AAPM REPORT NO. 270

Display Quality Assurance

The Report of AAPM
Task Group 270

January 2019

DISCLAIMER: This publication is based on sources and information believed to be reliable, but the AAPM, the authors, and the editors disclaim any warranty or liability based on or relating to the contents of this publication.

———————

The AAPM does not endorse any products, manufacturers, or suppliers. Nothing in this publication should be interpreted as implying such endorsement.

This page intentionally left blank

Display Quality Assurance

The Report of AAPM Task Group 270

Nicholas B. Bevins, Ph.D.* and Michael J. Flynn, Ph.D.*

Henry Ford Health System, Detroit, MI

Michael S. Silosky, M.S. and Rebecca M. Marsh, Ph.D.

University of Colorado Denver, Denver, CO

Alisa I. Walz-Flannigan, Ph.D.

Mayo Clinic, Rochester, MN

Aldo Badano, Ph.D.

U.S. Food and Drug Administration, Silver Spring, MD

---

* Task Group Co-chair

DISCLAIMER: This publication is based on sources and information believed to be reliable, but the AAPM, the authors, and the publisher disclaim any warranty or liability based on or relating to the contents of this publication.

The AAPM does not endorse any products, manufacturers, or suppliers. Nothing in this publication should be interpreted as implying such endorsement.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1   INTRODUCTION

1.1   Justification and Purpose

In 2005, AAPM published Online Report No. 03, "Assessment of Display Performance for Medical Imaging Systems," commonly referred to as the TG18 report. The stated intent of this report was "to provide standard guidelines to practicing medical physicists, engineers, researchers, and radiologists for the performance evaluation of electronic display devices intended for medical use." [TG18 2005, p. vii] While the report specifically states that it pays significant attention to both cathode ray tube (CRT) displays and liquid crystal displays (LCDs), CRTs were the dominant display technology at the time. As a result, the report focuses heavily on CRT displays, with many of the described quality assurance (QA) tests and performance criteria not directly applicable to LCDs. Since the report's publication, LCDs have supplanted CRT displays as the primary image display technology in medicine, and with the introduction of organic light-emitting diode (OLED) displays, it is appropriate to revisit the guidelines for the performance and QA evaluation of devices. This report is an update to TG18 that is specifically directed at LCD and OLED displays.

    The purpose of this task group report is to provide recommendations for the assessment of display quality for flat-panel displays used in medicine. This includes both LCDs and OLED displays used in the acquisition and review of medical images. The information provided in this report is intended to help design a QA program for flat-panel displays, as well as aid in purchasing decisions.

1.2   Report Structure

This report is broken into several sections to facilitate its use. The following section, entitled Performance Characteristics and Quality Control, provides an in-depth discussion of display classifications, performance characteristics, testing methods, and evaluation criteria. Each characteristic has a series of subsections addressing the rationale for testing, required equipment, test procedures, assessment of results, and suggested performance limits. It should be noted that several tests recommended in the TG18 report have been excluded from this document on the grounds that they either do not apply to flat-panel displays (LCD or OLED), or have been determined to be unnecessary for routine QA testing. Next, Summary of Tests and Recommendations provides a summary of tests and outlines recommendations for a routine QA program. After this summary, the Test Patterns section provides a more detailed description of the test patterns discussed throughout the document. Finally, sections providing a brief conclusion and a glossary are presented. The appendix of the document provides ImageJ [ImageJ 2016] macros for new test patterns introduced as part of this work.

2   PERFORMANCE CHARACTERISTICS AND QUALITY CONTROL

2.1   Display Categories

The TG18 report defined two display classes: primary displays were those used for interpretation of medical images, while secondary displays included all other displays used in medicine. This report further categorizes secondary displays based on their use, with varying requirements for performance and evaluation. In this report, all displays used in the review of medical images are separated into the following categories:

1. Diagnostic displays

2. Modality displays

3. Clinical specialist displays

4. Electronic health record (EHR) displays

The following subsections describe each of these categories in further detail. The ultimate goal is for all displays to maintain consistent presentation of images. This is accomplished by setting similar performance targets, while allowing different testing tolerances, testing mechanisms, and testing frequencies sufficient to ensure adequate performance for the intended clinical use. This report does not make special considerations for displays used in handheld devices. Many of the tests and categories described here can be directly applied to handheld viewing, but it is important to note that there are limitations with handheld viewing that are not discussed here. The reader should refer to the report from AAPM TG260 for a discussion of handheld image viewing considerations.

2.1.1   Diagnostic Displays

Referred to as primary displays in the TG18 report and other documents, diagnostic displays refer to displays used for the primary interpretation of medical images. While they are typically used in radiology, they may also be present in other departments and used by personnel in other medical specialties. Given the nature of the task for which they are used, diagnostic displays have the most stringent performance requirements. These displays often have substantially higher luminance, smaller pixel pitch, greater bit depth, improved noise characteristics, and built-in controls for long term luminance stability and self-testing. As a result of this increase in performance and capability, diagnostic displays are often substantially more expensive than standard business displays.

Dedicated diagnostic displays are typically used on workstations alongside non-diagnostic navigation displays. These navigation displays are used to access worklists, medical records, dictation software, e-mail, web browsers, etc. While it is usually possible to show medical images on these navigation displays, such practice should be avoided. If the navigation display must be used for medical image review, e.g., reviewing a color Doppler ultrasound study on a workstation with grayscale diagnostic displays, the display should meet the performance criteria for diagnostic displays.

2.1.2   Modality Displays

Modality displays refer to any display used during the acquisition and generation of medical images. These displays may or may not be located at the acquisition console for a given modality. Also included are displays used either partially or solely for image reconstruction and reprocessing, advanced image post-processing, drawing regions of interest, making screen captures, etc. Because the performance of these devices directly impacts image appearance on other displays, it is imperative that they meet minimum performance standards. It should be noted that displays used during image acquisition solely for the navigation of the modality interface, i.e., displays that do not display medical images, do not fall under this definition and are not addressed in this report.

Modality displays with direct feedback to the operator, e.g., displays for ultrasound or fluoroscopy, may be used for image acquisition, procedure guidance, or patient diagnosis. In these cases, the display performance criteria for diagnostic displays should be considered when evaluating the display. If diagnostic-quality displays are not available for the modality, the viewer should understand the limitations when making a medical diagnosis or patient care decision.

2.1.3   Clinical Specialist Displays

In certain clinical specialties, such as an emergency department, physicians often review patient images for the purpose of making healthcare management decisions before the interpreting radiologist provides a finalized report. Clinical specialist displays refer to displays used in this manner. Because images viewed on these displays may directly impact decisions, these displays should ideally perform similarly to diagnostic displays. However, given that these displays are generally of lower quality than diagnostic displays, the exact performance criteria and tolerances should be evaluated based on a display's specific usage and need for full diagnostic capabilities.

2.1.4   Electronic Health Record Displays

Images acquired during the course of a patient's care will often be viewed multiple times following the primary interpretation by a radiologist. This may occur in an exam room for patient education, at a separate workstation used by a referring physician reviewing images to better understand the diagnosis, or by a surgeon as part of pre-surgical planning. In these examples, the displays are categorized as Electronic Health Record (EHR) displays. A facility may find that standard business displays without specific calibration or luminance controls are sufficient for this need. However, when poor display performance creates barriers to efficiency or quality care, the QA guidelines in this report should be followed for luminance performance, grayscale calibration, uniformity, and ambient lighting. While optimizing all of these characteristics is needed for consistent image presentation, addressing any subset of these issues will improve image review.

2.1.5   Note on Display Categories

Manufacturers often categorize displays as medical-grade diagnostic displays, professional graphic displays, or standard business displays. The display category classifications in this report closely match the category names used by various display manufacturers. It is important to note that the decision on what type of display to use for a given image review

task should be based on both the clinical needs and the level of support. Displays that are marketed in one category may be sufficient for use in another category, especially if there is sufficient clinical support to set up and maintain the display. Any decision on the type of display should be made in consultation between the clinical user (e.g., physician), qualified medical physicist, and other relevant personnel.

2.1.6 Note on OLED Displays

At the time of writing this report, LCD flat-panels are the dominate technology across all display categories. As OLED technology continues to improve, it may someday supplant LCDs as the preferred display type. However, OLED displays currently suffer from certain limitations that limit their widespread implementation in medical image viewing. These limitations include differential aging of the color emitting materials, burn-in image residuals, luminance degradation with time, and luminance loading [IDMS 2012, p. 71]. The quality assurance methods recommended in this report may not fully characterize image quality in all these instances, and should be adapted as needed to accommodate future display technology and features.

2.2 Display Quality Assurance Tools and Techniques

As with other QA programs, certain tools and techniques are necessary to accurately evaluate and document the performance of medical displays. The following sections provide an introduction to these tools and techniques, and discuss usage, calibration, and other factors that may affect the accuracy and reliability of QA measurements.

2.2.1 Measurement Instrumentation

Quantitative evaluation of display performance requires the use of one or more measurement instruments, including photometers and/or colorimeters. The word "photometer" is used here as a general term for any device used to measure light with standard photometric units for luminance and illuminance. Luminance refers to the intensity of light being emitted from a surface and is measured in cd/m^2. Illuminance refers to the intensity of light incident on a surface and is measured in lux. Measurements of display luminance are used to evaluate the displays performance, while measurements of illuminance are used to characterize the environment in which the display is being used. Typically, illuminance meters are placed on the surface of the display with normal room-lighting conditions. The word "colorimeter" is used here as a term for photometers that additionally estimate color in standard CIE coordinates [CIE15:2004] using a small array of light sensors having different color filters. In addition to colorimeters, spectroradiometers may be used to measure the spectral characteristics of light emitted by a display. These devices are capable of reporting the light measurements as a function of wavelength. While expensive, they are useful for making highly accurate measurements of color coordinates and luminance [CIE15:2004; IDMS 2012]. Photometers and colorimeters are appropriate for use in clinical medical physics, although users should be aware of the potential error in the measurements compared with a spectroradiometer.

Quantitative evaluation of luminance is an essential part of display QA. Evaluating display luminance requires a basic knowledge of the types of luminance meters available and how they are used. Generally speaking, the meters used for display QA may be broken into two types: contact and telescopic.

Contact luminance meters are positioned with the sensitive portion or entry window of the detector in direct contact with the display. It should be noted that the direct contact between the meter and display excludes ambient light that would otherwise be reflected from the display toward the user. Therefore, the effects of ambient lighting must be specifically accounted for when using contact photometers to evaluate display performance. When operating these devices it is important to recognize that the detector has a finite measurement area, and that the measurement boundaries are not often visible to the user. Therefore, test patterns should be displayed such that the area being measured is uniform and falls well outside the sensitive area of the device. The accuracy of a measurement may be substantially affected if the sensitive portion of the detector falls outside of the intended portion of the test pattern.

In contrast, telescopic photometers are used away from the surface of the display (typically near where the eye of the user would be). Many telescopic photometers have eyepieces to allow the user to view the area to be measured. Because these measurements are made at a distance, ambient light reflected from the display will be directly included in the measurement. Unlike contact meters, the viewing direction and position of telescopic photometers are variable. If the user wants to characterize the viewing angle dependencies of a display, a telescopic photometer must be used. When either comparing displays or tracking the performance of a display over time, a telescopic photometer must be set up the same way for each measurement. A tripod can be used to reduce movement of the meter, and may help ensure robustness in telescopic photometer measurements. It should be noted that this concern is minimized with contact luminance meters since their positioning is easily reproducible.

An appropriate colorimeter is required for quantitative evaluation of display color. These devices measure the color of the display and may display results using several different color space coordinate systems. As with luminance meters, there are both contact and telescopic colorimeters. These devices are operated similarly, and are subject to similar limitations as their photometer counterparts. Colorimetery capabilities are often incorporated into photometers, resulting in a single device capable of both functions. Colorimeters can vary greatly in both cost and accuracy, and the reader should refer to the AAPM TG196 report on gray tracking in medical color displays when considering what type of colorimeter to use [TG196 2016]. As with all measurement devices, photometers and colorimeters should be used in accordance with the manufacturer's recommendations and guidelines.

While either contact or telescopic photometers and colorimeters may be used to evaluate luminance and color for all classes of displays, modern diagnostic displays are often equipped with built-in photometers and colorimeters for both calibration and verification of display output. Depending on the display design, these devices can either monitor the output of the backlight directly, or measure luminance emitted from the front panel. The output from built-in meters is typically accessible only through vendor-specific software. These devices can be useful for monitoring display performance for a large number of displays, especially when the output can be viewed remotely. It should be noted that neither design of built-in meter typically includes ambient light in its readout. To account for this, some vendors assume a fixed value for ambient luminance during calibration. As with all tools used for QA, the proper functioning of built-in photometers and colorimeters should be verified before use.

In addition to built-in photometers used to measure display luminance, some manufacturers install illuminance meters designed to measure ambient light. These measurements are often used in conjunction with energy-saving features that reduce the maximum lumi-

nance when the room is dark. Since maintaining a consistent luminance ratio and grayscale display function is essential to reproducible image display, this report recommends disabling energy-saving features that adjust the luminance characteristics.

2.2.2 Calibration and Intercomparison

Similar to dosimetry equipment and other devices used in the field of medical physics, calibration of instrumentation used for display QA helps ensure that measurements remain accurate over time. Meters that can be calibrated over the range of luminances typically measured from a medical display (0.1 to 1000 cd/m^2) are ideal. There are a number of meters available on the market, with a substantial range in prices. Some of these devices will come from the manufacturer with calibration documentation and an expectation of annual calibration, while others are not designed to be calibrated and cannot be corrected for changes over time. While calibrated devices should be the first choice, especially for facilities with only a single luminance meter, devices that cannot be calibrated may be appropriate if they are periodically compared against a recently calibrated meter. All meters that can be calibrated should follow the manufacturer's guidelines for calibration.

It has been shown that built-in photometer accuracy tends to degrade as a display approaches its backlight warranty [Silosky and Marsh 2013]. For front-panel sensors, this may be due to changes in display uniformity or white point, while for backlight sensors it may be due to changes to the optical properties of the display. An intercomparison with a calibrated luminance meter may be used to verify accuracy. This report suggests performing an intercomparison during acceptance of a new display, and again every 10,000 backlight hours. Different display models (with different built-in photometers) may age in different ways, so the person responsible for display QA should use their professional judgment regarding the frequency of intercomparison for these devices.

2.2.3 Display and Device Stabilization

When performing quantitative measurements on medical displays it is important to ensure that both the display itself and the measurement device have reached a steady state. Displays and measurement devices should be turned on for at least several minutes before taking measurements, and test patterns should be displayed long enough for the display output to stabilize. Further discussion of display luminance stability is provided in the Display Temporal Performance section. To ensure thermal stability, contact photometers and colorimeters should be put in place during the warm-up period. To verify stability, several serial recordings should be made to ensure that there is no trending in luminance or color measurements. If the measurement device has a constantly updating readout, the user should ensure that a steady state has been reached prior to recording the measurement.

2.2.4 Environmental Conditions

While measurements made using contact devices (both photometers and colorimeters) are relatively independent of environmental lighting conditions, measurements made using telescopic devices include both emitted and reflected light. When using telescopic devices, it is important that measurements are made using normal room-lighting conditions. The illuminance of the display surface can be affected by several factors including the display's position and orientation relative to ambient light sources. These sources may include lamps,

other displays, and even the clothing of the operator (white coats or shirts reflect ambient light toward the display, and strong colors may affect white point).

In addition to quantitative measurements, visual evaluation of test patterns can be strongly influenced by environmental conditions. As with measurement devices, the results of visual evaluations can be affected by ambient light reflected toward the user, display orientation and angle, and the clothing of the individual using the display. This is further complicated by the fact that the visual system of the viewer will adapt, not only to light emitted and reflected from the display surface, but also from other sources. Consequently, the effects of ambient lighting may have a greater effect on qualitative visual evaluations than for quantitative measurements.

2.2.5 Workstation Configuration and Documentation

Display acceptance and QA tests are often done for a workstation that may have multiple displays. The make, model, and serial number of all displays on the workstation, along with all display-related graphic configuration settings, should be documented and compared to information obtained at installation. It is not uncommon to find that a display has been moved from one workstation to another, or that the graphic configurations have changed. It is particularly important to verify that the row and column settings of the graphic driver match the native pixel rows and column of the display.

Workstation documentation can be easily done using an application that automatically retrieves the needed information. Modern displays provide internal storage for such specification information. The VESA Extended Display Identification Data (EDID) data structure standardizes the format and can be read by operating system drivers. The VESA DisplayID standard was recently released and is designed to eventually replace the EDID standard. Table I illustrates the table produced by the EDIDprofile utility application included with the open-source pacsDisplay package (pacsdisplay.org). The utility documents the workstation profile and display specifications for each active display on a Microsoft Windows workstation. Display manufacturers provide similar applications along with database services that are useful for tracking the performance of individual displays.

2.2.6 Use of Test Patterns

Test patterns are a vital part of display quality assurance, and the tests and procedures in this document rely on the proper display of appropriate test patterns. Various test patterns are mentioned throughout the remaining sections, and the reader is referred to Test Patterns in section 4 for a complete description of each pattern. It should be noted that when displaying any test pattern, the appropriate dynamic range of the display should be used. Typically, a display should be set to its full range (8-bit display set to window width 256, window level 128). While many vendors include test patterns with appropriate presets, some may require user adjustment.

2.3 Display Luminance

Display luminance refers to the intensity of visible light emitted from a unit area of the surface of the display. The term "luminance response function" refers to the changes in a display's luminance as image gray levels vary. Because an observer's perception of image

TABLE I. Example Display Configuration from EDID

System Profile

Hostname: hfh-ct-2-r6gy1
Date: 20150903

Manf: Hewlett-Packard
Model: HPZ640Workstation

CPU_0: Intel(R)Xeon(R)CPUE5-2667v3@3.20GHz
CPU_1: Intel(R)Xeon(R)CPUE5-2667v3@3.20GHz
CPU_2: Intel(R)Xeon(R)CPUE5-2667v3@3.20GHz
CPU_3: Intel(R)Xeon(R)CPUE5-2667v3@3.20GHz

Display Profile

getEdid request number	1	2	5	6
Adapter display ID	DISPLAY1	DISPLAY2	DISPLAY5	DISPLAY6
Adapter string	K2200	K2200	K2200	K2200
Display descriptor	MD211C3	MD211C3	EA234WMi	EA234WMi
Extended S/N	56003151TW	55003029TW	56313201TA	56313204TA
Week of manufacture	23	19	23	23
Year of manufacture	2015	2015	2015	2015
Max. horiz. image size (mm)	325	325	509	509
Max. vert. image size (mm)	433	433	286	286
Horiz. array size: Native	1536	1536	1920	1920
Horiz. array size: Current	1536	1536	1920	1920
Vert. array size: Native	2048	2048	1080	1080
Vert. array size: Current	2048	2048	1080	1080
Est. horiz. pixel size (um)	211.6	211.6	265.1	265.1
Est. vert. pixel size (um)	211.4	211.4	264.8	264.8

brightness and contrast of an image is determined in part by the luminance response function, quality control of both the display luminance and the luminance response function are an essential component of display quality assurance. Several characteristics are used when evaluating display luminance including ambient luminance (L_{amb}), minimum luminance (L_{min}), maximum luminance (L_{max}), luminance ratio, and luminance response function. The ACR–AAPM–SIIM Technical Standard for Electronic Practice of Medical Imaging [ACR-AAPM-SIIM 2017] has provided basic recommendations regarding these characteristics for different display classifications.

Historically, the maximum luminance, L_{max}, has been given the most attention as a display quality control metric. However, optimal display performance is best achieved by first considering the ambient lighting conditions. Once the ambient lighting conditions

are characterized, L_{min} and L_{max} may be set and evaluated before being used to determine appropriate luminance response at various image gray levels. The following section describes the rationale for this process, along with suggested methods for measurement and evaluation.

2.3.1   Ambient Lighting

2.3.1.1   Rationale

This report uses the term "ambient light" to refer to sources of visible light, other than the display itself, that reflect off the surface of the display. The reflected ambient light is superimposed on the user's view of the intended image, reducing the image contrast and interfering with the visibility of the displayed image. In addition, glossy displays can cause a specular (mirror-like) reflection of a light source to create a distinct image of the light source on the display, which may be distracting to the user. Both of these effects from ambient light impact how the displayed image is perceived.

The ambient luminance L_{amb} reflected from the display will depend on the characteristics of the illumination produced by the ambient light sources. The ambient luminance can be measured directly, but will be dependent on the illumination levels. As an alternative, one can determine how the display will reflect different sources of illuminance, and use that information to calculate the ambient luminance for any illumination level.

Light sources can be generally categorized as producing either hemispherical diffuse illumination or directed illumination. Diffuse light is typically present due to the background light from the walls, ceiling, and floor. Directed light comes from discrete light sources (e.g., lamps, windows, other displays). Indoor light will generally be a combination of multiple directed light sources and the hemispherical diffuse background light. Light from these sources scatter off the surface of the display at various angles. Light reflected from the surface of the display may be broadly characterized as either specular reflection or diffuse reflection.

It should be noted that different display panels may have very different reflection properties, and it is helpful to understand these factors for both purchase evaluation and in establishing settings for use. Additionally, users should be aware that additional protective panels modify the reflective properties of the display and may create internal reflections that degrade displayed contrast resolution [Ekpo and McEntee 2016].

Specular Reflection

Specular reflection is defined as light reflected from a surface at a definite angle. Smoother surfaces generally produce more specular reflection, giving them a reflective property like a mirror. Specular reflectance results in light emitters or reflectors forming virtual images on a surface. For example, when a user is sitting with a lamp to his/her back and viewing a computer display, the image of the lamp may be superimposed on the displayed image. The clarity of the reflected lamp will depend on the amount of specular reflectance of the display. Specular reflections are typically non-uniform and there is no display calibration to correct them. The management strategy for specular reflections is to minimize them as much as possible by

- Avoiding the use of glossy panels with high specular reflectance, especially for use in environments with high levels of ambient light;

- Using indirect lighting, or positioning displays to face away from any direct lighting sources (e.g., overhead lights and windows); and

- Minimizing reflective objects in front of the display (e.g., white lab coats) which create non-uniform reflected light sources.

Diffuse Reflection

Unlike specular reflection, where the image of the light emitter/reflector may be directly visible on the display, diffuse reflection causes incident light to be scattered in many directions. This creates a more uniform increase in reflected light, as opposed to a clear image of the light source. The net result is that a portion of the illuminance incident on the display is reflected toward the viewer as an increase in ambient luminance, L_{amb}. It should be noted that even with proper indirect lighting, there will still be some diffuse reflection from the display. Because the viewer's visual system perceives both L_{amb} and the display luminance, perceived image contrast is degraded if L_{amb} is not considered when establishing an appropriate luminance response function.

Ambient Luminance

Under all viewing conditions, the total luminance L' observed by a viewer can be expressed as

$$L' = L + L_{amb} \tag{2.1}$$

where L is the luminance emitted by the display, and L_{amb} includes the luminance from reflected room light which can have multiple components:

$$L_{amb} = L_{diff} + L_{dir1} + L_{dir2} + \dots, \tag{2.2}$$

where L_{diff} is the reflected luminance from hemispherical diffuse light, and L_{dir1} and L_{dir2} are the luminance contributions from various discrete light sources. It should be noted that specular reflection can occur from both diffuse and discrete light sources. In recent years, methods have been developed to characterize display reflectance and L_{amb} resulting from the specular and diffuse components of multiple light sources [International Committee for Display Metrology (ICDM) 2012]. Each of these sources has a specific illuminance or luminance, and a corresponding reflection coefficient that describes how the display scatters that light.

Historically, the difficulty in characterizing and correcting for the effects of direct light sources in a clinical environment lead to minimizing the amount of direct lighting, making diffuse sources the primary component of L_{amb}. Without specular reflections from directed light sources, L_{amb} can be estimated by measuring the illuminance from diffuse light sources. Traditionally, medical physicists have relied upon this simplified formalism for considering the effects of L_{amb}. This includes direct measurement of L_{amb} under normal usage conditions, or its estimation based on measurement of ambient illuminance E and the display's coefficient of diffuse reflection R_d [Chawla and Samei 2007; TG18 2005]. Measuring the coefficient of diffuse reflection is beyond the scope of this report, and the reader is referred to the TG18 report for guidance on estimating R_d.

2.3.1.2 Test Equipment

Ambient Luminance

A calibrated telescopic photometer is necessary to directly measure L_{amb}. Such a device can also be used to measure the reflection coefficients, typically using a calibrated diffuse reflectance standard [IDMS 2012]. Contact photometers are inappropriate for evaluating ambient luminance because their direct contact with the display will reduce or eliminate the visible light incident on the surface.

Ambient Illuminance

L_{amb} can be estimated from measurements of the illuminance E incident on the display. E can be measured directly by placing an illuminance meter on the display surface under normal usage conditions. For consistency, the meter should be aligned at the display center position in a way that does not require the operator to hold it (e.g., suspended via cable or mounted on a tripod). Illuminance may also be measured by using a diffuse reflectance standard and a telescopic photometer, but a description of this method is beyond the scope of this report [IDMS 2012].

2.3.1.3 Procedure

Directly Measure Ambient Luminance

To measure L_{amb} directly, the display to be tested should be turned off or otherwise put in a mode where any light leaving the display surface is due solely to reflection, and not the display backlight. For this measurement to be accurate, the room lighting should be maintained as it would be for clinical use (preferably avoiding any direct light sources). For the measurement, a telescopic photometer is positioned at the typical viewing direction to the display surface. Once the reading from the photometer has stabilized, the ambient luminance L_{amb} is recorded.

Determine Ambient Luminance from Ambient Illuminance

To measure E, an illuminance meter should be positioned in the center of the display with the room lighting maintained as it would be for clinical use. It should be noted that reflective clothing such as lab coats and other light sources directed at the display may significantly affect measurements of E. Once the meter has stabilized, the ambient illuminance is recorded. To estimate L_{amb}, multiply the measured ambient illuminance E by the diffuse reflection coefficient R_d:

$$L_{amb} = E \cdot R_d. \tag{2.3}$$

Note on the Diffuse Reflection Coefficient (R_d)

As mentioned above, a discussion of R_d measurements is beyond the scope of this work. If the value is unavailable from the manufacturer, and is not readily measurable, a typical value of 0.005 cd/m^2/lux may be assumed [Chawla and Samei 2007]. Typical values of R_d range from 0.002 to 0.010 cd/m^2/lux, though higher values are possible with glossy displays or protective panels. Users should be aware that the assumed R_d value greatly impacts calculated system characterization values. Table II illustrates this point by showing how

given the same reading environment, calculated display performance metrics can change considerably just by assuming a different R_d value.

TABLE II. Estimated Reflection Coefficient Comparison

	Case 1	Case 2
Ambient Illuminance (E)	40 lux	40 lux
Reflection Coefficient (R_d)	0.005 cd/m²/lux	0.010 cd/m²/lux
Ambient Luminance (L_{amb})	0.2 cd/m²	0.4 cd/m²
Minimum Luminance (L'_{min})	1.0 cd/m²	2.0 cd/m²
Maximum Luminance (L'_{max})	350 cd/m²	700 cd/m²

A relatively small absolute change in actual or estimated reflection coefficients results in a large change in required display performance. In Case 1, the recommended values are achievable with business-grade displays. However, in Case 2, only high-end diagnostic displays will be capable of achieving and maintaining the suggested level of maximum luminance. Careful consideration of display selection and reading environment can have a large impact on overall performance and evaluation.

Comment on R_d Formalism

As stated, the TG18 report provides a model for estimating L_{amb} as the simple product of the room illuminance E and R_d. While this methodology is relatively simple to implement, and many display manufacturers will provide R_d values for their display models, it should be recognized that modern developments in reflectance metrology rely on a different approach. Under this more modern formalism, $L_{amb} = E \cdot \rho_{si}/\pi$ where ρ_{si} is the hemispherical diffuse reflectance in the absence of direct light sources. A full description of the methodology can be found in the "Information Display Measurement Standard" published by the International Committee for Display Metrology [IDMS 2012]. The methodology suggested in the TG270 report (based on the TG18 method) provides for a simplification of the IDMS technique, and should not result in a meaningful difference in measurement values. Furthermore, the TG270 methodology is consistent with other standards for display evaluation in medical settings [IEC 62563-1 2016].

2.3.1.4 Assessment

The primary reason for determining L_{amb} is because it affects how a display's luminance response function should be calibrated and evaluated. Ultimately, the goal of ensuring that contrast is maintained across all gray levels must include accounting for L_{amb}, particularly at low luminance levels, where the effects of ambient light will be most apparent. Incorporating L_{amb} into the evaluation of the luminance response is discussed in the Display Luminance and Grayscale Function section.

In addition to incorporating L_{amb} into the evaluation of the luminance response, excessive L_{amb} values should be identified by comparing L_{amb} to L_{min}. The ambience ratio AR is defined as L_{amb}/L_{min}, and, in general, should be less than $1/4$. The AR limit of $1/4$ ensures that in darker regions of the image, at least 80% of the contrast that is observed in total darkness will be visible under ambient illumination. Deviating from this recommendation will degrade

contrast at the lower end of the luminance response function, unless L_{amb} is both relatively constant and explicitly considered in the luminance response function calibration. The AR of ¹/₄ recommendation is easily achievable on most modern flat-panel displays, where the diffuse reflection coefficient is fairly low. However, in situations where the ambient lighting is not controllable, or is very bright (e.g., ER areas, surgical suites), the user must be aware of the potential for image viewing degradation in the darker regions of the image.

It should be noted that eliminating all sources of ambient light, i.e., making the area as dark as possible, is typically not practical in a clinical setting. Newer research also suggests that the traditional goal of keeping a reading room as dark as possible is not ideal, and is unnecessary with modern flat-panel displays. Minimizing the need for visual adaptation may reduce visual strain, thus making it better to design viewing environments in which there is a consistent average perceived luminance over the visual field, from paper on a desk, to a worklist on a navigation display, to the medical image on a diagnostic display. For the reading room environment, it has been suggested that this can be achieved with ambient light conditions of 25–75 lux [Brennan et al. 2007; Chawla and Samei 2007; Pollard et al. 2012]. Whether set to this suggested range, determined by personal preference, or limited by clinical environmental conditions, understanding how to consider ambient lighting and display settings is necessary to ensure that visual information is not lost.

2.3.1.5 Suggested Limits

Diagnostic Displays

The suggested limit for the ambience ratio AR (L_{amb}/L_{min}) is ¹/₄. This is generally achievable in reading environments with ambient illuminance of 25–75 lux, where the increase over absolute darkness (0 lux) will reduce visual strain [Brennan et al. 2007; Chawla and Samei 2007; Pollard et al. 2012].

Modality, Clinical Specialist, and EHR Displays

For modality, clinical specialist, and EHR displays, this report recommends an AR of ¹/₄. Clinical environments outside of the reading room, such as fluoroscopy suites, may require higher ambient lighting conditions for the clinical task. In these brighter environments, maintaining the AR at or below the recommended ¹/₄ may be difficult. Viewing images in environments where the AR exceeds ¹/₄ may result in loss of contrast in darker regions of the image, with increasing loss for increasing ambient illumination. In these situations, it is important to consider factors that can help both mitigate the loss of visual information and provide sufficient and appropriate display for the clinical needs:

- Locally dim lights above and near displays, or temporarily dim lighting for viewing during image acquisition.

- Position or angle displays to reduce ambient reflection.

- Use high-luminance displays to allow increasing L_{min} and reducing AR, while also maintaining luminance ratio.

- Use displays with low diffuse reflection coefficients to reduce AR by reducing L_{amb}.

- Account for L_{amb} in the calibration of the display's luminance response function.

2.3.1.6 Qualitative Assessment

For more routine testing of displays (or if quantitative assessment is not possible), qualitative assessment should be considered. Many display test patterns include regions that contain low-contrast features in the darkest gray levels (e.g., TG270-sQC 0 square or SMPTE 0%/5% square). These features should be visible on a properly calibrated display under typical ambient lighting. To evaluate the appropriateness of the ambient lighting, one should first verify that all low-contrast features can be visualized when there is minimal or no ambient lighting (e.g., with all room lights turned off and windows covered). Once the test pattern visibility in near or complete darkness is confirmed, the ambient lighting should be set to level normal for that area. If the low-contrast features can no longer be visualized, it is likely the display's configuration is not appropriate under the normal ambient lighting.

In addition to the aforementioned patterns, the TG18 report defined the TG18-AD pattern for evaluating low-luminance, low-contrast display performance. The pattern contains 49 regions with modulating bar patterns with pixel values between black (gray level 0) and increasing gray levels (gray levels 1 through 49 on an 8-bit display). Physicists can use this pattern to evaluate the effects of ambient lighting on the ability to visualize features of the pattern under different lighting conditions for a given display. While a pattern such as the TG270-sQC pattern is useful for a pass/fail criteria for ambient lighting conditions, the TG18-AD pattern is useful both to provide a more-detailed qualitative analysis of the conditions and to give context to failures. This report does not make any specific pass/-fail recommendations, but recommends the user becomes familiar with this pattern when determining the clinical impact of potential display issues.

2.3.2 Display Luminance and Grayscale Function

2.3.2.1 Rationale

A display's luminance performance is generally considered its defining characteristic. As mentioned in a previous section, the maximum luminance of a display, L_{max}, is often given great consideration when describing the capabilities of a display. While L_{max} is indeed a critical characteristic, the luminance performance at all levels is of tremendous importance when considering how an image will appear to a user. In addition to L_{max}, the minimum luminance L_{min} and luminance response function must be evaluated as part of any display quality control program.

Minimum Luminance

L_{min} describes the luminance that a display will output when an image with the minimum pixel value is displayed. For calibrated display devices, the minimum luminance is typically on the order of 1.0 cd/m^2, though it may be as low 0.05 cd/m^2. For OLED display devices, the lowest possible minimum luminance may be closer to 0.00 cd/m^2. Given that the the recommended L_{amb} setting is $^1\!/_4$ of L_{min}, operating at levels close to absolute darkness (0.00 cd/m^2) is generally not desirable. At these levels, the effects of L_{amb} may obscure contrast present in the darkest regions of the image (Figure 1). The total combined minimum luminance of the display (both emitted and reflected) can be defined as $L'_{min} = L_{min} + L_{amb}$, which includes the effects of ambient luminance and the inherent luminance of the display. Evaluating this combined minimum luminance is important for maintaining an operating level that will reliably show contrast in the darkest regions of an image.
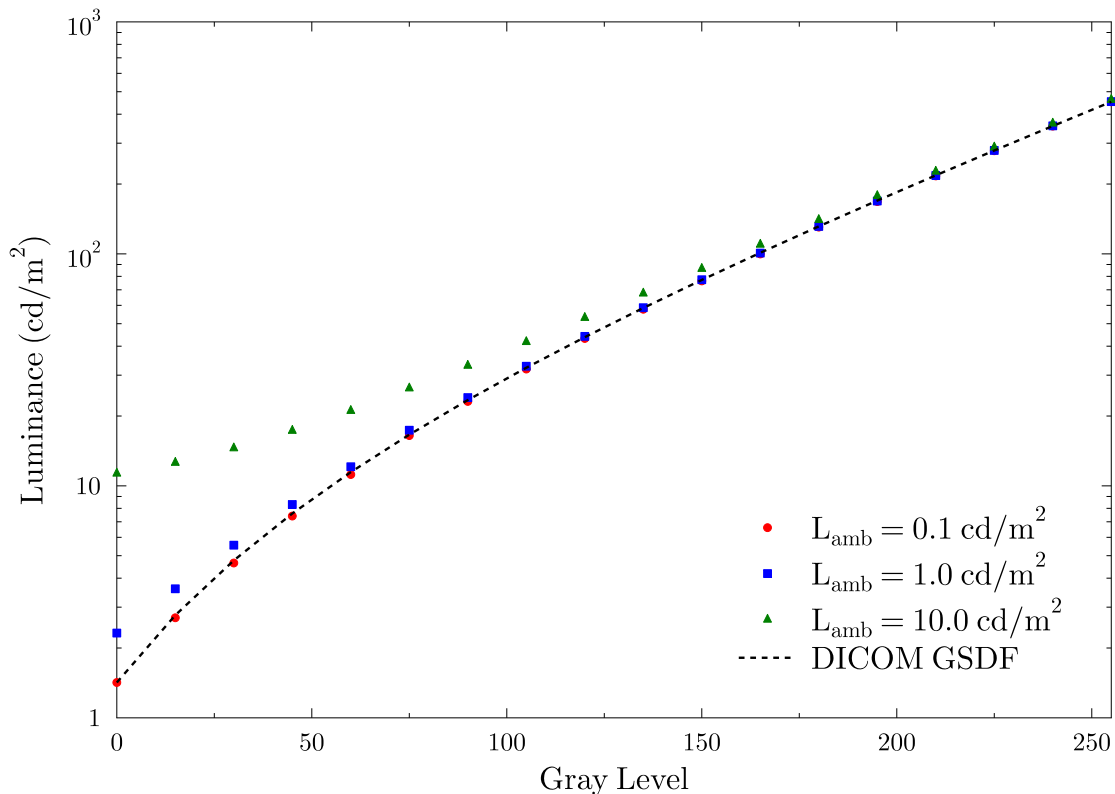
FIG. 1. Effect of different ambient lighting conditions on the perceived luminance. The effects of the ambient lighting are most significant at low luminance levels, where the constant ambient level is a much higher percentage of the total luminance reaching the observer. The display in this example was calibrated to the DICOM GSDF assuming $L_{amb} = 0.1$ cd/m^2.

As luminance decreases below 3 cd/m^2, the cones in the eye become increasingly unresponsive, while rods become increasingly responsive. This is referred to as mesopic vision. The foveal cones are responsible for many visual tasks, and their decreased response at very low luminance means that low-contrast objects can no longer be distinguished (Figure 2). In addition to minimizing the effect of L_{amb} obscuring contrast at the darkest operating levels of the display, setting L_{min} to a value higher than the absolute minimum of the display ensures the user's visual system avoids entering too far into mesopic vision.

Maximum Luminance

L_{max} describes the luminance that a display will output when an image with the maximum pixel value is displayed (255 on an 8-bit system). In the case of diagnostic displays, the operating L_{max} is typically less than the absolute maximum output of the display, which can be 1000 cd/m^2 or higher. The luminance capability is typically lower for modality, EHR, and clinical specialist displays, and L_{max} is often set to the maximum brightness possible on the display (generally closer to 250 cd/m^2). One major difference between diagnostic and most modality, clinical specialist, and EHR displays is that diagnostic displays will generally show the luminance settings in the on-screen display (OSD) menu or within the QC software with units of cd/m^2, while modality, clinical specialist, and EHR displays more commonly show

it in terms of percent brightness. This illustrates a major difference between these classes of displays and the difference between reporting an absolute measurement (units of cd/m^2) and a measurement relative to the maximum output of the display (units of percent brightness). While the luminance properties of all displays change with age and use, diagnostic displays often have built-in backlight sensors that automatically make adjustments to maintain a desired luminance. Modality, clinical specialist, and EHR displays, however, generally do not have this feature, resulting in different luminance measurements for the same "percent brightness" between displays of different ages. As a result of this change in output as the display ages, routine testing is necessary to ensure a display's performance is adequate.

Luminance Ratio

The ratio of maximum combined luminance, $L'_{max} = L_{max} + L_{amb}$, to the minimum combined luminance L'_{min} is defined as the luminance ratio (LR). LR describes how much "contrast" will be visible when viewing an image that maps across all gray levels. It is important that LR is large enough to ensure sufficient contrast across all gray levels but not so large that the range of the human visual system is exceeded [TG18 2005, Appendix II]. The luminance ratio should be similar for all displays to ensure similar image appearance [ACR-AAPM-SIIM 2017; TG18 2005]. It should be noted that the luminance ratio is not the same as the contrast ratio, which is often included in a display's technical specifications.

Luminance Response

The luminance response describes the points between L'_{min} and L'_{max}, and has a direct impact on the information available to the user. Proper calibration of a display's luminance response helps to ensure adequate contrast over a range of displayed gray levels. Additionally, the use of a luminance response function that is consistent between all displays in the image review process is a step toward providing similar image presentation, and therefore similar information, to all viewers. This report recommends the use of the DICOM Grayscale Standard Display Function (GSDF), which is described in Part 14 of the DICOM standard [DICOM Standard, Part 14]. The DICOM GSDF aims to provide similar visual contrast between adjacent gray levels over the full dynamic range of a display, and has been widely adopted for medical image viewing.

Despite its widespread use, users should understand the limitations of the DICOM GSDF. A primary difference between the DICOM GSDF theoretical viewing conditions and typical image viewing conditions is the adaptation state of the observer's visual system. The DICOM GSDF assumes varied visual adaptation, where the displayed image is almost completely uniform, with only a small target pattern to measure the contrast threshold. The image's background luminance is changed along with the target pattern and the observer adjusts to the new luminance state before determining the necessary contrast to detect the target. This process is continued (increase luminance, adapt to new state, measure contrast threshold) to create the contrast threshold vs. luminance plot in Figure 2 for varied visual adaptation. Typical viewing conditions for most users, however, operate under fixed adaptation, where the mean luminance of the image remains relatively fixed and images contain a mixture of bright and dark regions. Under these conditions, there is degraded contrast transfer as the luminance of the viewed region moves away from the mean luminance of the image [Baxter et al. 1981; Flynn et al. 1999; Tchou 2007]. Therefore, the DICOM GSDF is a good predictor for the contrast needed near the mean luminance of an image, but less so in the bright or dark regions (Fixed Visual Adaptation curve in Figure 2). This reduced contrast transfer

away from the mean luminance is further rationale for limiting the luminance ratio LR of a display. In addition to the visual adaptation model, the DICOM GSDF uses specific visual targets and other conditions, while the features typical to medical images are far more varied. The reason for using the DICOM GSDF is not because it is an optimal calibration for a given image. Instead, the DICOM GSDF provides a standardized function for display calibration that will aid in maintaining consistent image appearance on all calibrated systems.
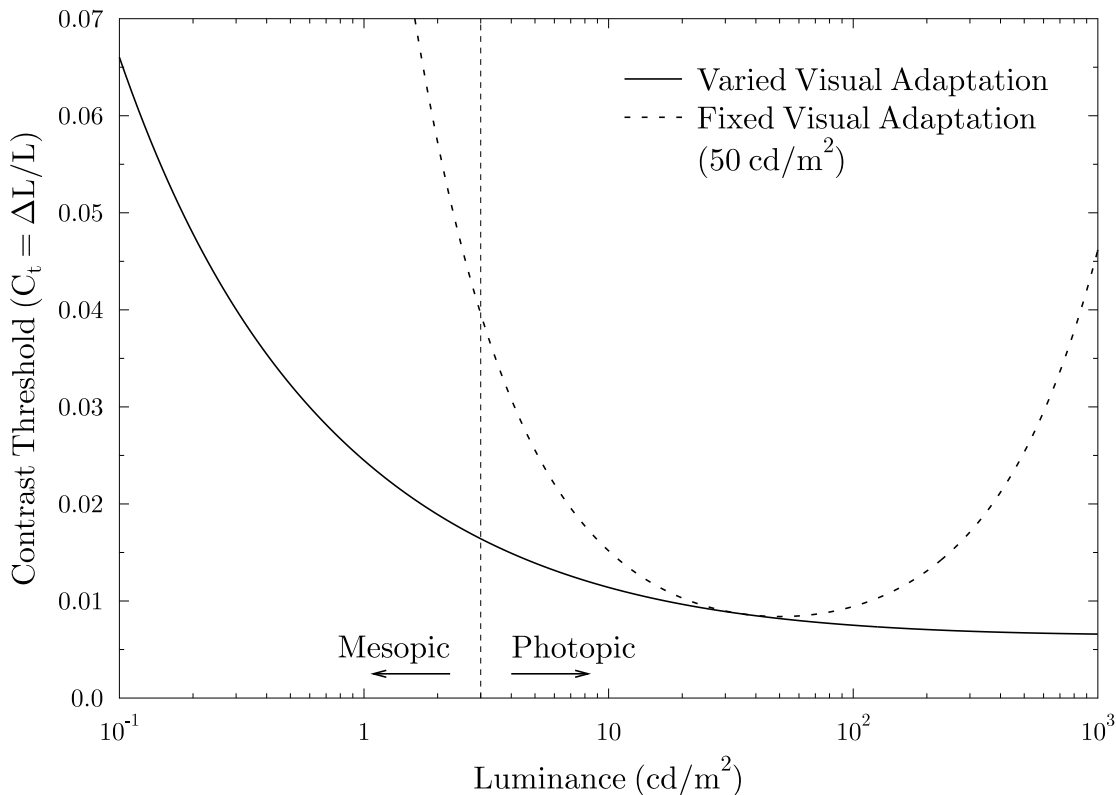
FIG. 2. Contrast threshold versus luminance for the standard observer for both varied and fixed visual adaptation. The contrast threshold is defined as the relative change in luminance required for a standard observer to perceive a difference. The DICOM GSDF uses a varied visual adaptation model, where the uniform luminance of the background changes for each contrast threshold measurement throughout the full luminance range. The varied visual adaptation curve can be adjusted by the biological contrast response to get a more realistic contrast threshold estimate for typical viewing conditions under fixed adaptation [Baxter et al. 1981; Flynn et al. 1999]. The fixed visual adaptation curve demonstrates that greater contrast is required to produce the same visual response as the luminance of a region within an image moves away from the mean luminance. Assuming a calibrated display has an appropriate luminance ratio LR (≈ 350), the DICOM GSDF is a good predictor for the contrast response of the human visual system. Within the mesopic region, the contrast threshold for both adaptation states rises dramatically, requiring a much larger relative change in luminance to produce the same visual response.

Measuring the luminance response of a display is essential for verifying that the palette of gray levels follows the DICOM GSDF. At the time of the TG18 report, CRT displays had a nonlinear luminance response function that was well-described by the relatively-smooth

power function. Consequently, the relationship between input voltage and luminance could be accurately characterized by sampling the luminance curve at various driving levels and then interpolating the other values. The data could then be described by a power law function with a characteristic "gamma" value. Following the TG18 methodology, the luminance response of an 8-bit system is measured at every 15 gray levels, resulting in a total of 18 measurements. In the case of LCD and OLED displays, a look-up table (LUT) maps each discrete driving level to a specific luminance, meaning that sampling only a subset of the luminance function may be insufficient to detect a significant deviation from a desired curve.

An example of the failure of an 18-point measurement to detect a deviation from the DICOM GSDF is presented in Figure 3. In this example, the display manufacturer software was used to calibrate the display to follow the DICOM GSDF, with specific minimum and maximum luminance values. After the calibration routine completed, independent testing software was used to evaluate all 256 gray levels. The results showed that while the levels measured by the manufacturer's software (including the standard 18 points) showed conformance to the DICOM GSDF, gray levels 1-14 did not. These results were forwarded to the display manufacturer, and a software patch was released to resolve the issue.

As a result of the potential for errors from assuming the luminance response function of an LCD or OLED display will follow a smooth, predictable function, the subsequent sections recommend additional test procedures beyond those described in the TG18 report. These testing procedures measure either every five gray levels (total of 52 measurements) or every gray level (total of 256 measurements). While only the 256-point measurement can fully ensure that there are no significant deviations from the desired luminance response, the 52-point measurement decreases the likelihood of a missed finding. In addition, if a visual inspection of a continuous grayscale gradient that spans the full luminance range is used in conjunction with the 52-point measurement, single outlying gray levels may be identified (see Figure 4). Given the widespread use of automated testing software for many displays, particularly diagnostic displays, the recommendation of increased measurements should not substantially increase total testing time from the older 18-point test.

2.3.2.2 Test Equipment

Measurement of luminance requires a photometer with appropriate read-out capabilities. Additionally, the display being tested must be able to display the appropriate test patterns as described in the individual test procedures below. Photometers vary in terms of their user interface. Some can be connected to and controlled by software loaded on the display workstation, while others work only with an external computer or other device. If the system being tested interfaces with the photometer, automated software from display manufacturers or third parties may be used to aid in the measurement and analysis of the luminance response.

Some displays are equipped with internal photometers that are used in conjunction with the vendor's QA software. These internal photometers may report luminance values that differ significantly (>10%) from those measured by a calibrated external photometer [Silosky and Marsh 2013]. Consequently, an intercomparison of the internal meter and external meter should be evaluated on acceptance testing and every 10,000 backlight hours. Failure of the display's internal meter to match an external meter could either cause the display's calibrated luminance response to exceed acceptable levels, or cause the display to fail adequately matching another display attached to the same workstation. Some vendors have included functionality that allows the user to recalibrate the internal meter based on an ex-
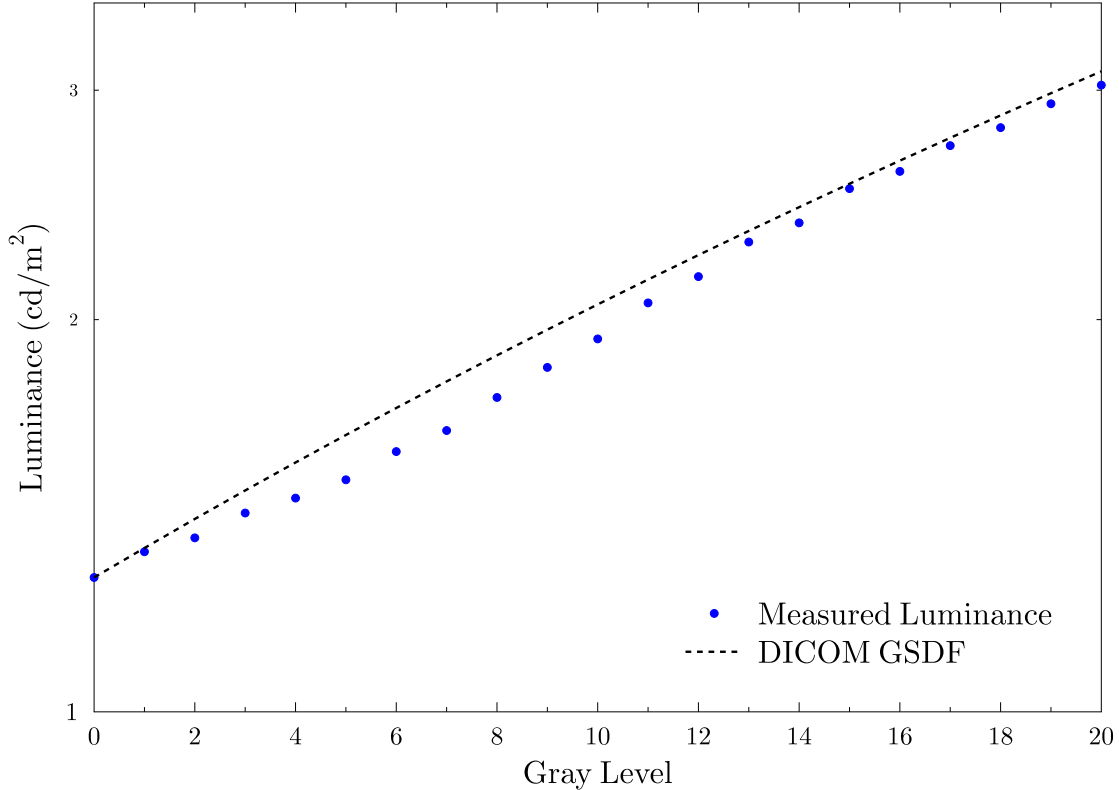
FIG. 3. Plot of a 256-point conformance measurement (focusing only on the first 20 points) following the manufacturer's calibration of a diagnostic display. The 256-point conformance measurement was performed using third-party software to verify the accuracy of a manufacturer's calibration software. The analysis revealed that the manufacturer's calibration failed to properly calibrate the display between gray levels 1-14. These points were missed by the manufacturer's own conformance check, which measured only every 15 gray levels (TG18 methodology). The average error across these points exceeded 20%.

ternal measurement. In these cases, the measurements of a calibrated external photometer may be used to recalibrate the internal meter. When the internal photometer cannot be recalibrated, the display's luminance response function should be evaluated using a calibrated external meter.

18-point Response (TG18 Method)

Characterization of the luminance response is often performed in steps of 15 gray levels (8-bit gray levels 0, 15, 30, ..., 240, 255), resulting in a total of 18 measurements. This number of measurements is easily performed by hand, and this test has gained widespread acceptance. However, for the reasons stated in the rationale section, coarse sampling of only 18 points may miss significant deviations from the expected luminance response function. As a result, finer sampling (52- or 256-point measurement) is suggested at acceptance testing or when using an automated measurement routine.

The TG18 report describes this 18-point test and suggests the use of the TG18-LN series of test patterns. The TG270-ULN series of patterns, an updated version of the TG18-LN
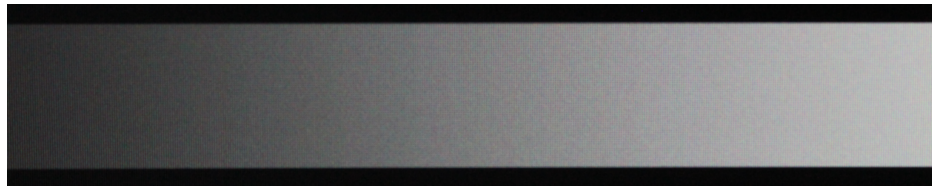
(a) No visual issues
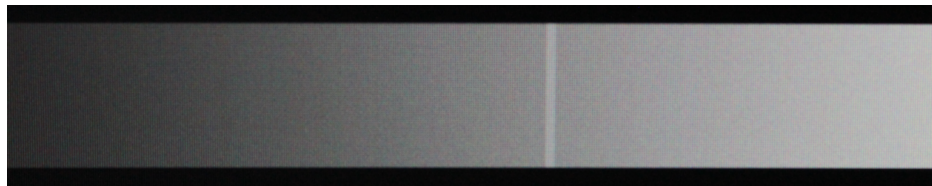
(b) Mis-calibrated gray level
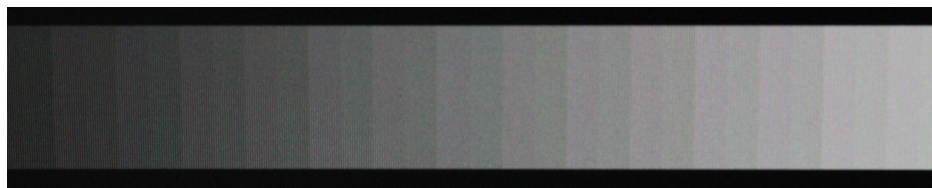
(c) Inappropriate bit-depth configuration

FIG. 4. Photographic examples of visible calibration and configuration issues. Visual inspection of a test pattern as part of a quantitative luminance assessment may be useful in identifying calibration errors that are not found as part of a quantitative measurement.

patterns, provides the ability to measure luminance over the entire display area in increments of 15 gray levels. In addition, the TG270-sQC and TG270-pQC test patterns include these same gray levels as visual steps for qualitative checks. If pan/zoom controls are available on the display, these qualitative patterns can be used to display the required luminance steps for an 18-point response measurement. The TG18-OIQ/TG18-QC pattern also has a series of 18 luminance patches, though their values do not increment in gray-level steps of 15. If the TG18-OIQ/TG18-QC pattern is the only one available, these steps could be used, but the analysis would need to be modified accordingly.

After deciding which test pattern or pattern series will be used, the luminance at each gray level should be measured and recorded. For each gray level step, the relevant portion of the pattern should be positioned in the center of the image viewing area (the area where clinical images appear on the display, also referred to as the useful area of the display). This may be different from the physical center of the display. Measurements should be made with the photometer directly over the center of the pattern. Once the reported luminance has stabilized for a given gray level, the measurement should be noted and the next pattern put in place. Often, longer integration time is required for darker levels compared with brighter levels. If there is concern over meter stability, multiple measurements should be made and averaged.

52-point Response

As described above, compared with the 18-point response, the 52-point response better characterizes the luminance response function. When coupled with a qualitative visual inspection of a continuous grayscale gradient that spans the full luminance range (found in the TG18-OIQ/TG18-QC, TG270-sQC, and TG270-pQC patterns), the user should have more confidence that a display's measured results accurately reflect its performance.

The 52-point response measurement can be performed by displaying gray levels in increments of five (8-bit gray levels 0, 5, 10, . . . , 250, 255). The TG270-ULN pattern series displays each of these levels, allowing manual measurement of each level. A process similar to that described in the 18-point response methodology should be followed, displaying each test pattern so that the photometer can be centered in the useful area of the display. If automated testing is available and the user can select the gray level interval, the total amount of time testing a display should not increase substantially from the 18-point technique. For this reason, the 52-point measurement is recommended whenever available with automated analysis. For measurements that are performed by hand, i.e., manually recording the luminance at each step, the qualified physicist should decide whether to use the 18- or 52-point method. Previous display performance and display class should be taken into consideration when deciding the appropriate level of testing.

256-point Response (Full Gray Palette)

The 256-point measurement will fully characterize the luminance response function. In general, this method is only necessary during equipment purchase evaluations and when troubleshooting errors found in either routine or acceptance testing. Because the 256-point response measurement requires the assessment of very small luminance changes, the result may be noisier than the 18- or 52-point methods. For this reason, interpretation of results from this analysis should be left to persons having low-noise measurement instruments.

To measure all 256 gray levels, test patterns such as the ones in the TG270-ULN series should be used. This series contains a uniform gray level pattern that fills the useful area of the display. Automated testing software from either the display manufacturer or an independent software provider is highly recommended due to the number of measurements that must be made.

11-point Response (SMPTE Method)

For displays attached to older systems with limited user access, the SMPTE pattern is often the only pattern available for quality assessment. If the pattern is large enough, or if pan/zoom controls are available, the user can measure the luminance response on the eleven increasing gray level squares in the central part of the phantom (0-100%, in increments of 10%). The previously described methods can be used to assess the luminance response, though with fewer points. If pan/zoom controls are available, the user should be sure to make the measurements within the central region of the image viewing area using the pan feature to move the various gray level boxes into place under the photometer.

2-point Response (L_{min} and L_{max})

In cases where the user is unable to accurately set or display intermediate gray levels or does not want to measure the luminance response, measurements of only the minimum and maximum luminance values should be made. These two measurements will still allow for

the evaluation of L'_{min}, L'_{max}, and LR values, but will not allow for an evaluation of DICOM GSDF conformance. In place of this quantitative measurement, qualitative evaluation of the display should be performed as described in the Qualitative Assessment section.

2.3.2.3 Assessment

L_{min}

The minimum luminance L_{min} is equal to the measured luminance at gray level 0. This value should be summed with the ambient luminance L_{amb} to determine the minimum combined luminance L'_{min}. Once L'_{min} is established, the maximum luminance of the display should be set such that the luminance ratio LR is maintained across all displays that are part of the image review chain. The suggested limits for minimum combined luminance are provided in Table VI.

LR, L_{max}

The maximum luminance L_{max} is equal to the measured luminance at gray level 255 (on an 8-bit system). This value is then added to the ambient luminance L_{amb} to determine the maximum combined luminance L'_{max}. This value should be set such that the luminance ratio, $LR = L'_{max}/L'_{min}$, is consistent across all displays part of the image review chain. The suggested limits for maximum combined luminance and luminance ratio are provided in Table VI.

Luminance Response

The measured luminance response should be evaluated by calculating the maximum deviation between the measurements and the DICOM GSDF. To do this, one must determine the DICOM GSDF response between the measured L'_{min} and L'_{max}. The first step in this process is to convert each measured luminance value to its corresponding DICOM GSDF JND index. A JND is a "just noticeable difference" in luminance, i.e., the difference in luminance required to be visibly noticed by an observer, under specific viewing conditions [Barten 1999]. The GSDF luminance response is linear in JND space, but it is nonlinear in luminance space. This transformation therefore provides a convenient means of comparing the measured values with the target luminance response. A comprehensive description of the theory behind and definition of the DICOM GSDF is beyond the scope of this work, and the reader is referred to the DICOM standard for a full explanation [DICOM Standard, Part 14].

The DICOM standard provides Equations 2.4 and 2.5 to convert between luminance L and JND index j:

$$\log_{10} \mathrm{L}\left(j\right) = \frac{a + c \cdot \ln\left(j\right) + e \cdot \ln^2\left(j\right) + g \cdot \ln^3(j) + m \cdot \ln^4\left(j\right)}{1 + b \cdot \ln\left(j\right) + d \cdot \ln^2\left(j\right) + f \cdot \ln^3(j) + h \cdot \ln^4\left(j\right) + k \cdot \ln^5\left(j\right)} \tag{2.4}$$

and

$$j\left(\mathrm{L}\right) = A + B \cdot \log_{10}\left(\mathrm{L}\right) + C \cdot \log_{10}^2\left(\mathrm{L}\right) + D \cdot \log_{10}^3\left(\mathrm{L}\right) + E \cdot \log_{10}^4\left(\mathrm{L}\right) +$$
$$F \cdot \log_{10}^5\left(\mathrm{L}\right) + G \cdot \log_{10}^6\left(\mathrm{L}\right) + H \cdot \log_{10}^7\left(\mathrm{L}\right) + I \cdot \log_{10}^8\left(\mathrm{L}\right). \tag{2.5}$$

The coefficients for Equations 2.4 and 2.5 are given in Table III.

TABLE III. DICOM GSDF Conversion Coefficients

	JND to L		L to JND
a	$-1.3011877E + 00$	A	$7.1498068E + 01$
b	$-2.5840191E - 02$	B	$9.4593053E + 01$
c	$8.0242636E - 02$	C	$4.1912053E + 01$
d	$-1.0320229E - 01$	D	$9.8247004E + 00$
e	$1.3646699E - 01$	E	$2.8175407E - 01$
f	$2.8745620E - 02$	F	$-1.1878455E + 00$
g	$-2.5468404E - 02$	G	$-1.8014349E - 01$
h	$-3.1978977E - 03$	H	$1.4710899E - 01$
k	$1.2992634E - 04$	I	$-1.7046845E - 02$
m	$1.3635334E - 03$		

Using Equation 2.5, the measured luminance at each gray level is converted to its corresponding JND value.

The next step in analyzing the measured luminance response is to calculate the mean ΔJND per gray level GL:

$$\text{mean } \Delta\text{JND/GL} = \frac{j\left(\text{L}'_{\text{max}}\right) - j\left(\text{L}'_{\text{min}}\right)}{\text{GL}_{\text{max}} - \text{GL}_{\text{min}}}, \tag{2.6}$$

where the denominator is equal the difference between the minimum and maximum gray levels (e.g., 255 on an 8-bit system). Since the goal of the DICOM GSDF is to have an equal number of JND indices between each gray level, the mean ΔJND/GL is the number of JND indices that would be present between each gray level on a DICOM GSDF conformant display. Next in the analysis, the ΔJND/GL between each measured gray level is compared to the mean ΔJND/GL to determine the error at each measured step. An example from an 18-point measurement is given in Table IV.

From the rightmost column in Table IV, the maximum relative error in ΔJND/GL from the DICOM GSDF is -4.4%. Note that this should be interpreted as an average error over the first 15 gray levels. For an analysis over 52 measurement points, there will be a total of 51 error calculations, each over 5 gray levels. Increasing the number of measurements from 18 to 52 will more finely sample the response curve, reducing the risk of missing any irregularities.

The third column of Table IV demonstrates the step of adding the assumed or measured value of L_{amb}. The L_{amb} value should be as close to the typical value of the display as possible. Figure 5 demonstrates the ΔJND/GL measurement accuracy as L_{amb} is varied.

In addition to analyzing ΔJND/GL, the luminance response can be evaluated using the contrast threshold model described by Barten 1999. In Equation 2.7, dL/L is the slope of the luminance response function in a plot of luminance versus gray level. The relative contrast dL/L per JND is then compared to the expected DICOM GSDF. According to the DICOM GSDF, the required relative contrast dL/L decreases nonlinearly as the JND index increases (Figure 2). This means that the relative change in luminance required to produce a visibly detectable change to the average observer decreases as the average luminance increases. In other words, at very low luminance, a high relative contrast is required to produce a change

TABLE IV. Sample ΔJND/GL Analysis of 18-point Luminance Measurement

Gray level (8-bit)	Measured L (cd/m^2)	L' (cd/m^2)[a]	Measured JND	ΔJND/GL	ΔJND/GL error[b]
0	1.32	1.42	86.91		
15	2.60	2.70	120.88	2.265	-4.4%
30	4.55	4.65	156.12	2.349	-0.8%
45	7.31	7.41	191.49	2.358	-0.5%
60	11.09	11.19	226.92	2.362	-0.3%
75	16.37	16.47	263.69	2.451	3.5%
90	23.02	23.12	298.72	2.335	-1.4%
105	31.78	31.88	334.20	2.365	-0.2%
120	43.09	43.19	369.66	2.364	-0.2%
135	57.62	57.72	405.17	2.367	-0.1%
150	76.37	76.47	441.01	2.390	0.9%
165	99.81	99.91	476.24	2.349	-0.9%
180	130.31	130.41	512.36	2.408	1.6%
195	168.00	168.10	547.61	2.350	-0.8%
210	216.66	216.76	583.62	2.401	1.3%
225	278.19	278.29	619.63	2.400	1.3%
240	355.31	355.41	655.37	2.383	0.6%
255	452.26	452.36	691.03	2.377	0.3%

[a] L_{amb} = 0.10 cd/m^2
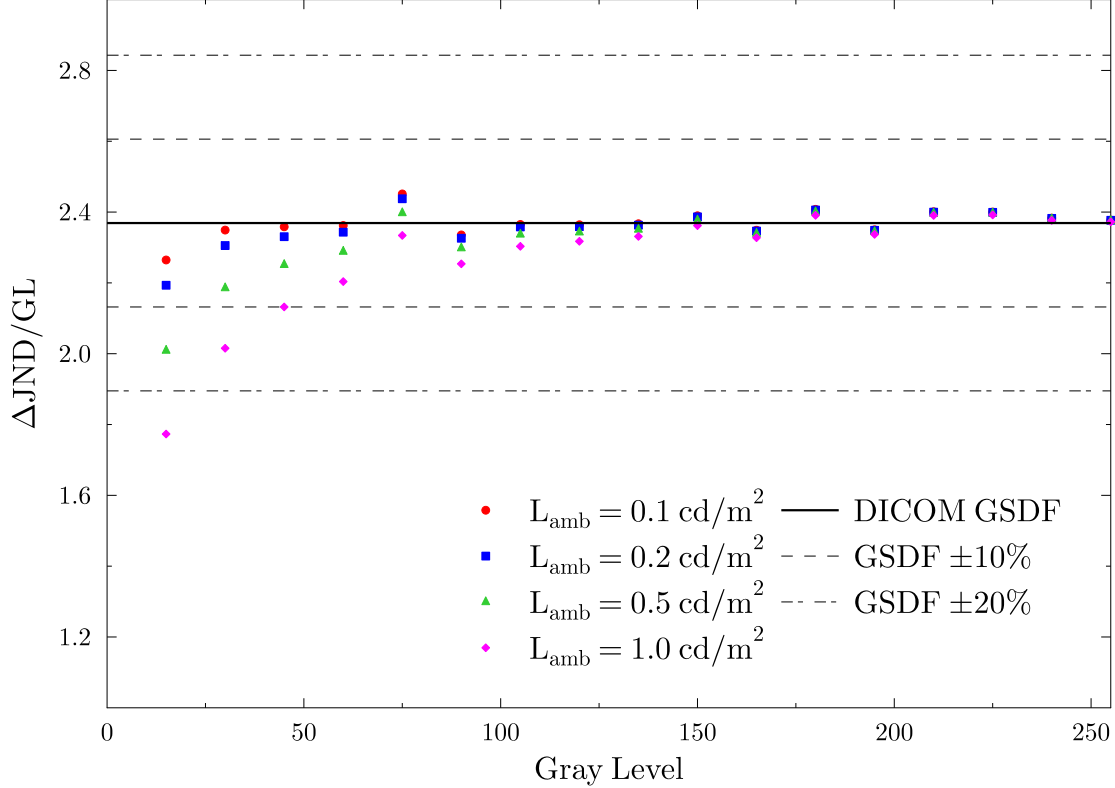
[b] Mean ΔJND/GL = (691.03-86.91)/255 = 2.369

24

FIG. 5. Plot of ΔJND/GL data from Table IV with varying $\mathrm{L_{amb}}$. In this example, the display was calibrated with an assumed operating $\mathrm{L_{amb}}$ of 0.1 $\mathrm{cd/m^2}$. Under this assumption, the display is well within the 10% limit threshold for all measurement points. As $\mathrm{L_{amb}}$ increases, the DICOM conformance of the display becomes increasingly inaccurate. Note that this effect is most pronounced as low luminance levels, as was the case in Figure 1.

that can be detected by the human visual system. At the same time, a smaller relative contrast will be visible at higher luminances.

The dL/L per JND can be calculated between two measured gray levels using Equation 2.7:

$$\mathrm{dL/L \; per \; JND} = \frac{2 \left(\mathrm{L'}_i - \mathrm{L'}_{i-1} \right)}{\left(\mathrm{L'}_i + \mathrm{L'}_{i-1} \right) \left(\mathrm{mean} \; \Delta \mathrm{JND/GL} \right) \left(\mathrm{GL}_i - \mathrm{GL}_{i-1} \right)}. \tag{2.7}$$

This calculation should be performed over the same luminance range for both the measured values and the target DICOM GSDF. Based on the target DICOM GSDF JND index for each gray level measurement, the DICOM GSDF luminance can be calculated using the transformation given in Equation 2.4. These luminance values can then be used in Equation 2.7 to determine the DICOM GSDF dL/L per JND for comparison. Similar to the ΔJND/GL analysis, the relative deviation between the measured and desired values should be compared. An example using the same measurements from Table IV is given in Table V.

The maximum relative error in dL/L per JND shown the right-most column of Table V is 3.8%. Over the same 15 measurement points (Gray levels 0-15) that gave the largest error in Table V, the relative error in dL/L per JND is -3.7%. Both analysis methods are

25

TABLE V. Sample dL/L per JND Analysis of 18-point Luminance Measurement

Gray level (8-bit)	Measured L (cd/m²) [a]	L' (cd/m²) [a]	Measured JND	GSDF JND [b]	GSDF L' (cd/m²)	Meas. dL/L per JND	GSDF dL/L per JND	dL/L per JND error
0	1.32	1.42	86.91	86.91	1.42			
15	2.60	2.70	120.88	122.45	2.77	0.0175	0.0182	-3.7%
30	4.55	4.65	156.12	157.98	4.77	0.0149	0.0149	-0.1%
45	7.31	7.41	191.49	193.52	7.60	0.0129	0.0129	0.2%
60	11.09	11.19	226.92	229.06	11.46	0.0114	0.0114	0.3%
75	16.37	16.47	263.69	264.59	16.62	0.0107	0.0104	3.8%
90	23.02	23.12	298.72	300.13	23.43	0.0095	0.0096	-1.1%
105	31.78	31.88	334.20	335.67	32.29	0.0090	0.0090	0.1%
120	43.09	43.19	369.66	371.20	43.74	0.0085	0.0085	0.0%
135	57.62	57.72	405.17	406.74	58.44	0.0081	0.0081	0.1%
150	76.37	76.47	441.01	442.27	77.20	0.0079	0.0078	1.0%
165	99.81	99.91	476.24	477.81	101.07	0.0075	0.0075	-0.8%
180	130.31	130.41	512.36	513.35	131.35	0.0075	0.0073	1.7%
195	168.00	168.10	547.61	548.88	169.65	0.0071	0.0072	-0.8%
210	216.66	216.76	583.62	584.42	218.00	0.0071	0.0070	1.4%
225	278.19	278.29	619.63	619.96	278.98	0.0070	0.0069	1.3%
240	355.31	355.41	655.37	655.49	355.77	0.0068	0.0068	0.6%
255	452.26	452.36	691.03	691.03	452.40	0.0068	0.0067	0.4%

[a] $L_{amb} = 0.10$ cd/m²
[b] GSDF JND assumes an equal number of JND indices per step; 2.369 JND/GL in this example

effective for assessment of DICOM GSDF conformance, and should produce similar results. Qualified personnel should evaluate measurements that fall near the acceptable limit, and fail one analysis method but not the other (e.g., 9.7% error in dL/L per JND and 10.1% error in ΔJND/GL). If possible, recalibration will likely bring both measurements back to within acceptable limits. If recalibration is not possible, the display should be monitored to ensure the performance does not degrade further by either metric.

2.3.2.4   Suggested Limits

The suggested limits in Tables VI and VII are recommendations based on the ACR-AAPM-SIIM technical standard [ACR-AAPM-SIIM 2017]. It should be noted that these are only suggested values, and the ultimate evaluation of the performance of a display should be left to the qualified physicist or other qualified personnel. The luminance recommendations in Table VI for L'_{min}, LR, and L'_{max} are sequentially dependent. The L'_{min} values are based on an assumption for typical working environments, response limitations of the human visual system, and technical capabilities for displays commonly used in these applications. From the L'_{min} values, an LR of 350 is recommended to maintain contrast, while not exceeding the response capabilities of the viewer's visual system. If any of these values are changed based on technical limitations or personal preference, the others should be adjusted accordingly.

TABLE VI. Recommended Display Luminance Levels

Display Type	Recommended Values[a] [Acceptable Ranges]
Diagnostic Displays (non-mammography)	$L'_{min} \geq 1.0 \text{ cd/m}^2$ $LR = 350\,[250\text{-}450]$ $L'_{max} = 350 \text{ cd/m}^2 \, [\geq 300 \text{ cd/m}^2]$
Diagnostic Displays (mammography)	$L'_{min} \geq 1.2 \text{ cd/m}^2$ $LR = 350\,[250\text{-}450]$ $L'_{max} = 420 \text{ cd/m}^2 \, [\geq 350 \text{ cd/m}^2]$
Modality Displays Clinical Specialist Displays EHR Displays	$L'_{min} \geq 0.8 \text{ cd/m}^2$ $LR = 350\,[250\text{-}450]$ $L'_{max} = 250 \text{ cd/m}^2 \, [\geq 200 \text{ cd/m}^2]$

[a] All recommended values assume an ambience ratio (L_{amb}/L_{min}) of 1/4 or smaller (refer to 2.3.1.5).

Display manufacturers often advertise mammography-specific displays with higher L_{max}. The higher brightness of these displays allows the user to operate at higher luminances with longer lifetimes. Table VI shows additional recommended operating levels for mammography diagnostic displays that take advantage of the improved technical capabilities of these displays. The mammography operating level increases L'_{min} to further avoid the mesopic region of the human visual response system, thereby resulting in a higher recommended L'_{max} to maintain the recommended LR.

TABLE VII. Recommended Display Luminance Response Accuracy

Display Type	Recommended Limits[a]
Diagnostic Displays	Deviation in either JND/GL or dL/L per JND from DICOM GSDF $\leq 10\%$
Modality Displays Clinical Specialist Displays EHR Displays	Deviation in either JND/GL or dL/L per JND from DICOM GSDF $\leq 20\%$

[a] Due to the noise associated with the 256-point analysis, these limits should serve only as general guidelines when analyzing every point in the luminance response.

2.3.2.5 Qualitative Assessment

Qualitative assessment should be considered when quantitative assessment is not possible, or for more frequent testing under the supervision of a physicist, e.g., daily or weekly quality assurance of a diagnostic review workstation by a technologist. Unfortunately, the efficacy of qualitative assessments is limited. Qualitative evaluation of the luminance response function is especially difficult because the human visual system is not as accurate as a photometer when evaluating the luminance response's adherence with the DICOM GSDF. For this reason, user knowledge and understanding of test pattern specifications and limitations can aid in determining effective methods for qualitative evaluation.

Test Patterns

The following descriptions are of test patterns selected to evaluate the luminance response of a display. A more complete description of these and other test patterns is provided in the Test Patterns section of this report.

SMPTE

The SMPTE pattern (formally the SMPTE Medical Diagnostic Imaging Test Pattern) has a long history of use for film printer and display QA. It provides the user with both 0%/5% and 95%/100% contrast objects, allowing the user to qualitatively evaluate the luminance response at the top and bottom of the grayscale range, but not between. On an 8-bit system, 5% contrast corresponds to approximately 13 gray levels. The 5% contrast can always be visualized, except on systems that either deviate significantly from the DICOM GSDF, or are in environments with poor ambient light conditions. The other features on the SMPTE pattern (geometric lines, scanning artifact windows, etc.) were designed to test properties of CRT displays, and are not relevant for modern LCD and OLED display technology. **As a result of the pattern's grayscale insensitivity and CRT-specific features, this report considers the SMPTE test pattern deprecated for qualitative display evaluation in favor of either quantitative measurement or updated test patterns.**

TG18-QC and TG18-OIQ

The TG18-QC pattern was introduced with the publication of the TG18 report. It is often available both on modality software and within display manufacturer software as an alter-

(a) 0%/5% and 95%/100% Luminance Patches

(b) QUALITY CONTROL Patches

(c) ±4 Contrast Corner Patches in three of the
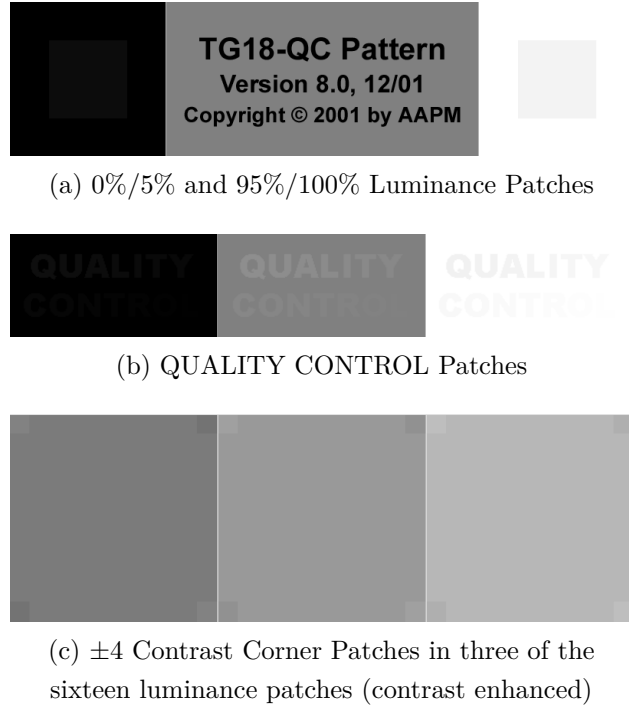sixteen luminance patches (contrast enhanced)

FIG. 6. Regions of interest for qualitative luminance assessment from the TG18-OIQ and TG18-QC test pattern.

native to the SMPTE test pattern. IEC 62563-1 provides TG18-OIQ as an updated version of TG18-QC with CRT-specific features removed [IEC 62563-1 2016]. Both TQ18-OIQ and TG18-QC contain three primary areas of interest for qualitative evaluation of low-contrast regions of the luminance response function.

The first areas of interest (Figure 6a) are the ends of the luminance patches, where there are 0%/5% and 95%/100% contrast objects, similar to those from the SMPTE pattern. The TG18 report specifies a 13-value difference on an 8-bit system for these patches, and like the SMPTE pattern, the 5% boxes should be visible on all displays. The patches may not be visible on displays that deviate significantly from the DICOM GSDF, or are in environments where the ambient lighting has not been properly considered.

The second area of interest (Figure 6b) contains larger patches with the words "QUALITY CONTROL" overlaid on three different background levels. The background levels are at the minimum, midrange, and maximum luminance levels, and the letters increment in contrast from the background from 1 to 14 gray levels (1 level per letter). The QUALITY CONTROL region of the pattern is of more utility than the SMPTE-like patches because it allows the user to more finely evaluate the display at three points along the luminance curve. Luminance levels between the minimum, midrange, and maximum luminance, however, still remain unevaluated with this region of the pattern.

The final area of interest (Figure 6c) is the series of luminance patches surrounding the central region of the pattern. Each luminance patch contains a smaller patch in each of its four corners that differs by ±4 from the patch background. Using these, the user is able to evaluate the visibility of a similar contrast level throughout the luminance range of the display.

(a) 0 Luminance Patch
(contrast enhanced)

(b) 120 Luminance Patch
(contrast enhanced)
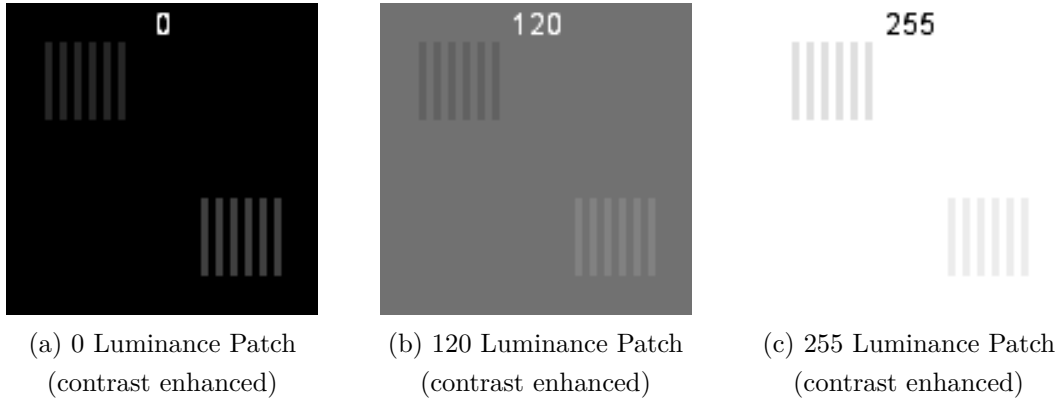
(c) 255 Luminance Patch
(contrast enhanced)

FIG. 7. Regions of interest for qualitative luminance assessment from the TG270-sQC test pattern.

TG18-OIQ and TG18-QC also contain luminance ramps along each side of the patterns to evaluate contouring and bit-depth artifacts. For LCD and OLED displays, the user can also use these ramps to evaluate single driving levels that have been improperly calibrated and may deviate significantly from the continuous nature of the ramps (Figure 4).

The TG18-OIQ and TG18-QC patterns remain a viable option for qualitative assessment of luminance response of a flat-panel display. The luminance patches with the low-contrast (± 4 levels) corners are preferred as the primary means of display evaluation, rather than the oversimplified SMPTE-like patches or the QUALITY CONTROL letters. In addition, the user should recognize that many of the other features of the TG18-QC pattern are CRT-specific, and can be ignored when evaluating flat-panel displays.

TG270-sQC

The TG270-sQC (simple quality control) pattern allows the user to evaluate luminance response characteristics around multiple gray levels over the entire luminance range. Modulating bar patterns are placed in the upper-left and lower-right corners atop different gray level patches (patch 8-bit levels of 0, 15, 30, . . . , 240, and 255, corresponding to every 15 gray levels on an 8-bit system, see Figure 7). These modulating bar patterns vary from the background of the patches by ± 5 gray levels, which allows for the user to assess the luminance response function over the entire range of the display. A user should see all of the ± 5 gray level patterns on a properly calibrated display. The 0 and 255 luminance patches (Figures 7a and 7c) contain 5 gray level and 3 gray level contrast patterns. Diagnostic displays should show both levels, and all other displays should show the 5 gray level pattern.

The modulating bar patterns are preferable to the "box-within-a-box" pattern (as is used in the SMPTE or TG18-OIQ patterns) because it more closely matches the Barten model of visualization (from which the DICOM GSDF is defined), which describes the contrast sensitivity of the human visual system at various spatial frequencies [Barten 1999].

In addition to the visual contrast array, a continuously increasing grayscale ramp is included along the bottom of the pattern. While a human observer would not likely detect small deviations from the DICOM GSDF, miscalibrated gray levels that deviate significantly from the ramp may be visible, as shown in Figure 4.

TG270-pQC

The TG270-pQC (physics quality control) pattern is a slightly-modified version of the TG18 printer QC pattern. It extends the rationale behind the TG270-sQC pattern to include additional spatial frequencies, contrast levels, and orientation to the modulating bar patterns over multiple gray levels covering the entire luminance range. The pattern is intended to help bridge the gap between qualitative and quantitative evaluation of a display's luminance response by providing physicists and other advanced users with more qualitative features. The TG270-pQC pattern can give context to any quantitative failures by demonstrating how visualization over the grayscale may be affected. The pattern is also useful for follow-up from abnormalities seen on other patterns. A complete description of TG270-pQC is provided in the Test Patterns section of this report.

2.4 Display Color

2.4.1 Rationale

For grayscale images, the "color" of a display is characterized by the white point of the light used to generate brightness on the display. Displays have a native white point, which describes the color of the white light without modification of either on-board adjustments in the on-screen display (OSD) menu, or the lookup table (LUT) on the graphics card or display. The native white point is known to vary with display model and type (LCD vs. OLED). Maintaining a consistent white point throughout the image viewing chain (from modality to review workstation to radiologist workstation to referring clinician) helps to ensure consistent image appearance.

As mentioned above, the white point of a display can be adjusted through either the OSD settings or a LUT on the graphics card (for more-sophisticated displays, the modified LUT can be directly loaded to the display). It is worth stating that while small adjustments in the white point of the display should not substantially affect the display capabilities, large changes may reduce overall brightness and could affect both performance and lifetime of the display. Any time the white point of a display is adjusted, the luminance response (and therefore L'_{min}, L'_{max}, LR, and AR) should be re-evaluated.

2.4.2 Test Equipment

Measuring the white point of a display uses much of the same equipment and setup as used in measuring the luminance response. As described in the Measurement Instrumentation section, a colorimeter is used to measure the color performance of a display. Some photometers also include colorimeter capabilities and are suitable for measuring chromaticity coordinates for white point determination. Colorimeters typically measure and report chromaticity values in one or more CIE color spaces. A detailed description of color space is beyond the scope of this document, but the reader is referred to the CIE report on colorimetry [CIE15:2004] for a further description.

To measure the white point, the display being tested must be able to display the appropriate test patterns, such as those used when evaluating luminance uniformity or luminance response. As with photometers, colorimeters vary in terms of their user interface and whether they can be operated by software loaded on the display workstation, or by an external computer or device. If the system being tested allows the user to access other programs and

interface with a connected photometer, automated software from display manufacturers or third parties may be used to aid in the measurement and analysis of the color performance.

2.4.3 Procedure

Full analysis of the color performance over the entire luminance range of a display may not be necessary for all display types. The physicist or other qualified personnel must decide what level of color accuracy is required for the types of tasks expected to be performed on a given display. In general, for routine display testing (more frequently than annual), the white point may be evaluated at a single gray level. The level could be at any luminance level (e.g., gray level 100, 200, or 255 on an 8-bit system), and the same level should be used each time for consistency. A test pattern with a uniform region of the desired gray level should be displayed in the central part of the useful area of the display (the area that shows the image), and the colorimeter correctly positioned to measure the color of the light emitted. This test can often be performed in conjunction with luminance response measurements. The same gray level should be tested across all displays and from year-to-year to ensure proper comparisons.

For acceptance or annual testing, the white point of the display should be measured over the full luminance range in 8-bit gray level steps of 15. These steps correspond to the 18-point response steps, and the same patterns or automated routine described for luminance measurements may be used. Finer sampling may be used (e.g., the 52-point response), though it may be unnecessary for the purposes of white point performance evaluation. The measurement should exclude gray levels with luminance less than 5 cd/m^2, where the color receptors of the eye are less sensitive to color differences.

2.4.4 Assessment

2.4.4.1 Target White Point

The recorded chromaticity coordinates should be converted to the CIE 1976 UCS diagram coordinates (u', v'). Once converted, the color distance Δ to a target $\left(u'_{\text{target}}, v'_{\text{target}}\right)$ value should be calculated for each gray level measurement according to Equation 2.8:

$$\Delta\left(u', v'\right) = \sqrt{\left(u'_1 - u'_{\text{target}}\right)^2 + \left(v'_1 - v'_{\text{target}}\right)^2}. \tag{2.8}$$

The values for the target $\left(u'_{\text{target}}, v'_{\text{target}}\right)$ in Equation 2.8 will depend on the type of test being performed. If the evaluation is meant to determine the difference between a display's white point and a standard white point, this report follows the use of the ACR–AAPM–SIIM Technical Standard for Electronic Practice of Medical Imaging recommendation that displays be set to the CIE Standard Illuminant D65 [ACR-AAPM-SIIM 2017]. D65 corresponds to color coordinates $u' = 0.19783$, $v' = 0.46833$. Note that D65 refers to a specific point in color space, and not a color "temperature." The use of correlated color temperature (CCT) is discouraged for evaluation of white point accuracy. For a further discussion of this recommendation, please see the Extended Discussion of White Point Rationale at the end of this section.

If the white point is measured over the entire grayscale, the distance from a target illuminant can be plotted to demonstrate how the white point changes over the luminance range
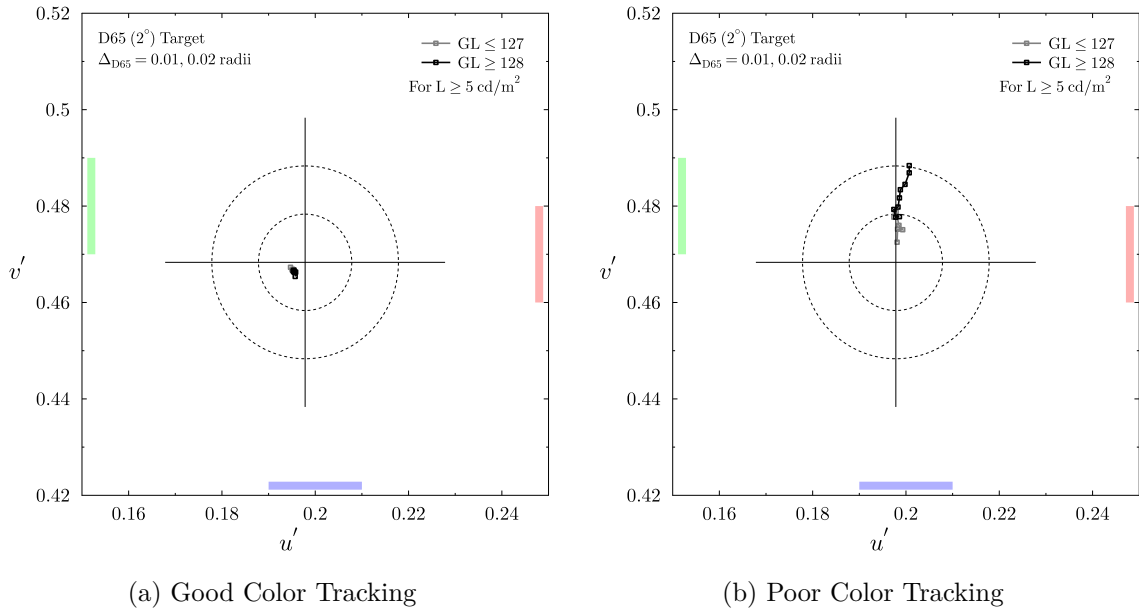
(a) Good Color Tracking   (b) Poor Color Tracking

FIG. 8. Plot of grayscale white point tracking over the entire luminance range (greater than 5 cd/m^2). Plot 8a shows the results from a display with a white point that is consistent over all luminance levels, and is well within the $\Delta = 0.01$ radius surrounding D65. Plot 8b shows the results from a display where the darker luminance levels hold steady around $\Delta = 0.01$ before beginning to drift away from D65 as the luminance increases to the maximum level. The display on the left would be preferable to the one on the right for consistent image presentation.

(Figure 8). This provides a way to visually assess color tracking over the full luminance range.

2.4.4.2 Display Comparison

In addition to comparing the measured color of each gray level to a standard white point (e.g., D65), the color may also be compared to the color of a corresponding gray level on another display. This is useful for color-matching displays that may be attached to the same workstation, or to compare the color of all displays that may show a given image throughout the image review chain (e.g., modality display, technologist PACS workstation, radiologist review workstation). For this comparison, the color distance between two measurements is given by Equation 2.9:

$$\Delta\left(u', v'\right) = \sqrt{\left(u'_1 - u'_2\right)^2 + \left(v'_1 - v'_2\right)^2}, \tag{2.9}$$

where (u'_1, v'_1) and (u'_2, v'_2) are the color coordinates for the first and second measurements, respectively. If a range of gray levels is evaluated, the color distance $\Delta\left(u', v'\right)$ may be calculated at each measured gray level. As was described earlier, gray levels with luminance less 5 cd/m^2 should be excluded.

Comparing the white point color performance across the full luminance range of two displays is most appropriate across two displays of the same class (e.g., diagnostic to diagnostic). The color difference between the displays should be recorded as the maximum

$\Delta\left(u', v'\right)$ calculated. When comparing displays of different classes, a single gray level should be chosen as the comparison point - e.g., compare full white (8-bit level 255) between a modality and diagnostic display.

2.4.5 Suggested Limits

The white point color accuracy will vary depending on the type of the display. When considering the distance $\Delta_{\mathrm{D65}}\left(u', v'\right)$ (other target illuminants may be used), the recommended values are in Table VIII.

TABLE VIII. Recommended Color Distance from a Standard Illuminant (e.g., $\Delta_{\mathrm{D65}}\left(u', v'\right)$)

Display Type	Optimal Limit	Acceptable Limit
Diagnostic Displays	$\Delta_{\mathrm{D65}}\left(u', v'\right) \leq 0.005$	$\Delta_{\mathrm{D65}}\left(u', v'\right) \leq 0.01$
Modality Displays Clinical Specialist Displays EHR Displays	$\Delta_{\mathrm{D65}}\left(u', v'\right) \leq 0.01$	$\Delta_{\mathrm{D65}}\left(u', v'\right) \leq 0.02$

When evaluating the color difference $\Delta\left(u', v'\right)$ between two displays attached to the same workstation or as part of the same image review chain (e.g., modality display, technologist PACS workstation, radiologist review workstation), Table IX provides recommended values.

TABLE IX. Recommended Color Distance Limits between Displays

Display Type	Optimal Limit	Acceptable Limit
Same Workstation	$\Delta\left(u', v'\right) \leq 0.005$	$\Delta\left(u', v'\right) \leq 0.01$
Same Image Review Chain	$\Delta\left(u', v'\right) \leq 0.01$	$\Delta\left(u', v'\right) \leq 0.02$

Tables VIII and IX provide both optimal and acceptable limits. The optimal limits are stricter, and ensure improved color tracking and matching for displays. Most displays should be able to achieve the optimal limits for their designated category, but these limits are not necessarily critical to overall performance. The acceptable limits provide a less-restrictive bound that will still enable sufficient color matching that should be achievable by most displays. As with all display performance metrics, a qualified physicist or other personnel should evaluate failures within the context of how the display is used and how the failure will affect the review of medical images.

2.4.5.1 TG196 Methodology

Along with accuracy of the white point of the display (either at a single luminance level, or over the luminance range), the color performance of a display may be evaluated using the metrics defined in TG196 and IEC 62563-1 [TG196 2016; IEC 62563-1 2016]. These metrics aim to describe the variation in color of a single display over the entire grayscale. This provides a more quantitative description of the differences that are only qualitatively visible

between the two plots presented in Figure 8. The first metric, τ_1, describes the average difference between grayscale chromaticity and the chromaticity of full white (i.e., gray level 255 on an 8-bit system). A second metric, τ_2, describes the average difference between the chromaticity of adjacent gray level measurements in the (u',v') plane. In addition, two values $\tau_{1,\max}$ and $\tau_{2,\max}$ describe the respective maximum values for each metric (rather than an average). Note that neither of these values uses a reference white point to determine color difference; both are measures of color precision over the range of luminance of a single display.

At the time of preparation of this report, no formal guidance on the use of the TG196 metrics has been published. As a result, this report cannot recommend performance limits. However, the TG196 metrics are recommended as part of any color performance analysis to establish baseline performance criteria that may aid in the interpretation of future measurements.

2.4.5.2 Note on Color Calibration

Much like grayscale calibration, some displays can be calibrated to specific white points by using display manufacturer software. This software often is capable of using either internal or external photometers, and the qualified medical physicist or other personnel should evaluate the use of the calibration software for the purposes of correcting display color. Prior to using automated software and internal photometer, the qualified medical physicist or other personnel should also confirm the accuracy of measurements made by an internal photometer.

2.4.5.3 Extended Discussion of White Point Rationale

The ACR–AAPM–SIIM Technical Standard for Electronic Practice of Medical Imaging recommends that displays be set to the CIE Standard Illuminant D65. Illuminant D65 is intended to represent average daylight, and is often cited as having a correlated color temperature (CCT) of about 6504 K (often rounded to 6500 K). However, it should be noted that not all displays with a CCT of 6500 K have a white point of D65. This is due to the fact that there are an infinite number of color coordinates with a CCT of 6500 K, ranging from the bluish-green to pinkish-purple (see the ends of the 6500 K isotherm in Figure 9).

As a result of the potential disparity in color accuracy from using CCT alone, a display's white point should be specified in terms of a standard illuminant (e.g., D65). D65 is defined using the CIE 1931 color space chromaticity coordinates of x=0.31271, y=0.32902 using the CIE 1931 2° Standard Observer, which transforms to the CIE 1976 UCS diagram coordinates of u'=0.19783, v'=0.46833.

2.5 Display Luminance Uniformity

2.5.1 Rationale

Luminance uniformity describes the variation in luminance output for a uniform image displayed over the entire display area. When any image is displayed, it is assumed that a given driving level will result in a consistent luminance response regardless of where on the display the image is shown. When this assumption is violated, it creates the potential for misinterpretation that the non-uniformity is part of the image, rather than a defect of the
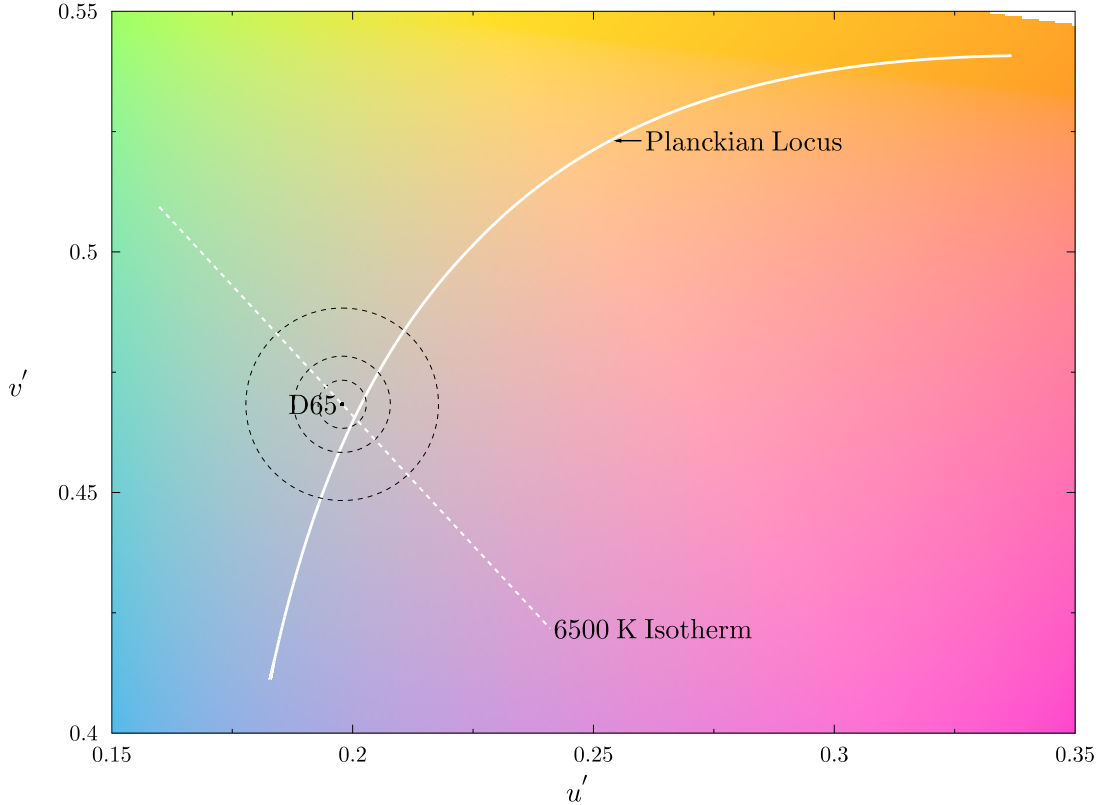
FIG. 9. D65 can be seen on this closeup of the chromaticity diagram of the 1976 CIE UCS. An isotherm (dashed white line) is drawn through the 6500 K point on the Planckian locus (solid white line), where the ends of the isotherm represent the maximum distance where the CCT is meaningful ($\Delta = 0.05$) [CIE15:2004]. Additionally, multiple circles drawn around D65 show increasing Δ, with radii of $\Delta = 0.005, 0.01$, and 0.02. Note that a change of $\Delta = 0.005$ (innermost circle) is perceptible for most viewers [Groth et al. 2001]. The implication of this plot is that two displays with a white point CCT of 6500 K may have different appearances to a typical human observer, especially when viewed side by side or when showing the same image.

display. Non-uniformities manifest in several ways, and the wide variation in size, shape, severity, and position results in the need for both qualitative and quantitative evaluation.

Global non-uniformities that span the entire display area are of less concern because of the insensitivity of the human visual system to these spatial frequencies [Barten 1999; TG18 2005]. Such non-uniformities are typically due to non-uniform illumination from backlights or non-uniform aging of display panel optical properties, and only need to be addressed if the level of variation is especially pronounced. The global uniformity of the entire display may be evaluated either quantitatively by luminance measurements at multiple locations on the display, or qualitatively with a series of uniform test patterns covering the full luminance range.

Small- to medium-sized, low-contrast, and irregularly-shaped patterns or regions called mura (from the Japanese "unevenness") occur for several reasons, and often require a qualitative assessment due to their variation in severity, size, shape, and position. Image latency or screen burn-in may also occur, though is less common than it was with CRT technology.

Finally, "bad" or "stuck" pixels may occur, when one or more pixels are permanently white ("hot") or black ("cold"). Non-uniformities that appear more locally (e.g., mura, burn-in, bad pixels), are of greater concern due to their potential similarity to the content of radiological images. These small non-uniformities require qualitative visual inspection of uniform test patterns, often over a range of luminances.

2.5.2 Quantitative Evaluation

2.5.2.1 Test Equipment

Quantitative assessment of luminance uniformity requires a photometer that measures luminance. Additionally, the display being tested must be able to display the test patterns as described in the test procedures below. Because the luminance level of the displayed test pattern may affect the appearance of non-uniformities, a series of uniform test patterns spanning the full luminance range is ideal. Test patterns such as the ones used for the 18-point luminance response evaluation in gray level increments of 15 (0, 15, 30, ..., 240, 255) on an 8-bit display should be considered for acceptance testing or product evaluation of display luminance uniformity. For more routine testing, fewer patterns, such as the set of TG270-ULN8-200, TG270-ULN8-100, and TG270-ULN8-020 could be used to demonstrate longitudinal consistency. Additional patterns should be considered for routine testing if there are specific levels of concern for a given display. If the system being tested allows the user to access other programs and interface with a connected photometer, automated software from display manufacturers or third parties may be used to aid in the measurement and analysis of the luminance uniformity.

2.5.2.2 Procedure and Assessment

TG18 5-point Maximum Luminance Deviation (MLD)

The TG18 report describes the Maximum Luminance Deviation (MLD) as the quantitative measure of display uniformity. Luminance is measured at the center and corners of the display for each test pattern, and the MLD is calculated by Equation 2.10:

$$\text{MLD} = 200 \cdot \frac{\text{L}'_{\max} - \text{L}'_{\min}}{\text{L}'_{\max} + \text{L}'_{\min}}, \tag{2.10}$$

where $\text{L}'_{\max}$ is the highest recorded luminance across the five measurements of the test pattern, and $\text{L}'_{\min}$ is the lowest, regardless of location on the display. All measurements include the ambient luminance contribution. As a result of changes in display technology, this report recommends an alternative method for quantitative evaluation of luminance uniformity, described in the following section.

TG270 9-point Luminance Deviation from the Median (LUDM)

Non-uniformities in CRT displays occur most often in the corners of the display. For flat-panel displays, luminance non-uniformities may be as likely along the edges of the display as at the corners. Consequently, this task group recommends that luminance uniformity measurements are made in the center, near the corners, and middle of the edges of the display. A photometer is used to measure the luminance at the center of each cell of a 3×3 grid. Each measurement is compared to the median reading, and the maximum luminance uniformity deviation from the median (LUDM) is determined by Equation 2.11:

$$\text{LUDM} = \max\left(100 \cdot \frac{|\text{L}'_n - \text{L}'_{\text{med}}|}{\text{L}'_{\text{med}}}\right), \hspace{2cm} (2.11)$$

where L'_n is the measured luminance at each of the nine locations, and L'_{med} is the median luminance. All luminance measurements should include the ambient luminance contribution. Typical defects for a flat-panel display are likely to affect only one of the nine measurements. As described in the rationale section, the human visual system is not sensitive to low-spatial-frequency non-uniformities. Therefore, comparing individual measurements is of less value than finding a single measurement that deviates significantly from the median, which is less affected by a single outlier.

2.5.2.3 Suggested Limits

TG18 recommends an MLD of less than 30%. As stated earlier, the TG18 report primarily addressed CRT technology, but commented that the MLD for flat-panel displays may be different than for CRT displays [TG18 2005].

Based on current flat-panel display performance, this task group recommends that all displays with a LUDM greater than 30% should be considered defective and replaced. Additionally, all displays with an LUDM greater than 15% should be reviewed to ensure that their overall performance is acceptable.

2.5.2.4 Note on Large-Format Displays

In recent years, the use of large format displays has become more common for diagnostic interpretation and other imaging applications. Anyone evaluating display uniformity on these displays should note that some large-format displays may not pass quantitative uniformity evaluations when evaluated as a single unit due to luminance variations across the entire display area. When evaluating uniformity on these devices, the user must consider that the large-format display may take the place of two, three, or more standard-sized displays. In this case of a single display replacing two smaller displays, it is appropriate to treat each half of the display as its own unit for uniformity evaluations. Similarly, displays used in interventional radiology or cardiac catheterization labs can be very large, with separate images being displayed over several display regions. Additionally, the individual areas of these displays maybe different in size depending on use. In all cases, the use of the display should be considered when determining how uniformity evaluation should be implemented.

2.5.3 Qualitative Evaluation

2.5.3.1 Procedure

As with quantitative evaluation, qualitative evaluation of uniformity requires a series of uniform test patterns spanning the range of luminance levels. At a minimum, qualitative evaluation should be performed using low-, medium-, and high-luminance test patterns such as TG270-ULN8-200, TG270-ULN8-100, and TG270-ULN8-020. On systems where the TG18 test pattern set is available, the TG18-UN80 and TG18-UN10 pair is also suitable.

Each test pattern should be evaluated at a typical viewing distance for the appearance of bright or dark mura, commonly appearing as spots or nebulous patches with a different luminance than the rest of the display. Additionally, the display should be evaluated for the existence of bad pixels. A raster-like search where the user moves a small uniform window
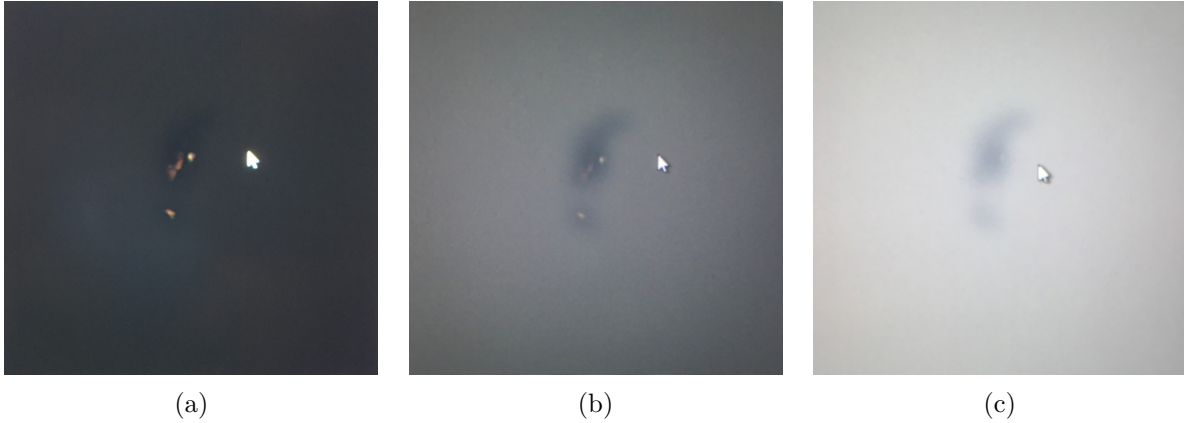
FIG. 10. Photographs of a display with a mura non-uniformity (mouse cursor included for scale). Note that as the luminance of the uniform image increases (from left to right: 8-bit gray levels 0, 75, 180), the presence and appearance of the mura changes.

across the display may aid in identifying small non-uniformities. It should be noted that for modality displays, images can only be displayed in the central area of the display with the other portions used for acquisition settings and other non-image data. These display only need to maintain uniformity within the area that images are displayed.

2.5.3.2   Assessment

Mura should be reviewed in context of the location and intensity with which they occur. Those appearing in the center of the image area should be given more thorough consideration than those that occur in areas where image information is rarely displayed. Similarly, when evaluating bad pixels, their clinical significance is largely dependent on the how the display is used. As an example, a bright bad pixel on an ultrasound display is not likely to be confused with actual image information, whereas on a display used for the primary interpretation of mammograms, it may be confused with small calcifications. Sample images of local non-uniformities are shown in Figure 10.

2.5.3.3   Suggested Limits

Ideally, no luminance variations should be visible when evaluating the luminance uniformity test patterns. However, the appearance of subtle non-uniformities in the form of mura or bad pixels is not uncommon as displays age. The specific limit or number of non-uniformities that one should expect on a display is typically defined as part of a manufacturer's warranty (and should be considered as part of a product evaluation). Once identified, their potential impact on clinical images should be determined by considering the contrast of the non-uniformity, the behavior over the full luminance range, and the position on the display. It is important to recognize that acceptable performance for visual inspection of uniformity should take into account the specific usage of the display. Consequently, a particular local non-uniformity visible on a diagnostic display might render the device unsuitable for clinical use, yet when visible on a display of another category, the performance may be considered acceptable. Final determination of whether small non-uniformities render a display unsuitable for clinical use often requires consultation with radiologists, technologists, and/or other clinicians.

2.6   Display Noise

2.6.1   Rationale

Noise, along with contrast and spatial resolution, is a fundamental property that determines the conspicuity of findings in an image. Noise can obscure the visibility of findings, as well as contribute to reader fatigue [Ikushima et al. 2013]. In addition to image noise, random variations in signal intensity of a displayed image can be introduced by noise caused by the display itself. Noise can arise from both temporal variation and stationary spatial variation (pixel-by-pixel), though stationary spatial variation has been reported to be the more substantial source of noise [Roehrig et al. 2003].

Spatial noise properties may vary between different models of flat-panel displays either because of liquid crystal non-uniformity, cell thickness, spacers, cell voltage, or from implementation of pixel-by-pixel uniformity corrections [Badano et al. 2004]. Generally, displays sold for professional graphics or medical imaging have panels that have been selected for high uniformity and may utilize pixel-by-pixel corrections for greater uniformity; however, this is not universal, and noise comparisons are worthwhile when selecting displays, especially from an unfamiliar vendor.

While display noise should be a relevant factor in evaluation for display selection, it has not yet been demonstrated to be a useful metric for quality control testing. Display noise is not a property that the end-user can control, nor is it anticipated to change significantly over the lifetime of a display. The noise assessment recommendations outlined here are intended to be useful for selecting a display with superior noise properties. Noise properties should be strongly considered as a differentiating purchase criteria because of the effect on image SNR, potentially impacting reader accuracy more than other factors such as luminance and dynamic range [Krupinski et al. 2002].

A truly quantitative assessment of spatial noise requires a high-quality photometric camera as described elsewhere [Fan et al. 2005; TG18 2005]. Analysis of the noise power spectrum is complicated by the detailed patterns seen in LCD pixel structures [Badano et al. 2004]. Furthermore, it does not directly correspond with the visibility of subtle image features, limiting its use in evaluating suitability for clinical use. Such quantitative measures are not presently practical to consider for acceptance testing or quality control.

For the evaluation of spatial noise, visual test patterns are recommended over quantitative assessment. This section will focus on describing the visual assessment for discrimination between noise properties of displays.

2.6.2   Test Patterns

To perform a qualitative comparison of noise, users should use a combination of a uniformity test pattern with a bright field (TG270-ULN8-200 or similar TG18-UN80), a contrast-detail test pattern (TG18-AFC), and representative clinical images (TG18-CH, TG18-KN, or TG18-MM2).

2.6.3   Procedure

When comparing two displays, all display aspects should be matched as closely as possible (e.g., luminance calibration, color calibration, ambient lighting, viewing angle, etc.).

In addition, displays are best compared side-by-side to help constrain environmental and perceptual factors.

For direct visual comparison of two displays, the person evaluating the display shows the same test pattern or image on both displays under identical configurations. The evaluation procedure is simply to see if one can perceive a difference in the image noise granularity between the displays (relatively uncorrelated noise with high spatial frequencies).

2.6.4 Assessment

As mentioned in the display noise rationale, assessment of display noise may be valuable when evaluating displays for purchase. The noise assessment is meant to differentiate between noise properties of different displays, i.e., which has better properties. While noise texture is often perceivable on flat-field images like TG270-ULN8-200 or TG18-UN80, it is difficult to assess the relative impact without direct, side-by-side visual comparison.

When directly visually comparing two displays, the person performing the evaluation should determine if there is perceptible difference in noise on a pixel-by-pixel basis. If a noise difference is seen in the test patterns (TG270-ULN8-200, TG18-UN80, TG18-AFC), but not the clinical images, the noise difference may not significantly impact clinical use.

For the TG18-AFC pattern, the TG18 report recommends that displays used for diagnostic purposes should be able to see all targets except those in the smallest-target quadrant. For other displays (modality, clinical specialist, EHR), all targets in the two largest-target quadrants should be visible. This report does not recommend strictly adhering to these limits, but does recommend using them as a basis for evaluation.

In general, regardless of the methodology for evaluation, if noise differences are perceptible, the least-noisy display should be the preferred selection.

2.7 Display Temporal Performance

2.7.1 Rationale

The temporal performance of a display directly impacts how images appear to the viewer [Liang et al. 2008], and measuring the effects of temporal performance depends heavily on the timescale. Long-term temporal performance, over the course of months or years, is typically characterized by luminance and uniformity degradation. These changes are related to aging of the display and are ideally documented by maintaining a routine QA program. The methodologies for measuring luminance and uniformity performance are discussed in sections 2.3 and 2.5, respectively. Over shorter time periods (milliseconds to hours), temporal performance metrics such as warm-up time, response time, input latency, and image retention should be understood and characterized to evaluate their potential impact on medical image viewing.

2.7.1.1 Medium-term Temporal Performance (minutes to hours)

The measurement of temporal performance on the order of minutes to hours can be performed either by visual inspection or by using the same equipment (contact or telescopic photometer) that is used for luminance and uniformity measurements. Medium-term temporal performance metrics that fit this description include warm-up time and image retention.

The display warm-up time is defined as the time required for a display to reach a luminance within 5% of the steady-state luminance [IDMS 2012, ch. 10]. Warm-up times can range from a few minutes to almost an hour. Warm-up time does not need to be part of routine display QA. However, the warm-up time of a display may be considered as part of product evaluation. Displays with substantial warm-up time (greater than 30 minutes) may not be appropriate in areas where the display is required to turn on quickly and display images.

Image retention typically results from high-contrast images displayed for long periods of time (hours or longer). This may include poorly-chosen desktop backgrounds or applications with toolbars or menus that are always displayed. Image retention may be on the order of seconds to permanent, depending on the specific display and the length of time the high-contrast object was displayed. Users should be aware of image retention, but like display warm-up, measuring image retention does not need to be included as part of routine QA. Latent images (transient or permanent) may be observed as part of uniformity analysis, and the effects of such artifacts should be evaluated in the context of medical image viewing (size, shape, severity, position, etc.).

2.7.1.2 Short-term Temporal Performance (seconds or less)

Measuring temporal performance on the order of seconds or less, e.g., response time and input latency, requires equipment with a faster measurement speed than that required for other tests described in this report. For routine display QA, quantitatively measuring short-term temporal performance is both unnecessary and well beyond the technical capabilities of most medical physics measurement equipment. For a description of such tests, the reader is referred to IDMS 2012, ch. 10. Qualitative evaluation of these metrics, however, may be of value when characterizing performance as part of a display's proposed-use evaluation or during acceptance testing. It can also be useful if there are concerns of underlying display latencies in the presenting and reviewing of dynamic medical images (e.g., viewing of fluoroscopic images). The remainder of this section describes a method for qualitative assessment of display short-term temporal performance.

2.7.2 Procedure

Qualitative evaluation of response time and latency requires a dynamic test pattern, preferably one with multiple gray-level transitions. The TG270-TR test pattern contains multiple contrast transitions at multiple background levels. See the Test Patterns section of this report for a complete description of the pattern. Once displayed, the user takes a photograph for evaluation of the display performance. By photographing the pattern with a short exposure time (e.g., 1/1000 sec), the user can determine if there are significant delays in gray-level transitions from frame-to-frame. The user should choose a pattern to match the display refresh rate, so if minimal artifacts are observed in a photograph of the pattern, one could assume that the luminance transitions are as fast as the display refresh rate. Modern display refresh rates are typically at least 30 or 60 Hz, which is beyond the refresh rates of most dynamic imaging.

The user should evaluate each background level of the TG270-TR pattern separately; this ensures the dynamic range of the camera is not exceeded, potentially limiting the assessment. Additionally, the evaluation should use multiple photographs of each background level to avoid potential artifacts that may occur with photographing a frame transition Figure 11c.
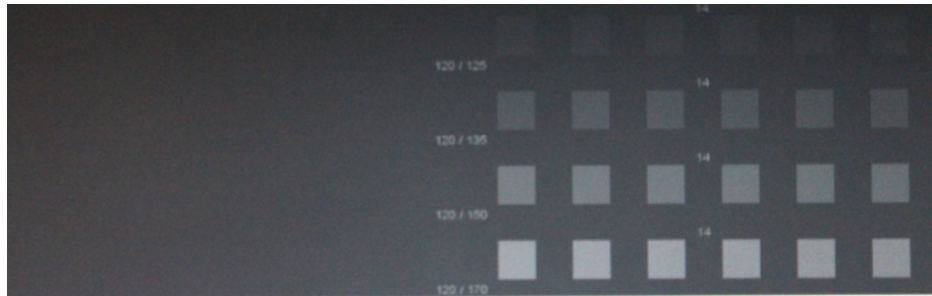
(a) No extra frames visible

(b) One extra frame visible

(c) Frame transition

FIG. 11. Photographic examples of the TG270-TR pattern. Each photograph is of a specific background section of the pattern (8-bit gray level 120 in this example). The camera used in these examples set the exposure time to 1/500 sec.

2.7.3 Assessment

The photographs of the dynamic test pattern give the user information on how the display performs gray-to-gray transitions at multiple levels of the display. The display has some inherent delay in changing between gray levels, and this delay is determined by both the start and end gray level. If the magnitude of the delay is on the order of the frame rate of the display, the photograph of the test pattern will not show more than one extra frame image Figures 11a and 11b. If there is a delay greater than the refresh rate of the display, additional frame images will appear in the photograph. The number of visible frames is determined by the severity of the response time delay. For example, if three additional boxes appear on the TG270-TR pattern, one can assume the delay is on the order of three frames. For a 60 Hz display, three frames is approximately 50 ms.

2.7.4 Suggested Limits

This report does not make any specific recommendations for limits when assessing a display's response time or input latency. However, in cases where there is a delay much greater than two display refresh frames, the user should assess the potential impact on clinical image assessment. For example, if a medical image is refreshed at a rate of 30 Hz (as in fluoroscopy), a delay in the display greater than 30 ms has the potential for impacting clinical image quality.

2.8 Display Spatial Resolution

2.8.1 Rationale

In modern flat-panel displays (LCD or OLED) used to view medical images, a small amount of light emitted from each pixel is dispersed to neighboring pixels. For LCD devices, ACR–AAPM–SIIM guidelines recommend that displays have advanced pixel structures such as "in plane switching" (IPS) or "vertical alignment" (VA) [ACR-AAPM-SIIM 2017]. These pixel structures typically have black material between the pixel elements which reduces diffuse reflection and helps minimize light dispersion. The guideline further recommends that only digital graphic interfaces be used (DVI-D, DisplayPort) and that the graphic controller driver be configured for the number of native rows and columns of the display.

If these guidelines are followed, it is not necessary to quantitatively measure spatial resolution. It is sufficient to verify the graphic driver settings and visually verify that there is only minimal emitted light dispersion. When evaluating displays on a workstation, the graphic driver settings should be reviewed and compared to the number of native rows and columns of each display. Displays store the native row and column size internally so that drivers can access this information. Utility programs to automatically obtain the graphic driver settings and the native row and column values can be used to quickly perform this verification (see Workstation Configuration and Documentation on page 7).

Most current medical displays have square pixels with red, green and blue sub-pixels of similar size. Horizontal and vertical line pair resolution patterns naturally align with the rows and column of square pixels. Display devices with significantly different arrangements of sub-pixels have recently been introduced. PenTile devices (Samsung) have green subpixels typically arranged in rows and columns. However they have half as many red and blue subpixels that are often arranged in a diagonal pattern. The device driver maps the rows and columns of an image to the sub-pixel layout. As a result, horizontal and vertical resolution test patterns do not naturally align to the sub-pixel structure. As new architectures are introduced, the pixel structure of a new display should be considered.

2.8.2 Procedure and Assessment

When the graphic driver settings are correctly set, a test pattern with alternative white and black lines with one line pair every two pixels can be used to verify minimal dispersion. The test region should be viewed with an optical loupe of about 5X magnification to verify high contrast. The dark lines should appear black as opposed to dark gray (Figure 12). If this is verified at the time of purchase it should not change over the lifetime of the device, although periodic visual verification can quickly detect improper graphic driver settings.

<center>(a)   (b)</center>
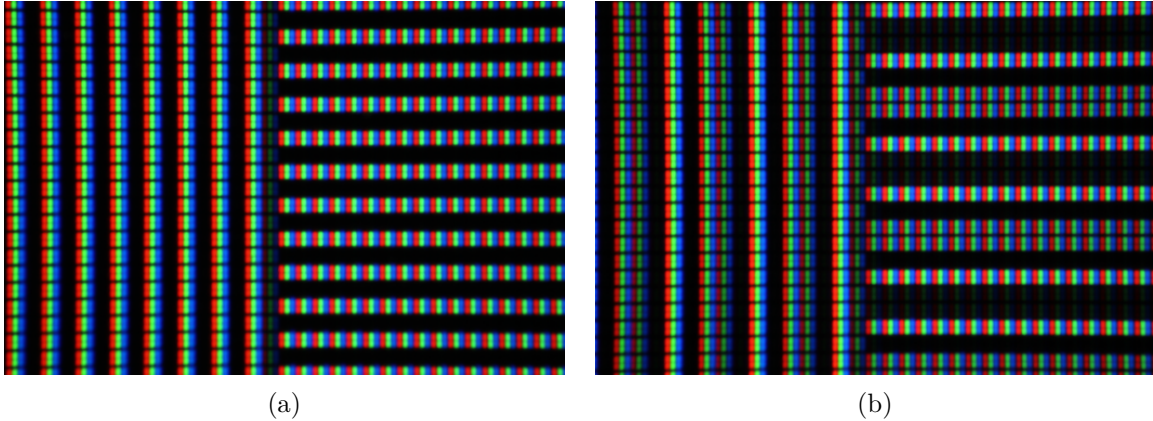
FIG. 12. Photographs of display pixels under magnification. In both images, the pixels (and red, green, and blue subpixels) are displaying the high-contrast, alternating black-and-white line pair region of the TG270-sQC pattern. The image on the left is from a display with a graphics driver set to the native resolution of the display and the pattern zoom set to 100%. Full white lines appear as bright red, green, and blue triplets in both the horizontal and the vertical direction. There is no visual evidence of luminance dispersion between the pixel columns or rows. On the right, the display's graphics have been set to a non-native resolution in both directions. As a result, the representation of the pattern is distorted with black, white, and gray (neither full brightness nor darkness RGB) lines.

2.8.3   Note on Pixel Pitch

In addition to assessing the driver settings and light dispersion, the pixel size should also be considered. When a display is viewed from a normal distance, the pixel pitch should be sufficiently small that the observer does not see pixel pattern structure. Rather, the viewer should see smooth tones and well defined edges. The recommended viewing distance for a radiologists workstation is 65 cm (about an arm's length) in order to reduce visual strain [Goyal et al. 2009; Harisinghani et al. 2004]. At this distance, a pixel pitch of 210 μm will keep the display free of pixel structure artifacts. Modality and clinical displays are often viewed while performing other functions, and the viewing distance is often larger. As a result, this report suggests a pixel pitch less than 250 μm for these devices. These recommendation are based on well-established properties of the human visual system and are consistent with guidelines for other markets (television, tablets, etc). From a quality assurance standpoint, users should carefully consider these guidelines when evaluating displays.

The above recommendation for a 210 μm pixel pitch for diagnostic displays suggests the use of a 3 megapixel (MP) display (or 6 MP for a wide-format display). Historically, there has been a preference to use 5 MP displays for the purposes of mammographic interpretation. This dates to the initial use of electronic imaging for medical interpretations when CRT technology was widespread. CRT displays demonstrate anisotropic spatial resolution, with particularly degraded resolution in the scanning direction [Blume 1999]. As a result of this characteristic, 5 MP CRT displays with quoted pixel sizes of approximately 150 μm will have degraded resolution due to blurring effects. Modern LCDs do not exhibit the same behavior [Blume et al. 2002], and so the pixel size appears closer to the physical dimension.

<center>45</center>

If the recommendation for a 210 μm pixel pitch is followed, the pixel structure will not be perceptible at the typical viewing distance. Modern 3 or 5 MP flat-panel diagnostic displays perform at levels sufficient for all diagnostic interpretation [Kamitani et al. 2011]. It is important to note that in the instance of mammography in the United States, there is no explicit Mammography Quality Standards Act (MQSA) requirement for the minimum number of pixels in a diagnostic breast imaging display for diagnostic review. Under MQSA, a display must simply meet the image receptor manufacturer, display manufacturer, or alternative standard quality assurance guidelines. Such guidelines do not generally specify the display array size.

3 Summary of Tests and Recommendations

Tables X and XI summarize the QA tests, equipment, test patterns, and passing criteria from the Performance Characteristics and Quality Control section of this report. Readers should use these tables as a quick reference, and refer to the relevant sections elsewhere in this report for a full description of procedures, rationales, and assessments. The reader should view the passing criteria presented in Tables X and XI only as suggestions, and medical physicist should use this document and its references to establish passing criteria for their sites. Different passing criteria may be appropriate where uses of displays differ from those described in this report. In addition, facilities must be sure to adhere to applicable laws and regulations, which may not necessarily agree with the limits presented in this document.

Table XII presents suggested display QA testing frequencies. The frequencies for the various tests are for a documented QA program, with either medical physicists or personnel under medical physicist supervision performing the QA tests. The suggested frequencies in Table XII are based on the relative importance of display performance as part of the image review chain. Table XII also groups related tests so that different tests may be performed at the same time using a single test pattern (e.g., Qualitative Luminance Response, Ambient Luminance, Uniformity, and Spatial Resolution may all be evaluated with the TG270-sQC pattern). In addition to the tests performed as part of a documented QA program, it is often useful to have users perform quick visual checks between tests. This may include a test pattern check at log in, or a display background with a test pattern.

The person responsible for display QA should establish a QA program with considerations for both personnel and testing capabilities. Facilities with a large deployment of displays, but limited testing resources, may find automated remote testing of displays to be the best solution for display QA. In an instance where remote testing is automated for the majority of the QA, frequencies for qualitative vs. quantitative testing in Table XII may be changed, with quantitative automated testing being performed more frequently, and qualified personnel performing qualitative testing less frequently. Regardless of the testing frequencies chosen for a QA program, Table XII should be used as a guide on what tests and display types are the most important parts of the image review chain.

TABLE X. Suggested Display QA Testing Criteria

Documented QA Tests	Procedure	Equipment	Patterns[a]	Suggested Passing Criteria	
				Diagnostic	Modality, CS, EHR[b]
Quantitative Ambient Luminance/Illuminance	TG270 2.3.1.3 (p. 11)	Photometer	NA (Display Off)	AR $(L_{amb}/L_{min}) \leq 1/4$ Illuminance 25–75 lux	AR $(L_{amb}/L_{min}) \leq 1/4$
Qualitative Ambient Luminance/Illuminance	TG270 2.3.1.6 (p. 14)	None	TG270-sQC TG270-pQC TG18-OIQ[c]	Low-contrast features in darkest region visible in both no-light and normal-light settings[d]	
Quantitative Min/Max Luminance	TG270 2.3.2.3 (p. 22)	Photometer	TG270-sQC TG270-ULN TG18-LN	$250 < LR < 450$ $L'_{min} > 1.0$ cd/m^2 $L'_{max} > 350$ cd/m^2	$250 < LR < 450$ $L'_{min} > 0.8$ cd/m^2 $L'_{max} > 250$ cd/m^2
Quantitative Luminance Response	TG270 2.3.2.3 (p. 22)	Photometer	TG270-ULN TG18-LN TG270-sQC	Deviation from DICOM GSDF $< 10\%$	Deviation from DICOM GSDF $< 20\%$
Qualitative Luminance Response	TG270 2.3.2.5 (p. 28)	None	TG270-sQC TG270-pQC TG18-OIQ[c]	All low-contrast features visible under typical conditions	TG270-sQC: All ± 5 gray level patterns visible TG270-pQC: All ± 4 gray level patterns visible
Quantitative Color Assessment	TG270 2.4.3 (p. 32)	Colorimeter	TG270-ULN TG18-LN	$\Delta_{D65} (u', v') \leq 0.01$ Difference between two displays: $\Delta (u', v') \leq 0.01$ (same workstation) $\Delta (u', v') \leq 0.02$ (different workstation)	$\Delta_{D65} (u', v') \leq 0.02$

[a] Though widely available, the SMPTE pattern should only be used if an updated pattern is unavailable.
[b] CS: Clinical Specialist; EHR: Electronic Health Record
[c] TG18-OIQ is the IEC-modified version of the TG18-QC pattern, removing the CRT-specific features. The TG18-QC pattern is also acceptable.
[d] The TG18-AD pattern may be used for qualitative assessment, but the passing criteria must be determined based on display use and physicist's evaluation of clinical impact.

TABLE XI. Suggested Display QA Testing Criteria

Documented QA Tests	Procedure	Equipment	Patterns[a]	Suggested Passing Criteria	
				Diagnostic	Modality, CS, EHR[b]
Quantitative Uniformity	TG270 2.5.2 (p. 37)	Photometer	TG270-ULN TG18-UL	LUDM $< 30\%$ (if $> 15\%$, evaluate qualitatively and determine clinical impact)	
Qualitative Uniformity	TG270 2.5.3 (p. 38)	None	TG270-ULN TG18-UL	No non-uniformities that impact clinical use	
Qualitative Noise	TG270 2.6.3 (p. 40)	None	TG18-AFC TG270-ULN	No noise effects that impact clinical use	
Qualitative Temporal Resolution	TG270 2.7.1.2 (p. 42)	Camera, Photometer	TG270-TR	No temporal effects that impact clinical use	
Qualitative Spatial Resolution	TG270 2.8.2 (p. 44)	Loupe	TG270-sQC TG270-pQC TG18-OIQ[c]	Pixel structure not visible at typical working distance One-to-one pixel mapping from graphics card	
Diffuse Reflection Coefficient (R_d)	TG18 4.2.4.1.2	Photometer (Telescopic)	None (Display Off)	R_d must be low enough for typical ambient lighting levels	

[a] Though widely available, the SMPTE pattern should only be used if an updated pattern is unavailable.
[b] CS: Clinical Specialist; EHR: Electronic Health Record
[c] TG18-OIQ is the IEC-modified version of the TG18-QC pattern, removing the CRT-specific features. The TG18-QC pattern is also acceptable.

TABLE XII. Suggested Display QA Testing Frequencies

Documented QA Test	Display Type			
	Diagnostic	Modality	Clinical Specialist	Electronic Health Record
Qualitative Luminance Response	Quarterly	Quarterly	Annually	Annually
Qualitative Ambient Luminance/Illuminance	Quarterly	Annually	Annually	Annually
Qualitative Uniformity	Quarterly	Annually	Annually	Annually
Qualitative Spatial Resolution	Quarterly	Annually	Annually	Annually
Quantitative Min/Max Luminance	Annually	Annually	Annually	Acceptance
Quantitative Luminance Response	Annually	Annually	Acceptance	Acceptance
Quantitative Color Assessment	Annually	Annually	Acceptance	Acceptance
Quantitative Ambient Luminance/Illuminance	Annually	Acceptance	Acceptance	Acceptance
Quantitative Uniformity	Acceptance	Acceptance	Acceptance	Evaluation
Qualitative Noise	Evaluation	Evaluation	Evaluation	Evaluation
Qualitative Temporal Resolution	Evaluation	Evaluation	Evaluation	Evaluation
Diffuse Reflection Coefficient (R_d)	Evaluation	Evaluation	Evaluation	Evaluation

4 TEST PATTERNS

Display test patterns are an essential tool for display quality assurance. Both qualitative and quantitative QA use test patterns, with a variety of patterns needed for the full complement of tests. Tables X and XI list test patterns for each of the recommended QA tests. This section provides a description of both existing and new test patterns. The new test patterns were developed to simplify, modify, or combine existing patterns specifically for the needs of QA for flat-panel displays.

The user must display a test pattern using the appropriate display settings (e.g., window and level). Test patterns are typically displayed using the full dynamic range of the display (e.g., window width 256, window level 128 on an 8-bit display).

4.1 Existing Patterns

The test patterns described in this section may be included with existing QA software and installed on modalities. The list should not be considered comprehensive for useful patterns, and the qualified physicist or display QA personnel is left to decide what patterns to use to complete any display QA procedure.

4.1.1 TG18-QC (qualitative quality control)

With the publication of the TG18 report, many new test patterns were created to aid in the evaluation of displays [TG18 2005]. The TG18-QC test pattern has been the mostly widely adopted pattern from the report, and has supplanted the SMPTE test pattern on some systems as the default qualitative assessment pattern. Readers should refer to the TG18 report for a full description of this pattern. Like the SMPTE pattern, TG18-QC has several features that are unique to the evaluation of CRT displays, and can now be considered unnecessary. The IEC 62563-1 report provides an updated version of TG18-QC (TG18-OIQ) where CRT-specific features have been removed. The TG18-QC and TG18-OIQ section of the qualitative luminance assessment provides a description of the qualitative pattern features of the TG18-OIQ pattern, along with considerations to make when using the pattern as part of qualitative luminance assessment.

4.1.2 TG18-LN (luminance response)

The TG18-LN test pattern series provides 18 patterns to assess the luminance response of the display [TG18 2005]. Each image in the series displays a uniform field at a pixel value of 20% of a display's maximum luminance, with a central test region occupying 10% of the image. The pixel value of the central region varies from 0 to 255 for the 8-bit version (TG18-LN8) and 0 to 4080 for the 12-bit version (TG18-LN12), with a 15 pixel value separation between each 8-bit pattern and a 240 pixel value separation between each 12-bit pattern. The user measures the display luminance response in the central region of each test pattern. While the TG18-LN test pattern series may still be used for this purpose, it should be noted that it does have a few drawbacks. As mentioned in the Luminance Response section, the luminance response for flat-panel displays may not follow a smooth, predictable function. Measurements of the luminance response have an increased likelihood

of missing discontinuities or calibration errors with fewer measurement points (Figure 3). Additionally, the background region of the TG18-LN series always has the same pixel value, which was specified to account for veiling glare in CRT devices. As a result, users cannot perform evaluations of luminance response or uniformity outside of the central region. The TG270-ULN pattern series addresses these shortcoming by providing a full grayscale palette of test patterns with the measurement region covering the entire pattern.

4.1.3   TG18-UN (uniformity)

The TG18 report introduced the TG18-UN10 and TG18-UN80 test patterns for the assessment of luminance uniformity, color uniformity, and angular response. These patterns display a uniform image at 10% (TG18-UN10) and 80% (TG18-UN80) of the maximum pixel value. Two additional uniformity patterns from TG18, TG18-UNL10 and TG18-UNL80, are identical to the UN10 and UN80 patterns but include low-contrast lines to help identify the corners and the center of the pattern for quantitative uniformity measurements. While these test patterns may still be useful for this purpose, it should be noted that they allow for quantitative measurement of uniformity at two gray levels. To address this shortcoming, this report defines the TG270-ULN patterns which allow for uniformity testing at any selection of gray levels.

4.1.4   TG18-AFC (display noise)

The TG18-AFC test pattern provides a visual test of the signal-to-noise characteristics of a display. Each quadrant of the pattern consists of a series of small squares with each having a small, low-contrast target feature randomly placed in a corner. Both the target feature size and the contrast differ for each quadrant. For the 12-bit, 1024 x 1024 pattern, the quadrants have targets with contrast values of +32, 48, 64, and 96 gray levels, and corresponding target sizes of 2, 3, 4, and 6 pixels. When displayed with a window width of 4096, the target contrast correspond to 2, 3, 4, and 5 gray levels for a 256-level graphic driver. The size and contrast are scaled accordingly for both the 2048 x 2048 and the 8-bit versions of the pattern. Five larger squares containing a contrast-detail Rose pattern with varying target sizes and contrasts are included in the center and the four corners.

4.1.5   TG18-AD (low-luminance, low-contrast)

The TG18-AD test pattern is used for the assessment of the impact of ambient lighting conditions on the ability to visualize low-contrast modulating bar patterns under low-luminance conditions. The pattern contains a seven-by-seven grid of modulating, horizontal bar patterns, varying in contrast from 1 to 49 from a background gray level of 0 (8-bit display values). The user evaluates the pattern under ambient lighting conditions to determine the change in visualization threshold from total-darkness conditions. Any reduction in the number of bar patterns that are visible represents the contrast reduction due to ambient lighting conditions. This report recommends the use of the TG18-AD pattern as an additional qualitative tool to evaluate the effects of ambient lighting.

4.1.6   SMPTE

The Society of Motion Picture and Television Engineers (SMPTE) Medical Diagnostic Imaging Test Pattern, commonly referred to as the SMPTE pattern, was first formally described

in the mid 1980s [Gray et al. 1985; SMPTE 1986]. At the time, the pattern quickly established itself as useful for evaluation and visual calibration of medical displays. Since its introduction, the SMPTE pattern and has become widely used as the default pattern for display QA and can be found as part of most display QA software tools and modalities. As discussed in the SMPTE section of the qualitative luminance assessment, the SMPTE pattern's features are primarily CRT-centric and have reduced sensitivity to the luminance response capabilities of modern flat-panel displays. While the pattern may still be of limited utility when no other patterns are available, **this report considers the SMPTE test pattern deprecated in favor of updated patterns.**

4.2   New Test Patterns

This report describes several new test patterns for display quality assurance. These patterns are intended to provide physicists and other display QA personnel with updated patterns, specifically designed for flat-panel displays. While existing patterns from the TG18 report (and elsewhere) may still be useful for display QA, many were designed in the CRT era, and either have CRT-specific features, or are not sensitive enough for the improved performance of modern displays. The two primary new patterns, TG270-sQC and TG270-pQC, provide general display QA patterns for both simple QC (sQC) and advanced physics QC (pQC). The appendix of this report provides ImageJ [ImageJ 2016] macros (`.ijm`) to generate these patterns.

4.2.1   TG270-sQC (simple QC)

The TG270-sQC (simple QC) test pattern (Figure 13) is for a quick visual evaluation of display performance. Table XIII provides a detailed description of each pattern feature. The pattern contains three rows of six incrementing grayscale patches (for a total of 18 patches) that cover the 8-bit grayscale range (0 to 255) in steps of 15 (identical to the gray levels used in the traditional 18-point TG18 luminance response measurement). Each patch contains a low-contrast bar pattern in the upper left and lower right corner. For a DICOM-conformant display, the user should be able to quickly scan the three rows of patches and assess if the bar patterns in each are equally visible across all 18 gray levels. The 0 and 255 patches contain two levels of low-contrast bar patterns, and the lowest contrast patterns may only be visible on the highest-performance displays.
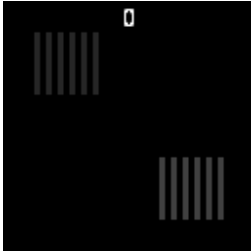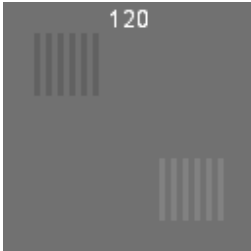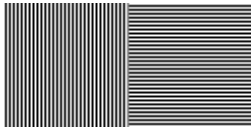
The TG270-sQC pattern also includes three larger patches for minimum and maximum luminance calculations (and therefore luminance ratio). These three patches (black, mid-gray, white) can be zoomed and panned around the display for both qualitative and quantitative uniformity assessment. Furthermore, any of the smaller, numbered luminance patches may be zoomed and panned for similar use.

The bottom of the pattern includes a grayscale gradient with continuous pixel value variation for evaluating bit-depth issues, contouring artifacts, and grayscale calibration errors. At the side of the gradient pattern, alternating black and white lines in the horizontal and vertical direction can be used to verify correct pixel mapping.

4.2.2   TG270-pQC (physics QC)

The TG270-pQC (physics QC) pattern (Figure 14) is for physicists and other advanced users to more thoroughly visually evaluate display performance. Table XIV provides a full

TABLE XIII. TG270-sQC Pattern Feature Descriptions

Image[a]	Description	Uses
	The 0 luminance patch has a background value of 0 with bar patterns in the upper left and lower right. The upper left bar pattern has an 8-bit pixel value of 3, while the lower right is 5. The bars are each three pixels wide, with a three pixel spacing.	Quantitative and qualitative luminance response assessment, qualitative ambient light assessment
	The 15-240 luminance patches (120 shown here) have 8-bit value backgrounds equal to their respective numbers. The bars have a spacing of three pixels and width of three pixels. The bars vary from the background by -5 in the upper left and +5 in the lower right.	Quantitative and qualitative luminance response assessment, quantitative and qualitative uniformity assessment, quantitative color assessment
	The 255 luminance patch has an 8-bit background value of 255 with bar patterns in the upper left and lower right. The upper left bar pattern has an 8-bit pixel value of 250, while the lower right is 252. The bars are each three pixels wide, with a three pixel spacing.	Quantitative and qualitative luminance response assessment, quantitative and qualitative uniformity assessment, quantitative color assessment
	Larger luminance patches at 8-bit gray levels 0, 128, and 255	L_{min} and L_{max} assessment, qualitative and quantitative luminance uniformity
	The high-contrast bar patterns have alternating white (8-bit pixel value 255) and black (pixel value 0) one-pixel-width bars in both the horizontal and the vertical directions.	Qualitative spatial resolution assessment (verify one-to-one pixel mapping)
	The continuous grayscale gradient along the bottom of the image shows all 8-bit gray levels (0-255).	Luminance response calibration errors, bit-depth artifacts, contouring artifacts

[a] Contrast may be enhanced to better show pattern features
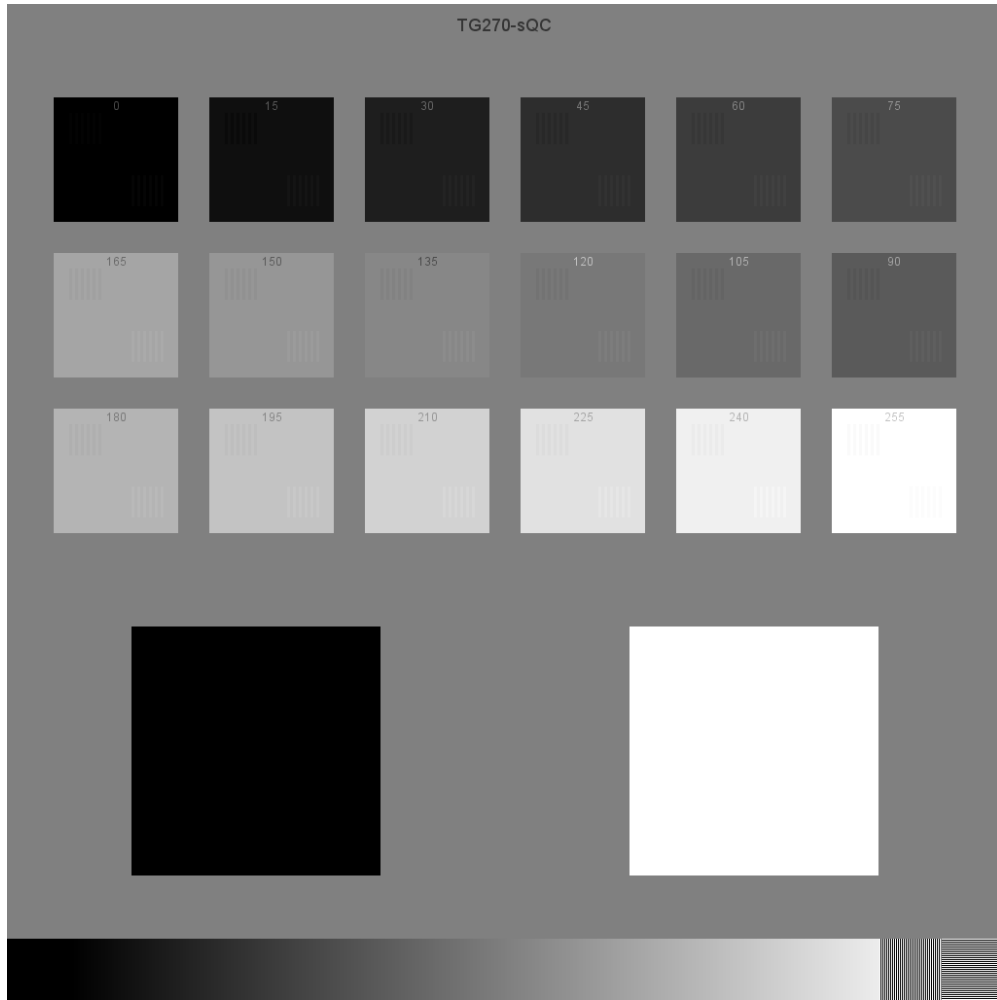
54

FIG. 13. TG270-sQC Quality Control Test Pattern

description of the pattern features. Like the TG270-sQC pattern, the TG270-pQC pattern contains rows that cover the 8-bit grayscale range (0 to 255) in steps of 15 (identical to the gray levels used in the traditional 18-point TG18 luminance response measurement). Within each gray level row, central square is a uniform patch for the respective gray level. Within the remainder of each row, the pattern shows multiple frequencies of bar patterns, and does so at two contrasts. The bar patterns are in both the vertical and the horizontal direction.

Along the sides of the pattern, two continuous grayscale ramps are provided for bit-depth, contouring, and grayscale calibration errors. Modulating sinusoids are overlaid atop the grayscale ramps, providing additional visual checks for luminance response calibration. At the top and bottom of the pattern, high-contrast bar patterns of multiple frequencies in both directions provide the user with spatial resolution and pixel mapping evaluation pattern features.

To use the pattern, the user visually scans the pattern up and down for a given frequency, direction, and contrast to determine the visibility of the pattern throughout the full luminance range. For displays that are calibrated to follow the DICOM GSDF, the visibility of a given frequency and contrast should be similar in appearance as the user visually scans ver-
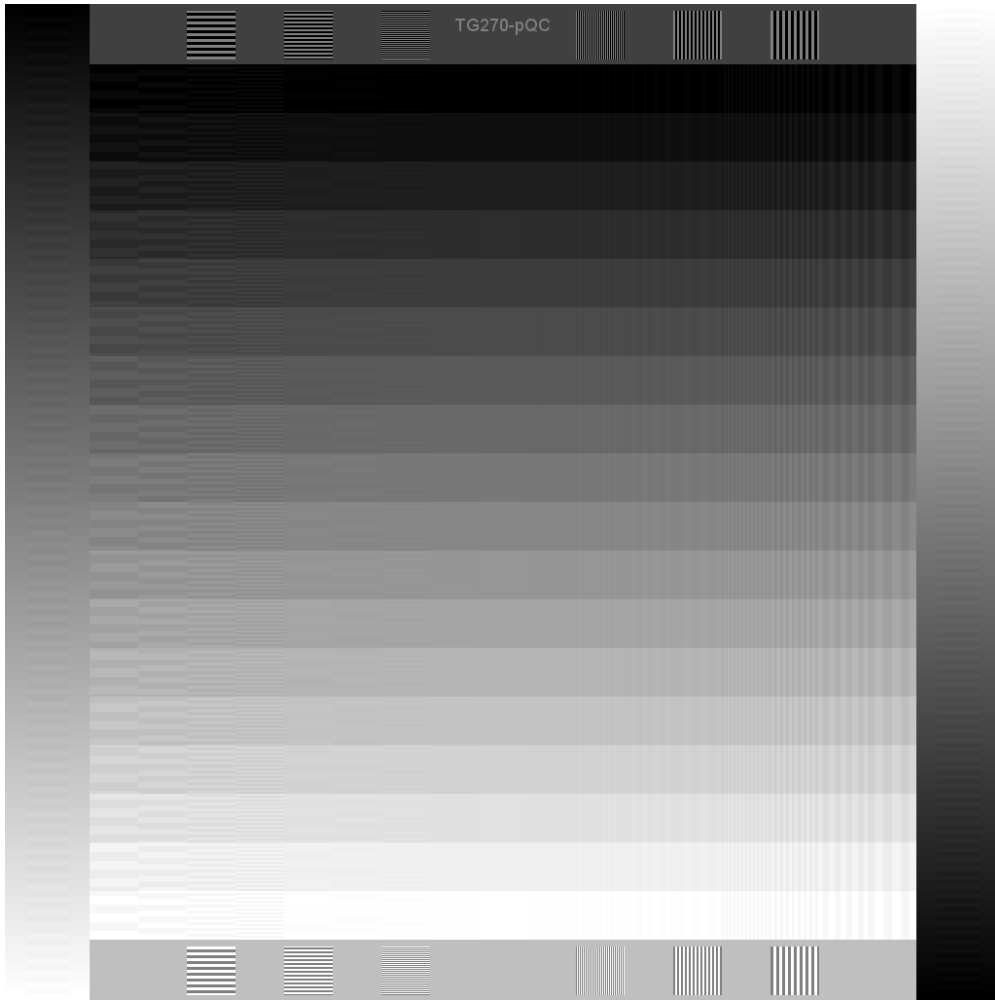
FIG. 14. TG270-pQC Quality Control Test Pattern

tically through gray levels. If a display is not calibrated, some rows may appear more visible than others. If the quantitative assessment of a display shows higher contrast in central gray levels than at the high or low end, the central row bar patterns should appear more visible than the top or bottom rows. Because the DICOM GSDF is developed using a varied visual adaptation model, and this pattern presents all luminance levels at once, the top and bottom rows may appear less visible than the central rows even for a correctly-calibrated display. Users are encouraged to view this pattern as part of a quantitative evaluation to gain an appreciation for how the pattern behaves under properly calibrated conditions. Users that are familiar with this pattern under optimal conditions can use the pattern to help assess display QA failures and their potential clinical impact.

4.2.3 TG270-ULN (combined uniformity and luminance response)

The TG270-ULN series of patterns provides an update to both the TG18-LN and the TG18-UN pattern series. The patterns are designated as TG270-ULNx-y, where x is the bit-depth and y is the gray level. For example, the pattern shown in Figure 15 is the TG270-ULN8-127 pattern, indicating that it is gray level 127 of an 8-bit display range. The TG270-ULN

TABLE XIV. TG270-pQC Pattern Feature Descriptions

Image[a]	Description	Uses
	Each row contains bar patterns of several contrasts and frequencies. Moving away from the central column, the inner four patches on either side have 8-bit contrast of 2 gray levels, and the outer four have contrast of 8 gray levels. The four frequencies are 4, 6, 12, and 18 pixels per line pair. The right side is vertically oriented, and the left is horizontal.	Qualitative luminance response assessment, qualitative ambient light assessment
	The central column of the pattern has 18 luminance patches covering the full luminance range (8-bit steps of 15 gray levels). The luminance patches are equal to the average gray level value of the bar patterns on either side.	Quantitative luminance response assessment, quantitative and qualitative uniformity assessment, quantitative color assessment
	High-contrast bar patterns are located at the top and bottom of the pattern (8-bit gray levels 0 and 128 on the bottom, 128 and 255 on the top). The frequencies of the patterns are 2, 4, and 6 pixels per line pair.	Qualitative spatial resolution assessment (verify one-to-one pixel mapping)
	The right and left sides of the pattern have gradients that run from minimum to maximum luminance. There is a sinusoidal pattern on each gradient, with an amplitude of 3 8-bit gray levels, and frequencies of 3π on the right and 4π on the left.	Luminance response calibration errors, bit-depth artifacts, contouring artifacts

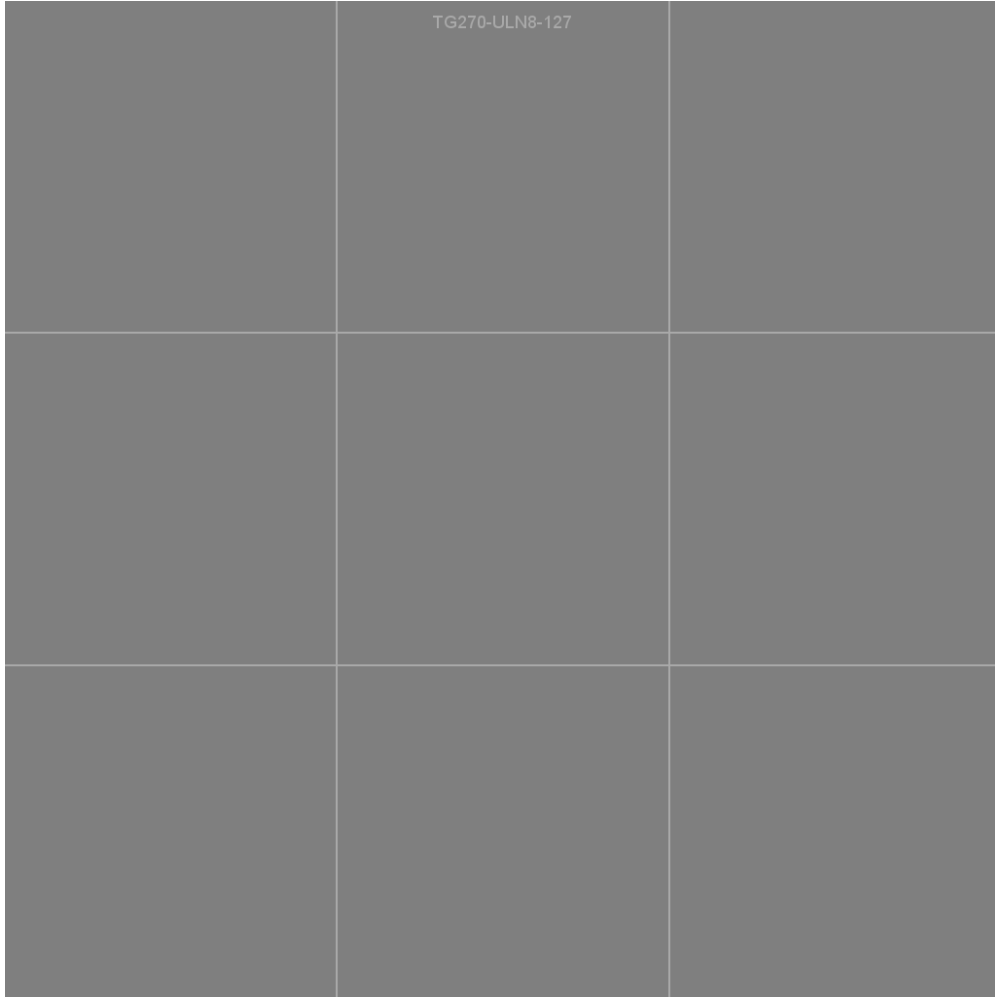[a] Contrast may be enhanced to better show pattern features

FIG. 15. TG270-ULN Uniformity and Luminance Test Pattern (TG270-ULN8-127 with background 8-bit gray level 127 shown)

pattern series uses the gray level of the pattern (e.g., 127 in the TG270-ULN8-127 pattern) as the value for the entire background of the pattern. Low-contrast lines separate the pattern into nine regions, which aid in the uniformity measurement process.

4.2.4   TG270-TR (temporal resolution)

The TG270-TR test pattern (Figure 16) is for evaluation of temporal performance of a display. It is multi-frame pattern, designed to be displayed so that the frame rate matches the refresh rate of the display (typically 30 or 60 Hz). It contains five background regions, each of which contains several contrast levels of temporal patches. The three top-most regions have 8-bit background values of 15, 120, and 245. Within each of these regions, four rows of six patches move across the pattern as the frames progress. The four rows within each region have contrasts of 5, 15, 30, and 50 8-bit gray levels. In the 15 and 120 region, the patches are higher in pixel value from the background, while in the 245 region the patches are of a lower value. The bottom two regions of the pattern have 8-bit background levels
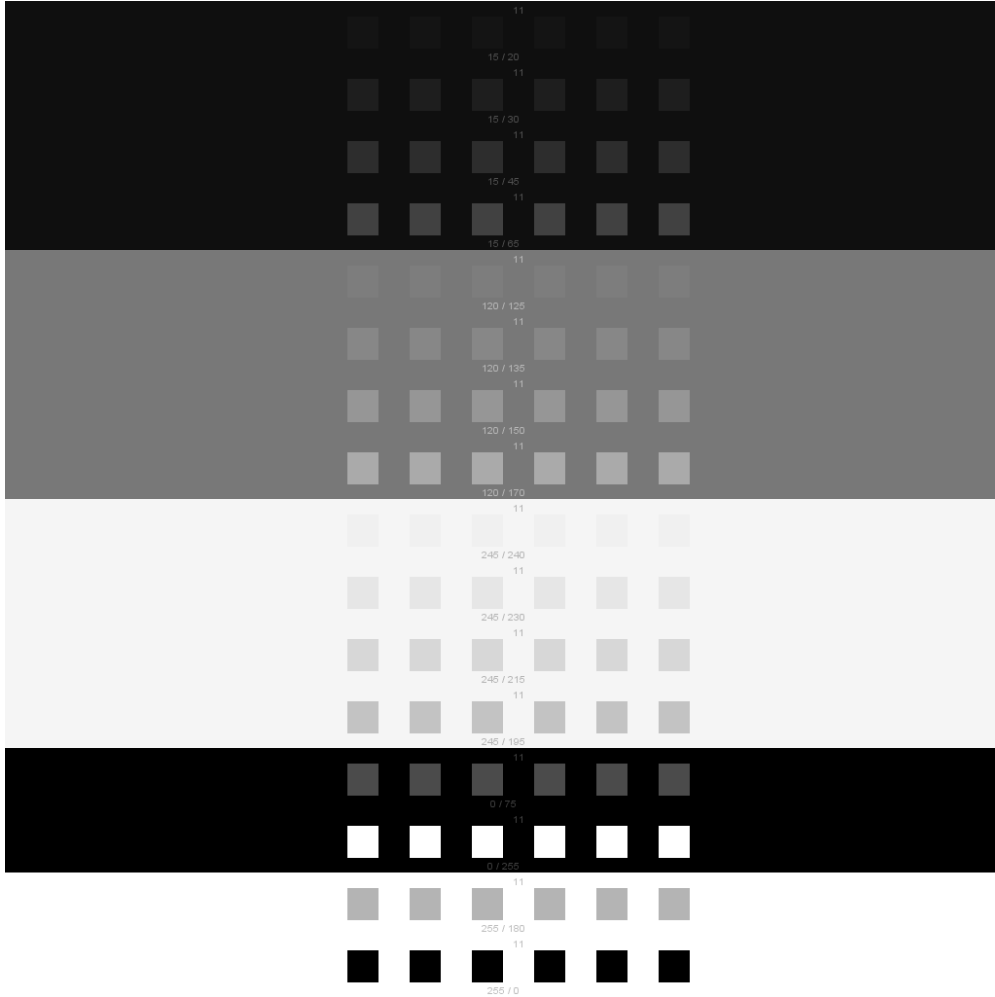
FIG. 16. Frame 11 of the TG270-TR Temporal Resolution Test Pattern

of 0 and 255, with contrasts of 75 and 255. For all regions, each row is labeled with its respective background and patch gray level (8-bit). In addition, each frame is labeled in each row just above the center of the six patches. The top three regions are used to evaluate the temporal performance of the gray-to-gray transitions at several luminance levels. The bottom two regions can be used to evaluate larger contrast transitions, which are more commonly quoted as part of a display's technical capabilities.

The pattern can be stored as a mutliframe object (e.g., DICOM) and displayed using software that can control the display frame rate. Animated gif versions of the pattern for both 30 Hz and 60 Hz displays provide a convenient means for displaying the patterns on a wide variety of systems that may not have software for controlling the frame rate. Once the moving pattern is displayed, the user should use a photographic camera with a short exposure time (e.g., 1/500 ms) to take several photographs of one region of the pattern (Figure 11) for evaluation of temporal performance.

5   CONCLUSIONS

The presentation of medical images on appropriately-calibrated display devices is essential for ensuring that information obtained from imaging exams provides the highest diagnostic value. A rigorous quality assurance program can ensure both that displays are functioning properly and that images have a consistent appearance throughout the imaging chain. To this end, the purpose of this report was to provide recommendations for the assessment of display quality for LCD and OLED medical displays. The quality assurance procedures, testing frequencies, and performance criteria provided here may be used as guidelines to establish a display quality assurance program. It should be noted that the implementation of these recommendations may vary depending both on how individual displays are being utilized and on the needs of specific facilities. A qualified medial physicist or other personnel with training specific to display assessment should use these recommendations in conjunction with a facility's specific needs to determine the most appropriate implementation when designing a quality assurance program for displays.

A. Badano, R. M. Gagne, R. J. Jennings, S. E. Drilling, B. R. Imhoff, and E. Muka. Noise in Flat-panel Displays with Subpixel Structure. *Medical Physics*, 31(4):715–723, 2004. doi: 10.1118/1.1656529.

A. Badano, J. Wang, P. Boynton, P. L. Callet, W.-C. Cheng, D. Deroo, M. J. Flynn, T. Matsui, J. Penczek, C. Revie, E. Samei, P. M. Steven, S. Swiderski, G. V. Hoey, M. Yamaguchi, M. Hasegawa, and B. V. Nagy. Technical Note: Gray tracking in medical color displays-A report of Task Group 196. *Medical Physics*, 43(7):4017–4022, 2016. doi:10.1118/1.4953186.

P. G. J. Barten. *Contrast Sensitivity of the Human Eye and its Effects on Image Quality*, volume PM72. SPIE Press, 1999. ISBN 9780819478498. doi:10.1117/3.353254.

B. Baxter, H. Ravindra, and R. Normann. Changes in Lesion Detectability Caused by Light Adaptation in Retinal Photoreceptors. *Investigative Radiology*, 17(4):394–401, 1981.

H. Blume. CRT-Based Display Systems in Radiology. *SID Symposium Digest of Technical Papers*, 30(1):968–971, 1999. ISSN 2168-0159. doi:10.1889/1.1834186.

H. Blume, P. Steven, M. Cobb, A. Ho, F. Stevens, H. Roehrig, and J. Fan. Characterization of Physical Image Quality of a 3-Million-Pixel Monochrome AMLCD: DICOM Conformance, Spatial MTF, and Spatial NPS. *SID Symposium Digest of Technical Papers*, 33(1):86–89, 2002. ISSN 2168-0159. doi:10.1889/1.1830920.

P. C. Brennan, M. McEntee, M. Evanoff, P. Phillips, W. T. O'Connor, and D. J. Manning. Ambient Lighting: Effect of Illumination on Soft-copy Viewing of Radiographs of the Wrist. *American Journal of Roentgenology*, 188(2):W177–W180, 2007. doi:10.2214/AJR.05.2048.

A. Chawla and E. Samei. Ambient Illumination Revisited: A New Adaption-based Approach for Optimizing Medical Imaging Reading Environments. *Medical Physics*, 34(1):81–90, 2007. doi:10.1118/1.2402583.

E. Ekpo and M. McEntee. An Evaluation of Performance Characteristics of Primary Display Devices. *Journal of Digital Imaging*, 29(2):175–182, 2016. doi:10.1007/s10278-015-9831-3.

J. Fan, H. Roehrig, M. K. Sundareshan, E. Krupinski, W. J. Dallas, and K. Gandhi. Evaluation of and Compensation for Spatial Noise of LCDs in Medical Applications. *Medical Physics*, 32(2):578–587, 2005. doi:10.1118/1.1855031.

M. J. Flynn, J. Kanicki, A. Badano, and W. R. Eyler. High-Fidelity Electronic Display of Digital Radiographs. *RadioGraphics*, 19(6):1653–1669, 1999. doi: 10.1148/radiographics.19.6.g99no081653.

N. Goyal, N. Jain, and V. Rachapalli. Ergonomics in Radiology. *Clinical Radiology*, 64(2):119 – 126, 2009. ISSN 0009-9260. doi:10.1016/j.crad.2008.08.003.

J. E. Gray, K. G. Lisk, D. H. Haddick, J. H. Harshbarger, A. Oosterhof, and R. Schwenker. Test Pattern for Video Displays and Hard-copy Cameras. *Radiology*, 154(2):519–527, 1985. doi: 10.1148/radiology.154.2.3966138.

D. S. Groth, S. N. Bernatz, K. A. Fetterly, and N. J. Hangiandreou. Cathode Ray Tube Quality Control and Acceptance Testing Program: Initial Results for Clinical PACS Displays. *RadioGraphics*, 21:719–732, 2001. doi:10.1148/radiographics.21.3.g01ma18719.

M. G. Harisinghani, M. A. Blake, M. Saksena, P. F. Hahn, D. Gervais, M. Zalis, L. da Silva Dias Fernandes, and P. R. Mueller. Importance and Effects of Altered Workplace Ergonomics in Modern Radiology Suites. *RadioGraphics*, 24(2):615–627, 2004. doi:10.1148/rg.242035089.

Y. Ikushima, H. Yabuuchi, J. Morishita, and H. Honda. Analysis of Dominant Factors Affecting Fatigue Caused by Soft-Copy Reading. *Academic Radiology*, 20(11):1448–1456, nov 2013. doi:10.1016/j.acra.2013.08.013.

International Commission on Illumination. *CIE15:2004: Colorimetry*. 3 edition, 2004.

International Committee for Display Metrology (ICDM). *Information Display Measurements Standard (IDMS), version 1.03*. 2012.

International Electrotechnical Commission. *IEC 62563-1:2009+AMD1:2016 Medical electrical equipment - Medical image display systems - Part1: Evaluation methods*. 2016.

T. Kamitani, H. Yabuuchi, Y. Matsuo, T. Setoguchi, S. Sakai, T. Okafuji, S. Sunami, M. Hatakenaka, N. Ishii, M. Kubo, E. Tokunaga, H. Yamamoto, and H. Honda. Diagnostic Performance in Differentiation of Breast Lesion on Digital Mammograms: Comparison among Hard-copy film, 3-megapixel LCD Monitor, and 5-megapixel LCD Monitor. *Clinical Imaging*, 35(5):341–345, 2011. doi:10.1016/j.clinimag.2010.08.015.

E. A. Krupinski, J. P. Johnson, H. Roehrig, and J. Lubin. Optimizing Soft Copy Mammography Displays Using a Human Visual System Model: Influence of Display Phosphor. volume 4686, pages 331–337, 2002. doi:10.1117/12.462694.

H. Liang, S. Park, B. D. Gallas, K. J. Myers, and A. Badano. Image Browsing in Slow Medical Liquid Crystal Displays. *Academic Radiology*, 15(3):370–382, 2008. ISSN 1076-6332. doi:10.1016/j.acra.2007.10.017.

National Electrical Manufacturers Association. NEMA PS3 / ISO 12052, Digital Imaging and Communications in Medicine (DICOM) Standard, Rosslyn, VA, USA. `http://dicomstandard.org/`.

B. J. Pollard, E. Samei, A. S. Chawla, C. Beam, L. E. Heyneman, L. M. H. Koweek, S. Martinez-Jimenez, L. Washington, N. Hashimoto, and H. P. McAdams. The Effects of Ambient Lighting in Chest Radiology Reading Rooms. *Journal of digital imaging*, 25(4):520–526, 2012. doi:10.1007/s10278-012-9459-5.

W. S. Rasband. ImageJ. `https://imagej.nih.gov/ij/`, 1997–2018. U.S. National Institutes of Health, Bethesda, Maryland, USA.

H. Roehrig, E. A. Krupinski, A. S. Chawla, J. Fan, and K. Gandhi. Spatial Noise and Threshold Contrasts in LCD Displays. In D. P. Chakraborty and E. A. Krupinski, editors, *Proc. SPIE*. SPIE, 2003. doi:10.1117/12.480358.

T. Ruckdeschel, C. Keener, J. Kofler, P. Nagy, E. Samei, K. Andriole, E. Krupinski, J. A. Seibert, A. Towbin, N. Bevins, and D. Lewis. ACR-AAPM-SIIM Technical Standard for Electronic Practice of Medical Imaging. *American College of Radiology*, 2017. `https://www.acr.org/~/media/ACR/Documents/PGTS/standards/ElectronicPracticeMedImg.pdf`.

E. Samei, A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M. J. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson, D. M. Moxley-Stevens, W. Pavlicek, H. Roehrig, L. Rutz, E. Samei, S. J. Shepard, R. A. Uzenoff, J. Wang, and C. E. Willis. Assessment of Display Performance for Medical Imaging Systems. *Online Only Report No. OR03 - Report of AAPM Task Group 18*, 2005. `http://www.aapm.org/pubs/reports/detail.asp?docid=153`.

M. Silosky and R. M. Marsh. Constancy of built-in luminance meter measurements in diagnostic displays. *Medical Physics*, 40(12):121902, Nov 2013. doi:10.1118/1.4829497.

Society of Motion Pictures and Television Engineers (SMPTE) Recommended Practice. RP-133:1986 Specifications for Medical Diagnostic Imaging Test Pattern for Television Monitors and Hard-Copy Recording Cameras. *SMPTE Journal*, pages 693–695, 1986. `http:`

//ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7259161.

P. M. Tchou. *Visual Performance in Medical Imaging Using Liquid Crystal Displays*. PhD thesis, University of Michigan, 2007. http://hdl.handle.net/2027.42/57628.

A GLOSSARY

Ambient Luminance (L_{amb})
The component of display luminance due to the diffuse reflection of room light incident on the display surface

CIE LUV
The color space adopted by the International Commission on Illumiance (CIE) in 1976 using the color coordinates L, U, and V

Coefficient of Diffuse Reflection (R_d)
The ratio of ambient luminance to illuminance

Color Space
A mathematical model representing colors as a combination of coordinates

Color Temperature
The temperature of an ideal black-body radiator that emits light of a color similar to that of a light source

Colorimeter
A device used to measure the color coordinates of light emitted by a display using a defined color space

DICOM Grayscale Display Function (GSDF)
The luminance response function recommended by DICOM for the display of medical images regardless of display device

Diffuse Reflection
Light reflected from a surface such that incident rays are reflected at many angles

Digital Driving Level (DDL)
A digital value that, given as input to a display system, produces a luminance

Digital Imaging Communications in Medicine (DICOM)
A standard for handling, storing, printing, and transmitting digital information in medical imaging

Display Noise
Random variations in luminance caused by display hardware and/or electronics

Display Spatial Resolution
The ability of a display to present images of clearly differentiated small objects

Gray Level
For images having only shades of gray (i.e. R=G=B), the gray level refers to the DDL associated with a specific luminance

Illuminance
The luminous flux incident per unit area of a surface

Liquid Crystal Displays (LCD)
Displays that rely on the light polarizing effects of liquid crystals, coupled with polarizing filters and some form of backlight, to achieve varying degrees of luminance

Luminance
The luminous flux emitted from a light source, divided by the unit area of an orthographically projected surface

Luminance Ratio
The ratio of L'_{max} to L'_{min}

Luminance Response Function
The relationship between display luminance and image gray level

Luminous Flux
The photometric quantity of light or luminous flux is represented in lumens, which correlate to the visual sensation of light

Maximum Luminance (L_{max})
The maximum luminance a calibrated display can emit. Corresponds to the most intense gray level, i.e., white

L'_{max}
The sum of L_{max} and L_{amb}

Minimum Luminance (L_{min})
The minimum luminance a calibrated display can emit. Corresponds to the least intense gray level, i.e., black

L'_{min}
The sum of L_{min} and L_{amb}

Organic Light Emitting Diode Displays (OLED Displays)
Displays that rely on the light emitting properties thin organic films to achieve varying degrees of luminance. These devices do not rely on a backlight.

Photometer
Any device used to measure photometric quantities of visible light, such as luminance and illuminance

Pixel Pitch
The center to center spacing between pixel structures

Specular Reflection
Light reflected from a surface such that the angle of incidence is equal to the angle of reflection

(u', v')
CIE 1976 UCS diagram color coordinates

White Point
The color coordinate of a displayed image measured for any gray level (R=G=B)

B IMAGEJ MACROS FOR TG270 TEST PATTERNS

The reader can generate the new test patterns in this document by saving the desired pattern's code as an ImageJ [ImageJ 2016] macro with a `.txt` or `.ijm` extension. Once saved, ImageJ can execute the file using the `Plugins`▸`Macros`▸`Run...` command. Users can edit the macros using either a basic text editor or the ImageJ `Plugins`▸`Macros`▸`Edit...` command.

B.1 TG270-sQC Macro

```
/*
  Use this code to generate the sQC (simple QC) TG270 pattern.

  This version of the script will assume an 8−bit grayscale.

  This pattern is intended to be used by non−physicists as a quick visual
  evaluation of display performance. It contains three rows
  of six incrementing grayscale squares (for a total of 18 squares)
  that cover from 0 to 255. Each square contains a smaller square in the
  upper left and lower right corner. These sub−squares contain a modulation
  pattern with a period of 6 pixels and 50% duty cycle. The patterns use
  vertical bars that differ from the background value of the square by 5 gray
  levels. The pattern in the upper left corner of each square has vertical
  bars with a lower pixel value than the square, while the pattern in the
  lower right has vertical bars with higher pixel values. For the first
  square (GL 0), the pattern in the upper left cannot be lower than the
  main square. Its bars have a contrast of 3 gray levels above the
  background. Similarly, the last square (GL 255) cannot have a lower right
  sub−square with higher pixel values, and so it has a modulation contrast
  of 3 gray levels below the background.

  For a DICOM−conformant display, the user should be able to quickly scan
  the three rows of squares and assess if the modulation patterns in each
  are equally visible across all 18 gray levels. The upper−leftmost and
  lower−rightmost patterns correspond to the ones with the lower contrast
  (3 gray levels vs. 5); these may only be visible on the highest−performance
  displays.

  The sQC pattern also includes three larger squares at the bottom of the
  pattern that allow for minimum and maximum luminance calculations (and
  therefore luminance ratio). These three patterns (black, mid−gray, white)
  can be zoomed and panned around the display for bad pixel and uniformity
  measurements. Furthermore, any of the squares in the three rows can be
  zoomed and panned for similar use. The squares in the three rows correspond
```

to the gray levels used in the traditional 18−point TG18 luminance
response measurement.

Along the bottom of the pattern, a grayscale gradient is included with
continuous pixel value variation for evaluating bit depth issues,
 contouring
artifacts, and grayscale errors. At the side of the gradient, line pairs in
the horizontal and vertical direction can be used to verify correct pixel
mapping.

```
*/

// Define the size of the image (1024 or 2048)
size=1024;
// Based on the size, set a scaler value
sc=size/1024;

// Create the image
newImage("TG270−sQC", "8−bit black", size, size, 1);

// Set the background
background=128;

// Set the period of the bar patterns (pixels/lp)
barper=6;

// Set the entire page to the background
run("Macro...", "code=v="+background+"");

// Set up fonts for labeling
setFont("SansSerif", 18, "antialiased");
setJustification("center");
color=background−75;
setColor(color,color,color);
drawString("TG270−sQC", size/2−1, 32);
setMetadata("Label","TG270−sQC");

// Create a for loop to draw the boxes and contrast squares
ioff=1; // Define i offset to move the group of boxes up/down

for(i=0;i<3;i++){
 for(j=0;j<6;j++){
  // Set the color based on the row, column (i,j) coordinates in steps of 15
  /*
   The power command allows the grays to be read continuously by switching
   the order of the second row. This way the eye can follow the grays from
```

```
left to right, move down, then go from right to left in the second row
before going left to right for the final row.
*/
color=(i*6+2.5-2.5*pow(-1,i)+j*pow(-1,i))*15;
setColor(color,color,color);
// Draw the main squares (128x128 for 1024)
fillRect((sc*160*j+sc*48),sc*128*(i+ioff)+(i-1)*sc*32,sc*128,sc*128);
// Label the squares (comment the drawSting command to remove labels)
setFont("SansSerif", 12, "antialiased");
setJustification("center");
tcolor=(i*6+2.5-2.5*pow(-1,i)+j*pow(-1,i))%9*15+75;
setColor(tcolor,tcolor,tcolor);
// Label the squares with gray levels
drawString((i*6+2.5-2.5*pow(-1,i)+j*pow(-1,i))*15,(sc*160*j+sc*48)+sc*64,sc
*128*(i+ioff)+(i-1)*sc*32+sc*16);
// Label the squares with indices
//drawString((i*6+2.5-2.5*pow(-1,i)+j*pow(-1,i))*15,(sc*160*j+sc*48)+sc*64,
sc*128*(i+ioff)+(i-1)*sc*32+sc*16);
// Set the contrast color to 2% of the main square and draw the modulation
pattern
// Create special conditions for the first and last box
if((i==0)&&(j==0)){
  for(k=0;k<((32/barper)*sc);k++){
    lowc=3;
    setColor(lowc,lowc,lowc);
    fillRect((sc*160*j+sc*48)+sc*16+k*barper,sc*128*(i+ioff)+(i-1)*sc*32+sc
*16,barper/2,sc*32);
    cont=5;
    setColor(cont,cont,cont);
    fillRect((sc*160*j+sc*48)+sc*80+k*barper,sc*128*(i+ioff)+(i-1)*sc*32+sc
*80,barper/2,sc*32);
  }
} else {
  if((i==2)&&(j==5)){
    for(k=0;k<(32/barper)*sc;k++){
      lowc=(i*6+2.5-2.5*pow(-1,i)+j*pow(-1,i))*15-3;
      setColor(lowc,lowc,lowc);
      fillRect((sc*160*j+sc*48)+sc*80+k*barper,sc*128*(i+ioff)+(i-1)*sc*32+sc
*80,barper/2,sc*32);
      cont=(i*6+2.5-2.5*pow(-1,i)+j*pow(-1,i))*15-5;
      setColor(cont,cont,cont);
      fillRect((sc*160*j+sc*48)+sc*16+k*barper,sc*128*(i+ioff)+(i-1)*sc*32+sc
*16,barper/2,sc*32);
    }
  } else {
    for(k=0;k<(32/barper)*sc;k++){
```

```
      cont=(i*6+2.5−2.5*pow(−1,i)+j*pow(−1,i))*15−5;
      setColor(cont,cont,cont);
      fillRect((sc*160*j+sc*48)+sc*16+k*barper,sc*128*(i+ioff)+(i−1)*sc*32+sc
   *16,barper/2,sc*32);
      cont=(i*6+2.5−2.5*pow(−1,i)+j*pow(−1,i))*15+5;
      setColor(cont,cont,cont);
      fillRect((sc*160*j+sc*48)+sc*80+k*barper,sc*128*(i+ioff)+(i−1)*sc*32+sc
   *80,barper/2,sc*32);
     }


   }
  }
 }


}


// Draw some big squares along the bottom for bad pixel testin' and luminance
      measurin'
koff=4;
for(k=0;k<3;k++){
  color=k*128; // Set the color
  setColor(color,color,color);
  fillRect(sc*(128+k*256),sc*128*(koff+1),sc*256,sc*256); // Draw the main
   squares (128x128 for 1024)
}


// Draw vertical gradient along the right side:
//run("Macro...", "code=[if(x>" + sc*(size−64) + "&&y<=" + size−sc*64 + "&&y
   >=" + sc*64 +
//   ") v=(y−" + sc*64 + ")/" + size−sc*64*2 + "*255]");
// Draw horizontal gradient along the bottom (either with or without line
   patterns):
// With line patterns at the sides
run("Macro...", "code=[if(y>" + (size−sc*64) + "&&x<=" + size−2*sc*64 + ") v
   =(x−" + sc*64 + ")/" + size−sc*64*2 + "*255]");
run("Macro...", "code=[if(y>" + (size−sc*64) + "&&x>" + size−2*sc*64 + "&&x<"
    + size−sc*64 +
 ") v=(x%2)*255]");
run("Macro...", "code=[if(y>" + (size−sc*64) + "&&x>" + size−sc*64 + ") v=(y
   %2)*255]");
// Without the line patterns
//run("Macro...", "code=[if(y>" + sc*(size−64) + ") v=(x−" + sc*64 + ")/" +
   size−sc*64*2 + "*255]");
```

B.2   TG270-pQC Macro

```
/*
  This code will generate a modified TG18PQC pattern following the same
  basic modifications used in the pacsDisplay iQC pattern. The user should
  set up the size variable for the desired image size. Typically this is
   either
  a 1024x1024 image or 2048x2048 image. In this script, a square pattern is
  generated, though it can be of arbitrary size. The size of everything
  is based on the TG18PQC definitions and will just be scaled accordingly
  (though it may have some errors).

  This version of the script will assume an 8-bit grayscale.
 */
size=1024;
newImage("TG270-pQC", "8-bit black", size, size, 1);


// Define some of the sizes (everything is normalized to 1024)
sidew=floor(size/1024*87);
barh=floor(size/1024*50);
toph=floor(size/1024*62);


// Create the horizontal bars and the contrast modulation within (varying
    contrast and frequency)
// 18 bars (0, 15, 30,..., 240, 255)
run("Macro...", "code=[if(y>=" + toph + "&&y<" + size-toph + ") v=floor((y-"
   + toph + ")/" + barh + ")*15]");
// Contrast modulation
// Horizontal bars
freqH=newArray(18,12,6,4);
for(i=0;i<4;i++){
 f=freqH[i];
 // Left contrast (8)
 run("Macro...", "code=[if(x>=" + sidew+i*barh + "&&x<"
  + sidew+barh+i*barh + "&&((y-" + toph + ")%"+ f +"<"+ f/2 +")) v=v-4]");
 run("Macro...", "code=[if(x>=" + sidew+i*barh + "&&x<"
  + sidew+barh+i*barh + "&&((y-" + toph + ")%"+ f +">="+ f/2 +")) v=v+4]");
 // Add additional four to the top row and subtract four from the bottom row
 run("Macro...", "code=[if(x>=" + sidew+i*barh + "&&x<"
  + sidew+barh+i*barh + "&&y<" + toph+barh + "&&((y-" + toph + ")%"+ f +">="+
    f/2 +")) v=v+4]");
 run("Macro...", "code=[if(x>=" + sidew+i*barh + "&&x<"
  + sidew+barh+i*barh + "&&y>=" + size-toph-barh + "&&((y-" + toph + ")%"+ f
   +"<"+ f/2 +")) v=v-4]");
 // Left contrast (2)
 run("Macro...", "code=[if(x>=" + sidew+(i+4)*barh + "&&x<"
  + sidew+barh+(i+4)*barh + "&&((y-" + toph + ")%"+ f +"<"+ f/2 +")) v=v-1]")
   ;
```

```
run("Macro...", "code=[if(x>=" + sidew+(i+4)*barh + "&&x<"
 + sidew+barh+(i+4)*barh + "&&((y-" + toph + ")%"+ f +">="+ f/2 +")) v=v
   +1]");
// Add additional one to the top row and subtract one from the bottom row
run("Macro...", "code=[if(x>=" + sidew+(i+4)*barh + "&&x<"
 + sidew+barh+(i+4)*barh + "&&y<" + toph+barh + "&&((y-" + toph + ")%"+ f
   +">="+ f/2 +")) v=v+1]");
run("Macro...", "code=[if(x>=" + sidew+(i+4)*barh + "&&x<"
 + sidew+barh+(i+4)*barh + "&&y>=" + size-toph-barh + "&&((y-" + toph + ")
   %"+ f +"<"+ f/2 +")) v=v-1]");
}


// Vertical bars
freqV=newArray(4,6,12,18);
for(i=9;i<13;i++){
 f=freqV[i-9];
 // Right contrast (8)
 run("Macro...", "code=[if(x>=" + sidew+(i+4)*barh + "&&x<"
  + sidew+barh+(i+4)*barh + "&&((x-" + sidew+i*barh + ")%"+ f +"<"+ f/2 +"))
    v=v-4]");
 run("Macro...", "code=[if(x>=" + sidew+(i+4)*barh + "&&x<"
  + sidew+barh+(i+4)*barh + "&&((x-" + sidew+i*barh + ")%"+ f +">="+ f/2 +"))
    v=v+4]");
 // Add additional four to the top row and subtract four from the bottom row
 run("Macro...", "code=[if(x>=" + sidew+(i+4)*barh + "&&x<"
  + sidew+barh+(i+4)*barh + "&&y<" + toph+barh + "&&((x-" + sidew+i*barh + ")
    %"+ f +">="+ f/2 +")) v=v+4]");
 run("Macro...", "code=[if(x>=" + sidew+(i+4)*barh + "&&x<"
  + sidew+barh+(i+4)*barh + "&&y>=" + size-toph-barh + "&&((x-" + sidew+i*
    barh + ")%"+ f +"<"+ f/2 +")) v=v-4]");
 // Right contrast (2)
 run("Macro...", "code=[if(x>=" + sidew+i*barh + "&&x<"
  + sidew+barh+i*barh + "&&((x-" + sidew+i*barh + ")%"+ f +"<"+ f/2 +")) v=v
    -1]");
 run("Macro...", "code=[if(x>=" + sidew+i*barh + "&&x<"
  + sidew+barh+i*barh + "&&((x-" + sidew+i*barh + ")%"+ f +">="+ f/2 +")) v=v
    +1]");
 // Add additional one to the top row and subtract one from the bottom row
 run("Macro...", "code=[if(x>=" + sidew+i*barh + "&&x<"
  + sidew+barh+i*barh + "&&y<" + toph+barh + "&&((x-" + sidew+i*barh + ")%"+
    f +">="+ f/2 +")) v=v+1]");
 run("Macro...", "code=[if(x>=" + sidew+i*barh + "&&x<"
  + sidew+barh+i*barh + "&&y>=" + size-toph-barh + "&&((x-" + sidew+i*barh +
    ")%"+ f +"<"+ f/2 +")) v=v-1]");
}
```

```
// Horizontal bars at the top and bottom of the pattern with high−contrast
    line pair phantoms
run("Macro...", "code=[if(y<" + toph + ") v=64]");
run("Macro...", "code=[if(y>=" + size−toph + ") v=191]");
// Line pairs
freqLP=newArray(6,4,2,2,4,6);
topM=toph/2;
for(i=0;i<3;i++){
 f=freqLP[i];
 run("Macro...", "code=[if(x>=" + sidew+(2*i+2)*barh + "&&x<"+ sidew+barh+(2*
    i+2)*barh +
  "&&y>" + topM−barh/2 + "&&y<=" + topM+barh/2 + "&&((y−" + (topM−barh/2) +
    ")%"+ f +"<"+ f/2 +")) v=128]");
 run("Macro...", "code=[if(x>=" + sidew+(2*i+2)*barh + "&&x<"+ sidew+barh+(2*
    i+2)*barh +
  "&&y>" + topM−barh/2 + "&&y<=" + topM+barh/2 + "&&((y−" + (topM−barh/2) +
    ")%"+ f +">="+ f/2 +")) v=0]");
 run("Macro...", "code=[if(x>=" + sidew+(2*i+2)*barh + "&&x<"+ sidew+barh+(2*
    i+2)*barh +
  "&&y>" + size−topM−barh/2 + "&&y<=" + size−topM+barh/2 + "&&((y−" + (topM−
    barh/2) + ")%"+ f +"<"+ f/2 +")) v=128]");
 run("Macro...", "code=[if(x>=" + sidew+(2*i+2)*barh + "&&x<"+ sidew+barh+(2*
    i+2)*barh +
  "&&y>" + size−topM−barh/2 + "&&y<=" + size−topM+barh/2 + "&&((y−" + (topM−
    barh/2) + ")%"+ f +">="+ f/2 +")) v=255]");
}
for(i=3;i<6;i++){
 f=freqLP[i];
 run("Macro...", "code=[if(x>=" + sidew+(2*i+4)*barh + "&&x<"+ sidew+barh+(2*
    i+4)*barh +
  "&&y>" + topM−barh/2 + "&&y<=" + topM+barh/2 + "&&((x−" + sidew+(2*i+4)*
    barh + ")%"+ f +"<"+ f/2 +")) v=128]");
 run("Macro...", "code=[if(x>=" + sidew+(2*i+4)*barh + "&&x<"+ sidew+barh+(2*
    i+4)*barh +
  "&&y>" + topM−barh/2 + "&&y<=" + topM+barh/2 + "&&((x−" + sidew+(2*i+4)*
    barh + ")%"+ f +">="+ f/2 +")) v=0]");
 run("Macro...", "code=[if(x>=" + sidew+(2*i+4)*barh + "&&x<"+ sidew+barh+(2*
    i+4)*barh +
  "&&y>" + size−topM−barh/2 + "&&y<=" + size−topM+barh/2 + "&&((x−" + sidew
    +(2*i+4)*barh + ")%"+ f +"<"+ f/2 +")) v=128]");
 run("Macro...", "code=[if(x>=" + sidew+(2*i+4)*barh + "&&x<"+ sidew+barh+(2*
    i+4)*barh +
  "&&y>" + size−topM−barh/2 + "&&y<=" + size−topM+barh/2 + "&&((x−" + sidew
    +(2*i+4)*barh + ")%"+ f +">="+ f/2 +")) v=255]");
}
```

```
// Vertical side bars with modulating gradient
strip=sidew/2
// Left side
run("Macro...", "code=[if(x<" + sidew + ") v=y/" + size-1 +
    "*1031/1024*255]");
run("Macro...", "code=[if(x<" + sidew/2+(strip/2) + "&&x>" + sidew/2-(strip
    /2) + ") v=v+3*sin(y/2)]");
// Right Side
run("Macro...", "code=[if(x>=" + size-sidew + ") v=-1*(y-" + size-1 + ")/" +
    size-1 + "*1031/1024*255]");
run("Macro...", "code=[if(x<" + size-sidew/2+(strip/2) + "&&x>" + size-sidew
    /2-(strip/2) + ") v=v+3*sin(y/1.5)]");


// Set up fonts for labeling
setFont("SansSerif", 18, "antialiased");
setJustification("center");
setColor(128,128,128);
drawString("TG270-pQC", size/2-1, 32);
setMetadata("Label","TG270-pQC");
```

B.3 TG270-ULN Macro

```
// Use this code to generate combination UN and LN test patterns (TG270-ULN)

// Define the number of images (18 (15 GL), 52 (5 GL), 86 (3 GL) or 256 (1 GL
    ))
images=256;
// Define the size of the images (1024 or 2048)
size=1024;

// Create the image stack (Re-define the image type for other bit-depths)
newImage("TG270-ULN8-", "8-bit black", size, size, images);

// Set the image values and draw the appropriate lines
step=255/(images-1);

// Set the entire page to the z*step value
run("Macro...", "code=[v=z*"+step+"] stack");

// Set up fonts for labeling
setFont("SansSerif", 18, "antialiased");
setJustification("center");

// Create a for loop to label the images and draw the grid lines
```

```
for ( i =0; i <images ; i++){
 setSlice ( i +1);
 color =(i%(images /2)+images /6)*step ;
 setColor ( color , color , color ) ;
 drawString ("TG270–ULN8–"+ IJ . pad ( i *step ,3) +"", size /2−1, 32);
 setMetadata ("Label" , "TG270–ULN8–"+IJ . pad ( i *step , 3 )+"") ;
 drawRect (340* size /1024 ,0 ,2 , size ) ;
 drawRect (682* size /1024 ,0 ,2 , size ) ;
 drawRect (0 ,340* size /1024 , size , 2 ) ;
 drawRect (0 ,682* size /1024 , size , 2 ) ;
}

// If this command is enabled , just select the
// option to use the slice labels as the file names.
//run ("Image Sequence ...  ",) ;
```

B.4   TG270-TR Macro

```
// Use this code to generate a series of temporal resolution images for an
    animated gif

// Define the size of the frame
vsize =1024;
hsize =1024;
// Define a scalar to adjust from the default size
vsc=vsize /1024;
hsc=hsize /1024;
// Define the number of rows ( powers of 2 work best )
rows=16;
// Define the size of the blocks ( assumes square size )
block =32;
// Define the number of blocks (4−7 recommended )
numbl=6;
// Define the number of frames ( based on block count )
frames=hsize /(2* block )+2+(numbl−1);
// Define frame rate
fps =60;

// Create the image stack
newImage ("TG270–TR2", "black", hsize , vsize , frames ) ;

// Set the gray level pairs. 2*rows pairs are required.
gl=newArray (
 015 ,
 020 ,
 015 ,
```

```
030,
015,
045,
015,
065,
120,
125,
120,
135,
120,
150,
120,
170,
245,
240,
245,
230,
245,
215,
245,
195,
000,
075,
000,
255,
255,
180,
255,
000);


setFont("SansSerif", 10, "antialiased");
setJustification("center");

// Creation the test pattern background. Loop through each gray level pair
    base.
for(i=0;i<rows;i++){
 setColor(gl[2*i],gl[2*i],gl[2*i]);
 run("Macro...", "code=[if(y>="+i*vsize/rows+"&&y<="+(i+1)*vsize/rows+") v="+
    gl[2*i]+"] stack");
 run("Macro...", "code=[if(y>="+i*vsize/rows+0.5*(vsize/rows-vsc*block)+"&&y
    <="+(i+1)*vsize/rows-0.5*(vsize/rows-vsc*block)+
  "&&x>="+2*block*hsc+"*z-"+hsize-(hsize-hsc*2*numbl*block)+"&&x<"+2*block*
    hsc+"*z&&x%"+2*block*hsc+">"+block*hsc-1+") v="+gl[2*i]+
  "+("+gl[2*i+1]+"-"+gl[2*i]+")] stack");
 // set text color
```

```
  tcolor=gl[2*i]+(64*pow(-1,floor(gl[2*i]/128)));
  setColor(tcolor,tcolor,tcolor);
  // set up labels
  for(j=0;j<frames;j++){
   setSlice(j+1);
   // label each frame
   drawString(j,j*hsc*block*2-(numbl-0.5)*hsc*block,i*vsize/rows+0.50*(vsize/
     rows-vsc*block));
   // label each pair
   drawString(gl[2*i]+" / "+gl[2*i+1],hsize/2,i*vsize/rows+(vsize/rows-vsc*
     block)+vsc*block);
  }
}

// set up the animation options, set to desired fps
run("Animation Options...", "speed="+fps+" first=1 last="+frames+" start");

//run("Image Sequence... ",); // If this command is enabled, just select the
    option to use the slice labels as the file names.
```