Development and validation of a psychometric scale for assessing PA chest image quality: a pilot study

Mraity H[12], England A[1], Akhtar I[1], Aslam A[3], De Lange R[4], Momoniat H[1], Nicoulaz S[5], Ribeiro A[6], Mazhir S[1], Hogg P[1].

University of Salford[1], United Kingdom.

University of Kufa[2], Iraq

Oslo and Akershus University College of Applied Sciences[3], Norway

Hanzehogeschool Groningen[4], The Netherlands

Haute École de Santé Vaud[5], Switzerland

Escola Superior de tecnologias da Saúde de Lisboa[6], Portugal

ABSTRACT

PURPOSE: To develop and validate a psychometric scale for assessing Image quality perception for chest radiographs.

METHODS: A review of the literature was undertaken to identify items/factors which could be used to evaluate image quality using a perceptual approach. A draft scale was then created (22 items) and presented to a focus group (student, qualified radiographers and radiology registers). Within the focus group the draft scale was discussed and modified. A series of seven postero-anterior chest images were generated using a phantom with a range of image qualities. Image quality perception was initially confirmed for the seven images using signal-to-noise ratio (SNR) and group consensus. Participants were invited to independently score each of the images using the draft image quality perception scale. Bandura's theory was used to guide scale development and Cronbach alpha was used to test interval reliability.

RESULTS: Fifty three participants used the scale to grade image quality perception on each of the seven images (SNR 17.2 to 36.5). Aggregated mean scale score increased with increasing SNR from 42.1 to 87.7 (r=0.98, P<0.001). For each of the 22 individual scale items there was clear differentiation of low, mid and high-quality images. A Cronbach alpha coefficient of >0.7 was obtained across each of the seven images.

CONCLUSION: This study represents the first development of a chest image quality perception scale based on Bandura's theory. There was excellent correlation between the image quality perception scores derived using the scale and the SNR and group consensus. Further research will involve a more detailed item and factor analysis.

Introduction

Chest radiography is one of the most frequently performed diagnostic radiographic examinations in the United Kingdom. A recent Health Protection Agency report (2010) showed that chest radiography represented 19.6% of all radiographic examinations [1]. Despite its dominance in the clinical assessment of a whole myriad of diseases the interpretation of a chest radiograph is notoriously difficult [2]. Interpretation can be improved by evaluating images of optimum diagnostic quality. For this to be possible there needs to be robust mechanisms for assessing image quality.

Primary perception of image quality can be measured by asking observers to assess contrast, spatial resolution and noise of an image. Physical assessment can be undertaken by measuring the detective quantum efficiency (DQE), modular transfer function (MTF) and signal-to-noise ratio (SNR) [3]. These parameters are, however, more of a measure of the system performance rather than the 'real' clinical image quality perception.

Before any realistic assessment of image quality perception can be made, the requirements need to be defined. These requirements include whether the necessary clinical information is contained within the image and whether this can be interpreted by the observer, rather than whether the appearance of the image is pleasing to the eye [4]. For measuring the actual clinical quality of an image (ie. assessing images from real human or human simulated phantom) an observer performance measurement is needed. These perceptual measurements can be undertaken by assessing an image using the receiver operating characteristic (ROC), [5] and visual grading analysis (VGA). The latter could be through use of a grading scale of criteria. Such VGA systems have been used for studies investigating radiographic image quality perception of the pelvis and lumbar spine [6-8].

A literature review highlighted that many articles relating to chest X-ray image quality perception investigated physical measurements (such as SNR) and dose reduction, rather than observer perception of image quality [9-10]. Articles which investigated perception of observer image quality have often been based on the Commission of European Communities (CEC) 1996 image criteria[11] or a modified version of those criteria. However, there are problems when using the CEC criteria, they are outdated and researchers tend to adjust the scale according to their needs because the criteria were developed for images acquired on film-screen systems. Presently there is no published and validated perceptual scale for assessing image quality. Lack of such a scale can lead to a lack

of consistency in evaluating image quality by perceptual means. There is, therefore, a need to create and validate scales for use by clinical and research staff.

Image quality perception assessment involves observers considering how much image detail (i.e. anatomical structures or abnormalities visualization) they can perceive. The approach often involves observers rating criteria using a Likert scale. This approach considers two issues: physical information within an image (stimulus or signal) and perceptual effect (psychological) that is related to human analysis of the perceived image [12, 13]. For the latter, the approach we have taken to scale creation and validation capitalizes on Bandura's theory for self-efficacy. This theory provides a robust theoretical background to guide the methodology of creating and validating a perception scale. This theory has been widely used in psychology and health fields. It gives a self-reported measure of an individual's perception of how confidently they feel in relation to performing a specific a task [14, 15]; equally it can give a self-reported measure of somebody's attitude toward a given topic. On this basis the theory can be adapted to developing and validating a scale for determining perception of image quality.

With specific reference to exposure factors the aim of this study was to create a psychometric scale to assess image quality perception of postero-anterior (PA) chest X-ray images.

Materials and methods

This study was conducted within a 3 week residential summer school. Limitations of time meant data collection could only occur over a one week period. The research consisted of two phases: Phase 1 - a literature review and focus group discussion to develop the image quality perception scale; Phase 2 - validate the scale using a series of 7 phantom images of known image quality.

Literature review and scale development

A literature review was conducted in order to identify the determinants of image quality perception of a PA chest radiograph (11, 16-18). Factors were developed into scale items suitable for creating the psychometric scale; this was guided by Bandura's theory of self-efficacy and the literature surrounding the construction of a psychometric scale [19,20]. Since all images were acquired using a phantom, factors related to positioning and

movement were excluded from the scale. The scale items generated from the literature review were presented to a focus group consisting of radiographers and students. Amendments were made to the scale based on the focus group feedback. Some of the scale items were negatively worded in order to avoid affirmation bias. A 5-point Likert scale was used to quantify the response. Items that had negatively worded statements were reversed so that all responses were unidirectional for scoring purposes (i.e. a score of 5 indicated a higher level of self-efficacy).

Chest phantom images

Images were acquired on a Wolverson Acoma X-ray unit (high frequency generator with VARIAN 130 HS standard X-ray tube with a total filtration of 3 mm Aluminum equivalent). An adult anthropomorphic chest phantom (LungMan) [21] was positioned in accordance with Clark's Positioning in Radiography [22] for a PA chest projection with a source-to-image receptor distance (SID) of 180 cm to mimic clinical conditions. The position of the phantom was kept constant in order to eliminate positioning errors. The primary x-ray beam was collimated to the edges of the image receptor (IR).

Images were acquired on the same 35cm x 43 cm Agfa Computed Radiography (CR) IR and image processing was undertaken using an Agfa 35-X digitizer. A secondary radiation grid was not used and all equipment quality control met the required specifications of Institute of Physics and Engineering in Medicine (IPEM) report 91[23]. To simulate clinical conditions all images were displayed using a chest look up table. The pre-set contrast and brightness settings were used to display the images. Exposure factors, dose area product (DAP) and exposure index (EI) were recorded for each image.

Determination of exposure factor combination

A preliminary investigation was conducted to determine the exposure factor combinations required to produce 7 images to validate the psychometric scale. During the investigation, the kVp was initially fixed at 85kVp. Images were then acquired at varying mAs values. The first image was acquired with an mAs of 1.6. Subsequent images were acquired by increasing the mAs by one increment allowed by the system.

Based on group consensus it was concluded that there was no visual change in image quality perception when mAs was increased by one increment. Therefore, the mAs was increased in increments of 2, then 3, and finally 4 - when a noticeable change in visual image quality was observed. At 25 mAs the images were over exposed and the study was terminated because the anatomical detail visualized on the image was inadequate. A similar technique was then used to determine the range and increments of kVp . Consequently, 35 possible exposure combinations were identified.

Image selection

SNR was calculated for the 35 images using four regions [24]. Seven images were selected, representing low to high SNR. SNR ranking agreed with group consensus ranking. Three of the seven images (poor, mid- and high-quality) are demonstrated in **Figure 1**.

Figure 1 - Low quality image (1), middle quality image (26) and highest quality image (28)
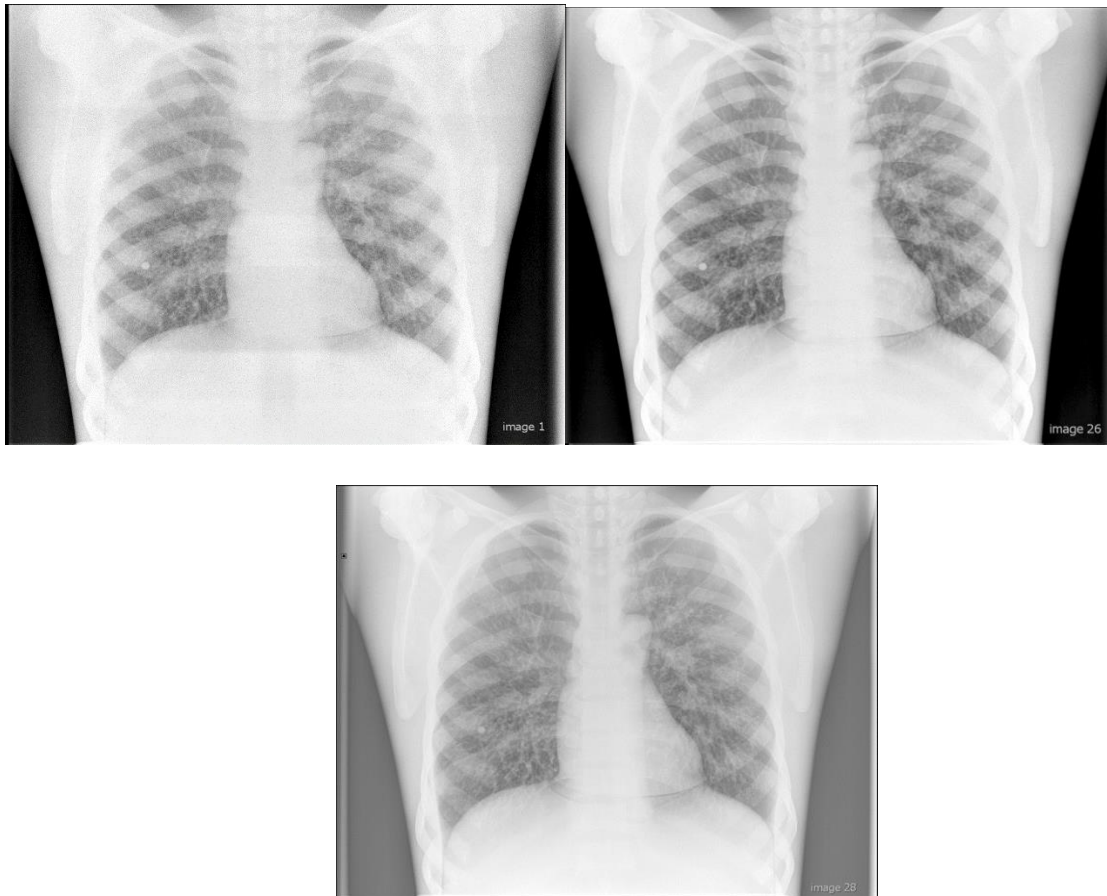
Validation of the scale

53 observers from 5 countries (student radiographers, qualified radiographers and radiology registrars) participated in scale validation. Observers were required to indicate their level of agreement with each scale item, where 1 was equivalent to strongly disagree and 5 was strongly agree. The draft scale consisted of 22 items (**Appendix A**).

Each participant viewed the images on a 22 inch Iiyama ProLite liquid crystal display (LCD) monitors (B2206WS) with a resolution of 3 megapixels. Monitors were calibrated to DICOM grayscale standard display function (GSDF) the ambient lighting conditions were kept constant and dimmed. Observers were trained to be able to complete the scale validation task. For each image the scores from each participant were aggregated to reflect the overall quality of each image.

It is necessary to say that each observer completed one scale for each of the 7 images. In total this means that 371 (53X7) scales were completed.

Statistical Analysis

All data were transferred to MS Excel 2010 (Microsoft Corp, Redmond, WA) where the mean values and the standard deviation were calculated. Using Excel mean SNR was plotted against the mean scale scores. A separate plot was also created for each of the 22 items. Following this, all data were transferred to SPSS Statistics for Windows, Version 20.0 (Armonk, NY: IBM Corp) where the Cronbach Alpha coefficients were calculated.

Results

Fifty three observers (final year undergraduate radiography students (n=28), postgraduates radiography students (n=2), radiography tutors (n=6), qualified radiographers (n=4) and 13 radiology registrars) completed the scale. Scores from each observer were aggregated to reflect the overall quality of each of the assessed images. The mean scale score was compared for each of the seven images against the SNR, DAP and EI.

The mean of scale scores was lower for images with a lower SNR, 42.1 (SD=10.8) when compared with an image with a high SNR, 86.7 (SD=10.8). The mean scale score also showed a strong positive correlation with SNR (Figure 2, $R=0.98$; $P<0.01$).

Figure 2. Changes in SNR trends against mean scale scores for the seven images (aggregated)
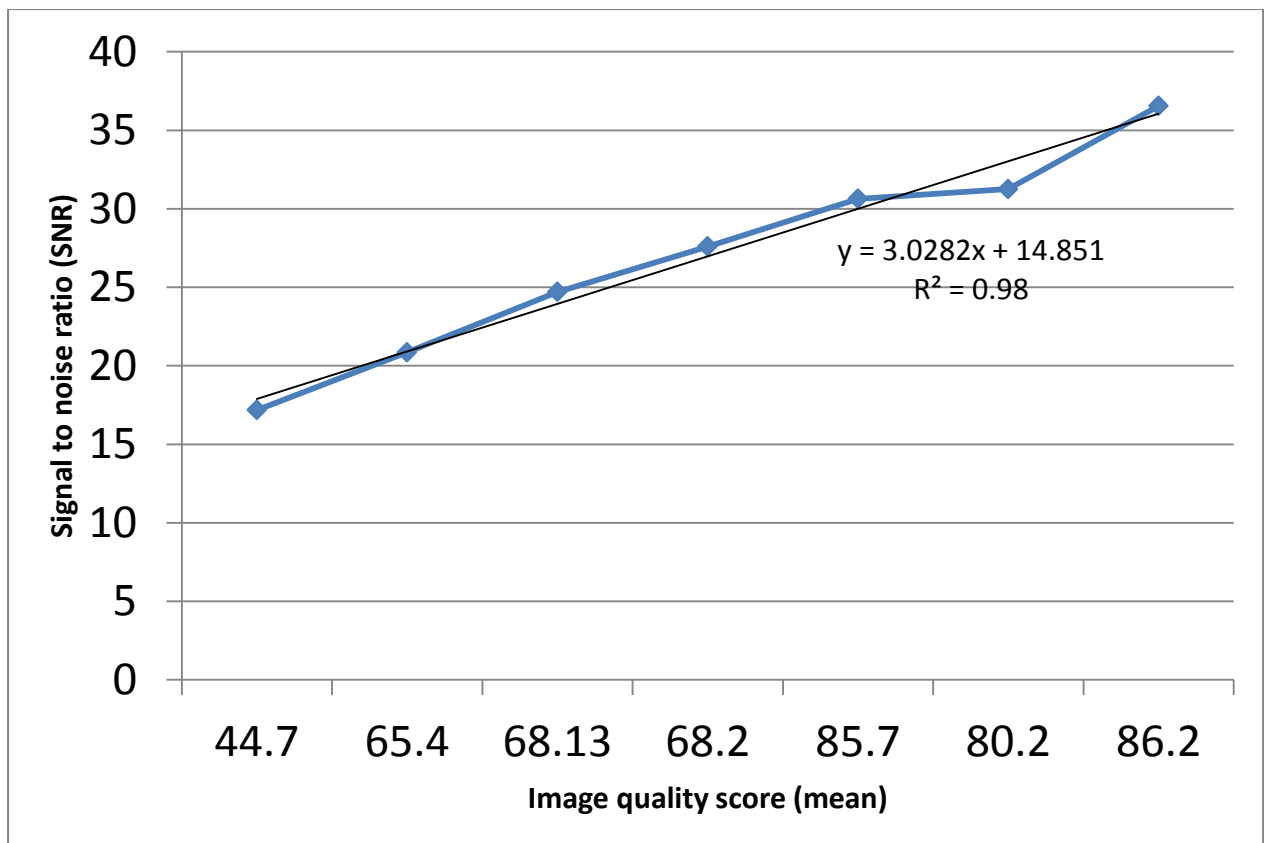
The next step was to consider the responses of each of the 22-items across a range of image qualities. To facilitate this analysis three images were chosen with the lowest (17.18), mid (27.59) and highest (36.54) SNR. This would indicate that scale principally works across a wide range of qualities and therefore suggest acceptable scale validity. Mean values for each of the 22-items were plotted for each of the three images (**Figure 3**).

Figure 3. Responses for individual scale items for three images of differing qualities (Numbers at the ends of the lines correspond to the SNR).
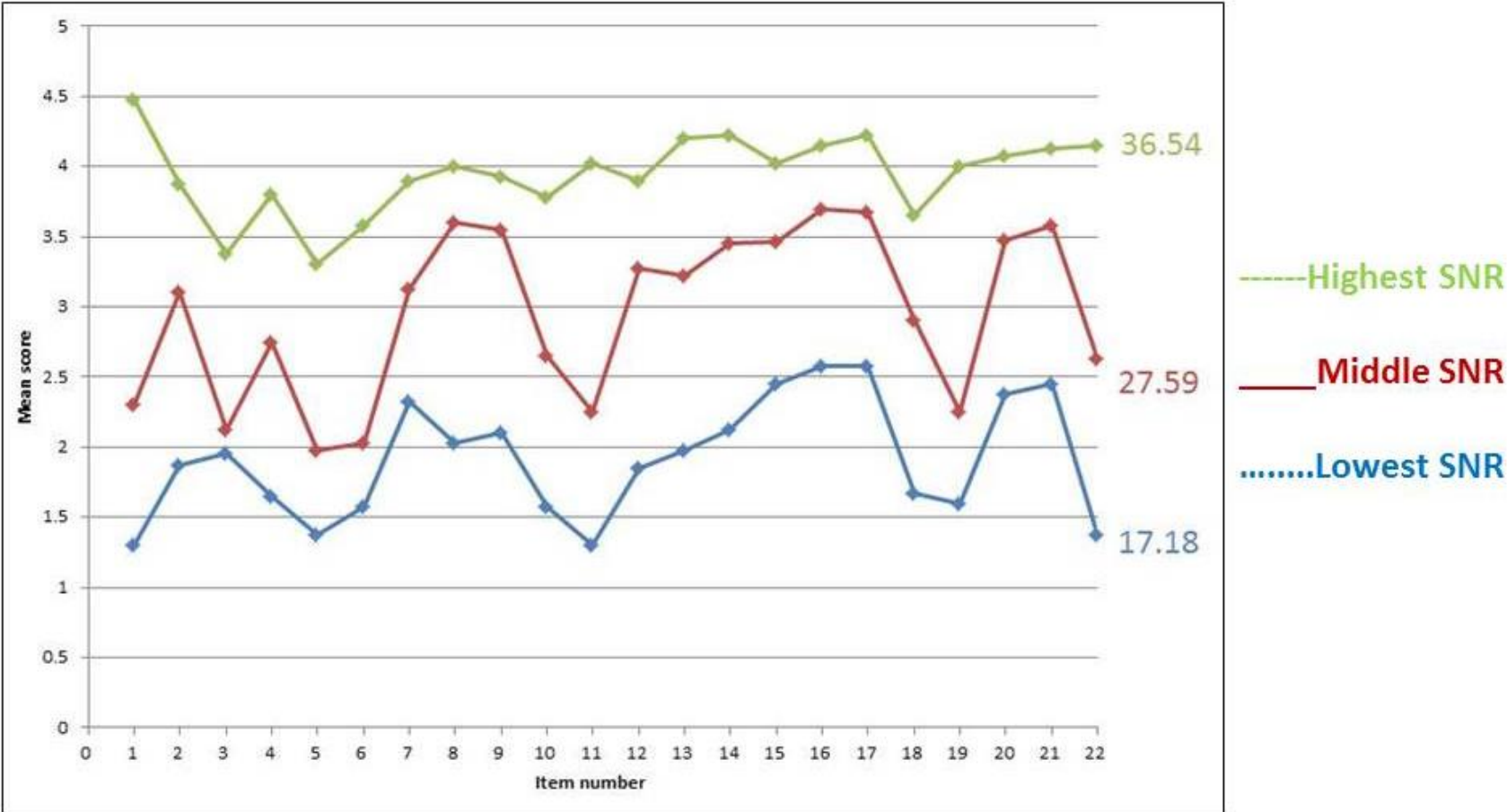
Next, internal reliability of the scale was assessed using Cronbach's alpha to measure how well the items correlated with each other and how each item correlated with the total score[25]. Cronbach (1951) suggested a value of 0.6 as a standard lenient cut off point for each extracted factor [26]. However other authors have recommended 0.7 as the acceptable value for internal reliability [27]. The Cronbach values for all 7 images fall above these values (**Table 4**), indicating good internal reliability. Image-4 indicated the highest internal reliability for this study, with a Cronbach value of 0.896.

Table 4. Demonstrates internal reliability coefficients for scale items across all image set.	
Image set No.	Alpha Coefficients
1	0.792
2	0.833
3	0.874
4	0.896
5	0.837
6	0.854
7	0.869

Discussion

The aim of this project was to use self-efficacy to develop and validate an image quality perception scale for PA chest radiography. For our 22 item scale, we found a strong relationship between SNR and the scale score. This suggests that the perceptual measure of image quality is relatively valid and reliable.

In all cases, Cronhbach alpha coefficients were greater than 0.7 for all images. This suggests the scale has good reliability. The scale items are therefore consistent in measuring perception of image quality.

Due to a lack of publications which discuss the creation and validation of a scale for assessing image quality perception of PA chest radiographs it is impossible to compare our work with similar studies. However, the scale does represent a more robust way of assessing image quality for chest x-rays than anything published so far, simply because the scale has validation data associated with it whereas others have only stated criteria for assessing image quality without actually testing whether they work (this includes the CEC criteria[10]).

Limitations of the scale and further work

Images were generated using a phantom; therefore, it was not possible to assess the effects of pathology. Other factors which could have been taken into account include problems with positioning, the use of high kVp, AEC and an anti-scatter grid. Further work should consider taking these factors into account.

The sample size of the participants could be considered a limitation. Other studies have used in excess of 150 participants for scale development [28, 29]. Adequate sample size would ensure reducing the random error that might be associated with human based studies. Since data could only be collected over a one week period we suggest that our research be extended to include a larger sample size. However, this was considered acceptable as a pilot to primarily testing the scale and obtaining a considerable feedback regarding the suitability of items for measuring what they are designed to measure and their reliability [30]

Conclusion

The development and validation of our image quality perception scale will allow the assessment of image quality perception in both clinical and academic environments. Although only a pilot, for the first time this study has taken the initial steps to create and validate a reliable scale for the assessment of PA chest radiographs. The scale, even in its current stage of development, could provide a valuable contribution to help standardize visual assessment of PA chest image quality. More work is needed to complete the development and validation process but this 22-item scale is a considerable improvement to the currently available scales.

REFERENCES

1. **D. Hart, B. F. Wall, M. C. Hillier, and P. C. Shrimpton, Frequency and Collective Dose for Medical and Dental X-ray Examinations in the UK , 2008, no. December 2010. 2008.**
2. **Van Ginneken, B., Ter Haar Romeny, B. M., Viergever, M. A. Computer-aided diagnosis in chest radiography: a survey. Medical Imaging, 2001;20(12):1228-1241.**
3. **Radiation exposure and Image quality perception in x-ray diagnostic radiology : physical principles and clinical applications H Aichinger (Horst), 1935-Berlin ; London : Springer 2nd ed. 2011.**
4. **Martin CJ, Sutton DG, Sharp FP. Balancing patient dose and image quality. Applied radiation and isotopes. 1999;50(1):1-19.**
5. **Thompson J, Hogg P, Thompson S, Manning D, Szczepura K. ROCView: prototype software for data collection in jackknife alternative free-response receiver operating characteristic analysis. The British journal of radiology. 2012 Sep;85(1017):1320-6. PubMed PMID: 22573294. Pubmed Central PMCID: 3487065.**
6. **Manning-Stanley, A.S., Ward, A. J., England, A. Options for radiation dose optimisation in pelvic digital radiography: a phantom study. Radiography, 2012;18:256-263.**
7. **Heath, R., England, A., Ward, A., Charnock, P., Ward, M., Evans, P., et al. Digital pelvic radiography: increasing distance to reduce dose. Radiol Tech 2011;83(1):20-8.**
8. **Mekis, N., McEntree, M. F., Stegnar, P. PA positioning significantly reduces testicular dose during sacroiliac joint radiography. Radiography 2010;16:333-8.**
9. **Neitzel, U., Maack, I., Gunther-Kohfahl, S. Image quality perception of a digital chest radiography system based on a selenium detector. Medical Physics 1994;21(509): http://dx.doi.org/10.1118/1.597389**
10. **Tingberg, A. M. Quantifying the quality of medical x-ray images: An evaluation based on normal anatomy for lumbar spine and chest radiography. Thesis(PhD). Lunds Universitet (Sweden), 2000.**
11. **Commission of European Communities. European guidelines on quality criteria for diagnostic radiographic images. EUR 16260EN. Brussels, Belgium, The European Commission, 1996.**
12. **Mansson LG. Methods for the evaluation of image quality . Radiation protection dosimetry. 2000;90(1–2):89–99.**
13. **Sharp FP. Quantifying image quality. Clinical Physics and Physiological Measurement. 1990;11:21-6.**
14. **Bandura A. Social foundations of thought and action: Asocial cognitive theory. New Jersey: Prentice Hall. ; 1986.**
15. **Niedermann K, Forster A, Ciurea A, Hammond A, Uebelhart D, de Bie R. Development and psychometric properties of a joint protection self-efficacy scale. Scandinavian journal of occupational therapy. 2011 Jun;18(2):143-52. PubMed PMID: 20450381.**
16. **Carver E, Carver B. Medical imaging: techniques , reflection & evaluation. Elsevier: Churchill Livingstone; 2006**
17. **Allen E, Hogg P, Ma WK, Szczepura K. Fact or fiction: An analysis of the 10 kVp 'rule' in computed radiography. Radiography. 2013;19(3):223-7.**
18. **. Ma WK, Hogg P, Tootell A, Manning D, Thomas N, Kane T, et al. Anthropomorphic chest phantom imaging – The potential for dose creep in computed radiography. Radiography. 2013;19(3):207-11.**
19. **Bandura A. In: Pajares F, Urdan T, editors. Self-efficacy Beliefs of Adolescents. New York: Information Age; 2006. p. 307-37.**
20. **Lev, E. L. Bandura's Theory of Self-Efficacy: Applications to Oncology. Research and Theory for Nursing Practice 1997;11(1):21-37.**
21. **Kyoto Kagaku Ltd. Multipurpose chest phantom N1 "LUNGMAN" http://www. kyotokagaku.com/products/detail03/ph-1.html [last accessed 18.05.13].**
22. **Whitley, A. S., Sloane, C., Hoadley, G., Moore A. D., Alsop, C. W. Clark's positioning in Radiography. 12th Ed. London, Hodder Arnold.**
23. **Institute of Physics and Engineering in Medicine (IPEM). Recommended standards for the routine performance testing of diagnostic X-ray imaging systems. Report No. 91. York, IPEM; 2005.**

24. **Lin Y, Luo H, Dobbins JT, Page McAdams H, Wang X, Sehnert WJ, et al. An image-based technique to assess the perceptual quality of clinical chest radiographs. Medical physics. 2012 Nov;39(11):7019-31. PubMed PMID: 23127093.**

25. **M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," International Journal of Medical Education, vol. 2, pp. 53–55, Jun. 2011.**

26. **Cronbach LJ. Coefficient alpha and the internal structure of tests. P sychometrka. 1951;16(3):297-334.**

27. E**. R. Waclawski, "Health Measurement Scales--A Practical Guide to Their Development and Use," Occupational Medicine, vol. 60, no. 2, pp. 156–156, Feb. 2010**

28. **J. Kitching, S. Cassidy, P. Eachus, and P. Hogg, "Creating and validating self-efficacy scales for students.," Radiologic technology, vol. 83, no. 1, pp. 10–9, 2011.**

29. **Champion V, Skinner CS, Menon U. Development of a self-efficacy scale for mammography. Research in nursing & health. 2005 Aug;28(4):329-36. PubMed PMID: 16028267.**

30. **Fliess-Douer O, van der Woude LH, Vanlandewijck YC. Development of a new scale for perceived self-efficacy in manual wheeled mobility: a pilot study. Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine. 2011 Jun;43(7):602-8. PubMed PMID: 21584482.**

Box : Image quality in Computed Radiography: PA chest projection

1. The spinous processes of the thoracic vertebrae T1-T4 are visualised adequately

2. The anterior ribs are adequately visualised through the heart

3. The vertebral bodies are not sufficiently visualised through the heart

4. The vascular patterns of the lungs are adequately visualised in the retro-cardiac region

5. The carena is not visualised adequately

6. The vertebral disc spaces are not visualised clearly

7. The mediastinum is visualised clearly

8. There is clear differentiation between soft tissue and bone

9. There is good contrast between air-filled structures and the surrounding tissue/structures

10. The trabecular patterns of the bones are visualised sufficiently

11. The trachea is visualised clearly

12. The vascular patterns of the lungs are not clearly visualised from the hila to the periphery

13. The left costophrenic angle is defined sharply

14. The right costrophrenic angle is defined sharply

15. The left cardiophrenic angle is defined sharply

16. The right cardiophrenic angle is defined sharply

17. The borders of the heart and aorta are visualised sharply

18. The proximal bronchi are visualised sharply

19. The trachea is not defined sharply

20. The left hemi-diaphragm is visualised sharply

21. The right hemi-diaphragm is visualised sharply

22. There is a significant amount of noise in this image.