Shades of Skin: Limitations of the Fitzpatrick Scale with CIELAB

Sunny Tian

Received May 29, 2024
Accepted September 09, 2024
Electronic access October 15, 2024

The Fitzpatrick Skin Classification (FST) system, despite its widespread use, fails to adequately represent the diversity of skin tones, particularly among individuals with darker complexions. This limitation can lead to misdiagnoses and suboptimal treatment recommendations. To address this issue, the CIELAB color space values (L*: lightness, a*: red/green, and b*: blue/yellow) were employed as a more objective method for skin color classification. Using a comprehensive database of skin images from males and females across four ethnic/racial groups (White, Asian, Latino, Black), we analyzed variations in L*, a*, and b* values. Results indicated significant variations in skin tones, especially within Latino and Black participants, with a variance of 353.35 for Black participants and 53.70 for Latino participants compared to just 11.37 for White participants. Asian participants were found to have a variance of 31.08. These findings highlight that FST may inadequately represent the real-world diversity of skin tones within Black and Latino groups. The observed discrepancies underscore the need for a more precise, data-driven approach. Integrating CIELAB values into skin classification could enhance diagnostic accuracy and treatment efficacy, ultimately improving healthcare equity. This shift towards a more nuanced classification system promises to advance dermatological care by addressing existing disparities and accommodating the diverse needs of all patients.

Keywords: Fitzpatrick scale, CIELAB, race, ethnicity, skin tone, dermatology, underrepresented groups

Introduction

"Historically, Black skin, brown skin, is not represented in our literature appropriately,"[1] points out Dr. Ginette Okoye, the Chair of Dermatology at Howard University. The issue is particularly evident in the Fitzpatrick Skin Classification System (FST), where individuals are classified into one of only six skin types—ranging from light color (Type I) to dark color (Type VI)—based on skin color, response to ultraviolet light, and phenotypes (hair and eye color). People of color are often underrepresented in this scale and face limitations when using this classification. Experts and health care practitioners have found that olive and darker tones when limited to only two types (III and IV), can be problematic in clinical settings[2]. The limitations of FST approach to assessing skin phototypes III to VI were addressed by referencing CIELAB color space values according to more diverse population. This paper explores the current limitations of FST for individuals with skin of color and aims to argue for changes to make it more representative and scientifically robust.

Background

FST, devised by Thomas B. Fitzpatrick in 1975, is a fundamental dermatologic classification system that predicts skin responses to ultraviolet light[3]. Initially encompassing types I to IV, its significance improved with subsequent modifications to include types V and VI, broadening its applicability to diverse ethnicities such as Asian, Indian, and African[4]. Despite its widespread adoption in clinical settings and recognition as the "golden classification" in medicine, the scale's portrayal of skin tones, particularly within types III to VI, poses limitations[5]. The scale's utility extends beyond mere classification it serves as a highly referenced tool in dermatology, used in many settings, including determining appropriate UVA exposure in PUVA treatment for conditions like psoriasis and guiding the development and assessment of dermatological devices. However, the scale's limitations become more apparent when it fails to accurately differentiate among various skin tones and responses to ultraviolet light, particularly within types III to VI. This lack of precision can lead to inappropriate treatment decisions and hinder the effectiveness of dermatological interventions, particularly for individuals with skin of color. In this context, the CIELAB color space was utilized to test the variation in skin tone of among racial and ethnic groups. Scientists believe that FST may not accurately assess response to UV light since skin color and tanning questions may not properly reflect real-world skin tone variation.

CIELAB values emerged as an important framework for quantifying color. The distance from the central axis represents the chroma (Cab*), or saturation of the color. While this colorspace is not exclusive to skin, originally used across various other settings including painting and design, its application to human skin is notable. The CIELAB color space was chosen for this

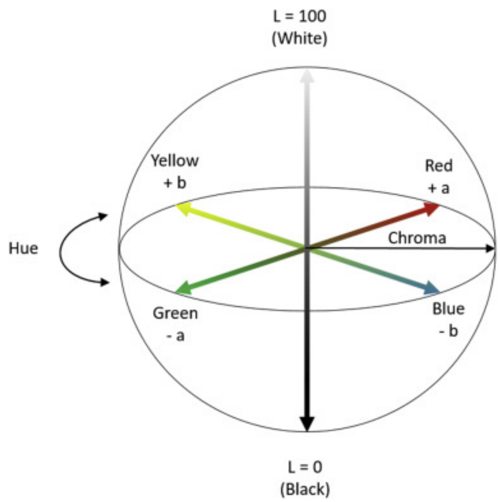

Fig. 1 The CIELAB color space diagram depicting L* on the vertical axis with a* and b* on the horizontal axis, indicating their relationship in reference to the central plane[6]. L* denotes lightness/darkness, a* corresponds to redness-greenness, and b* represents yellowness-blueness.

paper as it can be applied to human skin and has been used to describe injuries of the skin such as redness (increased a* value), or changes in the skin due to sun exposure (increased L* value in particular in the case of "tanning")[7]. The L*, a*, and b* values can be transcribed to dermatological parameters where the L* value may correlate to skin pigmentation, a* with erythema, and b* with pigmentation and tanning5.

Current Limitations of FST

As the current FST is widely used in dermatology to predict burn and tanning risk in the skin, it faces challenges in properly representing all skin types of ethnic/racial groups. Studies show that although FST is meant to predict the skin's response to UV light, many participants have reported that their skin type did not tan or burn, contrary to what FST stated. This resulted from mixed races, different L*a*b* values, and a lack of consideration for sunburn history[8]. Dr. Susan Taylor, Vice Chair for Diversity, Equity, and Inclusion for the Department of Dermatology, Perelman School of Medicine at the University of Pennsylvania, concurs that the current scale is outdated and inaccurate for people of non-European descent. Specifically, the scale states that there are just two skin types (III and IV) for people of olive and darker skin tones, without accounting for their many different shades and skin responses to UV[2]. In dermatology, such incorrect assessment of a person's skin tone could lead to miscalculated risk factors including cancer as well as inconsistencies for various skin risks and aesthetic procedures. With such a diverse range of skin tones falling under only a few categories, it is difficult to give precise and personalized treatments based on skin typing[2]. This means that FST may not be a fully accurate representation of the world's diverse skin tones. This could potentially lead to a bad diagnosis in a clinical setting, leading to possible harm to a patient. Additionally, the scale has been used by the Food and Drug Administration (FDA) as guidelines for products and over-the-counter treatments. Based on PDF summaries of 510k clearance results from the FDA database, it was found that 3596 devices use FST as a guideline, specifying which FST skin types the device should be used on. With many devices relying on FST classifications, it poses serious dangers to patients if the scale is too broad and unable to identify the diverse range of skin tones under FST types III and IV[9]. By establishing the limitations, it is clear that the scale does not adequately capture the range of melanin levels and genetics used to assess groups of various racial and ethnic backgrounds.

Recent Trends in Publications

This issue of FST is not unique; there has been a significant increase in submissions to journals addressing a broad spectrum of medical and surgical conditions in subjects grouped under the classification of "Skin of Color"[10]. To address these concerns, efforts have been made to update the FST to make it more accurate in stratifying various shades of skin color. For example, Santiago et al. (2023) proposed a modification of FST to address the growing interest in developing a skin classification system that captures a more diverse population[11]. Their study aimed to improve the utility of the scale by incorporating additional factors such as skin sensitivity and a history of hypertrophic scarring or keloids. Revised scales and an increase in journal submissions reflect a broader recognition within the dermatological community of the need for more accurate and inclusive skin color classification systems; however, a fully revised FST has yet to be developed and implemented in clinical settings.

False Sense of Security

The false sense of security led by FST poses significant risks for individuals with skin of color. The scale assumes that type V and VI skin always tans and rarely burns; however, recent studies have shown that sunburn in darker skin is more common than FST predicts. Although fair and delicate skin is normally associated with sensitive skin, darker skin with higher melanin levels may also be sensitive, contrary to the belief that FST has led people to assume. The current misconceptions expressed by most black individuals prevent sun-safe behaviors and may lead to more harm to the skin[12]. For example, a majority of the population in Ecuador are of Mestizo descent, who normally have darker skin but are still at great risk for skin cancer due to high UVRs in their country[13]. Even for groups living in milder climates within the US that experience reduced sun exposure intensity, intermittent sun exposure, especially in childhood, is a strong factor leading to melanoma. However, people of color, who feel a false sense of security due to this misconception, are less likely to use sunscreen, get skin examinations, or report skin concerns, sometimes leading to an even higher mortality rate compared to white populations because of a delayed diagnosis [14]. With immigration, intermarriage, and population migration, FST's limitations become apparent and its approach to capture groups with just one ethnic descriptor is no longer a viable predictor in dermatological practices.

While several modifications have been proposed, including the modification to include skin phototypes V and VI to better represent individuals of Asian, Indian, and African descent, there continues to be inaccurate assessment and poor consideration of the various shades of skin in the FST skin phototypes [14].

Building on this pressing need for a more inclusive and precise approach to skin classification, the CIELAB color space presents a framework that could enhance the accuracy and inclusivity of FST. By quantifying skin color variables through high-level instrumentation, such as a spectrophotometer, the CIELAB system allows for a more nuanced understanding of skin tone variations across different racial and ethnic groups. This method could address the current limitations of FST by providing a more detailed and accurate representation of the diverse spectrum of skin tones, particularly among individuals historically underrepresented in dermatology.

For example, CIELAB values measure lightness, red/green, and yellow/blue in skin, offering a more precise classification method than the categories currently used in FST. This approach has the potential to improve the scale's ability to differentiate between the various skin tones and responses to UV exposure within phototypes III to VI. By refining FST through the integration of CIELAB values, dermatologists could better assess skin cancer risk, tailor treatments for skin conditions, and ensure that all patients receive equitable care, regardless of their racial or ethnic background.

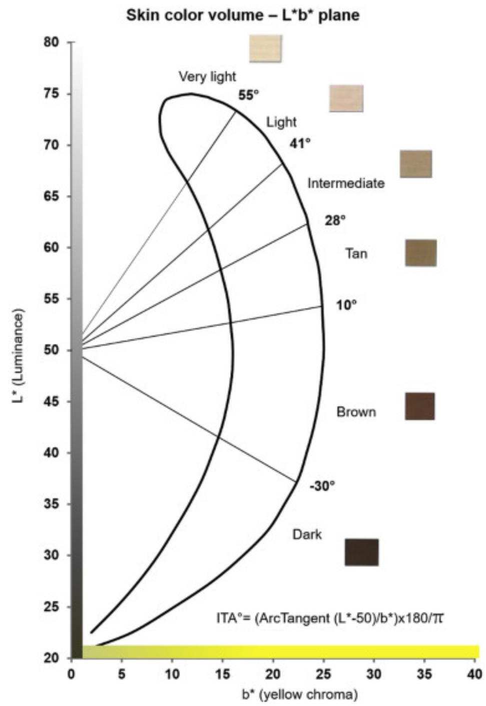

Fig. 2 Skin Color Volume - L*b* plane. Relationship between L* (vertical axis) and b* (horizontal axis) observing six different skin pigmentation groups [6].

The CIELAB Color Space

Using CIELAB values by quantifying skin color variables with high-level instrumentation, such as a spectrophotometer, may reveal improvements in FST's ability to discriminate skin tones in racial and ethnic groups. In a study by Everett et al. (2012), Spectrophotometry is employed to analyze a diverse female sample, highlighting common trends in African American participants, such as lower L* values but wider ranges of L* a* and b* values when compared to other groups [15].

In a related study, Ly et al. (2020) visually depict skin pigmentation trends using Adobe Photoshop, showcasing the relationship between the L* and b* values, as seen in Figure 1 [5]. The patterns of six groups: – very light, light, intermediate, tan, brown, and dark, –were quantified and graphed. "They used L* to measure luminescence/melanin, with the b* value plotted against the L* value, creating a "banana shape. Lighter and darker skin tones had a lower b* value (more blueness), while intermediate skin tones had a higher b* value (more yellowness). The color swatches from the data were representative of skin colors from the L'Oreal Skin Color Chart. The data revealed consistent patterns across ethnic and racial groups, suggesting a promising application of these trends to refine FST. Notably, the a* value, associated with redness/greenness, exhibits a correlation with sun exposure, being more prevalent in Caucasian participants than in African American participants. Moreover, the skin L* and b* values, sunburn history, and race/ethnicity significantly account for 72% of the variance in FST scores in a diverse sample of 466 women [16]. These findings underscore the potential of CIELAB values to contribute to the accuracy of the classification system. The studies collectively suggest that CIELAB values may help address existing limitations and allow FST to better represent populations historically underrepresented in dermatology.

The reference of CIELAB values to refine FST may enhance the accuracy of skin typing but also addresses the broader implications of underrepresentation in dermatological practice [13]. The existing limitations of FST have widely contributed to inequalities in healthcare, particularly for individuals with skin types III-VI. By referencing values derived from CIELAB, the revised classification system holds the potential to better represent the diverse range of skin pigmentation observed across different racial and ethnic groups. This inclusivity is crucial for ensuring equitable access to dermatological care and reducing disparities in skin cancer diagnosis and treatment among underrepresented populations.

Advanced clinical treatment

A refinement of FST with CIELAB color space values further reflects a larger trend toward personalized medicine in dermatology. By shifting away from one-size-fits-all approaches in skin classification, dermatologists can tailor interventions and treatment plans to meet the unique needs of all individual patients. FST is widely used in clinical settings for various purposes, including assessing an individual's risk of developing skin cancer, determining the most appropriate cosmetic procedures (such as chemical peels and laser treatments), cautioning certain groups regarding medication prescriptions related to photosensitivity reactions[8], categorizing study participants in clinical trials and epidemiological studies, and issuing cautions related to dermatological devices under the Food and Drug Administration, etc[17]. A revised and personalized approach holds the potential to improve the accuracy of risk assessment and diagnosis while also possibly enhancing patient outcomes and results. With the increasing diversity of the US population, dermatological practice must evolve to accommodate the unique characteristics and needs of all patients, regardless of their racial or ethnic background. The imperative shift towards personalized medicine not only enhances patient outcomes but also promotes equity and inclusivity in dermatological practice, ultimately advancing the field towards more effective and patient-centered care.

Changing phenotypes in the US

As the demographic landscape of the United States undergoes profound changes in diversity, the proportion of individuals of non-European descent increases. The US Census Bureau projects that the population of non-European descent is expected to surpass 50% by 2050. This demographic transformation renders the current classification system inadequate for accurately representing the diverse spectrum of skin tones[3].

This trend, coupled with the projected growth of mixed populations, Hispanics, and Asians, underscores the growth of the country's racial and ethnic composition. By 2020, it was estimated that fewer than 50% of children would be of non-Hispanic White descent, highlighting the impact of immigration on the nation's demographic makeup[18].

Furthermore, the diversity index was around 54.9%; however, in the recent 2020 Census Bureau report, the chance was raised to 61.1%. Specifically, the population aged over 18 had a score of 58.3% while the under 18 age had a score of 68.5%. The under-18 demographic, in particular, exhibits a substantially higher diversity index compared to adults, underscoring the increased pace of demographic change within younger generations[19].

Despite the shortcomings of FST, its widespread usage persists within clinical settings. In a study, by Ware et al. (2020), 140 respondents were included in the final analysis. Of these respondents, 41% of respondents agreed that FST should be included in clinical documentation. About 31% said they used FST to describe patients' race or ethnicity while 47% used it to describe their skin color. 22% used it in both scenarios. The results of the survey showed that almost 1/3 to 1/2 of dermatologists use FST in correlation with race/ethnicity or constitutive skin color[3]. A significant proportion of dermatologists continue to employ the scale with varying degrees of awareness and reliance on the scale knowing its limitations. The reluctance to adopt a revised classification system results in the need for more initiatives aimed at awareness among healthcare practitioners regarding the adoption of more inclusive approaches.

The revision of FST represents a crucial step toward achieving equity in dermatological care. By acknowledging the limitations of existing classification systems and embracing more inclusive approaches, healthcare practitioners can ensure that patients receive tailored and effective treatments regardless of their racial or ethnic background. As the demographic composition of the United States continues to evolve, dermatology must equally become more inclusive to fit the needs of an increasingly diverse population.

Counterclaim

The lack of representation within dermatology for those with skin color stands as a limiting factor14. Considering the trends in the United States, where phenotypes are becoming increasingly diverse, the imperative to address the limitations of FST becomes further pronounced. By referencing CIELAB values as skin color metrics, a proposed revision may hold a more accurate skin classification for populations historically underrepresented in dermatology in an increasingly diverse society. Undoubtedly, FST and the lack of research on dark skin as compared to light skin is an important issue that requires further action.

Methods

Hypothesis

An experiment was conducted to test the hypothesis that FST requires refinement due to A quantitative methodology was used to analyze skin color with CIELAB values using Adobe Photoshop. The independent variable was the ethnic/racial group, and the dependent variable was the Lab* value. Pre-existing digital images from a database were used. The Adobe Photoshop eye-dropper color sampler tool was set to a '5 x 5 Average' pixel sample size (total of 25 pixels) for consistencyTo ensure accuracy, the image was viewed as a clock. The sampling area was assessed starting at 12 o'clock and moved clockwise until a representative area was found and recorded for L*, a*, and b* values. This process was repeated nine times in multiple areas within the sampling area of the 5x5 pixel sample to ensure accuracy, with the L*a*b* values recorded from the color sampler tool in the properties tab[20]. A total of 360 trials were recorded, with nine trials per image, as this was found to provide a representative color analysis. Excel spreadsheets were employed to calculate the chroma of the skin using the CIELAB values (Chroma or Ca*b

This approach quantified skin color using CIELAB values to assess variation across multiple racial and ethnic groups. This experimental research included controlled sampling and quantitative measurement procedures, involving numerical data analysis. The research focused on quantitatively characterizing and understanding the distribution of skin tones, particularly among individuals of color. The ultimate objective is to support a refinement of FST, which may exhibit variability in skin tones among populations, particularly for Black and Latino communities.

Sampling

This study used a comprehensive database of skin images from the Skin Tone Identity and Inequalities Project (STiiP) at the University of Illinois-Chicago. The database was published in ICPSR and is publicly accessible. The STiiP research team, led by Principal Investigators Rachel A. Gordon and Amelia Branigan, focused on developing and validating techniques for gathering skin color data in social science research. The database included images gathered from two samples: "In-person/College Sample" and "Online Sample." For the "In-Person/College Sample," 230 screeners were collected from college students, with a cooperation rate of 62% where 141 students were contacted and 87 responses were received. This resulted in 82 scheduled visits to achieve the 50 participants selected for the study. Data collection from this sample involved assessing skin color using Massey-Martin and PERLA, two widely used skin tone rating scales, and devices to analyze L*a*b* values from the Nix colorimeter and Labby spectrophotometer. For the "Online Sample," participants were recruited via the online platform Prolific, with stringent pre-screening criteria to ensure eligibility and data quality. Sixty participants were selected in each of eight categories based on sex and racial-ethnic categories. The groups consisted of females and males in the following ethnic groups: Black, Latino, Asian, and White. In total, 40 digital images were selected to fit categories defined by the 2 x 4 x 5 cross-classification of binary gender (male or female), four-category race-ethnicity (Asian, Black, Latino, White), and five levels of skin tone (L1-L5 - darkest to lightest). They collectively provide a diverse representation of skin tones across different racial and ethnic groups, enhancing the reliability of the findings. Trust in the database is established through rigorous recruitment procedures, adherence to IRB-approved protocols, and transparent methodology[21].

Results

Data collection, including nine trials per photograph, included a diverse group of participants (Appendix A) representing a wide range of skin tones, grouped into distinct categories for analysis: white (Figure 3), Asian (Figure 4), Latino (Figure 5), and black (Figure 6).

Each category included both male and female participants. For White participants in (Figure 3), the distribution of L* (luminosity) and Cab* (chroma) values was the most constrained, indicating a relatively consistently uniform range of skin tones within this group. For Asian participants, (Figure 4) a broader distribution of L* and Cab* values compared to the white group. Although there was greater diversity in this category, the spread was still considerably moderate. Next, for Latino participants in Figure 5, the distribution of L* and Cab* expands, showing a greater variation in skin tones among Latino participants. Lastly, for Black participants, Figure 6 reveals the widest distribution of L* and Cab*, indicating the most diverse range of skin tones among black participants.

Calculated standard deviations and variances, seen in Table 1, confirm the trends observed in the visual dispersion of L* and Cab* values. Black participants possess the most diverse range of skin tones. Their L* standard deviation, a measure of spread around the average L* value, reached 18.80, signifying a much wider distribution compared to White participants with a L* standard deviation of 3.37. This translates to a variance of 353.35 for Black participants compared to just 11.37 for White participants. A similar pattern emerges with Cab* values where Black participants hold the highest standard deviation at 9.65 with a variance of 93.13, highlighting a broader spread in chroma compared to other groups.

White participants exhibit the most homogenous distribution. Their low standard deviations (3.37 for L* and 3.32 for Cab*) and variances (11.37 and 11.00, respectively) indicate a tightly clustered range of skin tones, as seen in Figure 3. The Latino and Asian groups fall between these two as their standard deviations and variances for both L* and Cab* are noticeably higher than White participants, reflecting a greater degree of diversity in skin tones. However, they still fall below the values observed in Black participants, suggesting a less extensive range.

These quantitative measures support the conclusions drawn from the visual distribution analysis. The considerable dispersion of L* and Cab* values among Latino and Black participants signifies the multifaceted nature of skin pigmentation

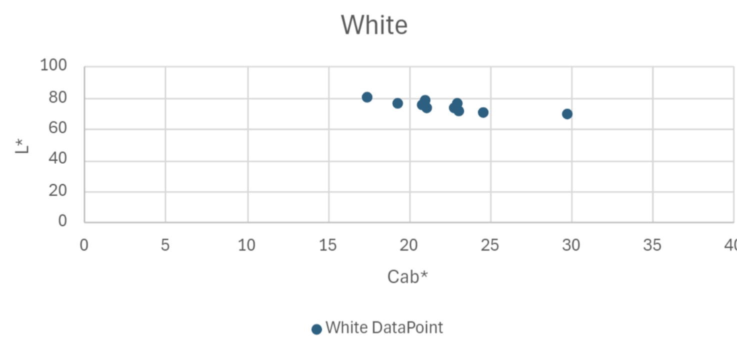

Fig. 3 White Racial Group: Chroma (Cab*) on the horizontal axis graphed against L* (lightness/darkness) on the vertical axis, depicting the distribution of quantified data from skin samples

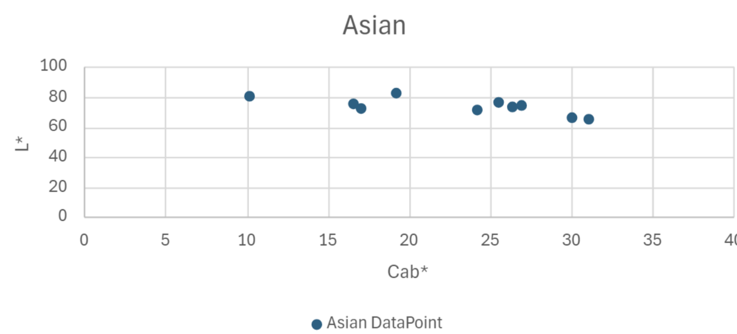

Fig. 4 Asian Racial Group: Chroma (Cab*) on the horizontal axis graphed against L* (lightness/darkness) on the vertical axis, depicting the distribution of quantified data from skin samples

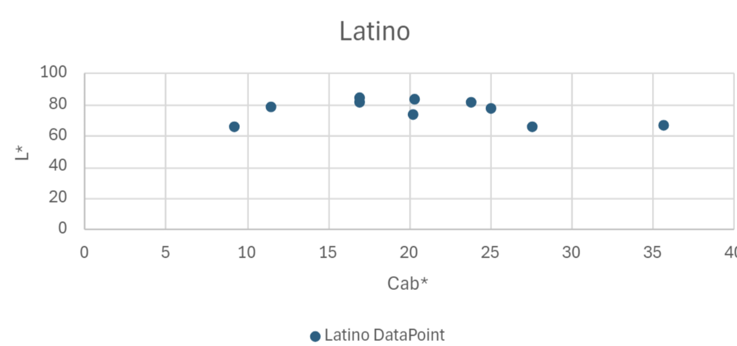

Fig. 5 Latino Racial Group: Chroma (Cab*) on the horizontal axis graphed against L* (lightness/darkness) on the vertical axis, depicting the distribution of quantified data from skin samples

within these groups. This variability, reflecting various genetic, physiological, and environmental factors, contributes to great differences in melanin production and distribution, which was observed in the experiment. The wide range of L* and Cab* values, specifically in the Latino and Black ethnic groups, underscores the limitations of classifying diverse skin tones in just six skin types under FST. By testing the variation in skin tone of racial and ethnic groups, it was observed that there is a much greater variation in certain groups, meaning the FST scale may not be an accurate way to assess response to UVL as its skin color and tanning questions may not be representative of the skin tones around the world.

The analysis demonstrates that the great variability, observed in Black and Latino groups, may show to be underrepresented with FST. They underscore the limitations of the existing classification system and advocate for a more nuanced approach that

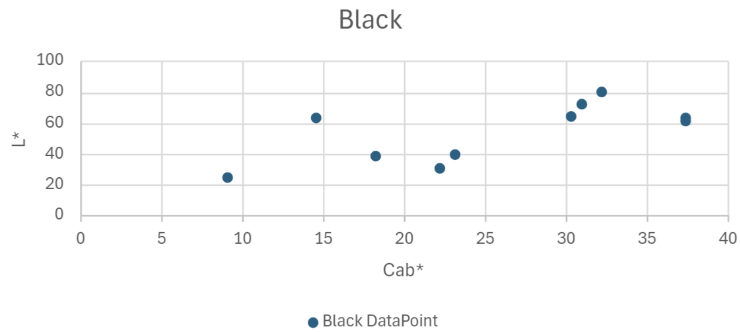

Fig. 6 Black Racial Group: Chroma (Cab*) on the horizontal axis graphed against L* (lightness/darkness) on the vertical axis, depicting the distribution of quantified data from skin samples

Skin Tone	L* Standard Deviation	L* Variance	Cab* Standard Deviation	Cab* Variance
White	3.37	11.37	3.32	11.00
Asian	5.57	31.08	6.70	44.84
Latino	7.33	53.70	7.79	60.73
Black	18.80	353.35	9.65	93.13

Table 1: Standard Deviations and Variances for L* and Cab* values observed under four racial groups: White, Asian, Latino, and Black.

acknowledges the remarkable variability in skin tones, particularly among individuals with darker skin complexions. The significant variance found in the L* and Cab* values among Latino and Black participants suggest that these groups encompass a broader spectrum of skin tones, requiring a change to clinical classification and treatment recommendations.

Statistical Analysis

Assessment of agreement of skin color metrics was conducted through intra-rater agreement. To evaluate the agreement of measurements for skin color metrics within individual raters, a series of intra-class correlational coefficients (ICC) were computed using nine trials for each photograph (nine observations from the color sampler technique). Data was recorded to an xsls and exported to .csv22 sorted by trial and L*, a*, and b* values. Following this, ICC values were computed pairwise for all trials (Appendix B). Specifically, based on the methodology used, involving two-way mixed effects and single rater/measurement (fixed set of raters judging a fixed set of targets), was used for assessing intra-rater reliability23. Additionally, single values were specified as the unit of analysis as opposed to averaged values. For these ICCs, agreement was computed in terms of consistency, rather than absolute agreement. Following this, values were computed and observed for each of the L*, a*, and b* variables from the color sampling technique.

Results of intra-rater agreement, determined by ICCs, showed an extremely high level of internal consistency and reliability within each rater. Rates were consistent for skin areas [ICC = 0.99596 for L* (light/dark), ICC = 0.96301 for a* (red/green), and ICC=0.98780 for b* (yellow/blue)]. Notably, the most pronounced intra-rater reliability occurred within the L* values of the skin, which play an important role in skin color analysis. This suggests that the data for the lightness/darkness measurement may confidently be considered reproducible with the color-sampler method used. Analysis of the ICCs for intra-rater reliability concerning the a* and b* values revealed similar outcomes, revealing a high level of agreement for the color-sampler method.

P-values associated with the ICCs were further calculated, underscoring the significance of the observed agreement. Specifically, the p-values for L*, a*, and b* were 3.28426e-86, 6.73936e-49, and 1.47620e-67, respectively. The small p-values (¡0.05), indicate strong evidence against the null hypothesis of no agreement or reliability between measurements. Therefore, the observed levels of intra-rater reliability were highly unlikely to occur by random chance alone, supporting the reliability of the findings.

This high level of intra-rater agreement, determined by ICCs and coupled with statistically significant p-values, supports that data collected from the nine trials per photograph resulted in reliable results. The consistency observed across raters and trials suggests the color-sampler technique facilitated precise and reproducible measurements of skin color metrics across multiple trials. Consequently, results obtained from this statistical analysis serve to support the conclusions drawn regarding skin tone variations among the study's sampling images.

Fitzpatrick Skin Types III and IV: Falling Short of Diverse Populations

Moreover, these implications extend beyond methodological validation, providing insights into the skin color variation within populations. FST traditionally associates types III and IV with individuals of Latino and Black descent, categorizing them primarily based on their phenotypes, including their olive and darker skin tones, respectively. However, the observed high variances in L* and Cab* values for these categories, particularly among Latino and Black participants, expose the limitations of a broader classification with such skin variation in individual groups. This indicates that FST may be inaccurate in representing the diverse skin tones among individuals with skin of color. As a result, medical practitioners relying solely on FST may struggle to provide optimal treatment and recommendations in response to UVL due to the lack of precise differentiation within these types, with FST failing to represent real-world skin tones with just skin color and tanning questions. For example, two individuals with skin type III may receive the same treatment because of their skin type, though their skin tones and reaction to the treatment may be entirely different. By integrating objective measures, such as the CIELAB color space, clinicians may better characterize and classify the wide range of skin tones. Resultantly, more accurate diagnoses, treatments, and monitoring may be observed.

Discussion

FST has been the gold standard for assessing skin types in clinical and research settings. However, the findings presented from this study highlight significant limitations, particularly in its inability to accurately categorize individuals with skin of color. The hypothesis—that FST should be revised by utilizing the CIELAB color space values could be applied to improve the representation of certain groups—appears to be supported by the findings. The significant variability in the L* (luminosity) and Cab* (chroma) values across ethnic groups indicates that the traditional classification system's broad categorizations do not effectively capture this diversity, particularly in groups with a wider range of skin tones.

This study demonstrates the advantages of employing objective measurements, such as the CIELAB color space. Unlike the subjective phenotypic descriptors utilized in FST, CIELAB contains values that offer a data-driven approach. This objectivity plays a crucial role in ensuring consistency and reducing risks of misclassification.

However, there are limitations to this research that must be acknowledged. The trials, consisting of a database of images, while inclusive of various ethnic groups, may not fully represent the entire spectrum of skin tones within each group due to the sample size. It consisted of four ethnic and racial groups, consisting of both female and male; however, it included only 40 images, resulting in a maximum of 360 trials with 9 trials per image. The sample size could further be expanded to enhance generalizability. Additionally, further research is needed to determine the most efficient and practical method for incorporating CIELAB measurements into clinical practice. While conducting 9 trials per image served to be representative of this study, a more efficient and reliable method could further be researched. Further studies may also explore the feasibility of a more clinician-friendly tool to apply this in a real clinical setting.

Additionally, while the CIELAB color space offers an objective measurement of skin color, its practical implementation in clinical dermatological practice requires further validation. The transition from digital image analysis to real-world application necessitates rigorous testing to ensure that CIELAB values obtained through tools like Adobe Photoshop accurately reflect real-world skin tones. This includes validating these values against physical measurements and standardized color references to account for potential discrepancies introduced by lighting conditions and camera settings. Further validation studies may focus on ensuring the reliability of digital measurements and addressing any limitations associated with digital imaging techniques.

While CIELAB provides a robust framework for color measurement, it is important to explore alternative approaches and their potential drawbacks. For instance, other color spaces or measurement techniques might offer different advantages or limitations. A future review of various methods and their comparative effectiveness will be essential in determining the most suitable approach for clinical use. Exploring a range of options and conducting validation studies will help in selecting a method that provides accurate, reliable, and clinically applicable results.

CIELAB's integration into clinical practice demands careful consideration and validation. Addressing these practical implementation challenges and exploring alternative approaches will be crucial in developing a more effective and inclusive dermatological classification system. Ongoing research and validation process will contribute to more accurate assessments and better patient outcomes across a diverse population.

Next, the potential refinement of FST using the CIELAB color space could significantly impact not only the assessment of skin response to ultraviolet light (UVL) but also the management of other dermatological conditions. For example, conditions such as melasma, hyperpigmentation, and vitiligo could benefit from a more nuanced classification system that accounts for the wide variability in skin tones. A more precise classification could lead to better-targeted treatments and more effective management strategies, as it would allow for customization based on individual skin color metrics rather than broad categories.

Integration of genetic markers for pigmentation could further enhance the accuracy of skin tone classification. Genetic factors play a crucial role in determining skin pigmentation and incorporating genetic data could provide a more comprehensive understanding of an individual's skin characteristics. This integration might involve correlating genetic markers with CIELAB values to develop a multifactorial model that combines both objective color measurements and genetic information. Such an approach could potentially improve the precision of dermatological assessments and contribute to more personalized treatment plans.

The path forward should involve a collaborative effort between dermatologists and researchers. These approaches hold the potential to yield a more comprehensive classification system that accounts for a wider range of skin tones. The possibility of a more updated approach to keep pace with the increasingly diverse population of the United States can lead to improved clinical skin classifications and treatments.

By improving skin color classification, healthcare professionals are able to tailor treatment plans with improved precision, leading to improved patient outcomes. The current limitations of the FST system raise important concerns regarding the treatment of individuals with darker skin complexions. With a nuanced classification system, a step towards equitable and effective dermatological care that better represents the real-world population may be made. This aligns to promote inclusive healthcare practices that cater to the needs of the diversifying population. By acknowledging the shortcomings of the current FST and embracing a data-driven approach, fosters a future of personalized dermatology, and ensures every patient receives precise care, regardless of the color of their skin. This shift signifies not just a scientific advancement, but a crucial step towards dismantling inequities within healthcare while fostering a more inclusive medical landscape for all.

Conclusion

Based on these findings, it is apparent that revising FST to include a stronger representation of individuals with skin of color by referencing values derived from the CIELAB color space would increase treatment and recommendations for individuals with skin of color in dermatology. The broader distribution of L* and Cab* in Latino and Black participants highlights the need for a more refined classification system that may account for the wider variety of skin tones in the real world. Assigning broad categories like the "olive" or "darker" skin tones under FST fails to capture the diverse range of skin tones within these populations. A revision would lead to improved accuracy in diagnosis, treatment, and research involving FST skin types, better serving a more diverse population.

References

1 E. Kwong, M. Sofia and T. Lu, *Meet The Dermatologists Advancing Better Care For Skin Of Color: Short Wave*, n.d.). NPR.org.

2 S. Taylor.

3 O. Ware, J. Dawson, M. Shinohara and S. Taylor, *Cutis*, **105**, 77–80.

4 T. Fitzpatrick, *Arch*, **Dermatol;124(6):869–871**, year.

5 S. Cancer, `https://link.gale.com/apps/doc/CV2643900115/SUIC?u=clar33415&sid=bookmark-SUIC&xid=5aecb632`.

6 B. Ly, E. Dyer, J. Feig, A. Chien and S., *The Journal of Investigative Dermatology*, **Jan;140(1):3-12**, year.

7 W. Huang, Y. Wang, K. Hung, P. Hsieh, K. Fu, L. Dai, N. Liou, K. Ma, J. Liu and N. Dai, *PeerJ*, **6**, 4815.

8 A. Sharma and B. Patel, *Laser Fitzpatrick Skin Type Recommendations*, `https://www.ncbi.nlm.nih.gov/books/NBK557626/`, Updated 2023 Mar 6]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from:.

9 Hijiri, `https://github.com/NyxIsBad/fdasearch`, NyxIsBad/fdasearch. GitHub.

10 W. Coleman, K. Mariwalla and P., *Grimes Updating the Fitzpatrick Classification: The Skin Color and Ethnicity Scale*, et al.],.

11 S. Santiago, R. Brown, K. Shao, J. Hooper and M. Perez, *Journal of drugs in dermatology: JDD*, **22**, 641–646.

12 D. Fajuyigbe and M. Verschoore, *Current problems in dermatology*, **55**, 62–71.

13 M. Fors, P. González, C. Viada, K. Falcon and S. Palacios, *Advances in skin wound care*, **33**, 1–5.

14 P. Goon, C. Banfield, O. Bello and N. Levell, *Skin health and disease*, **1**, 40.

15 J. Everett, M. Budescu and M. Sommers, *Clinical nursing research*, **21**, 495–516.

16 M. Sommers, J. Fargo, Y. Regueira, K. Brown, B. Beacham, A. Perfetti, J. Everett and D. Margolis, *Ethnicity disease*, **29**, 505–512.

17 F. Food and F. Drug Administration, *Fitzpatrick skin types*, https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma_pas.cfm?t_id=90246&c_id=104.

18 J. Vespa, L. Medina and D. Armstrong, *Demographic Turning Points for the United States: Population Projections for 2020 to 2060 Population Estimates and Projections Current Population Reports*, https://www.census.gov/content/dam/Census/library/publications/2020/demo/p25-1144.pdf.

19 A. Menchaca, B. Pratt, E. Jensen and N. Jones, *Examining the Racial and Ethnic Diversity of Adults and Children*.

20 M. Sommers, Personal communication [Email].

21 A. Rachel, *Gordon, and Amelia BraniganSkin Tone Identity and Inequalities Project (STiiP): 2018-2020, Journal Article "Measuring Skin Color: Consistency, Comparability, and Meaningfulness of Rating Scale Scores and Handheld Device Readings*, https://doi.org/10.3886/E142261V1, distributor], 2021-06-17.

Appendix - A

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Database	L*	a*	b*		L*	a*	b*		L*	a*	b*	L*	a*	b*	Inverse tan	Hue	Chroma
2	BF L1	30	15	17		31	15	16		32	15	16	31	15	16.33	47.43	47.43	22.17
3	BF L2	38	10	15		40	11	15		38	11	15	38.67	10.33	15	55.45	55.45	18.21
4	BF L3	62	13	16		64	14	14		65	13	14	63.67	21.33	30.67	55.18	55.18	37.36
5	BF L4	61	21	31		61	22	30		61	21	31	61	21.33	30.67	55.18	55.18	37.36
6	BF L5	64	15	27		65	14	27		64	14	26	64.33	14.33	26.67	61.75	61.75	30.28
7																		
8	BM L1	26	6	4		24	8	6		25	8	6	25	7.33	5.33	36.02	36.02	9.06
9	BM L2	39	13	18		40	14	19		39	14	19	39.33	13.67	18.67	53.79	53.79	23.14
10	BM L3	64	13	3		63	15	5		64	14	4	63.67	14	4	15.95	15.95	14.56
11	BM L4	73	15	25		70	16	26		73	16	26	72	15.67	26.67	59.56	59.56	30.93
12	BM L5	80	18	27		81	16	27		80	17	28	80.33	17	27.33	58.12	58.12	32.19
13	DATE: 4/3																	
14	LF L1	72	15	17		73	13	14		74	13	14	73	13.67	15	47.66	47.66	20.29
15	LF L2	66	20	29		66	21	30		67	23	27	66.33	21.33	28.67	53.35	53.35	35.73
16	LF L3	67	11	24		66	12	25		65	12	26	66	11.67	25	64.98	64.98	27.59
17	LF L4	76	15	22		78	12	20		77	14	21	77	13.67	21	56.94	56.94	25.06
18	LF L5	79	10	6		78	10	4		79	11	5	78.67	10.33	5	25.83	25.83	11.48
19																		
20	LM L1	66	9	5		66	8	5		66	7	4	66	8	4.67	30.27	30.27	9.26
21	LM L2	82	14	20		82	14	19		81	15	18	81.67	14.33	19	52.98	52.98	23.8
22	LM L3	84	10	14		83	10	14		85	10	13	84	10	13.67	53.81	53.81	16.94
23	LM L4	84	10	18		84	10	17		81	12	17	83	10.67	17.33	58.38	58.38	20.35
24	LM L5	81	14	9		81	14	9		82	15	9	81.33	14.33	9	32.13	32.13	16.92
25	DATE: 4/6																	
26																		
27	AF L1	65	15	27		65	16	27		66	15	27	65.33	15.33	27	60.41	60.41	31.05
28	AF L2	75	18	20		74	17	20		74	18	21	74.33	17.67	20.33	49.00	49.00	26.94
29	AF L3	72	14	23		75	14	23		72	14	21	73	14	22.33	57.91	57.91	26.36
30	AF L4	71	10	23		71	9	22		72	9	22	71.33	9.33	22.33	67.32	67.32	24.2
31	AF L5	83	9	17		83	9	17		83	9	17	83	9	17	62.10	62.10	19.24
32																		
33	AM L1	72	15	8		72	14	9		72	15	9	72	14.67	8.67	30.58	30.58	17.04
34	AM L2	67	20	23		65	20	22		66	19	23	66	19.67	22.67	49.05	49.05	30.01
35	AM L3	77	12	23		76	12	22		77	13	22	76.67	12.33	22.33	61.09	61.09	25.51
36	AM L4	74	11	13		75	10	12		76	11	13	75	10.67	12.67	49.90	49.90	16.56
37	AM L5	80	7	7		81	6	8		80	7	8	80.33	6.67	7.67	48.99	48.99	10.16
38	DATE: 4/6																	
39	WF L1	73	11	20		74	11	20		72	11	20	73	11	20	61.19	61.19	22.83
40	WF L2	76	10	21		76	10	22		77	10	19	76.33	10	20.67	64.18	64.18	22.96
41	WF L3	71	12	19		72	12	20		71	12	20	71.33	12	19.67	58.61	58.61	23.04
42	WF L4	75	12	17		75	12	17		75	12	17	75	12	17	54.78	54.78	20.81
43	WF L5	77	14	13		76	15	13		77	14	13	76.67	14.33	13	42.21	42.21	19.35
44																		
45	WM L1	70	19	22		69	19	22		70	19	22	69.67	20	22	47.73	47.73	29.73
46	WM L2	78	11	18		79	11	17		77	12	18	78	11.33	17.69	57.36	57.36	21.01
47	WM L3	70	16	19		71	16	18		71	16	19	70.67	16	18.67	49.40	49.40	24.59
48	WM L4	73	13	16		74	13	17		73	13	17	73.33	13	16.67	52.05	52.05	21.14
49	WM L5	79	11	13		80	12	13		81	12	13	80	11.67	13	48.09	48.09	17.47
50	DATE: 4/6																	

Data Collection Notebook: Quantitative Data Collected via Adobe Photoshop consisting of a total of 9 trials per Image/Sample Size

Appendix - B

```
C:\Users\sunny\OneDrive\Desktop\code>python pythonstuff.py
----------------------
L* results:
    Type              Description       ICC           F  df1  df2          pval         CI95%
0   ICC1   Single raters absolute  0.995993  746.643590   39   80  2.446177e-88  [0.99, 1.0]
1   ICC2     Single random raters  0.995993  739.689246   39   78  3.284256e-86  [0.99, 1.0]
2   ICC3      Single fixed raters  0.995955  739.689246   39   78  3.284256e-86  [0.99, 1.0]
3  ICC1k  Average raters absolute  0.998661  746.643590   39   80  2.446177e-88   [1.0, 1.0]
4  ICC2k    Average random raters  0.998661  739.689246   39   78  3.284256e-86   [1.0, 1.0]
5  ICC3k     Average fixed raters  0.998648  739.689246   39   78  3.284256e-86   [1.0, 1.0]
----------------------
a* results:
    Type              Description       ICC          F  df1  df2          pval         CI95%
0   ICC1   Single raters absolute  0.962592  78.195971   39   80  1.004845e-49  [0.94, 0.98]
1   ICC2     Single random raters  0.962597  79.101436   39   78  6.739362e-49  [0.94, 0.98]
2   ICC3      Single fixed raters  0.963009  79.101436   39   78  6.739362e-49  [0.94, 0.98]
3  ICC1k  Average raters absolute  0.987212  78.195971   39   80  1.004845e-49  [0.98, 0.99]
4  ICC2k    Average random raters  0.987213  79.101436   39   78  6.739362e-49  [0.98, 0.99]
5  ICC3k     Average fixed raters  0.987358  79.101436   39   78  6.739362e-49  [0.98, 0.99]
----------------------
b* results:
    Type              Description       ICC           F  df1  df2          pval          CI95%
0   ICC1   Single raters absolute  0.987833  244.577289   39   80  4.315122e-69  [0.98, 0.99]
1   ICC2     Single random raters  0.987833  243.951772   39   78  1.476202e-67  [0.98, 0.99]
2   ICC3      Single fixed raters  0.987802  243.951772   39   78  1.476202e-67  [0.98, 0.99]
3  ICC1k  Average raters absolute  0.995911  244.577289   39   80  4.315122e-69   [0.99, 1.0]
4  ICC2k    Average random raters  0.995911  243.951772   39   78  1.476202e-67   [0.99, 1.0]
5  ICC3k     Average fixed raters  0.995901  243.951772   39   78  1.476202e-67   [0.99, 1.0]
```

Calculated ICC3s (intraclass correlation coefficient 3) and P-values: Using pandas & python, parsed to a data frame and stripped for whitespace, with data sorted by trial and L*(lightness), a*(red/green), and b*(yellow/blue) values.