The Monk Skin Tone Scale: Cross-Industry Convergence on a Standard for Skin Tone Calibration

By Ellis P. Monk, Jr.

Earl Kage, former leader of the Color Photo Studio at Kodak Park and manager of Kodak Research studios (1960s-1970s), recounted a problem: two of Kodak's biggest professional accounts were not happy with Kodak. One, a chocolate company, complained that they were not getting the right shades of brown for their chocolates—photographs could not distinguish between dark and milk chocolate. The other, a furniture company, complained that stains and wood grains in their ads were not rendered faithfully enough. The problem was so acute that they struggled to distinguish between maple, oak, and dark woods—the photos were basically indistinguishable.[1]

What is striking is that *these* issues compelled Kodak to take action—not, at least according to Kage and others at Kodak, issues concerning how Kodak's products struggled to render all shades of human skin. For decades, Kodak used "Shirley cards," a reference card featuring a White woman, to calibrate and balance color: one woman stood in for everyone across the color continuum. Complaints about the consequences of this truncated approach to color calibration were rarely heard at Kodak in the form of unhappy parents remarking on how awful their Black children looked in graduation photos next to White children.[1] A similarly limited approach to skin tone calibration was also used in color television. As Taylor and Lent explain in the *SMPTE Journal* in 1978,[2] they used a "printed reproduction, with excellent color fidelity, of a live model's face," "BBC Test Card No. 61 (Flesh Tone Reference)," otherwise known as Cardboard Kate, a White woman with brunette hair. They took great care with the colorimetric characteristics of the card—its spectral reflectance, matched to measurements of the model's forehead and cheeks—and even with its surface finish. They also noticed the human range. The steep rise in reflectance around 600 nm, they wrote, "is typical of all flesh tones of whatever ethnic origin," a variation that "affects the luminance of the flesh tone far more than the chromaticity." Having seen the continuum, they classified it as a matter of brightness and built the reference on one face. Cardboard Kate was adopted on both sides of the pond, for both NTSC and PAL—in fact, it was used in over 30 countries around the world.[1]

In each case, the reference instrument used to calibrate color was truncated, limited, and lacked validation across the full spectrum of human skin tone, despite being used as a reference for human 'flesh' as a whole. This is the same issue the Macbeth ColorChecker (1976) suffered from, with its two skin patches, "light" and "dark." Furthermore, even in medicine, Thomas Fitzpatrick developed the Fitzpatrick "scale" in the mid-1970s to measure how "Caucasian skin" reacts to UV light for phototherapy for skin conditions. Though it was never designed or intended to be a "skin tone scale," the Fitzpatrick scale ended up becoming the default proxy for skin tone classification in clinical research, dermatology, pulse oximetry, and more recently, across myriad artificial intelligence & machine learning use-cases (e.g., biometrics, image sensors for still and motion photography, image search, etc.). Over a decade after its initial release, two more tones were added to the Fitzpatrick Scale, bringing its total to 6. Similarly, "multiracial" Shirley cards were developed in the mid-1990s to extend the tonal range of skin-tone calibration to each end of the color continuum. However, film was already giving way to digital, and biases migrated into sensor pipelines that inherited these same conventions.

The Monk Skin Tone Scale and the Regulatory Landscape

The Monk Skin Tone Scale (MST), a 10-point, perceptual skin tone classification system, sits in a permanent exhibit in the Sound and Vision Galleries of the National Science and Media Museum in the UK, beside a large reproduction of Cardboard Kate. The exhibit displays past biased references alongside the reference designed to supersede them by representing a much wider range of skin tones, spanning each extreme of the color continuum from light to dark. The exhibit explicitly names the MST "Beating Bias," and that is precisely what the scale has been adopted to do across a diverse range of industries seeking to mitigate bias in skin-tone calibration.

1545-0279/26©2026SMPTE



The Monk Skin Tone Scale (monkscale.com).

Google adopted the scale in 2022 as its skin tone standard and rebuilt its camera tuning, photo filters (i.e., RealTone), and image search engine around it, while also using it to "help AI see more skin tones." The MST was globally validated in a series of the largest validation studies ever run on a skin tone reference. The studies showed that the MST is not only as inclusive as a 40-point scale, but it is as easy to use as the 6-point Fitzpatrick Scale, with excellent inter-rater reliability under both optimal and suboptimal lighting conditions in annotation tasks using a global pool of annotators,[3,4]

Since its 2022 launch with Google, the MST has been adopted in medicine, biometrics, AI/ML fairness evaluation, and more. To the author's knowledge, the pace and breadth of the MST adoption are unprecedented. For decades, concerns have persisted about whether pulse oximeters work equally well across the skin-tone continuum. These concerns reached a fever pitch during the COVID-19 pandemic, when reliance on these devices, not only in hospitals but at home, heightened these long-standing concerns. The U.S. FDA was urged to tackle this issue and in 2025 released new guidance on pulse oximeter evaluation for skin tone, co-authored by the present author, which specifies the MST as its skin tone standard for characterizing study trial participant pools. The 3rd edition of ISO 80601-2-61, the international pulse oximeter standard, to which the author contributed the skin tone provisions, also specifies the MST as its skin tone standard. Recent studies already show the value of the MST: it was the strongest (and only) predictor of pulse oximeter skin tone bias in a study using seven other measures of skin tone and self-reported race/ethnicity.[5] Moreover, one of the first studies testing the new regulatory framework for pulse oximeter bias evaluation—34 oximeters, 155 patients, and 30,000 paired readings—found that ***only one device passed*** the new FDA guidelines.[6] A terrifying result, to be sure, but this is the power of a validated working reference—it shows what is broken.

Beyond pulse oximetry, the MST has also been adopted in clinical practice. This includes dermatology, where the Journal of the *American Academy of Dermatology* recommended the scale for dermatological research; the reduction of hospital-acquired pressure injuries (HAPIs), which are responsible for some 60,000 deaths annually and nearly $27 billion; and neonatal jaundice detection. A Gates Foundation-funded study has rolled out the Bili-ruler, a $9 device that uses the MST for skin-tone calibration. In Kenya, the device has been deployed to screen 930 infants, and there have been zero jaundice-related deaths since adopting the MST-powered Bili-ruler.

As a member of ISO/IEC JTC 1/SC 37 (Biometrics), the author is involved in revising their skin-tone standards. ISO/IEC 19795-10 already names the MST as a skin-tone reference: a fairness claim about face recognition, for example, is not meaningful unless one can point to what that claim rests on. The MST is also, already, the de facto skin tone standard in testing pedestrian safety in autonomous ve-

hicles (AV), where it is embedded in the CARLA simulator, which companies and governments around the world use to test autonomous vehicle safety—the question of whether skin tone impacts safety is expressed in MST units.[7] The author is currently involved in collaborative research to inform revisions to existing AV standards (e.g., IEEE P2020, National Highway Traffic Safety Administration (NHTSA) NCAP, and Euro NCAP).

The Monk Skin Tone Scale and Motion Imaging

While there certainly have been efforts to address the limitations of prior skin tone references such as "multiracial Shirley cards," adding tones to the Fitzpatrick Scale, and adding new patches to modern, digital ColorCheckers, the lack of standardization and, in many cases, validation has resulted in a haphazard and inadequate set of solutions to this industry-spanning problem. Without standardization and validation, every vendor can claim their product "handles skin well" according to their own chart. Coverage? Unknown. Failures? Potentially invisible (see earlier discussion). A pipeline whose components are tested against different, unvalidated "standards" has no real way of discovering if, how, and why certain skin tones may not render as well as others. References are tuning targets for sensors and image processors to be tested against—biased reference materials (e.g., scales and charts) engineer their biases into the very products they are intended to safeguard against.

The MST now exists as both a physical and a digital reference: Pantone has produced printed MST swatches, formulating new pigments to reach where existing print libraries fell short. Adoption in photographic test equipment is already underway—Imatest's *Diverse Skin Tone Face Targets*, a set of ten camera test charts calibrated to the spectral reflectance of human skin across the MST range, are in commercial use for camera and pipeline characterization. The accompanying research published at *Electronic Imaging* has used them to benchmark consumer cameras' skin tone accuracy and to argue for representing diverse skin tones in photographic test charts as a matter of image equity.[8] Accordingly, ISO/TC 42 describes the MST extensively in its new Technical Report on skin tone test charts for photography (ISO/DTR 26478), the first international standards document to set out what a skin tone reference must demonstrate for use in camera evaluation.

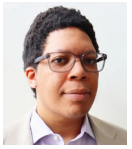
As in digital photography, where Google has adopted the MST for these purposes and its regulatory footprint in the ISO standard for photography is growing, the MST is positioned to play a large role in motion imaging, too. The list of potential applications is quite large so to name only a few: camera and sensor evaluation (e.g., skin rendition, color accuracy, noise, dynamic range headroom, detail, etc. reported at ten standardized and validated reference points), AI/ML in the pipeline (e.g., denoise, de-aging, relighting, auto-matching, etc.), display calibration, HDR and wide-gamut-tone and gamut-mapping, skin shaders, texture libraries, and coverage as metadata (e.g., mastering reports and camera tests stating the MST range across which skin tone performance was verified).

As Michael Johnston wrote in *The New Yorker*,[9] highlighting the MST as a critical intervention in photography, "getting the visual feeling of skin tone right respects a person's individuality." For far too long, the problem of rendering human skin in all its diversity was either ignored completely or inadequately addressed. Indeed, the cross-industry convergence on the MST as skin tone standard corroborates the concerns raised in the inaugural SMPTE "Power of Color" Symposium,[10] at which the author presented, and forcefully demonstrates that a viable means of correction is finally at hand.

References

1. L. Roth, "Looking at Shirley, the Ultimate Norm: Colour Balance, Image Technologies, and Cognitive Equity," *Can. J. Commun*. 34 (1): 111–136, 2009.
2. E. W. Taylor and S. J. Lent, "BBC Test Card No. 61 (Flesh Tone Reference): Colorimetric and Other Optical Considerations," *SMPTE J.* 87 (2), 1978. doi: 10.5594/J17432.
3. C. M. Heldreth, E. P. Monk, A. T. Clark, C. Schumann, X. Eyee, and S. Ricco, "Which Skin Tone Measures Are the Most Inclusive? An Investigation of Skin Tone Measures for Artificial Intelligence," *ACM J. Responsible Comput.* 1 (1): Article 7, 2024.
4. C. Schumann, F. O. Olanubi, A. Wright, E. Monk, C. Heldreth, and S. Ricco, "Consensus and Subjectivity of Skin Tone Annotation for ML Fairness," *Adv. Neural Inf. Process. Syst.* 36 (NeurIPS 2023), Datasets and Benchmarks Track, 2023.
5. S. Hao, K. Dempsey, J. Matos, C. E. Cox, V. Rotemberg, J. W. Gichoya, W. A. Kibbe, C. Hong, and A.-K. I. Wong, "Utility of Skin Tone on Pulse Oximetry in Critically Ill Patients: A Prospective Cohort Study," *Crit. Care Explor.* 6 (9): e1133, 2024. doi: 10.1097/CCE.0000000000001133.
6. C. Hughes, D. Chen, T. Law, P. Bickler, J. Feiner, L. Shmuylovich, E. Behnke, L. Ortiz, G. Leeb, R. V. Zamora, E. Monk, and M. S. Lipnick, "Pulse Oximeter Performance and Skin Pigment: Comparison of 34 Oximeters Using Current and Emerging Regulatory Frameworks," *Anesth. Analg.*, 2026.
7. B. Pathiraja, C. Liu, and R. Senanayake, "Fairness in Autonomous Driving: Towards Understanding Confounding Factors in Object Detection under Challenging Weather," *arXiv*, arXiv:2406.00219, 2024.
8. M. Borek, A. Limbocker, and E. Monk, "Validation of Skin Tone Test Charts with Real Human Data," *Electronic Imaging 2026*, Image Quality and System Performance (IQSP-247), Society for Imaging Science and Technology, 2026.
9. M. Johnston, "The Secret Art of the Family Photo," *New Yorker*, 2022. [Online]. Available: https://www.newyorker.com/culture/annals-of-inquiry/family-photos
10. E. P. Monk, "The Unceasing Significance of Colorism: From the Social to the Sociotechnical," presented at the inaugural SMPTE Power of Color Symposium, Clark Atlanta University, Atlanta, GA, Feb. 2024.

About the Author



Ellis Monk is Professor of Sociology at Harvard University and creator of the Monk Skin Tone Scale, embedded in FDA & ISO standards.

A contribution received from the author for the 2026 Progress Report.

Copyright © 2026 by SMPTE.

DOI: 10.5594/JMI.2026/ABPC1073
Date of publication: 15 November 2026