Patient Safety Starts with **Independence.**

With ever-increasing imaging and treatment variables, an independent approach to Quality Management ensures safety is never taken for granted.

At Sun Nuclear, we deliver proven independent QA solutions for Radiation Therapy and Diagnostic Imaging. More than 5,000 hospitals and clinics worldwide count on us to help:

- Mitigate errors
- Reduce inefficiencies
- Validate technologies and techniques
- Elevate clinical care

For user insights, key publications and product updates, **visit sunnuclear.com.**

See us at ASTRO in Booth #1029

SUN NUCLEAR corporation

Patient Safety Starts Here

Correlation between physical measurements and observer evaluations of image quality in digital chest radiography

Asena Yalcin
Institute of Nuclear Sciences, Ankara University, Ankara 06100, Turkey

Turan Olgar[a)]
Institute of Nuclear Sciences, Ankara University, Ankara 06100, Turkey
Faculty of Engineering, Department of Engineering Physics, Ankara University, Ankara 06100, Turkey

Tanzer Sancak
Department of Radiology, TOBB University of Economics and Technology Hospital, Ankara, Turkey

Gokce Kaan Atac
Faculty of Medicine, Department of Radiology, Ufuk University, Ankara, Turkey

Serdar Akyar
Department of Radiology, Ankara University School of Medicine, Ankara, Turkey

(Received 22 November 2019; revised 27 April 2020; accepted for publication 1 May 2020; published 16 July 2020)

Purpose: The aim of this paper was to investigate the relationship between the physical and subjective (observer) image quality metrics in digital chest radiography.
Methods: Five digital radiography systems, four with indirect flat panel detector and one with storage phosphor-based computed radiography system, were used in the study. The physical image quality assessments were carried out using effective detective quantum efficiency (eDQE) metric and subjective performance of the digital radiography systems was evaluated in terms of inverse image quality figure (IQF_{inv}) derived from the contrast-detail (CD) diagrams using CDRAD 2.0 phantom and CDRAD phantom analyzer software. All measurements were performed for different tube voltages (70, 81, 90, 102, 110, and 125 kVp) and polymethyl methacrylate (PMMA) phantom thicknesses. An anthropomorphic chest phantom and visual grading analysis (VGA) technique based on European image quality criteria for chest radiography were used for clinical image quality evaluation.
Results: The Spearman correlation coefficients were calculated for the investigation of the correlation between physical image quality and clinical image quality. The results showed strong positive correlation between the physical and clinical image quality findings. The minimum correlation coefficient was 0.91 ($p < 0.011$) for IQF_{inv} vs VGA scores and 0.92 ($p < 0.009$) for IeDQE vs VGA scores.
Conclusions: Our results confirm that clinical image quality can be predicted with both physical assessments and contrast-detail detectability studies. © *2020 American Association of Physicists in Medicine* [https://doi.org/10.1002/mp.14244]

Key words: integrated effective detective quantum efficiency (IeDQE), image quality figure (IQF), visual grading analysis (VGA)

1. INTRODUCTION

There are two main tasks in medical radiology: diagnose the patients (diagnostic) and treat them (interventional).[1] The image quality is an essential factor for the ability of the radiologists to interpret the image correctly. There are several methods for the assessment of image quality. These methods can be grouped as physical, psychophysical, and observer performance. They all have advantages and disadvantages.[2]

The modulation transfer function (MTF), noise power spectrum (NPS), and detective quantum efficiency (DQE) are the most widely used metrics for physical assessment of image quality.[1] The spatial resolution and noise transfer properties of the imaging system are given by MTF and NPS, respectively. The experimental measurement of DQE involves measurement MTF and NPS, and these measurement results are combined to give DQE with proper normalization.[3] Detective quantum efficiency indicates signal-to-noise transfer properties of the imaging systems and generally defined as the ratio of the squared output signal-to-noise ratio $(SNR_{out})^2$ to squared input SNR $(SNR_{in})^2$ of the imaging detector.[4] Although DQE is accepted as the fundamental performance parameter of digital x-ray detectors, it does not take the overall image quality distorting effects present in clinical conditions into account. These image quality distorting effects may include the presence of scattered radiation, the effect of anti-scatter grid, and focal spot blurring due to focal spot size and geometric magnification effect. In order to overcome this problem DQE has been improved to a new image quality metric called effective DQE (eDQE) by Samei et al.[5]

and generalized DQE (GDQE) by Kyprianou et al.[6] which provides the whole system performance. In addition, the integrated values of DQE (IDQE) were used by Fetterly et al to determine the efficiency of the computed radiography system for overall frequency range.[7] The same concept was used by some researchers for the evaluation of image quality of different digital radiography systems.[8,9]

Psychophysical evaluation contains contrast-detail (CD) analysis and visual assessment of an observer.[2] Although CD phantom imaging has the advantage of containing whole imaging chain (acquisition and display), it has some drawbacks. As the phantom has a homogeneous background, it does not reflect the real anatomical noise. Nevertheless, some researchers concluded that this method gave reasonable results according to clinical assessments.[2,10]

Clinical image quality based on observer performance can be evaluated by receiver-operating characteristics (ROC) or visual grading analysis (VGA) scoring.[11,12] ROC analysis is based on the signal detection theory where an observer detects a pathology in a normal background.[2] This method, however, is very time-consuming and it requires a large number of patients. VGA scoring is an easier method where the image quality is evaluated by means of anatomical structures according to the image quality criteria specified for a given task.

Physical measurements are essential for describing the quality of the imaging system itself.[2] Since the clinical assessment results cannot fully be explained by these physical measurements, the observer performance methods must be adopted to find the quality of the procedures applied in the clinic.[2] The requirement for combining the evaluation of the physical measurements and the clinical assessments is emphasized in the literature.[1,13,14] Sandborg et al.[15] have compared the signal-to-noise ratio (SNR) and VGA results in chest and pelvis radiography for different tube voltages. They found significant correlation between these two methods ($r^2 = 0.91$ for chest PA and $r^2 = 0.94$ for pelvis AP). Moore et al.[16] investigated the correlation between the quality of visual graded patient chest images and physical assessment results [contrast-to-noise ratio (CNR) and DQE]. They found a significant correlation ($r = 0.87$), but their findings were statistically limited to only one computed radiography (CR) system and clinical images were acquired without an anti-scatter grid which is not the case for chest imaging in general. Sund et al.[17] compared physical and clinical system performance for chest radiography using four digital systems. They used VGA for clinical evaluation and DQE for physical evaluation. De Crop et al.[10] established a good correlation ($r = 0.91$) between CD diagrams and clinical image quality for chest radiographs of embedded cadavers. However, the statistical significance of their results was limited with three cadavers and the results were case dependent. Al-Murshedi et al.[18] investigated the correlation between inverse image quality figure (IQF_{inv}) and relative visual grading analysis for image quality (IQ) and lesion visibility (LV). They found a strong positive correlation between the IQ and the IQF_{inv} ($r = 0.91$; $P < 0.001$) and a good positive correlation between the IQF_{inv} and the LV ($r = 0.68$; $P < 0.001$).

Yalcin et al.[19] performed eDQE measurements for different tube voltages, different PMMA thicknesses and different grid types in a single digital flat panel imaging system and investigated the correlation between the integrated effective DQE (IeDQE) values and the IQF_{inv} values obtained from the measurements taken with the CDRAD phantom under the same imaging conditions. They reported that eDQE decreased with increased grid ratio and observed a significant correlation between IeDQE and IQF_{inv} ($r > 0.96$). In the present study, the correlation between the clinical assessment and quantitative image quality measurements using multiple methods and different digital imaging systems for chest radiography was investigated. Integrated effective detective quantum efficiency (IeDQE) was used to provide the image quality for the whole frequency range for the physical assessment and visual graded anthropomorphic chest phantom images were used for clinical assessments. For psychophysical evaluation CD detectability measurements were carried out and IQF_{inv} values were calculated from the CD curves obtained from the CDRAD phantom images.

The concept of using IDQE and IeDQE across all frequencies has already been investigated in other studies for the evaluation of overall performance of digital detectors and whole imaging systems, respectively.[7–9] To obtain the total performance of an imaging system over the whole frequency range, the eDQE values were also integrated across all frequencies in the current study to compare the VGA and IQF_{inv} values obtained for each PMMA thickness and tube voltage with the effective DQE measurements. To the best of our knowledge, there are a few studies investigating the correlation and relationship between the clinical and physical image quality assessment results and the current study will be the first study investigating the correlation between integrated effective DQE, VGA, and IQF_{inv}.

2. MATERIALS AND METHODS

2.A. Radiological equipment

Five digital imaging systems by different manufacturers were used in this study; four of them were indirect digital radiography systems and one of them was a computed radiography system. System specifications are given in Table I. The images were acquired using a set of tube voltages (70, 81, 90, 102, 110, and 125 kVp) and large focus spot size (1.2 mm) was used for each imaging system. Focus-to-image distance (FID) was set to 180 cm for all measurements and no additional filter was used. The tube current-exposure time product was automatically selected by automatic exposure control (AEC) for all imaging systems.

2.B. Phantoms

Three different types of phantoms were used in the study. A phantom composed of 25 × 25 cm^2 PMMA slabs with

TABLE I. System specifications of the digital imaging systems used in the study.

	CARESTREAM DRX-1C	PHILIPS TRIXELL PIXIUM 4600	KONICA MINOLTA AERO DR	TOSHIBA FDX4343R	FUJI FCR PROFECT CS with IP CASSETTE TYPE CC
X-ray equipment	GE Silhouette VR	Philips Digital Diagnost VM	GE Silhouette VR	USX-RAY X3C	GE Silhouette VR
Detector type	Flat panel	Flat panel	Flat panel	Flat panel	Storage phosphor plate
Detector material	Pixelated CsI	Columnar CsI	Pixelated CsI	Pixelated CsI	Powder $BaFBr:Eu^{2+}$
Pixel area	3072 × 2560	3001 × 3001	1994 × 2340	3008 × 3072	1750 × 2150
Pixel size (mm)	0.139	0.148	0.175	0.143	0.200
Grid type	Stationary	Moving	Stationary	Stationary	Stationary
Grid ratio	10:1	8:1	10:1	12:1	10:1
Grid frequency	60 lp/cm	40 lp/cm	60 lp/cm	40 lp/cm	60 lp/cm
Grid focus distance (cm)	180	140	180	140	180
Focal spot size (mm)	1.2	1.2	1.2	1.2	1.2

different thicknesses (10, 15, 20, and 25 cm) was used to represent different patient sizes for physical measurements.

The CDRAD 2.0 phantom (Artinis Medical System, The Netherlands) was used to assess CD detectability. The phantom consists of a Plexiglass plate with cylindrical holes of various depths and diameters simulating different contrast and resolution, respectively.

For the clinical image quality assessments, an anthropomorphic phantom (Alderson RS-330 lung/chest phantom-Radiology Support Devices, Canada) and VGA technique were used.

2.C. Physical Image quality evaluations

2.C.1. eDQE estimation

eDQE was provided to estimate the total system performance. The eDQE measurements were carried out using PMMA slabs thicknesses of 10, 15, 20, and 25 cm. The eDQE was calculated as follows:[9]

$$eDQE(f') = \frac{eMTF^2(f') \cdot (1 - SF^2)}{q \cdot eNNPS(f') \cdot E_p \cdot TF} \quad (1)$$

where f' is the magnification-corrected spatial frequency (f' = f.m, f is the spatial frequency at the detector plane and m is the geometrical magnification), eMTF is the presampling MTF with scatter, SF is the scatter fraction, q is the number of quanta at the detector plane per unit air kerma per mm^2, eNNPS is the normalized NPS with scatter, Ep is the measured prephantom exposure free-in-air which is inverse square corrected to the detector plane and TF is the measured transmission fraction through the phantom using a narrow beam geometry.

For eMTF measurements, the edge response method was used. A tungsten edge tool with 1 mm thickness was placed on the PMMA slabs at the center of the x-ray beam at the tube side with a 3°–5° tilt to the detector pixel array. The edge spread function (ESF) was determined by acquiring the edge image. The ESF was differentiated to provide the line spread function (LSF). The MTF was determined in frequency domain by taking the Fourier transform of the LSF. The MTF was calculated by using the OBJ IQv2 software developed by NHSBSP.[20] Flat-field images acquired for the establishment of the signal transfer properties (STP) of the detectors were used for the calculation of NNPS. A region of interest (ROI) of 512 × 512 pixels was extracted from the center of the image and linearized to air kerma using the response curve. Subregions of 128 × 128 pixels were subdivided from the image. One dimensional eNNPS was determined from two dimensional eNPS by averaging central ±4 rows around each axis. Then the eNNPS was calculated by dividing eNPS to the square of the mean pixel value of the linearized image. The OBJ IQv2 software was used for the calculations.

For TF calculations, a calibrated ionization chamber (Radcal AcuPro, 6cc) was used. The PMMA slabs and the ionization chamber were positioned at 50 and 100 cm away from the focus, respectively. The output air kerma was measured with and without PMMA slabs. Three exposures were carried out for each condition (kVp and phantom thickness) and the TF was determined as the ratio of average air kerma with and without phantom. SF measurements were carried out by using the QRM beam stop phantom (QRM Quality Assurance in Radiology and Medicine GmbH, Germany). The phantom is composed of a 10 × 10 array of 100 lead cylinders, each 3 mm in diameter and 6 mm in thickness, separated by 25 mm and embedded in a 6 mm thick PMMA plate. For the SF measurements, the PMMA slabs were placed adjacent to the detector surface and the beam stop phantom was placed on the PMMA slabs at tube side. Sixteen beam stops in the center of the image field were evaluated using ten pixel-wide ROIs. The SF fractions were calculated as the ratio of mean pixel values behind the beam stop to the mean pixel values from background region around the beam stops.

2.C.2. Integrated eDQE

As small differences between eDQE curves for different beam qualities cannot be recognized easily and as the energy dependence of the systems effective detective quantum efficiency is better appreciated on integrated effective detective quantum efficiency (IeDQE) curves, the IeDQE over the frequency range was adopted as synthetic image quality index to provide the whole system performance as suggested by others.[7–9] Then the relationship between integrated values of eDQE and visual assessments was investigated. The integral values of eDQE in one dimension were calculated as follows:[7]

$$IeDQE = \int_0^{f'_{nyq}} eDQE(f')df' \tag{2}$$

where f'_{nyq} s the Nyquist frequency. The unit of IeDQE is mm^{-1}.

2.D. Contrast-detail measurements

In order to evaluate the CD detectability of the imaging systems, the CDRAD 2.0 phantom (Artinis Medical System, The Netherlands) was used. The CDRAD phantom images were acquired manually with the same exposure parameters selected by the AEC in physical measurements. The CDRAD phantom was placed in the middle of the PMMA slabs for each thickness. The phantom acquisitions were repeated three times for each tube voltage. The CD curves were determined for each condition. These curves show the visible hole depth against the hole diameter. Curves closer to the lower left corner, that is, closer to the origin, indicate a better detection and visualization of low-contrast objects. The CD curves obtained from CDRAD images were used for the calculation of image quality figure (IQF). The IQF is defined as the sum of the products of depth and diameters of the recorded visible objects. The IQF is expressed mathematically as follows:[21]

$$IQF = \sum_{i=1}^{15} C_i \cdot D_{i,th} \tag{3}$$

where i is the contrast-column number, C_i represents the object depth in the contrast columns i and $D_{i,th}$ denotes the corresponding smallest visible diameter (threshold diameter) in this column. The lower IQF corresponds to better image quality. To resolve this complexity relationship, the inverse of image quality figure (IQFinv) is introduced as:

$$IQF_{inv} = \frac{100}{\sum_{i=1}^{15} C_i \cdot D_{i,th}} \tag{4}$$

The higher values of the IQF_{inv}, indicate a better low contrast visibility. The images were evaluated via the CD Analysis Software (Artinis Medical System, The Netherlands).

2.E. Chest images and VGA scoring

A set of posteroanterior (PA) images of RS-330 lung/chest phantom (Radiology Support Devices, Canada) was collected for all imaging systems. FID was set to 180 cm which is commonly used in clinical chest examinations. Six different settings of tube voltages (70, 81, 90, 102, 110, and 125 kVp) were applied by using AEC without any additional filtration at the x-ray tube output. For all phantom acquisitions, each system's own postprocessing algorithm specified for chest PA examinations was used. After acquisition of chest images, all images were stored digitally, and soft-copy images were used for the evaluation. The monitor was calibrated to comply with the Digital Imaging and Communications in Medicine Part 3.14 gray-scale standard display function.[22] The readers were allowed to adjust the brightness and contrast levels of the images. The images were displayed in random order. An example of a chest radiograph acquired with lung/chest phantom is shown in Fig. 1.

Six experienced radiologists (five with 25 yr of experience and one with 6 yr of experience) evaluated the soft-copy chest images and scored the image quality according to the adopted European image quality criteria.[23]

An absolute VGA scoring was used. The image quality criteria and absolute rating criteria for each structure are listed in Tables II and III, respectively. For a given system, absolute VGA score (VGAS) was calculated as follows:

$$VGAS = \frac{\sum_{i=1}^{I} \cdot \sum_{s=1}^{S} \cdot \sum_{o=1}^{O} G_{(abs)\cdot i\cdot s\cdot o}}{I \cdot S \cdot O} \tag{5}$$

where $G_{(abs),i,s,o}$ is the absolute rating for a particular image, structure, and observer, I is the number of images, S is the number of structures, and O is the number of observers.

2.F. Statistical analysis

Data were analyzed using SPSS version 22.0 (IBM Inc, Armonk, NY). The correlations between the VGA scores,

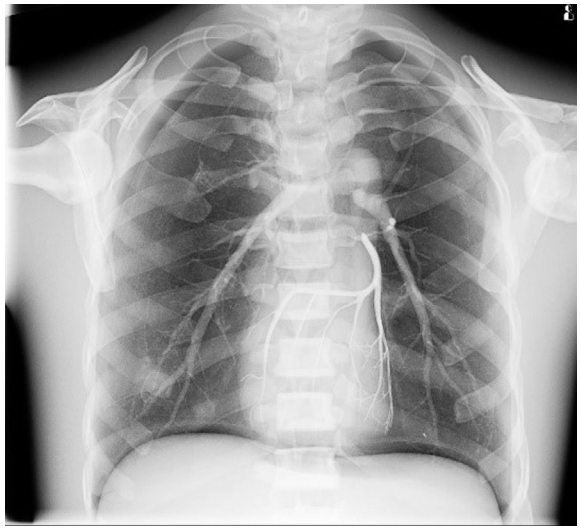

FIG. 1. Example of a chest radiograph of RS-330 lung/chest phantom.

TABLE II. Image quality criteria used in the study for chest radiography.

Criteria number	Description
1	Symmetrical reproduction of the thorax as shown by central position of the spinous process between the medial ends of the clavicles
2	Visually sharp reproduction of the vascular pattern in the whole lung, particularly the peripheral vessels
3	Visually sharp reproduction of the borders of the heart and aorta
4	Visualization of the spine through the heart shadow
5	Small round details in the whole lung, including the retrocardiac areas. High contrast: 0.7 mm diameter; low contrast: 2 mm diameter
6	Linear and reticular details out to the lung periphery. High contrast: 0.3 mm in width; low contrast: 2 mm in width

TABLE III. The absolute rating criteria used to evaluate the clinical images.

Absolute rating	Appearance of structure
1	Not visible (poor image quality)
2	Poorly reproduced (restricted image quality)
3	Adequately reproduced (sufficient image quality)
4	Very well reproduced (good image quality)

IQF_{inv} and IeDQE were calculated by using the Spearman correlation coefficient.[24] In order to determine the interobserver variation an intraclass correlation coefficient was calculated. The intraclass correlation coefficient of <0.8 was considered as poor reliability.[25] Statistical significance was calculated using the analysis of variance (ANOVA) in conjunction with Newman–Keuls test.[26] The confidence interval was selected as 95% for all statistical calculations.

3. RESULTS

The q values seen in the Eq. (1) were calculated with SpecCalc software for each kVp and phantom thickness and given together with SF and TF values as a supplementary file (Tables S1 and S2).[27] Representative eDQE results are given for each imaging system at different tube voltages and for 15 and 25 cm PMMA thicknesses in Fig. 2 and S1. The decrease observed in eDQE in Fig. 2(a) at 1.31 lp/mm was a result of the grid line pattern at 6.0 lp/mm being aliased in the sampled NNPS.[9,28] eDQE decreased with increasing phantom thickness. The lowest eDQE results were observed for the Fuji CR system as it consists of phosphor powder. The eDQE decreased with increasing tube voltage for DRX-1C, Philips, Fuji and Konica whereas the maximum values of eDQE were determined at 102 kVp for Toshiba. As a result, similar findings were determined for IeDQE values (Fig. 3).

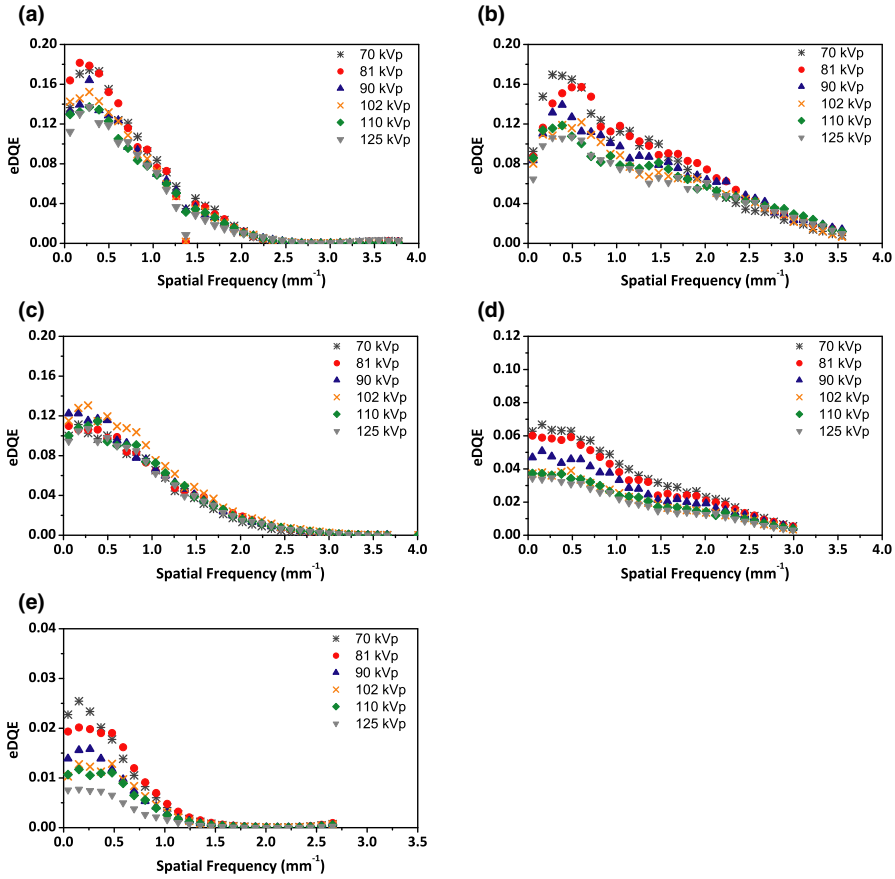

FIG. 2. Effective detective quantum efficiency results for each tube voltage and 15 cm polymethyl methacrylate thickness for (a) DRX-1C, (b)Philips Trixell Pixium 4600, (c)Toshiba FDX4343R, (d)Konica Minolta AERO DR, and (e) Fuji FCR. [Color figure can be viewed at wileyonlinelibrary.com]

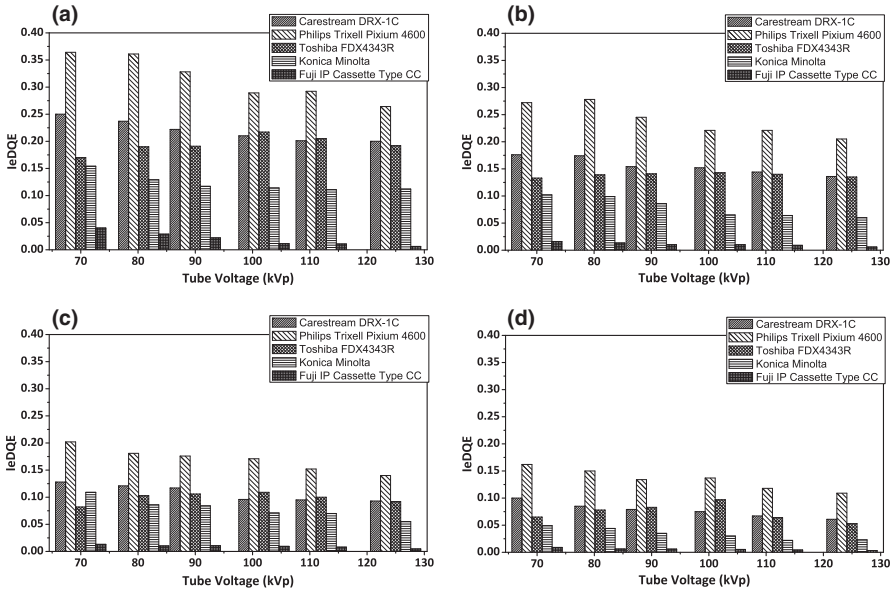

FIG. 3. Integrated effective detective quantum efficiency results for each imaging system at different tube voltages for (a) 10 cm polymethyl methacrylate (PMMA), (b) 15 cm PMMA, (c) 20 cm PMMA, and (d) 25 cm PMMA thickness.

The IeDQE results were reasonable when compared to Bertolini et al's findings for DRX-1C and Philips at similar tube voltage.[9] Their results for IeDQE at 120 kVp were 0.189 and 0.192 for DRX-1C and Philips, respectively and our findings at the same tube voltage and detector dose were 0.126 and 0.205, respectively for 15 cm PMMA thickness. The difference observed for DRX-1C systems may be due to different SF values resulting from different types of grids used (for DRX-1C system, the SF was 0.40 for CDRH Chest phantom in the reference study — and was 0.53 with 15 cm PMMA phantom for our study).

The mean IQF_{inv} values calculated from CDRAD phantom images recorded for different tube voltage and phantom thickness cases are compared for different digital radiography systems in Fig. 4. The IQF_{inv} values decreased with increasing tube voltage for DRX-1C, Fuji, Konica and Philips, whereas the maximum values were determined at 102 kVp for Toshiba, as in the physical assessments. Philips gave the highest IQF_{inv} results. The lowest values were determined for Fuji, in accord with the IeDQE results. Table IV illustrates the mean VGAS obtained from observer grading. Figure 5 represents the correlation between the mean IQF_{inv} and

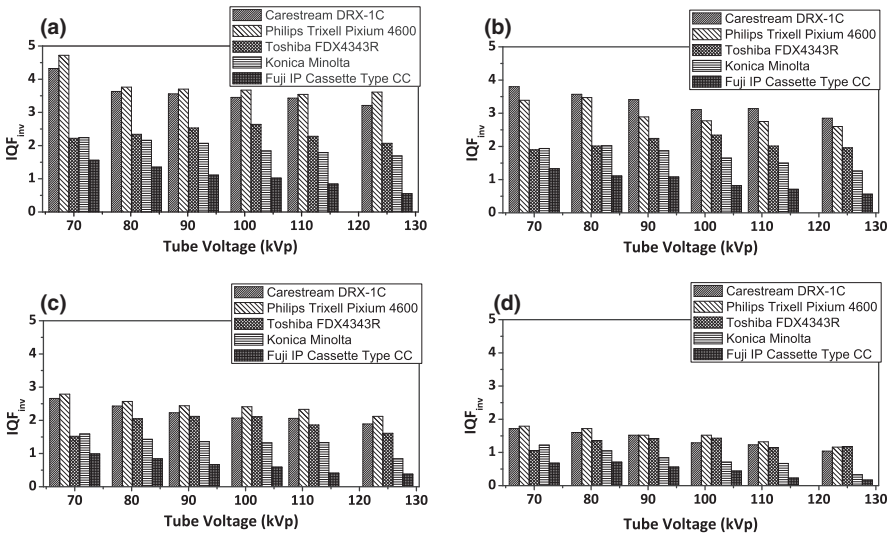

FIG. 4. The inverse image quality figure results for each imaging system at different tube voltages for a) 10 cm polymethyl methacrylate (PMMA), (b) 15 cm PMMA, (c) 20 cm PMMA, and (d) 25 cm PMMA thickness.

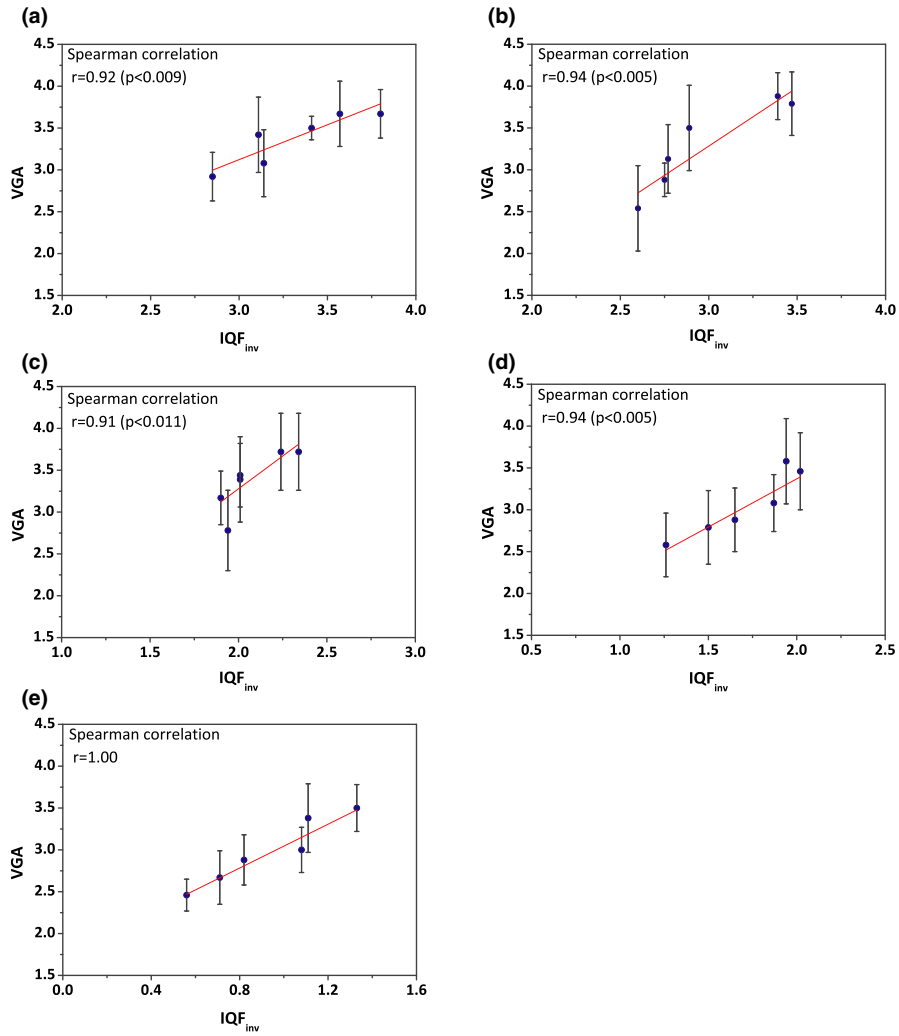

FIG. 5. The inverse image quality figure values vs mean visual grading analysis score for (a) DRX-1C, (b) Philips, (c) Toshiba, (d) Konica, and (e) Fuji.

TABLE IV. Mean visual grading analysis score results of six observers with different tube voltages for each imaging system.

kVp	Carestream DRX-1C	Philips Trixell Pixium	Toshiba FDX4343R	Konica Minolta	Fuji Ip Cassette type CC
70	3.67 (±0.29)	3.88 (±0.28)	3.17 (±0.32)	3.58 (±0.51)	3.50 (±0.28)
81	3.67 (±0.39)	3.79 (±0.38)	3.39 (±0.51)	3.46 (±0.46)	3.38 (±0.41)
90	3.50 (±0.14)	3.50 (±0.51)	3.72 (±0.46)	3.08 (±0.34)	3.00 (±0.27)
102	3.42 (±0.45)	3.13 (±0.41)	3.72 (±0.46)	2.88 (±0.38)	2.88 (±0.30)
110	3.08 (±0.45)	2.88 (±0.20)	3.44 (±0.38)	2.79 (±0.44)	2.67 (±0.32)
125	2.92 (±0.29)	2.54 (±0.51)	2.78 (±0.49)	2.58 (±0.38)	2.46 (±0.19)

VGAS at the same exposure settings. The Spearman correlation coefficient was used for the correlation and the results demonstrated a strong positive correlation (minimum r = 0.91, $P < 0.011$). Figure 6 represents the correlation between the IeDQE and mean VGAS at the same exposure settings. Correlation analysis demonstrated a strong positive correlation between IeDQE results and VGA score (minimum r = 0.92, $P < 0.009$). A significant intraclass agreement between six observers was observed (r = 0.92, $P < 0.008$), the mean score of all observers was used for the interobserver correlations.

4. DISCUSSION

Although digital imaging systems provide several advantages, they have some drawbacks. Due to their wide dynamic range, the digital radiographs provide a good image quality even if the patient is overexposed.[29] Therefore, it is essential to optimize the imaging parameters considering both image quality and radiation dose. Image quality evaluation is a more complicated issue than the radiation dose assessments.

In this study, the digital image quality was evaluated by using different methods. For physical image quality evaluations, the integrated values of eDQE results were investigated which takes the clinical conditions into account. The IeDQE

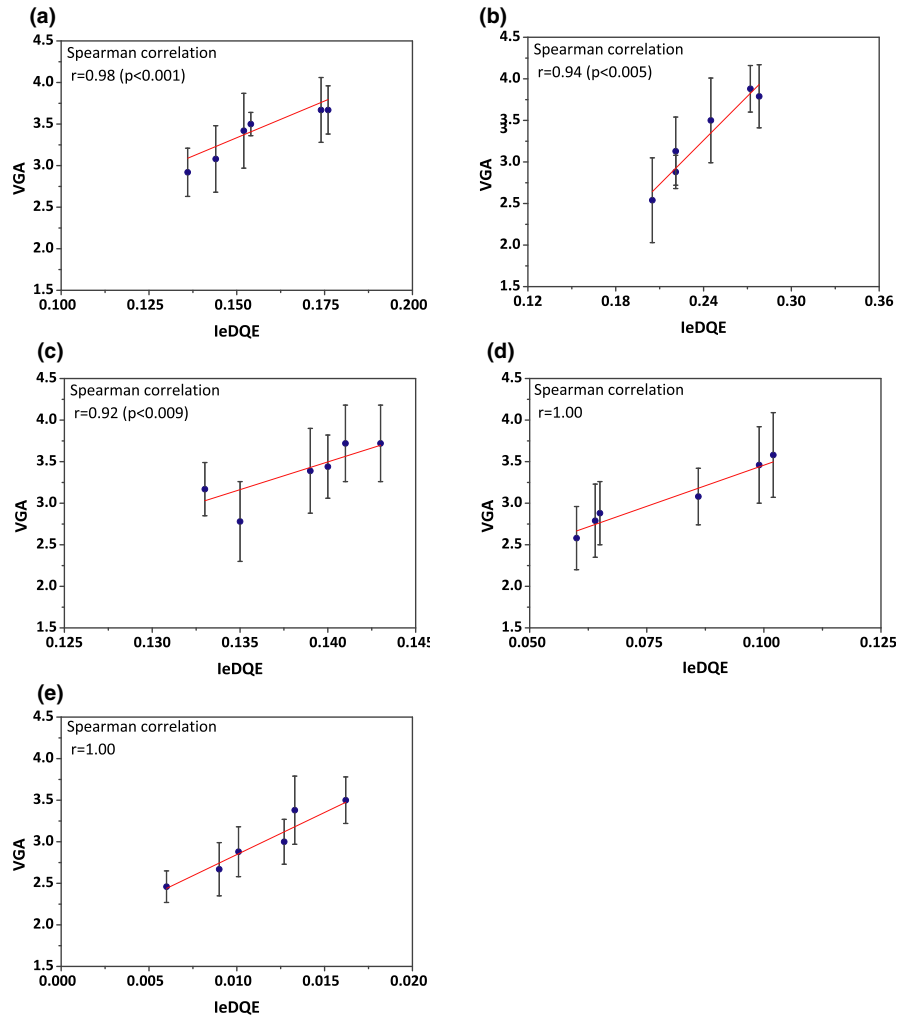

FIG. 6. The integrated effective detective quantum efficiency values vs mean visual grading analysis score for (a) DRX-1C, (b) Philips, (c) Toshiba, (d) Konica, and (e) Fuji. [Color figure can be viewed at wileyonlinelibrary.com]

results monotonically decreased with increasing tube voltage for all imaging systems used in the study, except for the Toshiba detector. The IeDQE results had a peak value at 102 kVp, and then the values decreased beyond this voltage. This can be explained by Lutfi et al's findings.[30] In this reference study, the researchers demonstrated that the electronic noise was dominant at low doses for the same imaging system. Similar results were observed for CD detectability and VGA evaluations for this detector. As Tingberg et al. have pointed out, noise levels and hence detector dose settings have a significant effect on the VGA and CDRAD score.[31]

In order to investigate the relationships of IeDQE and IQF_{inv} with VGAS, the IeDQE and IQF_{inv} results for 15 cm PMMA thickness were used. The calculated detector air kerma values (μGy) from phantom incident air kerma corrected to the detector plane and TF values (Ep.TF) for this PMMA thickness were 2.28, 2.14, 2.23, 1.83, and 2.13 μGy for DRX-1C, Philips, Toshiba, Konica, and Fuji, respectively. The results demonstrated significant correlation between IeDQE and clinical image quality evaluations [lowest $r = 0.92$ ($P < 0.009$)]. Similar results were obtained for IQF_{inv} vs VGAS [lowest $r = 0.91$ ($P < 0.011$)]. The results confirmed that clinical image quality can be predicted by both physical assessments and CD detectability studies.

Although high tube voltages (125 kVp) are mostly used in all radiology departments for chest radiography, these results indicated that IQF_{inv} and mean VGA score values decreased with increasing tube voltage in accordance with the IeDQE results. This was also concluded in some studies in the literature.[10,15,16]

Clinical image quality assessments using VGA scoring can be carried out with real patient images or alternatively with anthropomorphic phantoms. The results would be case dependent, and a large number of patient images should be evaluated for the former. In this study, an anthropomorphic chest/lung phantom was used, hence this case was eliminated.

For the digital radiography detectors with CsI scintillator, the IeDQE results were comparable, whereas the computed radiography system with Barium-based phosphor gave the lowest IeDQE values. The highest IeDQE results were observed with Philips. Although the detector doses were

comparable for the DRX-1C and Philips digital detectors, the IeDQE results for the DRX-1C were lower than those for Philips. As can be seen from Figs. 2(a) and 2(b), the performance of the Philips system at higher frequencies was better than DRX-1C. The eDQE results of the DRX 1C dropped quickly with increasing spatial frequency caused from the drop in the eMTF curve in comparison to those obtained with Philips. Toshiba had nearly the same pixel size and detector doses as the Philips (0.148 and 0.143 mm for Philips and Toshiba, respectively), but IeDQE values of Toshiba digital radiographic imaging system were lower than those for Philips. This may be explained by the different grid ratios used in these systems. In our previous study, it was demonstrated that the eDQE decreased with increasing grid ratio.[19] Toshiba had the grid with the highest grid ratio among the digital radiography systems and as a result, lower IeDQE results were determined for this system. Konica gave the lowest IeDQE results among four digital flat panel radiography systems. This is an expected result since the pixel size of that detector was the largest among all of the digital radiography systems studied. As mentioned before, the lowest IeDQE results were determined for Fuji computed radiography system. This is due to the detector material used in this detector which consists of granular-phosphor layer, where the light absorption and light scattering cause a large degradation in image quality. On the other hand, the flat panel digital radiography systems used in this study have columnar (needle-like) structured CsI scintillator.

Similar results were determined for CD detectability evaluations. The highest IQF_{inv} values were observed for Philips, whereas the lowest values were observed for Fuji digital radiography system. Significant correlation was determined between VGA and IQF_{inv} for all imaging systems. The lowest correlation coefficient was 0.91 ($P < 0.011$).

Significant correlation coefficients were determined for intraclass relationship between methods for each system. The results demonstrated a strong correlation between physical image quality measurements and clinical assessments.

The comparison results between imaging systems indicated that the order of systems with respect to clinical image quality from the highest to the lowest was found to be Philips, DRX-1C, Toshiba, Konica, and Fuji, in agreement with both physical image quality and CD detectability results.

It is not easy to compare the results of each image quality measurement technique directly. The VGA studies are commonly used to assess clinical image quality on real patient images or by using special anthropomorphic phantoms designed for this purpose. Implementation of the VGA method is not feasible within a routine quality assurance program because it requires a substantial additional workload for the radiologists and large patient numbers or phantom images.[10,16,18] The psychophysical measurements based on CD images analysis are more practical than VGA method because no patient data are required and images can be analyzed by the medical physicist or dedicated image analyzer software. The eDQE measurement technique is a new metric when compared with VGA and CD analysis methods and provides overall performance of the digital radiographic imaging system in clinical conditions.[5,9] However, the eDQE method also requires special phantoms and software developed for the mathematical processing of phantom images and should be performed by an experienced medical physicist. For the determination of the correlation between ordinal data (e.g., Likert scale) and continuous (interval) data Spearman correlation method was suggested in other studies.[10,18,32] Therefore, we used this statistical method to determine the correlation between VGA score (ordinal data), IQFinv and IeDQE (continuous data), respectively. But Keeble et al. have demonstrated that this should not be the way to analyze correlations between VGA (noncontinuous) and numerical metrics (continuous) and there is considerable debate within the statistics community on this subject.[33–35] Therefore, using the Spearman correlation method to determine the correlation between continuous and discontinuous data is one of the main limitations of this study, considering the discussions in the literature.

5. CONCLUSIONS

In this study, the image quality was evaluated in terms of the physical measurements and observer evaluations for digital chest radiography. The correlation between the physical and clinical image quality as quantified by VGA score was investigated over a wide range of tube voltage. Strong positive correlations were found between physical and clinical image quality findings. This showed that clinical image quality can be predicted with both physical assessments and CD detectability studies. The established significant correlation between physical and clinical image quality in the current study shows that observed changes in image quality measurement results implemented for optimization studies in digital radiography using CDRAD phantom can help to predict changes in the clinical image quality.

CONFLICT OF INTEREST

The authors have no conflict to disclose.

[a)]Author to whom correspondence should be addressed. Electronic mail: olgar@eng.ankara.edu.tr; Telephone: +90 (312) 203 3516, +90 (533) 5187768 (Mob.); Fax: +90 (312) 212 7343.

REFERENCES

1. Tapiovaara M. Relationships between physical measurements and user evaluation of image quality in medical radiology-a review. Report STUK-A219 (Helsinki: STUK-Radiation and nuclear safety. authority); 2006.
2. Mannson LG. Methods for the evaluation of image quality: a review. *Radiat Prot Dosimetry*. 2000;90:89–99.
3. Neitzel U, Günther-Kohfahl S, Borasi G, Samei E. Determination of the detective quantum efficiency of a digital x-ray detector: comparison of three evaluations using a common image data set. *Med Phys*. 2004;31:2205–2211.
4. Bushberg JT, Seibert JA, Leidholdt EM Jr, Boone JM. *The Essential Physics of Medical Imaging*, 3rd edn. Philadelphia: Lippincott Williams & Wilkins; 2012.

5. Samei E, Ranger NT, MacKenzie A, Honey ID, Dobbins JT, Ravin CE. Detector or system? Extending the concept of detective quantum efficiency to characterize the performance of digital radiographic imaging systems. *Radiology*. 2008;249:926–937.
6. Kyprianou IS, Rudin S, Bednarek DR, Hoffmann KR. Generalizing the MTF and DQE to include x-ray scatter and focal spot unsharpness: application to a new microangiographic system. *Med Phys*. 2005;32:613–626.
7. Fetterly KA, Hangiandreou NJ. Effect of x-ray spectra on the DQE of a computed radiography system. *Med Phys*. 2001;28:241–249.
8. Borasi G, Nitrosi A, Ferrari P, Tassoni D. On site evaluation of three flat panel detectors for digital radiography. *Med Phys*. 2003;30:1719–1731.
9. Bertolini M, Nitrosi A, Rivetti A, et al. A comparison of digital radiography systems in terms of effective detective quantum efficiency. *Med Phys*. 2012;39:2617–2627.
10. De Crop A, Bacher K, van Hoof T, et al. Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study. *Radiology*. 2012;262:298–304.
11. Leppert AG, Prokop M, Schaefer-Prokop CM, Galanski M. Detection of simulated chest lesions: comparison of a conventional screen-film combination, an asymmetric screen-film system, and a storage-phosphor radiology. *Radiology*. 1995;195:259–263.
12. van Heesewijk HP, van der Graaf Y, de Valois JC, Feldberg MA. Effects of dose reduction on digital chest imaging using a selenium detector: a study of detection simulated diffuse interstitial pulmonary disease. *AJR*. 1996;167:403–408.
13. ICRU Report 54. Medical imaging - the assessment of image quality, International Commission on Radiation Units and Measurements (International Commission on Radiation Units and Measurements). J ICRU 1996 ICRU Report 54 (Bethesda, MD: ICRU).
14. ICRU (International Commission on Radiation Units and Measurements). Image quality in chest radiography. ICRU Report 70 (Bethesda, MD: ICRU); 2003.
15. Sandborg M, Tingberg A, Ullman G, Dance DR, Carlsson GA. Comparison of clinical and physical measures of image quality in chest and pelvis computed radiography at different tube voltages. *Med Phys*. 2006;33:4169–4175.
16. Moore CS, Wood TJ, Beavis A, Saunderson JR. Correlation of the clinical and physical image quality in chest radiography for average adults with a computed radiography imaging system. *Br J Radiol*. 2013;86:1–12.
17. Sund P, Bath M, Kheddache S, Mansson LG. Comparison of visual grading analysis and determination of detective quantum efficiency for evaluating system performance in digital chest radiography. *Eur Radiol*. 2004;14:48–58.
18. Al-Murshedi S, Hogg P, England A. An investigation into the validity of utilizing the CDRAD 2.0 phantom for optimization studies in digital radiography. *Br J Radiol*. 2018;91:1–8.
19. Yalcin A, Olgar T. Characterizing the digital radiography system in terms of effective detective quantum efficiency and CDRAD measurement. *Nuclear Inst Methods Phys Res A*. 2018;896:113–121.
20. Marshall NW. Calculation of quantitative image quality paramaters, NHSBSPEquipment Report 0902 2009; NHS Breast Screening Programme, Sheffield.
21. Thijssen M, Bijkerk K, van der Burgth R. Manual contrast-detail phantom CDRAD type 2.0. project quality assurance in radiology. St Radboud, the Netherlands: Department of Radiology, University Hospital Nijmegen; 1998.
22. Samei E, Badano A, Chakraborty D, et al. Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. *Med Phys*. 2005;32:1205–1225.
23. European Commission. European Guidelines on Quality Criteria for Diagnostic Radiographic Images: Eur 16260 EN. European Commission, 1996.
24. Weaver KF, Morales V, Dunn SL, Godde K, Weaver PF. *An Introduction to Statistical Analysis in Research: With Applications in the Biological and Life Sciences*, 1st edn. Hoboken: John Wiley & Sons Inc; 2017.
25. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163.
26. Miller RG Jr. *Simultaneous Statistical Interference*, 2nd edn. Berlin, Heidelberg: Springer, New York; 1980.
27. Poludniowski G, Landry G, Deblois F, Evans PM, Verhaegen F. SpekCalc: a program to calculate photon spectra from tungsten anode X-ray tubes. *Phys Med Biol*. 2009;54:N433–N438.
28. Samei E, Ranger NT, MacKenzie A, Honey ID, Dobbins JT 3rd, Ravin CE. Effective DQE (eDQE) and speed of digital radiographic systems: an experimental methodology. *Med Phys*. 2009;36:3806–3817.
29. Alsleem H, Davidson R. Quality parameters and assessment methods of digital radiography images-a review. *Radiographers*. 2012;59:46–55.
30. Ergun L, Olgar T. Investigation of noise sources for digital radiography systems. *Radiol Phys Technol*. 2017;10:171–179.
31. Tingberg A, Båth M, Håkansson M, et al. Evaluation of image quality of lumbar spine images: a comparison between FFE and VGA. *Radiat Prot Dosimetry*. 2005;114:53–61.
32. Khamis H. Measures of association how to choose? *JDMS*. 2008;24:155–162.
33. Keeble C, Baxter PD, Gislason-Lee AJ, Treadgold LA, Davies AG. Methods for the analysis of ordinal response data in medical image quality assessment. *Br J Radiol*. 2016;89:20160094.
34. Norman G. Likert scales, levels of measurement and the "laws" of statistics. *Adv Health Sci Educ*. 2010;15:625–632.
35. Sullivan GM, Artino AR. Analyzing and interpreting data from Likert-Type Scales. *J Grad Med Educ*. 2013;5:541–542.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Measured scatter fractions for different tube voltages and polymethyl methacrylate thicknesses.

Table S2. Phantom transmission fractions and q-values for different tube voltages and polymethyl methacrylate thicknesses.

Figure S1. Effective detective quantum efficiency results for each tube voltage and 25 cm polymethyl methacrylate thickness for (a) DRX-1C, (b) Philips Trixell Pixium 4600, (c) Toshiba FDX4343R, (d) Konica Minolta AERO DR, and (e) Fuji FCR.