Article

Variability and Subjectivity in the Grading Process for Evaluating the Performance of Latent Fingermark Detection Techniques

P. Fritz [1]
A. A. Frick [1]
W. van Bronswijk [2]
S. W. Lewis [1]
A. Beaudoin [3]
S. Bleay [4]
C. Lennard [5]

Abstract: When assessing latent fingermark development methods, forensic researchers commonly evaluate treated samples using a grading scale. However, the subjective nature of these evaluation methods leaves the results of such investigations open to criticism for potential grader bias. Assessment of fingermark development quality is ultimately dependent on an individual's background and experience.

A pilot study was conducted as a preliminary stage of a large-scale international collaboration. A set of 80 fingermark samples was developed with 1,2-indanedione-zinc chloride. Grades for photographic images of the developed fingermarks were assigned independently by 11 fingermark researchers. Sixty-seven percent of the scores given to each individual sample were the same as the median grade, and

[1] Nanochemistry Research Institute and Department of Chemistry, Curtin University, Perth, Western Australia, Australia
[2] Department of Chemistry, Curtin University, Perth, Western Australia, Australia
[3] Service de la criminalistique, Sûreté du Québec, Québec, Canada
[4] Centre for Applied Science and Technology, Home Office Science, Sandridge, St Albans, Hertfordshire, United Kingdom
[5] School of Science and Health, University of Western Sydney, Richmond, NSW, Australia

Received October 17, 2014; accepted May 7, 2015

99% of the scores were within 1 grade. The researchers were also assessed on their consistency by including 20 duplicate images to be graded. Seventy-eight percent of the grades given were identical to their original scores.

These results indicate that a small group of independent fingermark graders is sufficient to produce reliable and consistent data in projects requiring the assessment of fingermark quality.

Introduction

A significant aspect of fingermark research is the investigation of new latent fingermark development methods, or the improvement of existing methods. The evaluation of the effectiveness and applicability of these novel techniques is often complicated because of a wide range of variables. Factors such as personnel, laboratory equipment (including reagents and substrates), and prevailing climate often differ greatly between institutions, in addition to the exceedingly variable nature of latent fingermark deposits themselves [1]. In addition to these issues, the evaluation of the performance of a development method is a relatively unsophisticated process that may introduce further uncertainties. Recent guidelines from the International Fingerprint Research Group (IFRG) highlight the requirement for standardized research and validation methods [1]. More specifically, the use of grading scales is highly recommended to assess the quality of developed fingermarks for all comparison and validation experiments.

There are several fingermark grading schemes currently in use by researchers and industry professionals alike, tailored to suit specific investigations as recommended by the IFRG [1–9]. Commonly used assessment methods generally categorize fingermark development along a scale ranging from "good" to "poor" ridge detail. The main issue with such ranking systems is their subjectivity; assessment of fingermark quality relies on human observation and, therefore, is subject to bias stemming from an individual's own experience and personal notions as to what constitutes "good" fingermark development. It is known that these differences in personal opinion cannot be completely controlled by assessment protocols, with similar issues having been noted in the fingermark identification process [10, 11]. Related studies have indicated that experience and training play a pivotal role when an assessor encounters a fingermark that is incomplete or in some way distorted [10].

This study forms part of a larger, ongoing investigation into the effects of time since deposition and donor traits, such as age and gender, on fingermark development. A preliminary study found statistically significant variation, correlated to amino acid content, because of the age of the developed sample, the age of the donor, and the washing of hands [12]. Further, ongoing investigations using a number of amino acid- and lipid-sensitive reagents to develop fingermarks from a large donor population will provide complementary data, with results to be published in due course.

The consistent evaluation of developed latent fingermarks is necessary to this investigation to obtain valid statistical information, and much concern lies with the uncertainty associated with the subjectivity of fingermark evaluations. Because statistical means will be employed to determine what, if any, correlations exist between donor traits and the quality of developed fingermarks, the reliability and robustness of the grading method has to be established.

The grading of treated fingermarks is usually done by a single individual, for the purposes of method comparison or evaluation. As long as the grading of fingermark quality is consistent, any conclusions drawn are more relevant than the actual assigned grades. It is less common for the same fingermark to be given a score by two or more individuals, except in the case of collaborations. In this scenario, consensus between the graders must also be reached to eliminate bias or error. In such instances, a large group of assessors would be ideal to provide an average to compensate for subjectivity; however, this is rarely feasible because of individual graders' workloads, geographical locations, and other commitments. A smaller, localized group of graders is more practical, provided that this approach can be assumed to accurately represent the performance of a larger group.

The aim of this present study is to investigate the variation in fingermark assessment by several researchers from different research institutions, geographical locations, and varying familiarity with latent fingermarks. These graders were required to assess a large number of developed fingermark images to discern intergrader variation, as well as intragrader consistency. From the obtained data, it can be ascertained whether a small, localized group of fingermark graders can produce reliable data, or whether a larger group is necessary, despite practical constraints.

Materials and Method

Chemicals

1,2-Indanedione (Optimum Technology, Australia), anhydrous zinc chloride (Sigma-Aldrich, U.S.A.), ethyl acetate (Univar Analytical, Australia), glacial acetic acid (Lab-Scan, Thailand), absolute ethanol (CSR Chemicals, Australia), and HFE-7100 (1-methoxynonafluorobutane; 3M Novec, Australia) were all used as received and were of analytical reagent grade.

Preparation of Reagent Solutions

1,2-Indanedione-zinc chloride (Ind-Zn) reagent was prepared as recommended by the National Center of Forensic Studies (NCFS) [13]. The working solution consisted of 2 mL zinc chloride stock solution (8 g zinc chloride dissolved in 200 mL absolute ethanol) and 50 mL stock solution (4 g 1,2-indanedione dissolved in 450 mL ethyl acetate with 50 mL glacial acetic acid) added to 450 mL HFE-7100 solvent.

Collection of Latent Fingermarks

Depletion series consisting of 10 latent fingermark impressions (5 three-finger impressions from each hand) were collected on white copy paper templates (Fuji Xerox Professional, 80 g/m^2) from 4 donors. Donors had not washed their hands, consumed food, or handled chemicals for at least 30 minutes before providing samples. The templates were divided in half, creating 20 individual squares (Figure 1), producing a total of 80 fingermark samples, and treated with Ind-Zn within 48 hours following deposition.

Development of Latent Fingermarks Using Ind-Zn

Ind-Zn treatment was carried out as described by the NCFS [13]. Samples were developed by dipping briefly in the working solution and were allowed to air-dry before being heat-treated for 10 seconds with an Elna laundry press set at 160 °C.

Photography of Samples

Samples were photographed using a Nikon D300 camera mounted on a Firenze Mini Repro tripod and connected to a computer using Nikon Camera Control Pro Version 2.0.0. The samples were photographed in the luminescence mode only, where illumination was achieved using a Rofin Polilight PL500 (Rofin, Australia), with an excitation wavelength of 505 nm (40 nm bandwidth) and an orange long-pass barrier filter on the

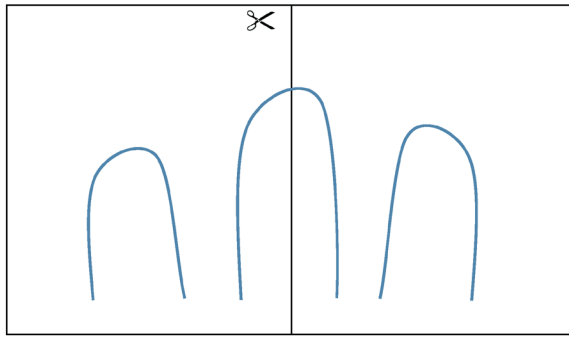


Figure 1

Schematic representation of finger placement for sample collection, showing 2 sample squares for each three-finger impression.

camera (Foster + Freeman Schott OG550). The shutter speed that was used was 1 second and the aperture was set to f/11 at a sensitivity of ISO 200. The focal length was 60 mm and the automatic white balance was used. Images were resized using FastStone Photo Resizer (v.3.1) for distribution purposes, but were not otherwise changed or enhanced.

Data Distribution and Assessment of Developed Latent Fingermarks

Sample images were assessed by 11 graders, who could be broadly classified into one of the following groups: experienced fingermark researchers (4), fingermark research students (3), and research students with no previous experience with fingermarks (4).

From the 80 sample images, 20 were randomly selected to be duplicated, producing a total of 100 images to be graded. The duplication of images was done without the graders' knowledge. To reduce the effects of exhaustion, stress, and so forth and to make the process less time-consuming, the samples were distributed to the graders in 5 batches, each containing 20 images. This was implemented by numbering the samples and then randomly assigning them to one of the 5 batches using a random number generator in Excel Professional Plus 2010 (Microsoft). The images were distributed to fingermark assessors via an online cloud program, Dropbox (v.1.4.8).

Samples were graded using a five-point system based on that used by the Home Office Police Scientific Development Branch, United Kingdom [6], modified by the addition of representative images of developed marks to aid in classification (Table 1) [14]. The results were recorded and evaluated using Excel Professional Plus 2010 (Microsoft).

Statistics

The intragrader agreement was assessed by comparing the grades given by each grader to the 20 duplicated samples. The Cohen's kappa test was used to determine the level of agreement for these paired grades [20, 21]. The intergrader consistency was assessed using the intraclass correlation coefficient (based on the 80 sample images) [22]. Statistical analyses were performed using the SPSS version 2.0 statistical software, and a p value<0.05 was taken to indicate a statistically significant association in all tests.

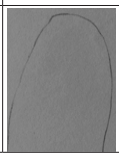
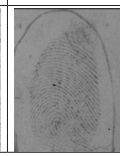
Grade	0	1	2	3	4
Friction Ridge Detail Development	No development	Signs of contact, but less than 1/3 of fingermark continuous ridges	1/3–2/3 of fingermark continuous ridges	More than 2/3 of fingermark continuous ridges, but not quite a 'perfect' fingermark	Full development; whole fingermark, continuous ridges
Contrast of Ridge Detail and Background	No contrast	Poor contrast	Moderate contrast	Good contrast	Very good contrast
Photographic Representation					

Table 1

The fingermark grading scale provided to the fingermark graders.

Results and Discussion

The aims of this investigation were (1) to examine any differences or inconsistencies in the performance of the 11 fingermark graders in evaluating the 100 fingermark samples, and (2) to determine whether the disparate backgrounds or expertise of the graders had any impact on the assessments. It is important to note that the ability to examine treated fingermarks for identification purposes was not investigated, but rather the assessment of fingermark development quality. An absolute scale, adapted

from Bandey [6], is used routinely at Curtin University to assess reagent performance. This requires an individual to take into account contrast, clarity, and ridge continuity. For the purpose of this study, detailed descriptions of each grade, with examples, were provided to the graders to reduce bias and encourage more consistent results.

Intragrader Variation

It is imperative that an individual grader can perform consistently. If individual graders are unable to dependably assign a grade to identical fingermark images, then no meaningful conclusions can be drawn from any study involving subjective evaluation methods. To investigate this effect, the grades given to the replicate images were examined for each fingermark grader. It was found that 172, or 78.2%, of the replicate grades were identical to their original scores (Figure 2). Forty-eight (21.8%) of the replicates showed a difference of 1 grade between the samples, and none of the replicate samples showed a difference of 2 or more grades. From this it can be concluded that fingermark grades can be viewed as reliable and reproducible data. The relative experience of each grader with latent fingermarks appeared to have no significant impact on each individual's ability to grade fingermarks consistently.

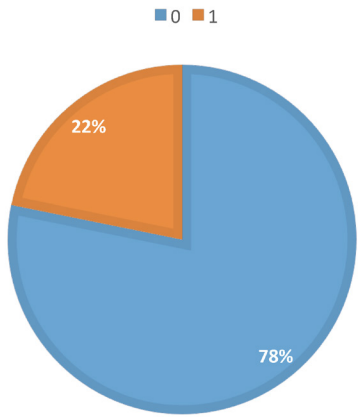
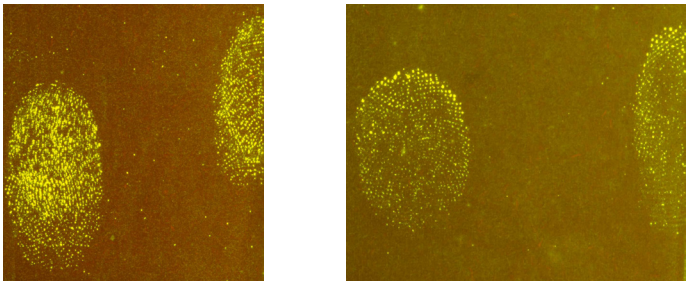


Figure 2
Differences between 2 grades assigned to replicated images.

Inconsistent grades for half of the replicate images were assigned by at least 2 graders, indicating that these may be samples that are borderline (in between 2 categories) or otherwise difficult to categorize. Overall, it was found that very low and very high fingermark grades were the most easily reassigned. The majority of disagreements were found to occur with images that were assigned a 2 or 3 rating in at least one instance. These samples, for example, may have exhibited good contrast or detail but also had smudged portions or did not show continuous ridges (Figure 3). The more ambiguous nature of these samples in terms of their quality may cause graders to rely more on their own individual idea of a fingermark grade than adhering to the grading scale provided.



(a) *(b)*

Figure 3

Examples of replicated fingermark images graded inconsistently (a) and consistently (b) by the same individual.

Cohen's kappa tests were performed on the grades assigned by each grader to the duplicated image pairs to determine the level of agreement (Table 2). The average Cohen's kappa value was found to be 0.684 and in all cases $p<0.0005$, which can be classified as showing good agreement between the replicate grades [20, 21, 23, 24].

Grader	1	2	3	4	5	6	7	8	9	10	11
Cohen's Kappa Value	0.744	0.839	0.672	0.704	0.732	0.732	0.538	0.794	0.659	0.565	0.603

Table 2

Statistical values gained from the Cohen's kappa test, where the original scores given by each grader were compared to the duplicated ones (n=20 and p<0.0005).

Intergrader Variation

Having established that the 11 graders performed relatively consistently within the circumstances of this investigation, comparisons between grader performances were conducted to examine any general trends. Figure 4 gives an overview of the absolute distribution of grades assigned for the 100 fingermark images by the 11 graders. For 42 images, the grades given differed by 1 amongst the graders, and for a further 53 samples, there was a difference of 2 grades. For 3 images, there was unanimous agreement on a grade between all graders. The remaining 2 images were graded the most inconsistently between the 11 graders, with a total difference of 3 grades assigned to these images. Agreement between all graders was most frequent when assessing fingermarks that exhibited very strong or very weak development (Figure 5). At first glance, the frequent disagreements between graders appeared to indicate that the grading scale does not appear to be a reliable indication of fingermark development quality. However, this data only accounts for the absolute distribution of grades, rather than any consensus reached between the graders.

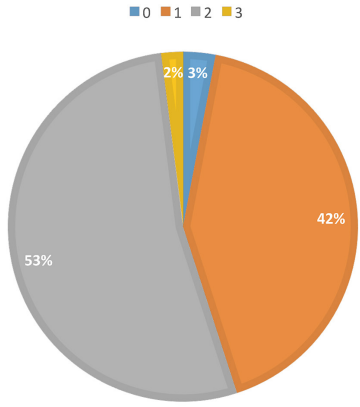


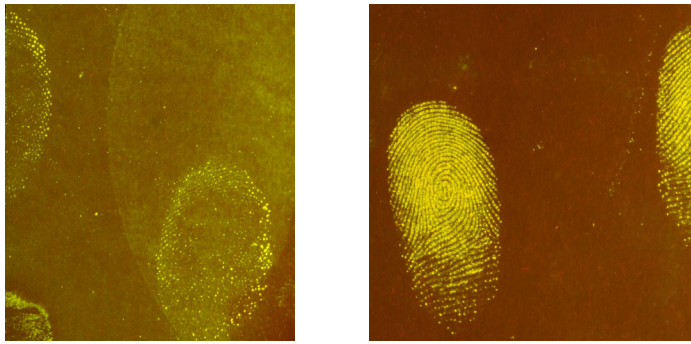
Figure 4

Absolute distribution of grades assigned to the 100 treated fingermark samples.

When grades were compared to the median grade for each image, there was better agreement between all 11 graders (Figure 6). Total agreement between assessors' grade and the median occurred in 66.9% of the 1100 grades assigned in total, and a difference of 1 between the grade and the median occurred in 32.0% of cases. Therefore, 98.9% of all grades provided differed by only 1 less from the median score assigned to each sample. The remaining 1.1% of grades differed by 2 from the median. No instance occurred where there was a disagreement of 3 or 4. In light of these statistics, the grading scale appears to be much more robust.

Although there was general agreement between all graders and the median, there were significant differences in how individual graders performed regarding how frequently they agreed with the median. In some cases, graders agreed with over 85% of the median grade for each fingermark image, whereas other graders only agreed with 40 to 50%. Obviously, this has a large impact on the above results because considerably more samples would be in agreement, as a percentage, without these graders. This is also reflected in the mean of each individual sample rather than the median. Because the median reflects the grade most commonly given for each exhibit, it is less affected by individual outliers. The mean, however, does display the skewing of these samples because of differences in grading. On the whole, there was strong agreement between the median and the mean values, where skewing indicated that there was no trend of either over- or underscoring. It is interesting to note that, although these assessors may not agree with the median grades as strongly as other examiners, their grading consistency was no different.

Furthermore, the performance of each grader appeared to have no correlation to the institution or geographical location; however, graders with greater experience in latent fingermark research tended to give grades that disagreed with the median more often than the more inexperienced graders. It may be that, having greater experience, these graders have formed their own standards for fingermark quality, and that these opinions unconsciously influenced their performance even while using the provided grading scale. Conversely, the less experienced graders, some of whom were completely unfamiliar with fingermarks, were more likely to rely almost solely on the grading scale as a guide [10, 11]. However, given that a pool of only 11 graders was used, it is difficult to draw any definitive conclusions in these regards.



(a) (b)

Figure 5

Examples of fingermark images unanimously assigned a grade of 1 (a) and 4 (b).

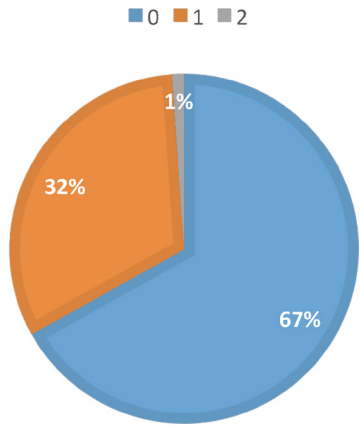


Figure 6

Differences between grades given to fingermark images and the median grade for each sample.

To ascertain whether a smaller, local subgroup of fingermark graders gave a similar response to the grades given overall, the graders were split into 3 groups. Group 1 consisted of the experienced fingermark researchers, group 2 consisted of the experienced fingermark research students, and group 3 consisted of inexperienced students. The medians of all the grades given were found to be 3, 2, and 2, respectively. The means, which can show the direction of the skew of results, were shown to be 2.49, 2.34, and 2.42, respectively. The similarity of these means very strongly indicates that there is no significant difference between the 3 groups when looking at the overall spread of grades given.

The intergrader variation was also evaluated using intraclass correlation coefficients, which establish the consistency between 2 or more measurements [17, 22]. The intraclass correlation coefficient was calculated as a two-way mixed model with absolute agreement, giving a result of 0.768, indicating strong agreement between all 11 graders. The lower and upper confidence intervals show that 95% of the time, the assigned grade will give a correlation between 0.706 and 0.826. Overall, use of this grading scale appears to be a method that offers consistent and robust results for the assessment of fingermark samples and is therefore seen as a feasible approach for use in a pending large-scale donor study.

Conclusion

The present study investigated the robustness and consistency of a fingermark grading method. The purpose was to evaluate its suitability in an ongoing large-scale donor project and to assess whether this grading scale is acceptable for fingermark development comparisons in general.

It was found that 67% of the assessed fingermark images were graded consistent with the median score, and 99% within 1 grade. Additionally, all fingermark graders were demonstrated to assign grades consistently for 78% of duplicated images. The margin of error for the remaining duplicates was 1 grade. The Cohen's kappa test and intraclass correlation coefficients show that no significant difference was found between the median grades of samples as a function of the intra- or intergrader variation, respectively.

Overall, the grading scale was deemed an appropriate and consistent technique to acquire absolute values for developed latent fingermark samples, which can be used on its own or in

combination with statistical methods to procure further knowledge. Furthermore, it was found that in this study a smaller subgroup of graders was representative of the larger group in their assessment. This is of importance in future work to avoid practical constraints in international collaborations.

Acknowledgment

The authors would like to thank all fingermark donors and fingermark graders at Curtin University, Australia. They also wish to thank Kieran Lim (Deakin University, Australia) for helpful discussions in the development of this investigation and Richard Parsons (Curtin University, Australia) for statistical advice. A.A. Frick and P. Fritz are supported by Australian Postgraduate Awards. This study has been approved by the Curtin University Human Research Ethics Committee (Approval Numbers SMEC-94-11 and SMEC-08-13).

For further information, please contact:

Simon W. Lewis
Department of Chemistry
Curtin University, GPO Box U1987
Perth, Western Australia 6845
S.Lewis@curtin.edu.au

References

1. International Fingerprint Research Group (IFRG). Guidelines for the Assessment of Fingermark Detection Techniques. *J. For. Ident*. **2014**, *64* (2), 174–200.
2. Almog, J.; Springer, E.; Wiesner, S.; Frank, A.; Khodzhaev, O.; Lidor, R.; Bahar, E.; Varkony, H.; Dayan, S.; Rozen, S. Latent Fingerprint Visualization by 1,2-Indanedione and Related Compounds: Preliminary Results. *J. For. Sci*. **1999**, *44* (1), 114–118.
3. Jasuja, O. P.; Toofany, M. A.; Singh, G.; Sodhi, G. S. Dynamics of Latent Fingerprints: The Effect of Physical Factors on Quality of Ninhydrin Developed Prints–A Preliminary Study. *Sci. Just*. **2009**, *49* (1), 8–11.
4. Wood, M. A.; James, T. ORO. The Physical Developer Replacement? *Sci. Just*. **2009**, *49* (4), 272–276.
5. McLaren, C.; Lennard, C.; Stoilovic, M. Methylamine Pretreatment of Dry Latent Fingermarks on Polyethylene for Enhanced Detection by Cyanoacrylate Fuming. *J. For. Ident*. **2010**, *60* (2), 199–222.

6. Bandey, H. L. The Powders Process, Study 1: Evaluation of Fingerprint Brushes for Use with Aluminium Powder. *PSDB Fingerprint Dev. and Imag. News.: Spec. Ed.* **2004**, *54* (4), 1–12.
7. Becue, A.; Moret, S; Champod, C.; Margot, P. Use of Quantum Dots in Aqueous Solution to Detect Blood Fingermarks on Non-Porous Surfaces. *For. Sci. Int.* **2009**, *191* (1–3), 36–41.
8. Sears, V. G.; Bleay, S. M.; Bandey, H. L.; Bowman, V. J. A Methodology for Finger Mark Research. *Sci. Just.* **2012**, *52* (3), 145–160.
9. Kent, T. Standardizing Protocols for Fingerprint Reagent Testing. *J. For. Ident.* **2010**, *60* (3), 371–379.
10. Schiffer, B.; Champod, C. The Potential (Negative) Influence of Observational Biases at the Analysis Stage of Fingermark Individualisation. *For. Sci. Int.* **2007**, *167* (2–3), 116–120.
11. Dror, I. E.; Champod, C.; Langenburg, G.; Charlton, D.; Hunt, H.; Rosenthal, R. Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a 'Target' Comparison. *For. Sci. Int.* **2011**, *208* (1–3), 10–17.
12. Fritz, P.; van Bronswijk, W.; Patton, E.; Lewis, S. W. Variability in Visualization of Latent Fingermarks Developed with 1,2-Indanedione-Zinc Chloride. *J. For. Ident.* **2013**, *63* (6), 698–713.
13. Stoilovic, M.; Lennard, C. *NCFS Workshop Manual: Fingermark Detection & Enhancement*, 6th ed.; National Centre for Forensic Studies: Canberra, 2012.
14. Fritz, P.; van Bronswijk, W.; Dorakumbura, B., Hackshaw, B.; Lewis, S. W. Evaluation of a Solvent-Free P-Dimethylaminobenzaldehyde Method for Fingermark Visualization with a Low-Cost Light Source Suitable for Remote Locations. *J. For. Ident.* **2015**, *65* (1), 67–90.
15. Zimmerman, D. W. Comparative Power of Student *T* Test and Mann-Whitney *U* Test for Unequal Sample Sizes and Variances. *J. Exp. Ed.* **1987**, *55* (3), 171–174.
16. Koch, G. G. Intraclass Correlation Coefficient. In *Encyclopedia of Statistical Sciences*; Kotz, S., Johnson, N. L., Read, C. B., Eds.; John Wiley & Sons: New York, 1982.
17. Müller, R.; Büttner, P. A Critical Discussion of Intraclass Correlation Coefficients. *Stat. in Med.* **1994**, *13* (23–24), 2465–2476.
18. Miller, J. N.; Miller, J. C. *Statistics and Chemometrics for Analytical Chemistry*, 6th ed.; Pearson Education Limited: Harlow, England, 2010.
19. Hibbert, D. B.; Gooding, J. J. *Data Analysis for Chemistry: An Introductory Guide for Students and Laboratory Scientists*. Oxford University Press: New York, 2005.

20. Fleiss, J. L.; Levin, B.; Paik, M. C. *Statistical Methods for Rates and Proportions,* 3rd ed.; Wiley: Hoboken, New Jersey, 2013.
21. Altman, D. G. *Practical Statistics for Medical Research*. Chapman & Hall: London, England, 1990.
22. McGraw, K. O.; Wong, S. P. Forming Inferences about Some Intraclass Correlation Coefficients. *Psych. Meth*. **1996**, *1* (1), 30–46.
23. Al-Khawari, H.; Athyal, R. P.; Al-Saeed, O.; Sada, P. N.; Al-Muthairi, S.; Al-Awadhi, A. Inter- and Intraobserver Variation Between Radiologists in the Detection of Abnormal Parenchymal Lung Changes on High-Resolution Computed Tomography. *Annals of Saudi Med*. **2010**, *30* (2), 129–133.
24. Viera, A. J.; Garrett, J. M. Understanding Interobserver Agreement: The Kappa Statistic. *Fam. Med*. **2005**, *37* (5), 360–363.

Appendix

Raw data of each grade given to the original (n=80) and replicated (n=20) sample images by each grader.

Sample Number	Grader 1	Grader 2	Grader 3	Grader 4	Grader 5	Grader 6	Grader 7	Grader 8	Grader 9	Grader 10	Grader 11
1	1	1	2	3	3	2	1	2	2	1	2
2	3	4	3	4	4	4	4	4	4	4	2
3	3	4	4	4	4	3	4	4	4	3	3
4	2	2	3	3	2	3	2	2	2	2	2
5	3	4	4	4	4	4	3	4	4	3	4
6	3	4	4	4	4	4	4	4	4	4	3
7	2	1	2	3	2	2	2	2	1	2	2
8	1	1	2	2	1	1	2	1	1	1	1
9	1	1	2	2	1	1	1	1	1	1	1
10	4	4	4	4	4	4	4	4	4	3	4
11	2	2	1	2	3	2	3	2	2	2	2
12	3	3	3	3	3	3	3	3	3	3	4
13	4	4	4	4	4	3	4	4	4	4	3
14	4	4	3	4	3	3	3	3	4	3	3
15	3	4	4	4	4	4	4	4	4	3	4
16	1	1	1	2	1	1	1	1	1	1	0
17	4	4	3	4	4	4	4	4	4	4	4
18	3	2	3	3	3	3	3	2	3	2	3
19	3	4	4	4	4	4	3	4	4	3	4
20	4	4	4	4	4	4	4	3	4	4	4
21	4	4	4	4	4	4	4	3	4	3	2
22	4	3	3	4	3	3	4	3	4	2	4
23	4	3	3	3	3	3	3	2	3	3	3
24	1	1	1	3	2	2	2	2	2	2	2
25	1	1	2	3	2	2	1	2	2	1	1
26	1	1	1	3	2	2	1	2	1	1	2
27	1	1	2	3	2	3	1	2	2	1	1
28	1	1	1	3	1	2	1	1	2	1	1
29	4	3	4	4	3	2	4	3	3	4	3
30	4	3	3	3	3	4	4	3	3	3	4
31	4	2	3	3	3	2	3	3	3	3	2
32	2	2	1	3	2	3	2	2	2	2	2
33	3	2	3	3	3	2	2	2	2	2	1
34	2	1	1	3	2	3	2	2	2	1	2
35	1	1	1	3	1	2	1	2	2	1	1
36	1	1	2	3	2	2	1	2	1	2	2
37	1	1	1	2	1	1	1	1	1	1	1
38	4	4	4	4	4	4	4	4	4	4	4
39	1	1	1	2	1	1	1	1	1	1	1
40	2	4	3	4	4	3	3	4	4	2	3
41	2	3	3	3	4	3	3	3	4	4	3
42	1	1	1	2	1	2	1	1	1	1	1
43	1	1	2	3	2	2	1	2	1	1	2
44	2	3	3	3	3	3	3	3	4	3	4
45	1	1	1	2	1	2	1	1	1	1	2
46	1	1	1	3	2	2	2	2	1	2	2
47	4	4	2	4	3	3	4	4	4	3	4

48	4	3	3	3	3	3	4	3	3	3	4
49	4	3	2	3	3	3	3	2	3	3	3
50	1	1	2	3	2	2	1	2	2	1	2
51	4	2	3	3	2	3	3	3	3	2	3
52	4	2	3	3	3	2	3	3	3	3	2
53	3	2	2	3	2	2	2	2	2	2	2
54	3	4	3	4	4	4	4	4	4	2	3
55	1	4	4	4	3	3	3	4	3	1	2
56	2	4	4	4	3	4	2	4	3	3	3
57	2	4	4	4	3	4	3	4	3	2	3
58	3	2	3	3	3	3	2	2	2	3	2
59	1	1	2	2	2	1	2	2	2	2	2
60	2	1	3	3	2	3	1	2	2	2	3
61	1	1	2	3	2	2	1	2	2	1	2
62	3	3	4	4	3	4	2	3	3	2	3
63	2	1	2	3	2	2	1	2	2	2	2
64	2	1	3	3	2	3	2	2	2	2	2
65	1	1	2	3	2	2	2	2	2	1	2
66	3	2	3	3	2	2	2	2	2	2	3
67	2	2	3	3	3	2	2	3	2	2	3
68	1	1	2	3	2	2	2	2	2	2	2
69	1	1	2	3	2	2	2	2	2	1	1
70	1	1	2	3	2	3	1	2	1	1	2
71	1	1	2	3	2	3	1	2	2	1	1
72	1	1	2	3	2	1	1	2	1	1	1
73	4	4	4	4	4	4	4	4	4	4	4
74	4	3	4	4	4	3	4	3	4	3	3
75	1	1	1	2	1	1	1	1	1	1	1
76	1	1	1	1	1	1	1	1	1	1	0
77	1	1	0	1	1	1	1	1	1	1	0
78	1	1	1	2	1	1	1	1	1	1	1
79	1	1	2	3	2	2	1	2	1	1	1
80	1	1	3	3	2	3	2	2	1	1	3
Replicate 1	1	1	2	3	2	2	1	2	1	2	3
Replicate 9	1	1	2	2	1	1	1	1	1	1	1
Replicate 12	4	2	3	3	3	3	4	3	3	3	4
Replicate 16	1	1	0	1	1	1	1	1	1	1	1
Replicate 20	4	4	4	4	4	4	4	4	4	4	4
Replicate 23	4	3	3	4	3	3	4	3	3	3	3
Replicate 26	1	1	2	2	2	3	1	2	1	1	2
Replicate 30	4	3	3	4	3	3	3	3	4	3	4
Replicate 35	1	1	1	3	2	2	2	2	1	2	2
Replicate 37	1	1	1	2	1	1	1	1	1	1	1
Replicate 38	4	4	4	4	4	4	4	4	4	3	4
Replicate 43	1	1	2	3	2	2	1	2	2	1	1
Replicate 47	4	3	3	4	4	4	4	3	4	4	4
Replicate 50	1	1	2	3	2	2	1	2	2	2	2
Replicate 55	2	4	3	4	3	3	4	4	3	1	3
Replicate 60	1	1	3	3	2	2	2	2	2	2	2
Replicate 66	3	2	3	3	3	2	2	2	3	2	3
Replicate 73	4	4	4	4	4	4	4	4	4	3	4
Replicate 75	1	1	2	2	1	1	1	1	1	1	1
Replicate 78	1	1	1	2	1	1	1	1	1	1	1