Radiography 20 (2014) 306–311

Contents lists available at ScienceDirect

Radiography

journal homepage: www.elsevier.com/locate/radi

Review article

Developing and validating a psychometric scale for image quality assessment

H. Mraity [a,b,*], A. England [a], P. Hogg [a]

[a] University of Salford, United Kingdom
[b] University of Kufa, Iraq

ARTICLE INFO

Article history:
Received 7 February 2014
Accepted 5 April 2014
Available online 9 May 2014

Keywords:
Pelvis
Image quality
Psychometric scale

ABSTRACT

Purpose: Using AP pelvis as a catalyst, this paper explains how a psychometric scale for image quality assessment can be created using Bandura's theory for self-efficacy.
Background: Establishing an accurate diagnosis is highly dependent upon the quality of the radiographic image. Image quality, as a construct (i.e. set of attributes that makes up the image quality), continues to play an essential role in the field of diagnostic radiography. The process of assessing image quality can be facilitated by using criteria, such as the European Commission (EC) guidelines for quality criteria as published in 1996. However, with the advent of new technology (Computed Radiography and Digital Radiography), some of the EC criteria may no longer be suitable for assessing the visual quality of a digital radiographic image. Moreover, a lack of validated visual image quality scales in the literature can also lead to significant variations in image quality evaluation. Creating and validating visual image quality scales, using a robust methodology, could reduce variability and improve the validity and reliability of perceptual image quality evaluations.

Introduction

Psychometric scale development and validation for image quality applications is not well understood by the wider radiography audience and it was felt appropriate to give context to another paper in this special issue of *Radiography* that has used this method.[57]

Radiographic images of the human body continue to provide a fundamental source of information that can assist clinicians with diagnosis and management. However, diagnostic accuracy is highly dependent upon the quality of information within the image and subsequently the quality of an image can affect how a patient will be managed.[1,2] Image quality plays an essential role in radiographic dose optimisation. Optimisation involves producing an image with acceptable image quality and with low patient radiation dose (ALARP principle).[3,4] Estimates of radiation doses received by patients are relatively easy to make. By contrast, image quality assessments can be difficult and time consuming.[5]

Image optimisation is generally concerned with making an image which is fit for purpose. Fit for purpose is rarely defined adequately within clinical papers. Ultimately, the quality of an image is a descriptor of the subjective analysis of the visual data contained within it.[6,7] When defining the quality of an image the purpose of the image should also be considered. The reason for this is that a variety of factors can affect the quality of an image. It is widely agreed that image quality can be defined in terms of its acceptability for answering the primary clinical question.[8–10]

Apart from the physical and technical parameters, it is important to take into account human perception and cognition. Image perception refers to the unified realisation of the content of the displayed image (image signal), whereas human cognition can be defined as the ability to determine the meaning of observed data in the context of the medical problem that led to the production of a particular image. Accordingly, this may significantly affect the process of image quality evaluation.[1,11]

Image quality evaluation

The utility of radiographic images and the precision of image interpretation are highly dependent upon the image quality and observer competency. Several approaches can be used to measure the quality attributes of an image.[12] These approaches include

* Corresponding author. University of Salford, Salford M54WT, United Kingdom. Tel.: +44 (0) 7741756986.
E-mail addresses: h.a.a.mraity@edu.salford.ac.uk, hussien_zahed@yahoo.com (H. Mraity).

http://dx.doi.org/10.1016/j.radi.2014.04.002
1078-8174/

physical, psychophysical and clinical/diagnostic performance (Fig. 1). Examples of physical methods include signal to noise ratio (SNR), modulation transfer function (MTF), detector quantum efficiency (DQE) and noise. The physical methods can be used to evaluate system performance.

Psychophysical methods represent visual evaluation of physical parameters (e.g. line pairs/mm-spatial resolution testing) and contrast detail analysis. It must be stated that by using a visual/clinical method such as Visual Grading Analysis (VGA) or Receiver Operating Characteristic (ROC) when quantifying image quality then this may be more relevant than physical measures, since visual methods focus on how clearly each anatomical structure/pathology can be visualised by an observer.[13] Within ROC and VGA the observer is required to set their agreement/confidence on how clearly the anatomical structure or pathology is seen within a given image. This approach reflects observer opinion and therefore it is highly susceptible to inter-observer variability.[14] A further limitation of ROC and VGA techniques is that the individual anatomical structures under evaluation must be pre-specified. No formal guidelines on this exist; it is also likely that these will be highly variable between studies.

Variability in assessing image quality

System performance may not only be the sole reason behind expected diagnostic variations, rather, observer variability could also have a significant contribution on the overall diagnostic accuracy.[15] This issue could influence the reliability of results that are obtained from visually based image quality assessment methods. Variability in determination of image quality has been investigated since the 1940s. This variation in image quality assessment may result from a lack of standards – such as a common visual evaluation scale. In this context, Krupinski and Jiang[16] have identified two important issues which need to be considered when addressing variability: systems are required to minimise observer interpretation variation; approaches are required to assess the systems and their influence on observer interpretation.

With due consideration of the problems of assessing visual image quality, several years ago it was felt that new standards would be needed.[17] This led the Council of European Communities (EC) to establish a list of image quality criteria for a number of common radiographic procedures.[18] The purpose behind these criteria was to standardise practice and also reduce the variability in the radiation dose and most importantly in the evaluation of image quality. Since the guidance was published several articles have highlighted deficiencies with the criteria.[19,20] For instance, some of the criteria have been difficult to use and in some cases anatomical areas are missing (e.g. iliac crests for the pelvis). Two main conclusions can be drawn: the EC criteria are not perfect and the health care professionals may interpret them differently; the EC image criteria were designed for X-ray film and with the advent of new technology new criteria are likely to be required to better fit digital images.[21]

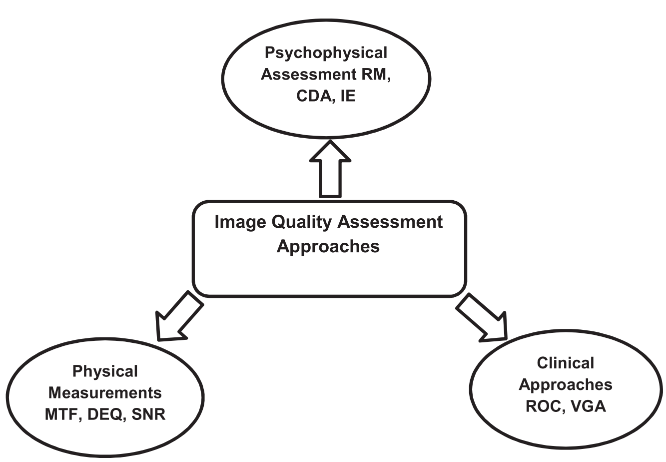

Figure 1. Different methods for image quality evaluation.[12]

Variations in visual quality criteria

Within this paper we have used the AP pelvis projection as the basis for explaining how Bandura's theory of self-efficacy can be applied to develop and validate a visual image quality scale.

A review of the literature revealed that there are significant variations in how image quality for AP pelvis can be assessed. For example, several authors measured the overall diagnostic quality in terms of image noise and contrast whilst others used EC criteria.[22,23] In other cases the researchers found that these approaches were not appropriate and opted to assess diagnostic usefulness, to give an impression of how much noise was present in the image and whether this affected image clarity. Two recent studies[24,25] attempted to develop an optimisation framework for AP pelvis radiography. In one study the researchers used the EC criteria with the addition of three more physical inclusion criteria (i.e. contrast, noise and sharpness). For the other, the authors completely depended on the EC criteria. There appears to be a complete lack of common standards for assessing image quality, and this will no doubt result in a large variation of image quality assessments with potentially conflicting results.[10,26]

In addition to the variability of quality criteria for assessing AP pelvis there are other issues that need to be considered. The process of image quality evaluation in diagnostic imaging is of a psychometric/psychophysical nature. Psychophysics in radiology refers to the way in which the relationship between the visual stimuli and human response can be studied quantitatively.[11] This process comprises of two fundamental elements, namely the human observer and the displayed image. An observer is required to perceive the information carried by an image and therefore analyse it to institute the required decision (Fig. 2).[27]

It would not be practical to separate the physical effects from psychological ones, in the sense of recognising image contents. For this purpose, it could be concluded that achieving appropriate clinical judgement for the visual quality of an image would necessitate the appreciation of those attributes/factors that may impact on observer response.[27] These factors can be represented by anatomical and physical landmarks of the visualised body parts in the X-ray image. This may greatly contribute to the reduction of observer variability through focussing their attention upon certain features in the image.[28–30]

It has become evident that there is no validated published visual image quality scale for AP pelvis. Developing a psychometric scale for visual image quality comprising appropriate factors which have a significant influence on image quality should promote improvements in image quality assessment using visual means.

Measurement scale

Measurement of an image can either be made directly (e.g. physical measurement of image attributes – e.g. SNR) or indirectly (e.g. visual). Visual assessment of image quality can be done simply (e.g. no criteria and just asking for an impression from a clinician) or through use of validated image quality criteria used within a scale. In this context a scale can be defined as an instrument whose main function is to measure a variety of visual characteristics/