



# UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES

POSTGRADUATE DENTAL COLLEGE  
SOUTHERN REGION OFFICE  
2787 WINFIELD SCOTT ROAD, SUITE 220  
JBSA FORT SAM HOUSTON, TEXAS 78234-7510  
<https://www.usuhs.edu/pdc>



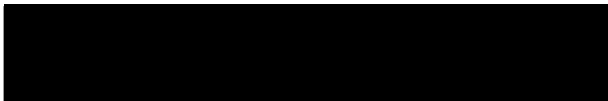
## THESIS APPROVAL PAGE FOR MASTER OF SCIENCE IN ORAL BIOLOGY

Title of Thesis: "The Reliability of Interpreting a CBCT Image"

Name of Candidate: Joshua Williams  
Master of Science Degree  
May 23, 2018

THESIS/MANUSCRIPT APPROVED:

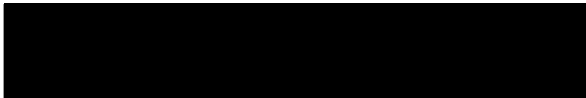
DATE:



Joseph M. Dutner, MAJ, DC  
DEPARTMENT OF ENDODONTICS, FORT GORDON  
Committee Chairperson

11 June 2018

08 June 2018



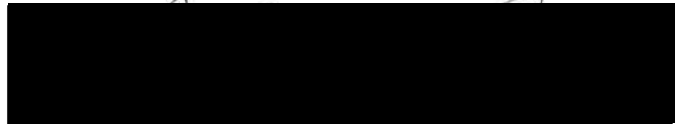
Stephanie J. Sidow, COL, DC  
DEPARTMENT OF ENDODONTICS, FORT GORDON  
Committee Member

The author hereby certifies that the use of any copyrighted material in the thesis/  
dissertation manuscript entitled:

"The Reliability of Interpreting a CBCT Image"

is appropriately acknowledged and, beyond brief excerpts, is with the permission of  
the copyright owner.

Joshua Williams  
Endodontic Residency  
Uniformed Services University  
23 May 2018



# The Reliability of Interpreting a CBCT Image

*Joshua Williams, DMD, Stephanie J. Sidow, DDS, Joseph M. Dutner, DDS,*

## Abstract

**Introduction:** Cone-beam computed tomography can be an important aid in the diagnostic process. The purpose of this study is to determine if CBCT allows the dental provider to interpret the images in a reliable manner.

**Methods:** CBCT images of 120 permanent teeth were interpreted by 2 board certified endodontists, 7 endodontic residents, and 2 general dentists. The observers determined whether or not a specific tooth had normal periapical tissue, widened periodontal ligament (PDL) space, or a periapical radiolucency (PARL). Scans were viewed using i-Dixel (Kyoto Japan) software multi-planar reconstructed volumes. The images were examined on a Dell CPU with the same 27 inch screen monitor for all evaluators. To assess the agreement between providers' ratings of 120 images, two measures were calculated. First, an agreement index (Pi) was calculated for each image, and an average of these scores was obtained. Since this average includes agreement due to chance, Fleiss' Kappa statistic (a chance-adjusted index of agreement for multi-rater categorization of nominal variables) was also calculated.

**Results:** An agreement index was calculated for each image and an average of these scores was obtained. Due to chance a Fleiss kappa statistic was used to measure agreement among multiple observers. 51.4% indicates intermediate/good agreement. The average of the individual image agreement index scores was 68.8%. Out of 120

total images only 27.5% had perfect agreement. 80% of images had at least 7 out of 11 providers agreeing.

**Conclusion:** The results demonstrate that the interpretation of a CBCT volume is subjective. However, inter-observer agreement with CBCT images is higher than two-dimensional film or digital periapical radiography.

## Introduction

There are many tools that general practitioners and endodontists use when forming a diagnosis that leads to a corresponding treatment. One that has been in use for over a century is radiography. This technology allows the provider to see through soft tissue and view contrasting hard tissues. This is important because it allows the provider to view the results of disease where bone or odontogenic structures are resorbed or destroyed by infection (1). However, one of the key limitations with the images created by radiography is they convert a three dimensional structure into a two dimensional view. This means there is a certain level of overlap between various levels of hard tissue that can create some confusion when interpreting these images. Consequently in order for the periapical radiolucency to appear on a digital radiograph there needs to be at least 23-35% cortical bone loss or at least 7% mineralized bone loss, due to the creation of overlap in the 2D image. (2,3) Greater sensitivity in the detection of bone loss would therefore be advantageous.

Cone-beam computed tomography (CBCT) offers a unique adjunct to a provider's armamentarium. The 3D images generated for a defined FOV (Field of View) can be seen from multiple angles and different vantage points. In CBCT, a cone-shaped X-ray beam is passed through the tissues of the FOV onto a flat panel detector plate while rotating 360° in unison around the patient. This captures a cylindrical volume of data made of square shaped voxels (a voxel being the minimum unit of 3D data). Depending on the FOV or the size of the voxel, a CBCT image can be made of up to a million voxels. Software then reconstructs this dataset into viewable images that can be examined in three different spatial planes. Recently, the AAE recommends specific scenarios in endodontics where CBCT should be the modality of choice instead of traditional or digital radiography (14).

Two previously published studies examined the reliability of the interpretation of conventional and digital periapical radiographs. These studies aimed at measuring the reliability and consistency of interpretation of traditional radiographs between providers. The examiners only agreed on less than half of the cases (12). More recently, in 2011, another study demonstrated that interpreting digital radiography is also subjective. Between 6 observers only 25% agreement was observed (13). In 2017, another paper was published that compared CBCT observer variability between endodontic faculty, residents, and dental students. They showed that a clinician's experience correlated with their ability to accurately diagnose periapical pathosis using CBCT. However, they only used 22 images and 9 subjects, making it difficult to draw any conclusions on true reliability due to small sample size (15). The purpose of this study is to determine if CBCT allows the dental provider to interpret the images in a reliable and consistent manner. There have been no studies with similar methodology as previous reliability studies using conventional and digital radiography to CBCT images.

### **Materials and Methods:**

This study was reviewed for compliance with the applicable human subject protection regulations by the Regional Health Command-Atlantic Institutional Review Board. The study was granted approval because it is minimal risk and there are no outstanding human research protections issues.

A total of 120 images in which at least one of the following conditions listed below were selected from patient archives: normal periapical tissue, widened periodontal ligament which will be determined as a CBCT PAI score of 1 or 2, or presence of periapical radiolucency.

In 2008, a new periapical index (PAI) score system was established that measures the quantitative bone alterations in mineral structures. This PAI score ranges from 0-5 and includes two variables E (expansion of cortical bone) and D (destruction of cortical bone). This index is as follows: 0, intact periapical bone structures; 1, diameter of periapical radiolucency >0.5-1mm; 2, diameter of periapical radiolucency >1-2 mm; 3, diameter of periapical radiolucency >2-4 mm; 4, diameter of periapical radiolucency >4-8 mm; 5, diameter of periapical radiolucency >8 mm (5). The CBCT PAI score was used in this study to identify a widened PDL space using a score of 1 or 2.

A total of 11 observers including seven endodontic residents, two endodontists, and two general practitioners evaluated these images to determine which conditions existed. One of the authors selected from archived patient records 120 small FOV 40x40mm Hi-Res CBCT scans. J. Morita 3D Accuitomo was used to capture all the images viewed. These scan consisted of 40 images to exhibit normal periapical bone, periodontal ligament, and lamina dura, 40 with defined periapical radiolucency, and 40 with a widened PDL space defined as 0.5-2mm based upon the CBCT PAI score by Estrela. Images were selected so that only one condition would be present.

Scans were viewed using i-Dixel (Kyoto Japan) software multi-planar reconstructed volumes. All images were examined on a Dell CPU with the same 27 inch screen monitor for all evaluators. Each scan was viewed by the evaluators as multi-planar reconstructed sections in the axial, sagittal, and coronal planes. The evaluators were permitted to scan through all 3 planes in the entire scan, toggle back and forth between any views as needed, and ask questions about the operation of the software throughout the screening. Precautions were taken to mask personal identifiable

information (PII) throughout the observation period. As the CBCT images were loaded on the i-Dixel software PII disappeared because each image was saved under a new name in a separate folder not containing any PII.

Volunteers were recruited to participate in the study from the Tingay Dental Clinic Endodontic residency program and from Snyder Dental Clinic located at Fort Gordon, GA. Each candidate evaluator was informed about the study, and given the opportunity to enroll, indicated by signing an informed consent. Participants were free to drop out at any time of their own volition. Participation or withdrawal would have no effect on the evaluation of the resident within the program. Each participant received a standardization tutorial both written and verbal regardless of previous experience. The tutorial gave an overview of the i-Dixel software and how many of the features functioned.

Each observer was given a spreadsheet with the image numbers listed and a checkbox to fill out containing the above parameters. The observers were not permitted to take the test at the same computer monitor at the same time to avoid compromised results. Each subject was instructed not to discuss findings and results with anyone during or after the test.

### **Statistical Analysis**

SAS 9.4 was used for all statistical analyses, unless otherwise noted. To assess the agreement between providers' ratings of 120 images, two measures were calculated. First, an agreement index ( $P_i$ ) was calculated for each image, and an average of these scores was obtained. Since this average includes agreement due to



chance, Fleiss' Kappa statistic (a chance-adjusted index of agreement for multi-rater categorization of nominal variables) was also calculated.

## **Results**

The average of the individual image agreement index scores (weighted for frequency) was 68.8%. However, since this number includes agreement due to chance, Fleiss' kappa was also calculated, and the value of 51.4% indicates intermediate/good agreement.

Out of 120 images: 33 images (27.5%) had perfect agreement. 18 images (15.0%) had 10/11 providers agreeing on the scoring. 18 images (15.0%) had 9/11 providers agreeing on the most common score. 13 images (10.8%) had 8/11 providers agreeing on the most common score. 14 images (11.7%) had 7/11 providers agreeing on the most common score. 80% of images had at least 7 providers agreeing.

Of the 33 images that had 100% agreement 21/33 were labeled as periapical lesions and 12/33 were labeled as normal apical tissue.

## **Discussion**

With the increasing use of cone-beam computed tomography in the field of endodontics, identifying and differentiating the anatomy and pathology becomes imperative. However, the interpretation of 3D CBCT images has some differences from the interpretation of traditional radiographs. As the prevalence of CBCT grows within endodontics the validity of interpretations of images becomes an important consideration. Whilst CBCT imaging has proven to be more sensitive in identifying various lesions (4,6,7), the reliability with which images can be interpreted at current

levels of clinician training shows some improvement when compared to film and digital periapical radiographs. When previous studies evaluated conventional film periapical radiographs they showed that all 6 of their examiners agreed only 47% of the time (12). Years later when a similar study was repeated using digital periapical radiographs they showed that all 6 of their examiners agreed less than 25% of the time (13). In our study we had 11 observers instead of 6 and all 11 examiners only agreed 27.5% of the time however 80% of time 7/11 of the observers agreed with each other.

The Fleiss kappa for the 11 observers was 51.4% indicating intermediate/good agreement. In the previous study using digital periapical radiographs they had an overall Fleiss kappa of 34% which indicated a fair range of agreement (13). The intermediate/good agreement obtained in this study demonstrates that the observers could agree with each other enough to provide moderate consistency and reliable results the majority of the time. Whereas, in the other study the fair range of agreement showed that the observers could not agree with each other enough to provide reliable results (13).

Although the CBCT shows improvement over digital periapical radiography in the diagnosis and interpretation of periapical tissues there is obviously still some level of subjectivity. Considerable effort was made in this study to homogenize the testing process. Each examiner evaluated the CBCT images from the same type monitor that was also the same size. Unlike previous studies that examined periapical radiographs or still 2D images this study allowed the observers to scroll through slices of the CBCT images and enhance each volume in the same manner a clinician would to aid in diagnosis of the periapex. Each observer was given the freedom to adjust contrast, utilize magnification, apply measurement tools, and rotate and scroll through different

angles of each image. Although this method undoubtedly created some level of variability between observers the results are much more clinically applicable.

**Conclusion:**

The results of this study demonstrate that the interpretation of a CBCT volume is subjective. However, inter-observer agreement with CBCT images is higher than two-dimensional film or digital periapical radiography. On average there is a 68.8% chance of agreement on any given CBCT image when labeling the periapex.

## References

1. American Dental Association Council on Scientific Affairs. "The use of dental radiographs: update and recommendations." *The Journal of the American Dental Association* 137.9 (2006): 1304-1312.
2. Hadley, David L., et al. "A comparison of five radiographic systems to D-speed film in the detection of artificial bone lesions." *Journal of endodontics* 34.9 (2008): 1111-1114.
3. Bender IB, Seltzer S. Roentgenographic and direct observation of experimental lesions in bone: part I. *J Am Dent Assoc* 1961;62:152-60.
4. Estrela, Carlos, et al. "Accuracy of cone beam computed tomography and panoramic and periapical radiography for detection of apical periodontitis." *Journal of endodontics* 34.3 (2008): 273-279.
5. Estrela, Carlos, et al. "A new periapical index based on cone beam computed tomography." *Journal of endodontics* 34.11 (2008): 1325-1331.
6. de Paula-Silva, Francisco Wanderley Garcia, et al. "Accuracy of periapical radiography and cone-beam computed tomography scans in diagnosing apical periodontitis using histopathological findings as a gold standard." *Journal of endodontics* 35.7 (2009): 1009-1012.
7. Dutra, Kamile Leonardi, et al. "Diagnostic accuracy of cone-beam computed tomography and conventional radiography on apical periodontitis: a systematic review and meta-analysis." *Journal of endodontics* 42.3 (2016): 356-364.

8. Brynolf, I. "Roentgenologic periapical diagnosis. IV. When is one roentgenogram not sufficient?." *Svensk tandlakare tidskrift. Swedish dental journal* 63.6 (1970): 415-423.
9. Venskutonis, Tadas, et al. "The importance of cone-beam computed tomography in the management of endodontic problems: a review of the literature." *Journal of endodontics* 40.12 (2014): 1895-1901.
10. Patel, S., et al. "Cone beam computed tomography in endodontics—a review." *International endodontic journal* 48.1 (2015): 3-15.
11. Goldman, Melvin, Arthur H. Pearson, and Nicholas Darzenta. "Reliability of radiographic interpretations." *Oral Surgery, Oral Medicine, Oral Pathology* 38.2 (1974): 287-293.
12. Goldman, Melvin, Arthur H. Pearson, and Nicholas Darzenta. "Endodontic success—who's reading the radiograph?." *Oral Surgery, Oral Medicine, Oral Pathology* 33.3 (1972): 432-437.
13. Tewary, Shalini, Joseph Luzzo, and Gary Hartwell. "Endodontic radiography: who is reading the digital radiograph?." *Journal of endodontics* 37.7 (2011): 919-921.
14. Fayad MI, Nair M, Levin MD, et al. AAE and AAOMR joint position statement: use of cone beam computed tomography in endodontics 2015 update. *Oral Surg Oral Med Oral Pathol and Oral Radiol Endod* 2015;120:508–12.
15. Parker, Jeffrey M., et al. "Cone-beam Computed Tomography Uses in Clinical Endodontics: Observer Variability in Detecting Periapical Lesions." *Journal of endodontics* 43.2 (2017): 184-187.

**Table 1. Summary of Image Scores**

Image Score	# of scores	% of scores
1 (no lesion)	587	44.47%
2 (Widened PDL)	296	22.42%
3 (lesion present)	443	33.56%
<b>Total</b>	<b>1320</b>	<b>100%</b>

**Table 2a. Provider Scores and Score Counts with Individual Agreement Index ( $P_i$ ) for Image #s 1-40**

Image #	Provider #											Score Counts			$P_i^*$
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	
1	1	1	1	2	1	3	2	3	2	1	1	6	3	2	34.55%
2	1	2	2	1	2	3	2	2	2	2	2	2	8	1	52.73%
3	1	1	1	1	1	1	1	1	1	1	2	10	1	0	81.82%
4	1	1	1	1	1	1	1	1	1	1	2	10	1	0	81.82%
5	1	1	1	1	1	1	1	1	1	2	1	10	1	0	81.82%
6	1	1	2	1	1	2	1	1	1	2	2	7	4	0	49.09%
7	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
8	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
9	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
10	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
11	1	2	2	1	1	1	1	2	2	3	2	5	5	1	36.36%
12	1	1	1	1	1	1	1	2	1	1	1	10	1	0	81.82%
13	2	2	2	3	2	2	2	2	3	3	2	0	8	3	56.36%
14	2	2	2	2	2	2	2	2	2	3	3	0	9	2	67.27%
15	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
16	1	3	1	3	1	1	1	1	3	1	1	8	0	3	56.36%
17	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
18	1	1	1	1	1	2	2	1	2	2	1	7	4	0	49.09%
19	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
20	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
21	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
22	1	2	2	1	1	3	2	1	1	2	2	5	5	1	36.36%
23	1	1	1	1	1	1	1	1	1	2	1	10	1	0	81.82%
24	2	2	1	1	1	2	1	1	2	2	1	6	5	0	45.45%
25	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
26	2	1	1	2	2	1	2	1	1	2	2	5	6	0	45.45%
27	3	3	3	1	3	3	1	3	3	3	3	2	0	9	67.27%
28	2	2	1	2	2	2	3	1	2	3	2	2	7	2	41.82%
29	1	1	1	1	1	2	1	1	1	1	1	10	1	0	81.82%
30	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
31	3	2	2	1	3	3	2	1	1	2	1	4	4	3	27.27%

32	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
33	1	1	1	1	1	2	2	1	1	2	1	8	3	0	56.36%
34	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
35	3	3	3	3	3	3	3	3	2	3	3	0	1	10	81.82%
36	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
37	3	3	3	3	3	3	3	1	3	3	3	1	0	10	81.82%
38	2	1	3	2	1	2	2	1	1	3	3	4	4	3	27.27%
39	3	3	3	3	3	3	3	1	3	3	3	1	0	10	81.82%
40	1	1	1	1	1	2	1	1	1	2	2	8	3	0	56.36%

*\*Agreement index score that measures the extent that raters agree for a single image*

**Table 2b. Provider Scores and Score Counts with Individual Agreement Index ( $P_i$ ) for Image #s 41-80**

Image #	Provider #											Score Counts			$P_i^*$
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	
41	1	1	1	1	1	2	1	1	1	1	2	9	2	0	67.27%
42	1	2	1	1	2	2	2	2	1	2	2	4	7	0	49.09%
43	1	1	1	1	1	1	1	2	1	1	2	9	2	0	67.27%
44	1	1	1	1	1	3	1	1	1	2	1	9	1	1	65.45%
45	3	3	3	3	3	3	3	1	3	3	3	1	0	10	81.82%
46	1	1	1	1	1	2	1	1	1	1	1	10	1	0	81.82%
47	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
48	3	3	3	2	3	3	3	3	3	3	3	0	1	10	81.82%
49	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
50	1	2	1	2	1	2	2	1	1	2	2	5	6	0	45.45%
51	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
52	1	1	1	1	1	1	1	1	1	1	2	10	1	0	81.82%
53	2	2	3	3	2	3	2	2	1	2	2	1	7	3	43.64%
54	1	1	1	1	1	2	2	1	1	1	2	8	3	0	56.36%
55	2	3	3	2	2	3	3	3	2	3	3	0	4	7	49.09%
56	2	3	2	1	3	3	2	3	2	3	2	1	5	5	36.36%
57	1	1	1	2	1	3	1	1	1	2	1	8	2	1	52.73%
58	1	1	2	2	2	2	1	2	1	1	2	5	6	0	45.45%
59	2	2	2	3	1	2	1	1	2	2	1	4	6	1	38.18%
60	1	1	1	1	1	2	1	1	1	2	1	9	2	0	67.27%
61	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
62	3	2	2	1	1	2	2	3	1	2	2	3	6	2	34.55%
63	2	2	3	2	1	2	2	2	2	3	2	1	8	2	52.73%
64	1	2	2	2	1	3	2	2	1	3	2	3	6	2	34.55%
65	1	1	1	1	1	1	2	1	1	1	2	9	2	0	67.27%
66	3	3	3	3	3	3	3	3	2	3	3	0	1	10	81.82%
67	3	1	2	2	1	3	1	1	1	2	1	6	3	2	34.55%
68	1	2	2	2	1	2	1	1	1	2	1	6	5	0	45.45%

69	1	1	1	1	1	1	1	1	1	2	1	10	1	0	81.82%
70	1	1	3	1	1	1	1	1	1	3	1	9	0	2	67.27%
71	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
72	3	3	3	3	3	3	3	1	3	3	3	1	0	10	81.82%
73	1	2	1	2	1	2	2	1	2	2	2	4	7	0	49.09%
74	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
75	1	1	1	1	1	2	1	1	1	2	1	9	2	0	67.27%
76	3	2	2	2	2	3	3	2	2	3	3	0	6	5	45.45%
77	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
78	1	1	1	1	1	2	2	1	1	2	2	7	4	0	49.09%
79	1	2	2	1	2	2	3	2	1	3	3	3	5	3	29.09%
80	3	3	2	3	3	3	3	2	2	3	3	0	3	8	56.36%

*\*Agreement index score that measures the extent that raters agree for a single image*

**Table 2c. Provider Scores and Score Counts with Individual Agreement Index ( $P_i$ ) for Image #s 81-120**

Image #	Provider #											Score Counts			$P_i^*$
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	
81	1	1	1	1	1	1	2	1	1	2	2	8	3	0	56.36%
82	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
83	3	3	3	3	3	1	3	3	3	3	3	1	0	10	81.82%
84	1	1	2	1	1	2	2	1	2	2	2	5	6	0	45.45%
85	2	2	2	1	2	3	1	1	2	3	1	4	5	2	30.91%
86	1	2	1	2	2	3	2	3	2	3	2	2	6	3	34.55%
87	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
88	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
89	1	1	1	1	3	3	3	3	3	3	3	4	0	7	49.09%
90	1	1	1	1	1	1	1	1	2	2	2	8	3	0	56.36%
91	1	1	1	1	1	2	2	2	1	1	1	8	3	0	56.36%
92	1	2	1	2	1	1	1	1	1	1	1	9	2	0	67.27%
93	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
94	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
95	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
96	1	2	1	2	2	2	2	2	1	3	2	3	7	1	43.64%
97	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
98	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
99	1	1	1	1	1	2	1	2	1	1	1	9	2	0	67.27%
100	1	2	1	1	1	2	1	2	1	2	1	7	4	0	49.09%
101	2	2	2	2	1	2	1	2	2	3	2	2	8	1	52.73%
102	1	1	1	1	1	2	1	1	1	3	1	9	1	1	65.45%
103	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
104	1	1	1	1	1	1	2	1	1	1	2	9	2	0	67.27%
105	1	2	2	1	1	3	3	3	1	3	3	4	2	5	30.91%



106	3	2	2	3	3	3	3	3	3	3	3	0	2	9	67.27%
107	1	3	3	3	3	3	3	3	3	3	3	1	0	10	81.82%
108	2	2	2	1	1	2	2	2	1	3	2	3	7	1	43.64%
109	1	1	1	1	1	1	2	1	1	1	2	9	2	0	67.27%
110	1	2	2	1	1	2	2	1	2	3	2	4	6	1	38.18%
111	1	1	1	1	2	1	2	1	1	2	2	7	4	0	49.09%
112	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
113	1	3	1	1	1	1	2	2	1	3	1	7	2	2	41.82%
114	3	2	3	2	3	3	3	3	3	3	3	0	2	9	67.27%
115	1	2	1	1	1	1	2	2	1	2	2	6	5	0	45.45%
116	1	1	1	1	2	1	1	1	1	2	1	9	2	0	67.27%
117	1	1	1	1	1	2	1	1	1	2	1	9	2	0	67.27%
118	2	2	2	1	2	1	3	2	2	3	3	2	6	3	34.55%
119	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
120	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%

*\*Agreement index score that measures the extent that raters agree for a single image*

**Table 3a. Provider Scores and Score Counts for Images that had 100% Agreement between providers**

Image #	Provider #											Score Counts			$P_i^*$
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	
7	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
8	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
9	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
10	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
15	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
17	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
19	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
20	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
21	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
25	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
30	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
32	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
34	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
36	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
47	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
49	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
51	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
61	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
71	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
74	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
77	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
82	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%

87	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
88	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
93	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
94	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
95	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
97	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
98	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
103	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%
112	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
119	1	1	1	1	1	1	1	1	1	1	1	11	0	0	100.00%
120	3	3	3	3	3	3	3	3	3	3	3	0	0	11	100.00%

*\*Agreement index score that measures the extent that raters agree for a single image*

**Table 3b. Provider Scores and Score Counts for Images that had 67-82% Agreement between providers**

Image #	Provider #											Score Counts			$P_i^*$
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	
3	1	1	1	1	1	1	1	1	1	1	2	10	1	0	81.82%
4	1	1	1	1	1	1	1	1	1	1	2	10	1	0	81.82%
5	1	1	1	1	1	1	1	1	1	2	1	10	1	0	81.82%
12	1	1	1	1	1	1	1	2	1	1	1	10	1	0	81.82%
23	1	1	1	1	1	1	1	1	1	2	1	10	1	0	81.82%
29	1	1	1	1	1	2	1	1	1	1	1	10	1	0	81.82%
35	3	3	3	3	3	3	3	3	2	3	3	0	1	10	81.82%
37	3	3	3	3	3	3	3	1	3	3	3	1	0	10	81.82%
39	3	3	3	3	3	3	3	1	3	3	3	1	0	10	81.82%
45	3	3	3	3	3	3	3	1	3	3	3	1	0	10	81.82%
46	1	1	1	1	1	2	1	1	1	1	1	10	1	0	81.82%
48	3	3	3	2	3	3	3	3	3	3	3	0	1	10	81.82%
52	1	1	1	1	1	1	1	1	1	1	2	10	1	0	81.82%
66	3	3	3	3	3	3	3	3	2	3	3	0	1	10	81.82%
69	1	1	1	1	1	1	1	1	1	2	1	10	1	0	81.82%
72	3	3	3	3	3	3	3	1	3	3	3	1	0	10	81.82%
83	3	3	3	3	3	1	3	3	3	3	3	1	0	10	81.82%
107	1	3	3	3	3	3	3	3	3	3	3	1	0	10	81.82%
14	2	2	2	2	2	2	2	2	2	3	3	0	9	2	67.27%
27	3	3	3	1	3	3	1	3	3	3	3	2	0	9	67.27%
41	1	1	1	1	1	2	1	1	1	1	2	9	2	0	67.27%
43	1	1	1	1	1	1	1	2	1	1	2	9	2	0	67.27%
60	1	1	1	1	1	2	1	1	1	2	1	9	2	0	67.27%
65	1	1	1	1	1	1	2	1	1	1	2	9	2	0	67.27%
70	1	1	3	1	1	1	1	1	1	3	1	9	0	2	67.27%
75	1	1	1	1	1	2	1	1	1	2	1	9	2	0	67.27%

92	1	2	1	2	1	1	1	1	1	1	1	9	2	0	67.27%
99	1	1	1	1	1	2	1	2	1	1	1	9	2	0	67.27%
104	1	1	1	1	1	1	2	1	1	1	2	9	2	0	67.27%
106	3	2	2	3	3	3	3	3	3	3	3	0	2	9	67.27%
109	1	1	1	1	1	1	2	1	1	1	2	9	2	0	67.27%
114	3	2	3	2	3	3	3	3	3	3	3	0	2	9	67.27%
116	1	1	1	1	2	1	1	1	1	2	1	9	2	0	67.27%
117	1	1	1	1	1	2	1	1	1	2	1	9	2	0	67.27%

*\*Agreement index score that measures the extent that raters agree for a single image*

**Table 3c. Provider Scores and Score Counts for Images that had 41-66% Agreement between providers**

Image #	Provider #											Score Counts			$P_i^*$
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	
44	1	1	1	1	1	3	1	1	1	2	1	9	1	1	65.45%
102	1	1	1	1	1	2	1	1	1	3	1	9	1	1	65.45%
13	2	2	2	3	2	2	2	2	3	3	2	0	8	3	56.36%
16	1	3	1	3	1	1	1	1	3	1	1	8	0	3	56.36%
33	1	1	1	1	1	2	2	1	1	2	1	8	3	0	56.36%
40	1	1	1	1	1	2	1	1	1	2	2	8	3	0	56.36%
54	1	1	1	1	1	2	2	1	1	1	2	8	3	0	56.36%
80	3	3	2	3	3	3	3	2	2	3	3	0	3	8	56.36%
81	1	1	1	1	1	1	2	1	1	2	2	8	3	0	56.36%
90	1	1	1	1	1	1	1	1	2	2	2	8	3	0	56.36%
91	1	1	1	1	1	2	2	2	1	1	1	8	3	0	56.36%
2	1	2	2	1	2	3	2	2	2	2	2	2	8	1	52.73%
57	1	1	1	2	1	3	1	1	1	2	1	8	2	1	52.73%
63	2	2	3	2	1	2	2	2	2	3	2	1	8	2	52.73%
101	2	2	2	2	1	2	1	2	2	3	2	2	8	1	52.73%
6	1	1	2	1	1	2	1	1	1	2	2	7	4	0	49.09%
18	1	1	1	1	1	2	2	1	2	2	1	7	4	0	49.09%
42	1	2	1	1	2	2	2	2	1	2	2	4	7	0	49.09%
55	2	3	3	2	2	3	3	3	2	3	3	0	4	7	49.09%
73	1	2	1	2	1	2	2	1	2	2	2	4	7	0	49.09%
78	1	1	1	1	1	2	2	1	1	2	2	7	4	0	49.09%
89	1	1	1	1	3	3	3	3	3	3	3	4	0	7	49.09%
100	1	2	1	1	1	2	1	2	1	2	1	7	4	0	49.09%
111	1	1	1	1	2	1	2	1	1	2	2	7	4	0	49.09%
24	2	2	1	1	1	2	1	1	2	2	1	6	5	0	45.45%
26	2	1	1	2	2	1	2	1	1	2	2	5	6	0	45.45%
50	1	2	1	2	1	2	2	1	1	2	2	5	6	0	45.45%
58	1	1	2	2	2	2	1	2	1	1	2	5	6	0	45.45%
68	1	2	2	2	1	2	1	1	1	2	1	6	5	0	45.45%

76	3	2	2	2	2	3	3	2	2	3	3	0	6	5	45.45%
84	1	1	2	1	1	2	2	1	2	2	2	5	6	0	45.45%
115	1	2	1	1	1	1	2	2	1	2	2	6	5	0	45.45%
53	2	2	3	3	2	3	2	2	1	2	2	1	7	3	43.64%
96	1	2	1	2	2	2	2	2	1	3	2	3	7	1	43.64%
108	2	2	2	1	1	2	2	2	1	3	2	3	7	1	43.64%
28	2	2	1	2	2	2	3	1	2	3	2	2	7	2	41.82%
113	1	3	1	1	1	1	2	2	1	3	1	7	2	2	41.82%

*\*Agreement index score that measures the extent that raters agree for a single image*

**Table 3d. Provider Scores and Score Counts for Images that had 27-39% Agreement between providers**

Image #	Provider #											Score Counts			$P_i^*$
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	
59	2	2	2	3	1	2	1	1	2	2	1	4	6	1	38.18%
110	1	2	2	1	1	2	2	1	2	3	2	4	6	1	38.18%
11	1	2	2	1	1	1	1	2	2	3	2	5	5	1	36.36%
22	1	2	2	1	1	3	2	1	1	2	2	5	5	1	36.36%
56	2	3	2	1	3	3	2	3	2	3	2	1	5	5	36.36%
1	1	1	1	2	1	3	2	3	2	1	1	6	3	2	34.55%
62	3	2	2	1	1	2	2	3	1	2	2	3	6	2	34.55%
64	1	2	2	2	1	3	2	2	1	3	2	3	6	2	34.55%
67	3	1	2	2	1	3	1	1	1	2	1	6	3	2	34.55%
86	1	2	1	2	2	3	2	3	2	3	2	2	6	3	34.55%
118	2	2	2	1	2	1	3	2	2	3	3	2	6	3	34.55%
85	2	2	2	1	2	3	1	1	2	3	1	4	5	2	30.91%
105	1	2	2	1	1	3	3	3	1	3	3	4	2	5	30.91%
79	1	2	2	1	2	2	3	2	1	3	3	3	5	3	29.09%
31	3	2	2	1	3	3	2	1	1	2	1	4	4	3	27.27%
38	2	1	3	2	1	2	2	1	1	3	3	4	4	3	27.27%

*\*Agreement index score that measures the extent that raters agree for a single image*

**Table 4. Individual Image Agreement Score ( $P_i$ ) with Number (%) of Images and Scoring Distribution**

$P_i^*$	# of Images with $P_i^*$	% of images with $P_i^*$	Cumulative % of images with $P_i^*$	Scoring Distribution**
100.00%	33	27.5%	27.5%	11, 0, 0
81.82%	18	15.0%	42.5%	10, 1, 0
67.27%	16	13.3%	55.8%	9, 2, 0
65.45%	2	1.7%	57.5%	9, 1, 1
56.36%	9	7.5%	65.0%	8, 3, 0
52.73%	4	3.3%	68.3%	8, 2, 1
49.09%	9	7.5%	75.8%	7, 4, 0

45.45%	8	6.7%	82.5%	6, 5, 0
43.64%	3	2.5%	85.0%	7, 3, 1
41.82%	2	1.7%	86.7%	7, 2, 2
38.18%	2	1.7%	88.3%	6, 4, 1
36.36%	3	2.5%	90.8%	5, 5, 1
34.55%	6	5.0%	95.8%	6, 3, 2
30.91%	2	1.7%	97.5%	5, 4, 2
29.09%	1	0.8%	98.3%	5, 3, 3
27.27%	2	1.7%	100.0%	4, 4, 3

*\*Agreement index score that measures the extent that raters agree for a single image*

*\*\*Distribution of the three scores given by the 11 providers. (For example, 8, 2, 1 means 8 providers gave the same score, 2 gave another score, and 1 gave the remaining score.)*

**Table 5. Individual Image Agreement Index Score ( $P_i$ ) and corresponding images #s**

$P_i^*$	Image #s
100.00%	7, 8, 9, 10, 15, 17, 19, 20, 21, 25, 30, 32, 34, 36, 47, 49, 51, 61, 71, 74, 77, 82, 87, 88, 93, 94, 95, 97, 98, 103, 112, 119, 120
81.82%	3, 4, 5, 12, 23, 29, 35, 37, 39, 45, 46, 48, 52, 66, 69, 72, 83, 107
67.27%	14, 27, 41, 43, 60, 65, 70, 75, 92, 99, 104, 106, 109, 114, 116, 117
65.45%	44, 102
56.36%	13, 16, 33, 40, 54, 80, 81, 90, 91
52.73%	2, 57, 63, 101
49.09%	6, 18, 42, 55, 73, 78, 89, 100, 111
45.45%	24, 26, 50, 58, 68, 76, 84, 115
43.64%	53, 96, 108
41.82%	28, 113
38.18%	59, 110
36.36%	11, 22, 56
34.55%	1, 62, 64, 67, 86, 118
30.91%	85, 105
29.09%	79
27.27%	31, 38

*\*Agreement index score that measures the extent that raters agree for a single image*