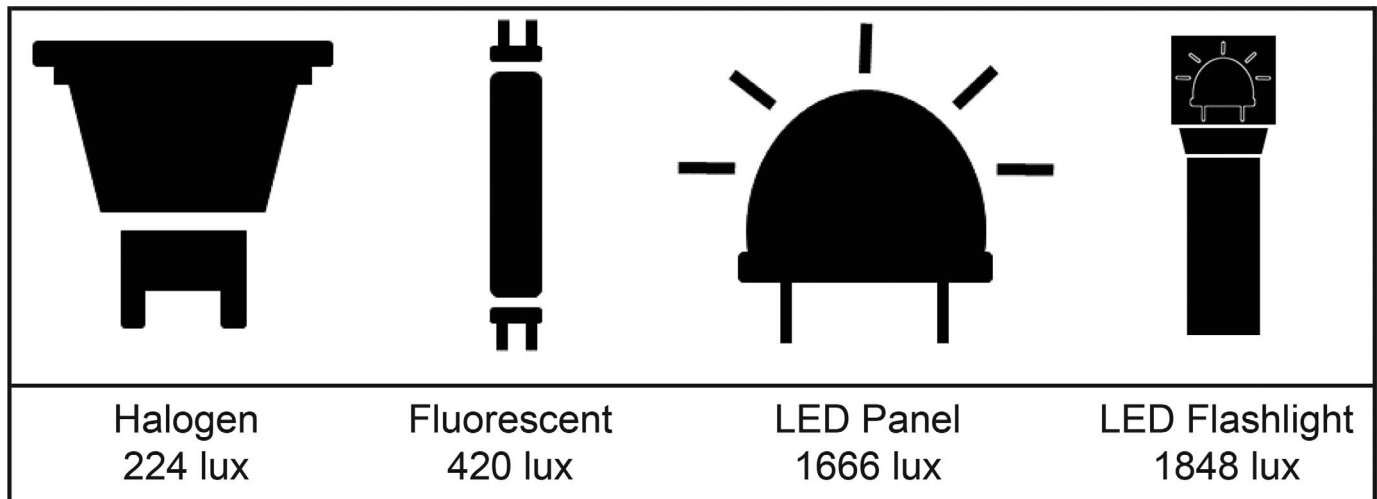


Figure 1. Urine color (Uc) scored under 4 light conditions with and without a Uc scoring box. A total of 178 samples were scored by 3 technicians. Samples were scored by sliding the sample or being placed in a Uc scoring box. Abbreviation: LED, light-emitting diode.

**Halogen Light.** The restroom halogen light was part of the testing facility. The lights were built into the ceiling 180-cm (6-ft) apart in a square, at the corners of all ends of the 2 tables that created the testing station. The light intensity for this condition was 224 lux, measured at the center of the testing station, with the surface of the tables 180 cm from the ceiling.

**Fluorescent Light.** The laboratory space had fluorescent office light. The two 3-light fluorescent parabolic troffers were built into the ceiling 180 cm apart, matching the long end of the testing stations, ensuring that the measurements were obtained directly under the light source. The light intensity for this condition was 402 lux, measured at the center of the testing station, with the surface of the tables 180 cm from the ceiling.

**The LED Panel.** For the sample-over-chart method, the 28-W LED panel light set to full white (model NL480; Neewer), providing 1666 lux, was placed at 12 cm (7.5 in) on the left side of the sample. For Uc box scoring, the light was placed on the left side of the 30-mL urine sample and the scorer's perspective. The LED panel analysis was conducted in the restroom facility with the halogen ceiling lights switched on.

**The LED Flashlight.** The flashlight contained 6 LEDs (Ozark Trail), providing 1848 lux when covered with a single layer of white masking tape to create a filter. The light was projected directly from underneath the 30-mL centrifuge tube for both scoring methods. During the sample-over-chart method, the technicians wore blue laboratory gloves and held the urine sample directly on the flashlight. For the color box scoring, the flashlight was built into the box to light the sample from underneath when it was placed in the center of the box on top of the flashlight. The LED flashlight analysis was performed in the office space with the fluorescent ceiling lights switched on.

### Data Analysis

The urine concentration from the freshly measured samples was reported as the median and interquartile range (IQR). To address aim (a), to evaluate the scored Uc using

different lights and techniques, we reported the median, IQR, mean, and SD for each scoring condition. Differences were assessed using the Friedman and Wilcoxon signed rank tests.

To address aim (b), investigating the diagnostic ability under the 8 conditions to distinguish between low and high urine concentrations based on the correct classification of Uc, we calculated receiver operating characteristic (ROC) curves. The Uc scores were optimally fitted against the USG values that were initially measured in fresh urine samples. The best Uc cutoff value for distinguishing urine samples with a low versus high urine concentration (<1.020 USG cutoff value) was determined from the area under the curve (AUC) using the max approach for sensitivity and specificity.

To address aim (c), assessing the generalizability of the study outcomes based on frozen samples before analysis, we compared data from a subsample ( $n = 78$ ) measured before and after freezing. The original USG value of the freshly measured samples was reported and correlated against the frozen samples using the Spearman correlation, including 95% CIs, using the Fisher Z transformation. The mean difference between the frozen and fresh samples was analyzed using the Wilcoxon signed rank test and Spearman correlation. Further, a Bland-Altman plot was produced to assess the agreement between scores for individual samples, comparing the Uc scores from 1 to 7 to evaluate the outcomes of freshly scored urine samples against frozen samples. To assess whether the Bland-Altman results were biased for scoring Uc lighter, similar to, or darker between fresh and frozen urine samples, we calculated an additional Spearman correlation coefficient. This was done by correlating the Bland-Altman results from the y axis (the difference between the reported Uc outcomes) against the results of the x axis (the means of both outcomes); reporting bias was present when the correlation was significant. The level of correlation provided more information about the direction of this bias. Finally, the agreement level was calculated, defined as  $M_{(\text{difference})} \pm 1.96 SD_{(\text{difference})}$ . Statistical significance was set for all analyses at  $P \leq .05$ .

## RESULTS

The original median USG value measured in the 178 fresh samples was 1.018 (IQR = 1.012–1.027), with 42% of the samples at or above the suggested USG cutoff value of 1.020. The outcome of aim (a), scoring Uc using different lights and techniques, is reported in Table 1. The average of the combined median values per scoring technique resulted in a difference for Uc among all light types ( $P < .001$ ). The conditions with lower light intensity, such as halogen and fluorescent, scored 1.0 to 1.5 shades darker than the LED light conditions. The average Uc for fluorescent light was 3.5 (2.0–6.0); for halogen light, 3.0 (2.0–6.0); the LED panel, 2.0 (1.0–5.5); and the LED flashlight, 2.0 (1–4). Despite the higher light intensity, Uc was darkest under fluorescent light, before halogen and the brighter LED conditions. Scoring technique, ie, the sample over chart versus in scoring box, influenced the reported Uc under each light type ( $P \leq .01$ ), with the exception of halogen light ( $P = .91$ ).

The AUC, calculated to investigate the diagnostic ability of the 8 measured conditions as formulated for aim (b), slightly increased with brighter light conditions. The lowest AUC for the sample-over-chart and Uc scoring-box techniques, respectively, was 0.82 and 0.84 for fluorescent light, followed by halogen light (0.84 and 0.86), LED light panel (0.87 and 0.84), and LED flashlight (0.86 and 0.87; Table 1).

The accuracy of correctly classified urine samples for low or high urine concentration increased incrementally, similar to the reported AUC. Starting with fluorescent light (76% for both scoring techniques), followed by halogen light (77% for the sample-over-chart and 76% for the scoring box), LED light panel (79% for the sample-over-chart and 78% for the scoring box), and ending with the highest values for the LED flashlight (79% for the sample-over-chart and 80% for the scoring box). The rate of false-positive (FP) values (as a result of a low Uc when concentration was actually above the selected USG cutoff) was inversely associated with light intensity: the highest percentage was for halogen light with the lowest light intensity (20% and 17% for the sample-over-chart and box-scoring techniques, respectively) versus LED flashlight with the highest light intensity (12% and 15% for the sample-over-chart and the Uc box, scoring techniques, respectively). The ROC-based Uc cutoff value for fluorescent light was 1 shade darker ( $Uc \leq 4$ ) than the suggested best fit for Uc cutoff for all other light types ( $Uc \leq 3$ ) for classifying low versus high urine concentration.

### Difference Between Fresh and Frozen Urine Samples

Aim (c) was to investigate the generalizability of the outcomes between fresh and frozen urine samples. The median USG for the 78 freshly measured samples was 1.018 (IQR = 1.012–1.023). The median Uc score was 2, with a larger IQR in the frozen condition (+2 Uc shades) than in the freshly measured samples ( $P = .004$ ), indicating a larger number of darker-scored urine samples in the frozen condition.

Correlations between Uc and USG were somewhat stronger for frozen samples ( $r = 0.74$ ; 95% CI = 0.61, 0.83;  $P < .001$ ) than fresh samples ( $r = 0.59$ ; 95% CI = 0.42, 0.85;  $P < .001$ ) as shown in Table 2, but the samples

**Table 1. Urine Color, Area Under the Curve, Accuracy, and Cutoff Values for Each of the 4 Light Conditions (N = 178)**

Light Color-scoring Box?	Urine Color				Receiver Operating Curve Analysis							
	Median	Interquartile Range	Mean±SD	Area Under the Curve	Sensitivity, %	Specificity, %	True Positives <sup>a</sup>	True Negatives	False-Positives	False-Negatives	Accuracy, %	Cutoff Value
Fluorescent light												
No	4	2–6	3.7 ± 2.1	0.82	72.1	56.4	60	75	35	8	76	≤4
Yes	3	2–6	3.6 ± 2.0	0.84	83.8	55.6	57	79	31	11	76	≤4
Halogen light												
No	3	2–6	3.6 ± 2.2	0.84	86.8	57.7	59	78	32	9	77	≤3
Yes	3	2–6	3.6 ± 2.1	0.86	91.2	57.5	62	73	37	6	76	≤3
LED panel												
No	2	1–6	3.5 ± 2.4	0.87	85.3	59.8	58	82	28	10	79	≤3
Yes	2	1–5	2.9 ± 2.0	0.84	72.1	53.0	49	89	21	19	78	≤3
LED flashlight												
No	2	1–4	2.7 ± 1.6	0.86	85.3	63.1	53	88	22	15	79	≤3
Yes	2	1–4	2.6 ± 1.6	0.87	72.1	57.9	59	84	26	9	80	≤3

Abbreviation: LED, light-emitting diode.

<sup>a</sup> True positive, true negative, false-positive, and false-negative values are provided as numbers; the combined outcome equals  $n = 178$  for each line. A panel of 3 research associates scored all urine samples; per sample, the median score was used for further calculation. Each sample was randomly scored 8 times using 2 techniques, ie, with the sample held directly over the chart or positioned in a color-scoring box for comparison with the 7-color chart for each of the 4 light conditions (fluorescent light [office light], halogen light [restroom light], LED flashlight, and LED panel light, such as is common in US restrooms).

**Table 2. Urine Color Scores of Fresh and Frozen Urine Samples (n = 73) Using a 7-Color Chart and Color-Scoring Box<sup>a</sup>**

Sample	Urine Concentration Markers				Receiver Operator Characteristics								
	Urine Color, Median (IQR)	Minimum-maximum	Specific Gravity, Median (IQR)	r (95% CI)	Area Under the Curve	Sensitivity, %	Specificity, %	True Positives <sup>a</sup>	True Negatives	False-Positives	False-Negatives	Accuracy, %	Cutoff Value
Fresh	2 (1–2)	1–6	1.018 (1.012–1.023)	0.59 <sup>b</sup> (0.42, 0.85)	0.77	47.7	89.7	21	26	3	23	64.4	1
Frozen	2 (1–4)	1–7		0.74 <sup>b</sup> (0.61, 0.83)	0.88	88.6	75.9	39	22	7	5	83.6	≤2

Abbreviation: IQR, interquartile range.

<sup>a</sup> This comparison was based on a randomly selected subset of fresh urine samples (n = 78) versus frozen samples collected during a previous study.

<sup>b</sup> Significant Spearman correlation ( $P < .001$ ).

were correlated moderately for Uc when fresh versus frozen Uc scores were compared ( $r = 0.64$ ; 95% CI = 0.48, 0.76;  $P < .001$ ).

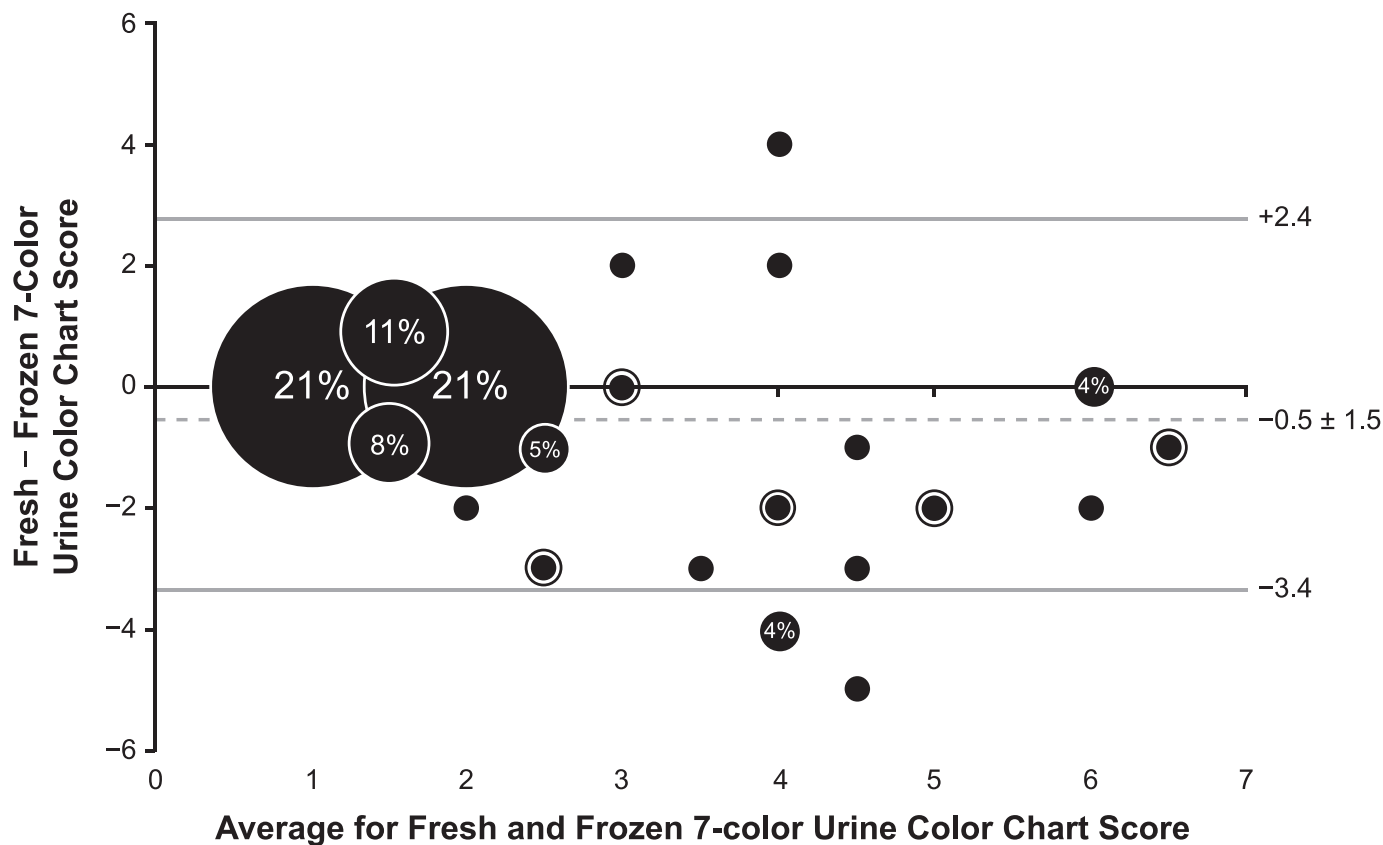
Based on the Bland-Altman plot comparing the Uc scores of fresh and frozen urine samples on an individual level, 48% of the reported Uc scores were similar for both conditions (Figure 2), and 75% of the reported Uc scores stayed within a  $\pm 1$  Uc shade difference. When correlating the results of the y and x axes of the Bland-Altman plot, we detected a small amount of reporting bias ( $r = -0.45$ ,  $P < .001$ ). The nature of this bias was expressed at the midsection of the 7-color Uc chart, with slightly darker Uc scores (~13%) in the frozen condition, indicating a higher level of underreporting in comparison with freshly scored samples.

The AUC was higher for frozen samples (0.88) than for fresh samples (0.77). This resulted in less accuracy in correctly classifying fresh samples (64.4%) versus frozen samples (83.6%). The ROC-based Uc cutoff value was 1 shade darker (Uc  $\leq 2$ ) than the suggested best fit for freshly measured urine samples (cutoff = 1) used to classify low versus high urine concentration.

## DISCUSSION

The key insight of our study was that under brighter LED light conditions, the diagnostic ability to discriminate urine samples with low versus high urine concentration was greater than in the halogen and fluorescent light conditions. Light intensity influenced perceived Uc, but despite differences for Uc among all light conditions, the 1-shade Uc cutoff value difference between the bright (1660–1848 lux) and darker (224–420 lux) light conditions was the most important practical finding. Additionally, Uc was reported differently between the scoring techniques, ie, assessing the sample while holding it against the chart versus assessing the sample using a standardized Uc scoring box, for the fluorescent and the LED light conditions but not for halogen. Although Uc was reported differently, no practical differences were evident for the AUC and accuracy of correctly classified urine samples or Uc cutoff value between scoring techniques.

Regarding objective (a), to evaluate the scored Uc using different lights and scoring techniques, the median scores ranged from 2 to 4, with the lowest and highest IQR values (1 and 6, respectively) fitting well within earlier reported data. Authors of previous studies described average Ucs of  $3 \pm 1$ <sup>10</sup> and  $3 \pm 2$  up to  $6 \pm 1$ .<sup>9</sup> Earlier researchers who addressed Uc scoring validation tested the urine samples in a well-lit room<sup>3,10,18</sup> and may have used different container sizes, both of which may affect Uc scoring. The question is which of our conditions matched the definition of a *well-lit room*, as the light intensity varied substantially among the fluorescent, halogen, and LED lights. We suggest that future publications should specify the type of light, as wavelength may influence how Uc is perceived and intensity and the container volume, as both may affect Uc scoring accuracy. Despite the significant difference between scoring techniques under fluorescent and LED light conditions, no clear differences were seen for reported Uc or the diagnostic ability of the scoring technique, suggesting that more rigid scoring conditions, such as using a separate Uc scoring box, do not necessarily result in



**Figure 2.** Bland-Altman plot for fresh versus frozen comparison based on 78 urine samples scored on a 7-color urine color chart. The bubble size indicates the percentage of urine color scores. Small black dots represent 1% values, and circled dots represent 2% to 3% values. The gray dotted line indicates the average reporting bias, and the limits of agreement are indicated by the gray lines.

better performance than the traditional sample-over-chart method.

Objective (b) was to investigate the diagnostic ability of the 8 measurement conditions. We showed fair to good diagnostic ability based on an overall high AUC ranging from 0.77 to 0.87, which was comparable with values (0.73–0.82) previously reported by athletes,<sup>6</sup> but other scores have been similar<sup>3</sup> or higher, ranging from 0.85 to 0.92.<sup>22</sup> On average, regardless of the scoring technique, the 2 brighter light conditions resulted in a Uc cutoff value that was at least 1 shade lower than under the standard fluorescent and halogen lights found in traditional office and restroom spaces. The LED flashlight conditions also resulted in 3% to 4% better accuracy for classifying low versus high urine concentration and a lower number of FP-classified urine samples. When the goal is to detect high urine concentration, a substantial ~5% fewer FP cases in the brighter light conditions is important; a reported low Uc when the actual urine concentration is above the selected 1.020 USG value will not prompt an individual to adjust fluid intake, which defeats the sole purpose of the hydration assessment.

Interestingly, the ROC-based best-fit Uc cutoff value was similar for halogen and the 2 LED conditions, whereas under fluorescent light (with a higher light intensity than halogen light), the optimal Uc cutoff value was 1 shade darker. The visual spectrum of light ranges from 360 to 830 nm.<sup>23</sup> Light types overlap in color range but differ in color tone. For example, LED light covers a blueish range of 395 to 530 nm; fluorescent light, a green to yellow range of 480

to 570 nm; and halogen, a range of 650 to 950 nm, starting as more reddish and eventually exceeding human perception. Although our study was not designed to differentiate among different intensities within light conditions, it seems likely that wavelength, in addition to light intensity, can influence how Uc is perceived. This is especially consequential because spectrophotometric analysis also demonstrated that, while Uc was reported as a darker color with increasing concentration, individual values of tristimulus colorimetry (CIE L\*a\*b\*) scores changed, indicating a notable polynomial color trend along the green-red axis, with a green hue in slightly dehydrated urine.<sup>24</sup> Therefore, although Uc tends to darken as concentration increases, subtle differences in how the color is constructed in combination with the selected light condition may influence how the urine is perceived by the human eye.

The goal of aim (c), comparing the results of frozen versus originally scored fresh outcomes, was to generalize these findings to freshly scored urine samples. Freezing tends to result in Uc that is 0.5 to 0.6 (95% CI = 0.3, 0.8) shades darker,<sup>25</sup> possibly as a result of the freeze-thaw method,<sup>26</sup> consistent with our protocol. At the same time, we noticed sedimentation,<sup>25</sup> which dissolved when the tubes were slowly inverted. Despite the mean Uc difference between fresh and frozen urine samples, the correlation between Uc scores was fair. The reporting bias in the Bland-Altman plot was caused by Uc samples that would likely have been classified above the suggested Uc cutoff value for underhydration in both fresh and frozen conditions. Therefore, the actual result for classifying urine

samples of low versus high urine concentration would have been the same for fresh and frozen samples, suggesting that our results can be generalized toward Uc scoring of fresh samples.

The strengths of this study were the relatively large sample size and multiple technicians, ensuring objective Uc scoring under the 4 light conditions. Scoring under the LED flashlight condition, which was the most accurate, could be easily accomplished at low cost with only slight modifications, ie, applying a single layer of white masking tape or painter's tape to block out the visual blue light, making the method accessible for a large population. However, our work also had limitations. The sensitivity analysis was based on only 1 light condition (LED panel) and the Uc scoring-box technique because this was the only method used to score the subsample of fresh urine samples before freezing. Another limitation was that the Uc scoring was compared with USG and not with urine osmolality, although USG has correlated strongly (up to  $r = 0.97$ ) with urine osmolality.<sup>20</sup> An additional factor was our use of randomly selected spot urine samples provided at different times of the day. However, previous data from our laboratory indicated that, when urine collections were stratified by morning versus other times of the day, Uc scoring accuracy was not affected.<sup>6</sup> As we did not control for activity, Uc and urine concentration could have been influenced by acute rehydration, leading to a mismatch between Uc and urine concentration due to acute dilution of the urine. The results of this analysis are based on urine samples from a healthy active population and may not be generalizable to patient populations with illnesses that affect their Uc. Finally, we used 30-mL tubes, whereas for urine collection, 90-mL containers are often used. Based on Uc comparisons in our laboratory, Uc differs a full shade between those volumes, resulting in a 1-shade-darker Uc when scoring the larger container.

The practical importance of these findings for athletic trainers and other sport medicine practitioners lies in the fact that the accurate classification of samples with high versus low urine concentration can be improved by modifying light conditions. Normal light conditions in well-lit rooms will probably be around 200 to 400 lux. We observed that scoring urine samples under very bright light (>1600 lux) resulted in greater accuracy of correctly classifying urine samples and fewer FP scores. A smaller number of FP scores is vital, as a reported light Uc when the actual urine concentration is high is unlikely to result in increased fluid intake.

The assessment of hydration status via Uc can be an important educational tool for athletes,<sup>11</sup> but even in applied settings, when laboratory analysis is not strictly necessary, athletes should use the most accurate method available. We suggest that 30-mL urine samples lit from underneath by a small LED flashlight displayed the highest number of correctly classified samples while using a modified Uc cutoff value. Previous researchers<sup>20</sup> identified the cutoff value for fresh urine samples scored using bright LED light was 1; therefore, samples with a Uc score  $\geq 2$  indicated a USG  $\geq 1.020$ .

In conclusion, the accuracy of Uc scoring was affected by light conditions, while the scoring techniques we evaluated did not seem to lead to practical differences in Uc scoring accuracy. These results can be translated to an everyday

suggestion that Uc cutoff values need to be adjusted to light conditions to optimize scoring accuracy.

## ACKNOWLEDGMENTS

We thank student helpers Margeret Matzinger and Josh Boeckman.

## REFERENCES

1. Kavouras SA. Hydration, dehydration, underhydration, optimal hydration: are we barking up the wrong tree? *Eur J Nutr*. 2019;58(2):471–473. doi:10.1007/s00394-018-01889-z
2. Perrier ET, Buendia-Jimenez I, Vecchio M, Armstrong LE, Tack I, Klein A. Twenty-four-hour urine osmolality as a physiological index of adequate water intake. *Dis Markers*. 2015;2015:231063. doi:10.1155/2015/231063
3. Perrier ET, Bottin JH, Vecchio M, Lemetais G. Criterion values for urine-specific gravity and urine color representing adequate water intake in healthy adults. *Eur J Clin Nutr*. 2017;71(4):561–563. doi:10.1038/ejcn.2016.269
4. Armstrong LE, Pumerantz AC, Fiala KA, et al. Human hydration indices: acute and longitudinal reference values. *Int J Sport Nutr Exerc Metab*. 2010;20(2):145–153. doi:10.1123/ijsnem.20.2.145
5. Hew-Butler TD, Eskin C, Bickham J, Rusnak M, VanderMeulen M. Dehydration is how you define it: comparison of 318 blood and urine athlete spot checks. *BMJ Open Sport Exerc Med*. 2018;4(1):e000297. doi:10.1136/bmjsem-2017-000297
6. Wardenaar FC, Thompsett D, Vento KA, Bacalzo D. A lavatory urine color (LUC) chart method can identify hypohydration in a physically active population. *Eur J Nutr*. 2021;60(5):2795–2805. doi:10.1007/s00394-020-02460-5
7. Baron S, Courbebaisse M, Lepicard EM, Friedlander G. Assessment of hydration status in a large population. *Br J Nutr*. 2015;113(1):147–158. doi:10.1017/S0007114514003213
8. American College of Sports Medicine; Sawka MN, Burke LM, et al. American College of Sports Medicine position stand. Exercise and fluid replacement. *Med Sci Sports Exerc*. 2007;39(2):377–390. doi:10.1249/mss.0b013e31802ca597
9. Armstrong LE, Maresh CM, Castellani JW, et al. Urinary indices of hydration status. *Int J Sport Nutr*. 1994;4(3):265–279. doi:10.1123/ijsn.4.3.265
10. Kavouras SA, Johnson EC, Bougatsas D, et al. Validation of a urine color scale for assessment of urine osmolality in healthy children. *Eur J Nutr*. 2016;55(3):907–915. doi:10.1007/s00394-015-0905-2
11. McDermott BP, Anderson SA, Armstrong LE, et al. National Athletic Trainers' Association position statement: fluid replacement for the physically active. *J Athl Train*. 2017;52(9):877–895. doi:10.4085/1062-6050-52.9.02
12. Kostelnik SB, Davy KP, Hedrick VE, Thomas DT, Davy BM. The validity of urine color as a hydration biomarker within the general adult population and athletes: a systematic review. *J Am Coll Nutr*. 2021;40(2):172–179. doi:10.1080/07315724.2020.1750073
13. Chevront SN, Ely BR, Kenefick RW, Sawka MN. Biological variation and diagnostic accuracy of dehydration assessment markers. *Am J Clin Nutr*. 2010;92(3):565–573. doi:10.3945/ajcn.2010.29490
14. Ellis LA, Yates BA, McKenzie AL, Muñoz CX, Casa DJ, Armstrong LE. Effects of three oral nutritional supplements on human hydration indices. *Int J Sport Nutr Exerc Metab*. 2016;26(4):356–362. doi:10.1123/ijsnem.2015-0244
15. Martin WF, Cerundolo LH, Pikosky MA, et al. Effects of dietary protein intake on indexes of hydration. *J Am Diet Assoc*. 2006;106(4):587–589. doi:10.1016/j.jada.2006.01.011
16. Zhang N, Du S, Zheng M, et al. Urine color for assessment of dehydration among college men students in Hebei, China—a cross-

- sectional study. *Asia Pac J Clin Nutr.* 2017;26(5):788–793. doi:10.6133/apjcn.052017.09
17. Adams WM, Hevel DJ, Maher JP, McGuirt JT. Racial and sex differences in 24 hour urinary hydration markers among male and female emerging adults: a pilot study. *Nutrients.* 2020;12(4):1068. doi:10.3390/nu12041068
  18. Armstrong LE, Soto JA, Hacker FT Jr, Casa DJ, Kavouras SA, Maresh CM. Urinary indices during dehydration, exercise, and rehydration. *Int J Sport Nutr.* 1998;8(4):345–355. doi:10.1123/ijns.8.4.345
  19. Wardenaar F, Ortega-Santos CP, Vento K, et al. Reliability of 3 urine specific gravity meters for measuring Brix and urine solutions at different temperatures. *J Athl Train.* 2021;56(4):389–395. doi:10.4085/1062-6050-0142.20
  20. Wardenaar FC, Thompsett D, Vento KA, Pesek K, Bacalzo D. Athletes' self-assessment of urine color using two color charts to determine urine concentration. *Int J Environ Res Public Health.* 2021;18(8):4126. doi:10.3390/ijerph18084126
  21. Olzinski S, Beaumont J, Toledo M, Yudell A, Johnston CS, Wardenaar FC. Hydration status and fluid needs of Division I female collegiate athletes exercising indoors and outdoors. *Sports (Basel).* 2019;7(7):155. doi:10.3390/sports7070155
  22. Arnaoutis G, Kavouras SA, Angelopoulou A, et al. Fluid balance during training in elite young athletes of different sports. *J Strength Cond Res.* 2015;29(12):3447–3452. doi:10.1519/JSC.0000000000000400
  23. Sliney DH. What is light? The visible spectrum and beyond. *Eye (Lond).* 2016;30(2):222–229. doi:10.1038/eye.2015.252
  24. Belasco R, Edwards T, Munoz AJ, Rayo V, Buono MJ. The effect of hydration on urine color objectively evaluated in CIE L\*a\*b\* color space. *Front Nutr.* 2020;7:576974. doi:10.3389/fnut.2020.576974
  25. Adams JD, Kavouras SA, Johnson EC, et al. The effect of storing temperature and duration on urinary hydration markers. *Int J Sport Nutr Exerc Metab.* 2017;27(1):18–24. doi:10.1123/ijsnem.2016-0098
  26. Daniels RL, Smith CR, Venn-Watson S. Effects of freeze-thaw cycle on urine values from bottlenose dolphins (*Tursiops truncatus*). *Aquat Mamm.* 2013;39(4):330–334. doi:10.1578/AM.39.4.2013.330

---

Address correspondence to Floris Wardenaar, PhD, College of Health Solutions, Arizona State University, 425 N 5th Street, Phoenix, AZ 85004. Address email to [floris.wardenaar@asu.edu](mailto:floris.wardenaar@asu.edu).