



Consistency of using two dental staging systems between undergraduate and graduate students – a pilot study

Pinto, Roberta ¹; Causs, Maria Eduarda ¹; Silva, Laís ¹; Moreira, Débora ²; Angelakopoulos, Nikolaos ^{1,3*}; Franco, Ademir ^{1,4}

¹Division of Forensic Dentistry, Faculdade São Leopoldo Mandic, Campinas/SP, Brazil.

²Division of Oral Radiology, Faculdade São Leopoldo Mandic, Campinas/SP Brazil.

³Department of Orthodontics and Dentofacial Orthopedics, University of Bern, Bern, Switzerland.

⁴Department of Therapeutic Dentistry, Institute of Dentistry, Sechenov First Moscow State Medical University, Moscow, Russian Federation.

*Corresponding author: Nikolaos Angelakopoulos nikolaos.angelakopoulos@unibe.ch; Department of Orthodontics and Dentofacial Orthopedics, University of Bern, Freiburgstrasse 7, Bern 3010, Switzerland

KEYWORDS

Age;
Dental;
Dental students;
Forensic dentistry;
Radiograph;
Reproducibility

ABSTRACT

Background: Radiographic dental staging remains one of the most frequently applied approaches for subadult dental age assessment in forensic odontology. In this preliminary study, we compared the consistency across observers with different academic level applying two staging systems. Methods: Forty-six participants composed four groups according to academic level: undergraduate (n = 16), specialization (n = 13), master's (n = 10), and doctoral students (n = 7). Each participant assessed 10 panoramic radiographs from individuals aged 6.5–13.5 years (mean: 9.45 ± 2.45 years), staging seven left permanent mandibular teeth (International Dental Federation: 31–37) using both Demirjian's and Moorrees' systems. A total of 920 radiographic assessments and 6,440 dental stage allocations were generated. Observer consistency was assessed separately for each tooth, staging system, and academic level-based group using weighted Fleiss' kappa coefficients with ordinal weights and 95% confidence intervals. Results: Staging performance using Demirjian's and Moorrees' systems by observers with different academic levels showed similar consistency. Absolute differences between systems remained below 0.10 in 92.9% of the comparisons. Teeth 31 and 36 generally presented the lowest consistency estimates across staging systems and observer groups, whereas tooth 34 frequently showed among the highest consistency values. Despite the discrete differences, the largest absolute differences between Demirjian's and Moorrees' systems varied according to the observers' educational background, being more evident for teeth 31, 36, and 37, while the remaining teeth showed even more discrete differences between staging systems. Conclusion: Despite the discrete differences observed between staging through Demirjian's and Moorrees' systems, the results of this study should be carefully interpreted since it presented a preliminary point-specific assessment of consistency and not accuracy. Future studies should implement comparisons with gold standard observers and with chronological age to enable accuracy estimates.

1. INTRODUCTION

Dental age assessment methods applicable to living individuals, namely children, adolescents and adults, usually require radiological examination [1]. This is because these methods are considered non-invasive [2], despite the ionizing biological effect of X-ray emission in human tissues [3]. The analytical techniques applied in radiological methods are metric and non-metric [4]. Among the non-metric ones, dental staging is the most common [5-10]. This approach is useful when teeth have not reached complete apical closure, which occurs about the age of 21.5 years, after third molar full formation [11].

Several radiological methods for dental age estimation rely on staging techniques [5-10]. Examples include the systems proposed by Nola [5] (1960), Moorrees et al. [6] (1963), Demirjian et al. [7] (1973), and the modification by Kohler et al. [8] (1994) of Gleiser and Hunt's [9] (1955) system. The rationale behind staging systems varies between (i) predictive assessments of the remaining dental formation [6, 8] and (ii) proportional assessments of the visible ratio between crown and root [7]. The former is usually represented by schematic illustrations with fractions of crown and root formation (e.g. labeled R ¼, ½ and ¾) alternated with other representative stages [6], such as Rc for root complete [8]. The latter is more descriptive towards the radiographic aspect of the tooth and normally addresses proportions of root length based on the visible crown length [7]. This

DOI: <https://doi.org/10.14710/actodont.33953>

Submitted: 29/07/2026 Revised: 25/08/2026

Accepted: 31/08/2026 Available Online: 25/09/2026

Published: 01/12/2026

is illustrated by the following sentence from Demirjian's seminal system: Stage F (uniradicular teeth) – “The root length is equal to or greater than the crown height” [7].

In addition to the different presentations, labels and level of description that a staging system can have, there is also variability in the number of stages per system. Demirjian's approach [7], for instance, is referred to as a popular method worldwide [12], which is based on the allocation of eight distinctive stages per tooth position [7]. On other hand, Moorrees' approach includes 13 stages for single-rooted teeth and 14 for multi-rooted teeth [6]. When it comes to reproducibility, a lower number of stages can lead to higher consistency in intra- and inter-observer assessment [13], but in practice dental age assessment may not be as accurate as it could because of the inherent large gaps left between stages. Conversely, a higher number of stages can increase complexity and decrease consistency because of the higher granularity for stage allocation [13].

Another aspect encompassed in this scenario is the examined tooth position [13]. Authors have demonstrated that staging consistency is not uniform across different tooth types [13]. It has been demonstrated, for instance, that mandibular permanent lateral incisors had a lower consistency rate compared to other permanent teeth in the third quadrant after the application of Demirjian's system [13]. In this context, recent dental age assessment models have prioritized specific teeth for the quantification of age scores [14]. This is the case of the mandibular permanent second molar that received more weight in a statistical model for the Brazilian population [14].

Who is the person conducting dental staging is also a variable to be considered in dental age assessment due to a potential subjectivity, even with the standard presentation set by each staging system. According to the American board of Forensic Odontology (ABFO) Standards and Guidelines for Dental Age Assessment and the American Dental Association (ADA) Technical Report No. 1077 for Age Assessment by Dental Analysis [15] “The practitioner (...) is responsible for being familiar with and utilizing current age assessment scientific methodology.” The instruction does not specify that a certain degree of training is necessary to perform such analyses, but it endorses the importance of familiarization with current methods [15], which includes knowledge of radiographic dental staging systems. Understanding how different groups of observers perform when staging with two distinct systems into the same radiographs can be fundamental to learning where major attention is needed during dental age assessment.

Based on the above, the present study proposes a preliminary observational pilot investigation of the consistency, reproducibility rather than the accuracy, of two dental developmental staging systems widely used in forensic age estimation, Demirjian's and Moorrees'. The primary objective was to quantify observer consistency across groups of observers with different levels of formal academic training. Because dental staging is an observer-dependent procedure, the level and nature of formal training may represent an important source of variability in stage allocation. Academic training was therefore selected as the principal stratification variable to investigate whether consistency is maintained across different levels of formal exposure to dental age assessment. Additionally, consistency was evaluated according to tooth position to identify the teeth that may present greater challenges during stage allocation. The findings are intended to provide preliminary evidence to support the methodological design of future studies.

2. MATERIALS AND METHODS

2.1 Study design and ethical aspects

This study was reported according to the Enhancing the Quality and Transparency of Health Research (EQUATOR) network [16]. More specifically, EQUATOR's checklist for observational analytical cross-sectional designs was consulted. Ethical approval was obtained from the Institutional Committee of Ethics in Human Research of Faculdade São Leopoldo Mandic (protocol: 95980626.1.0000.5374). The study was carried out under the ethical standards laid down by the Declaration of Helsinki (Finland) and its later amendments [17]. No additional data were collected for this study, and no subjects were exposed to radiation specifically for research purposes.

2.2 Participants and sample

This study involved human subjects in two distinct capacities, namely as participants performing the dental staging task and as patients represented in a radiographic database. The participants of the present study were recruited composing four groups, namely undergraduate, specialization, master's and doctoral students. Hence, the eligibility criteria behind the participants' sampling were formal enrolment as undergraduate bachelor of dentistry or in specialization, master's or doctoral training in forensic odontology. For undergraduate students, recruitment was accomplished in the last academic year, after the forensic odontology subject of dental age estimation. For graduate students, recruitment was accomplished during the training process on the topic. Since enrolment in graduate programmes is open to dentists from diverse professional backgrounds and specialties, the groups of graduate students included participants with varied experience. In general, all participants had at least two years of clinical experience. Academic training was selected as the principal variable for observer stratification because it provides a structured framework for comparing different levels of formal exposure to forensic odontology and dental age assessment. Professional experience alone was considered less suitable because graduate students may have heterogeneous clinical backgrounds, specialties, and practical involvement in dental age assessment. Participants that have not completed the

task providing full staging to the radiographic dataset were excluded as well as those that have not delivered the analysis within seven days.

The total sample consisted of 46 individuals (35 females and 11 males, all Brazilian nationals), being 16 undergraduate, 13 specialization, 10 master's and 7 doctoral students. Regarding their background, undergraduate students of the last academic year had an exposure time of 10 hours of subadult dental age assessment, while graduate student students from specialization, master's and doctoral courses had 40 hours each, including lectures and practice of subadult and adult dental age estimation. Their minimum course length was 1.5, 2 and 3 years, respectively. It must be noted that graduate students came from different undergraduate institutions and were exposed to a varied set of dental curricula, including or not dental age estimation as part of the forensic odontology discipline.

For the radiographic sample, sets of 10 panoramic radiographs (five females and 5 males) were presented to each student. The radiographs were from individuals between 6.5 to 13.5 years (mean: 9.45 ± 2.45 years) all obtained retrospectively from an existing institutional image database for maxillofacial diagnosis, meaning that no radiographic exposure was performed for research purposes. The inclusion criteria consisted of digital images displaying developing permanent teeth, from individuals with known sex, date of birth and date of image acquisition, and proper spatial resolution. Exclusion criteria considered individuals with bilaterally missing, decayed, endodontically treated or largely restored permanent teeth in the mandible, presence of pathological lesions such as tumors and cysts associated with mandibular teeth, visible mandibular malformation, mandibular orthodontic, and orthopedic or surgical appliances involving the permanent teeth.

2.3 Settings, variables and assessment

Each student received a set of 10 digital panoramic radiographs, the original articles published by Demirjian et al. [7] and Moorrees et al. [6], followed by a Microsoft Excel (Microsoft Corp., Redmond, WA, USA) data collection form with instructions on how to conduct the staging and subsequent registration into the data collection form. Individually, they had to assess each radiograph, which were anonymized and numbered from 1 to 10. Their assessment consisted of staging the seven mandibular permanent teeth on the left side (International Dental Federation): central incisor (tooth 31), lateral incisor (tooth 32), canine (tooth 33), first premolar (tooth 34), second premolar (tooth 35), first molar (tooth 36) and second molar (tooth 37). The staging process should be presented in two different data collection sheets within the data extraction form, one for Demirjian's stages (presented from A to H) and one for Moorrees' stages (presented from 1 to 13 for single-rooted teeth and from 1 to 14 for multi-rooted teeth). Considering that each student ($n = 46$) had to assess 10 radiographs twice (one for each staging system), the complete number of expected observations for the dataset was 920, representing 6.440 dental stages (seven teeth per radiograph). Instructions were made so the participants should not base their stage allocation solely on the schematic representations available in each article, but also on stage description and full article content. Native Windows (Microsoft Corp., Redmond, WA, USA) or MacOS (Apple, Cupertino, CA, USA) operating systems' image viewers were allowed, as well as image magnification preferably up to 200%. Task deadline was set to seven days, starting immediately after initial instructions. A supervisor, forensic odontologist, was available to support and solve potential doubts.

Prior to data analysis, it must be noted that the primary objective of the present study was to evaluate the reproducibility and consistency of dental developmental stage allocation rather than fitness in comparison to a gold standard observer or accuracy compared to chronological age. These are distinct methodological constructs. While these metrics reflects how closely an observer's assessment corresponds to a reference standard or the known chronological age, consistency evaluates the extent to which developmental stages are assigned across observers. Accordingly, the present study focused exclusively on the consistency of stage allocation, which represents a fundamental prerequisite for any stage-based age estimation method, as reliable stage assignment must precede meaningful investigations of diagnostic accuracy and age estimation performance.

2.4 Statistical analysis

Agreement among observers in the classification of dental development stages was assessed separately for each of the seven teeth, considering the dental staging systems by Demirjian et al. [7] and Moorrees et al. [6]. To quantify consistency, weighted Fleiss' kappa for multiple raters was calculated using ordinal weights in order to penalize disagreements proportionally to the distance between assigned categories. This method is appropriate for the present study because dental development stages constitute an ordinal scale and the number of observers exceeded two. Quadratic weighting was chosen to reflect the clinical assumption that larger disagreements between stages carry disproportionately greater consequences than smaller ones. While Fleiss' weighted kappa is the most widely used statistic for inter-observer agreement among multiple raters on an ordinal scale, it is susceptible to the kappa paradox, whereby highly asymmetric rating distributions can yield spuriously low values that underestimate true agreement. Results were expressed as weighted kappa coefficients (κ) with their respective 95% confidence intervals (95% CI). Kappa values were interpreted according to the criteria proposed by Landis and Koch [17] (1977): $\kappa < 0.20$ (slight agreement), 0.21–0.40 (fair), 0.41–0.60 (moderate), 0.61–0.80 (substantial), and $\kappa > 0.80$ (almost perfect). In parallel, Gwet's AC2 was applied as a supplementary approach, also considering the quadratic modelling. This approach was chosen since it is a more stable alternative, as it defines expected chance agreement based on a uniform distribution rather than observed marginal frequencies, making it robust to distributional asymmetry. For both statistics, standard errors and 95% confidence intervals were estimated using the bootstrap method with 1,000 replications. All analyses were performed in Stata (StataCorp) using the user-written command kappaetc.

3. RESULTS

For undergraduate students, weighted kappa values ranged from 0.60 to 0.91 for Demirjian's staging system and from 0.69 to 0.95 for Moorrees' system, indicating moderate to almost perfect agreement across the evaluated teeth. The largest absolute difference between the two systems was observed for tooth 31 (0.14), whereas similar agreement estimates were obtained for teeth 32, 33, 36, and 37. Regardless of the staging system, teeth 31 and 36 consistently yielded the lowest kappa coefficients, whereas teeth 34 and 37 showed the highest levels of agreement (Table 1). Gwet's AC2 values were high for both systems, ranging from 0.90 to 0.96 for Demirjian's and from 0.89 to 0.99 for Moorrees' system.

Tooth	Demirjian			Moorrees			Absolute difference Demirjian - Moorrees
	Number of stages	κ	CI95%	Number of stages	κ	CI95%	
31	3	0.60	0.21; 0.99	6	0.74	0.44; 1.00	-0.14
32	5	0.87	0.70; 1.00	7	0.87	0.72; 1.00	0.00
33	6	0.86	0.68; 1.00	9	0.86	0.72; 1.00	0.00
34	5	0.88	0.69; 1.00	9	0.95	0.90; 1.00	-0.07
35	5	0.79	0.54; 1.00	8	0.88	0.78; 0.98	-0.09
36	3	0.69	0.45; 0.93	5	0.69	0.50; 0.88	0.00
37	6	0.91	0.84; 0.98	12	0.90	0.81; 1.00	0.01

Tooth	Demirjian			Moorrees			Absolute difference Demirjian - Moorrees
	Number of stages	Gwet's AC2	CI95%	Number of stages	Gwet's AC2	CI95%	
31	3	0.92	0.83; 1.00	6	0.95	0.88; 1.00	-0.03
32	5	0.96	0.92; 1.00	7	0.94	0.87; 1.00	0.02
33	6	0.95	0.92; 0.99	9	0.96	0.92; 1.00	-0.01
34	5	0.94	0.86; 1.00	9	0.99	0.97; 1.00	-0.05
35	5	0.93	0.85; 1.00	8	0.97	0.95; 1.00	-0.04
36	3	0.90	0.80; 1.00	5	0.89	0.81; 0.98	0.01
37	6	0.96	0.94; 0.98	12	0.96	0.93; 0.99	0.00

Table 1. Outcomes for dental staging accomplished by undergraduate students considering Demirjian's and Moorrees' systems for each of the seven permanent mandibular teeth of the left side. K: Weighted Kappa; CI: Confidence interval; 31-37: International Dental Federation's notation system.

For participants enrolled in specialization programs, weighted kappa values ranged from 0.53 to 0.80 for Demirjian's system and from 0.51 to 0.81 for Moorrees' system. The largest absolute difference between the two systems was observed for tooth 37 (0.12), followed by tooth 31 (0.07). Tooth 36 consistently exhibited the lowest kappa coefficient, whereas teeth 32 and 34 showed the highest agreement estimates (Table 2). Gwet's AC2 demonstrated high agreement for both systems, ranging from 0.85 to 0.93 for Demirjian's and from 0.89 to 0.93 for Moorrees' system.

Tooth	Demirjian			Moorrees			Absolute difference Demirjian - Moorrees
	Number of stages	κ	CI95%	Number of stages	κ	CI95%	
31	3	0.63	0.44; 0.81	6	0.70	0.57; 0.83	-0.07
32	5	0.79	0.60; 0.98	7	0.81	0.70; 0.92	-0.02
33	6	0.72	0.49; 0.95	9	0.77	0.61; 0.92	-0.05
34	5	0.80	0.63; 0.97	9	0.81	0.63; 0.98	-0.01
35	5	0.67	0.53; 0.82	8	0.72	0.57; 0.87	-0.05
36	3	0.53	0.32; 0.74	5	0.51	0.12; 0.90	0.02
37	6	0.61	0.27; 0.96	12	0.73	0.52; 0.94	-0.12

Tooth	Demirjian			Moorrees			Absolute difference Demirjian - Moorrees
	Number of stages	Gwet's AC2	CI95%	Number of stages	Gwet's AC2	CI95%	
31	3	0.91	0.82; 1.00	6	0.93	0.85; 1.00	-0.02
32	5	0.85	0.70; 0.99	7	0.93	0.85; 1.00	-0.08
33	6	0.89	0.77; 1.00	9	0.89	0.79; 1.00	0.00
34	5	0.91	0.86; 0.96	9	0.90	0.83; 0.97	0.01
35	5	0.93	0.86; 1.00	8	0.91	0.84; 0.98	0.02
36	3	0.90	0.79; 1.00	5	0.92	0.82; 1.00	-0.02
37	6	0.87	0.73; 1.00	12	0.90	0.82; 0.97	-0.03

Table 2. Outcomes for dental staging accomplished by specialization students considering Demirjian's and Moorrees' systems for each of the seven permanent mandibular teeth of the left side. K: Weighted Kappa; CI: Confidence interval; 31-37: International Dental Federation's notation system

For master's students, weighted kappa values ranged from 0.63 to 0.92 for Demirjian's system and from 0.54 to 0.88 for Moorrees' system. Demirjian's system generally yielded higher kappa values, with the largest absolute difference observed for tooth 36 (0.18), followed by tooth 31 (0.09). Tooth 31 showed one of the lowest agreement estimates, whereas tooth 34 consistently demonstrated the highest agreement (Table 3). Gwet's AC2 values ranged from 0.82 to 0.95 for Demirjian's and from 0.71 to 0.94 for Moorrees' system.

Tooth	Demirjian			Moorrees			Absolute difference Demirjian - Moorrees
	Number of stages	κ	CI95%	Number of stages	κ	CI95%	
31	3	0.63	0.38; 0.88	6	0.54	0.28; 0.80	0.09
32	5	0.72	0.43; 1.00	7	0.73	0.51; 0.95	-0.01
33	6	0.83	0.63; 1.00	9	0.76	0.54; 0.99	0.07
34	5	0.92	0.83; 1.00	9	0.88	0.76; 1.00	0.04
35	5	0.81	0.72; 0.90	8	0.74	0.56; 0.92	0.07
36	3	0.76	0.60; 0.92	5	0.58	0.15; 1.00	0.18
37	6	0.73	0.44; 1.00	12	0.72	0.43; 1.00	0.01

Tooth	Demirjian			Moorrees			Absolute difference Demirjian - Moorrees
	Number of stages	Gwet's AC2	CI95%	Number of stages	Gwet's AC2	CI95%	
31	3	0.83	0.65; 1.00	6	0.71	0.43; 0.98	0.12
32	5	0.82	0.62; 1.00	7	0.82	0.67; 0.96	0.00
33	6	0.95	0.88; 1.00	9	0.91	0.82; 1.00	0.04
34	5	0.95	0.92; 0.98	9	0.94	0.90; 0.98	0.01
35	5	0.92	0.87; 0.97	8	0.87	0.79; 0.94	0.05
36	3	0.86	0.75; 0.97	5	0.82	0.60; 1.00	0.04
37	6	0.89	0.76; 1.00	12	0.87	0.72; 1.00	0.02

Table 3. Outcomes for dental staging accomplished by master's students considering Demirjian's and Moorrees' systems for each of the seven permanent mandibular teeth of the left side. K: Weighted Kappa; CI: Confidence interval; 31-37: International Dental Federation's notation system.

For doctoral students, weighted kappa values ranged from 0.60 to 0.87 for Demirjian's system and from 0.64 to 0.87 for Moorrees' system. Agreement estimates were highly comparable between the two systems, with differences not exceeding 0.07 for any tooth. Tooth 31 showed the lowest kappa coefficient, whereas teeth 33 and 34 consistently exhibited the highest agreement (Table 4). Gwet's AC2 values ranged from 0.83 to 0.95 for Demirjian's and from 0.81 to 0.95 for Moorrees' system.

Tooth	Demirjian			Moorrees			Absolute difference Demirjian - Moorrees
	Number of stages	κ	CI95%	Number of stages	κ	CI95%	
31	3	0.60	0.40; 0.80	6	0.64	0.40; 0.89	-0.04
32	5	0.77	0.57; 0.98	7	0.70	0.45; 0.96	0.07
33	6	0.87	0.74; 1.00	9	0.87	0.75; 0.98	0.00
34	5	0.86	0.72; 1.00	9	0.86	0.76; 0.96	0.00
35	5	0.70	0.56; 0.84	8	0.77	0.61; 0.94	-0.07
36	3	0.72	0.56; 0.88	5	0.71	0.52; 0.90	0.01
37	6	0.77	0.55; 1.00	12	0.78	0.44; 1.00	-0.01

Tooth	Demirjian			Moorrees			Absolute difference Demirjian - Moorrees
	Number of stages	Gwet's AC2	CI95%	Number of stages	Gwet's AC2	CI95%	
31	3	0.83	0.64; 1.00	6	0.86	0.73; 0.99	-0.03
32	5	0.86	0.67; 1.00	7	0.81	0.60; 1.00	0.05
33	6	0.95	0.89; 1.00	9	0.95	0.91; 0.99	0.00
34	5	0.94	0.89; 1.00	9	0.94	0.91; 0.97	0.00
35	5	0.91	0.80; 1.00	8	0.91	0.87; 0.95	0.00
36	3	0.91	0.85; 0.97	5	0.87	0.76; 0.98	0.04
37	6	0.91	0.82; 1.00	12	0.91	0.73; 1.00	0.00

Table 4. Outcomes for dental staging accomplished by doctoral students considering Demirjian's and Moorrees' systems for each of the seven permanent mandibular teeth of the left side. K: Weighted Kappa; CI: Confidence interval; 31-37: International Dental Federation's notation system.

Overall, both staging systems demonstrated broadly comparable patterns of inter-observer consistency across the four groups of academic levels. Weighted kappa values were generally moderate to almost perfect, while Gwet's AC2 consistently indicated high agreement. Across groups, teeth 31 and 36 generally showed the lowest consistency estimates, whereas tooth 34 was consistently among the teeth with the highest agreement. The most notable difference between the two staging systems was observed among master's students for tooth 36, where the absolute difference in weighted kappa was 0.18. The overall patterns of consistency for Demirjian's and Moorrees' systems across groups of academic levels and tooth positions are illustrated in Figures 1 and 2.

Figure 1. Overview of consistency rates for dental staging with Demirjian's system across the seven tooth positions and the four groups of participants allocated based on their academic level, being undergraduate (grey), specialization (blue), master's (pink) and doctoral (red) students.

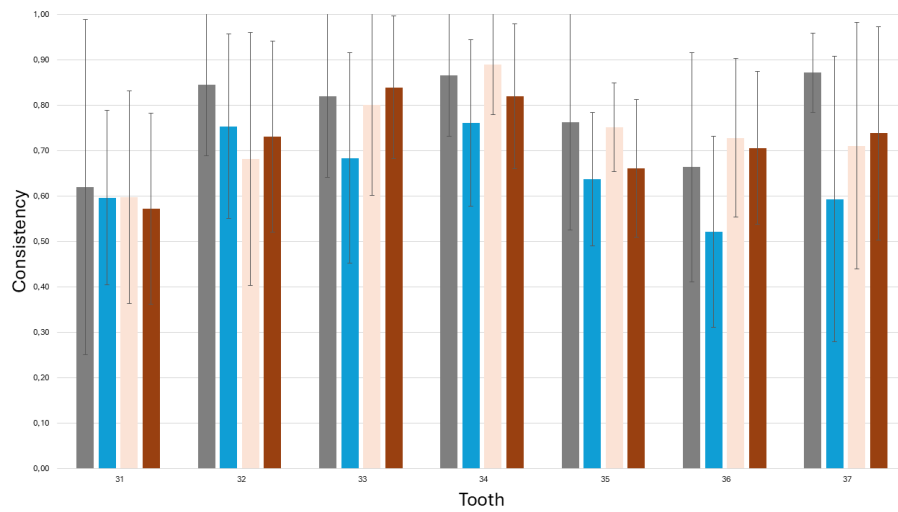
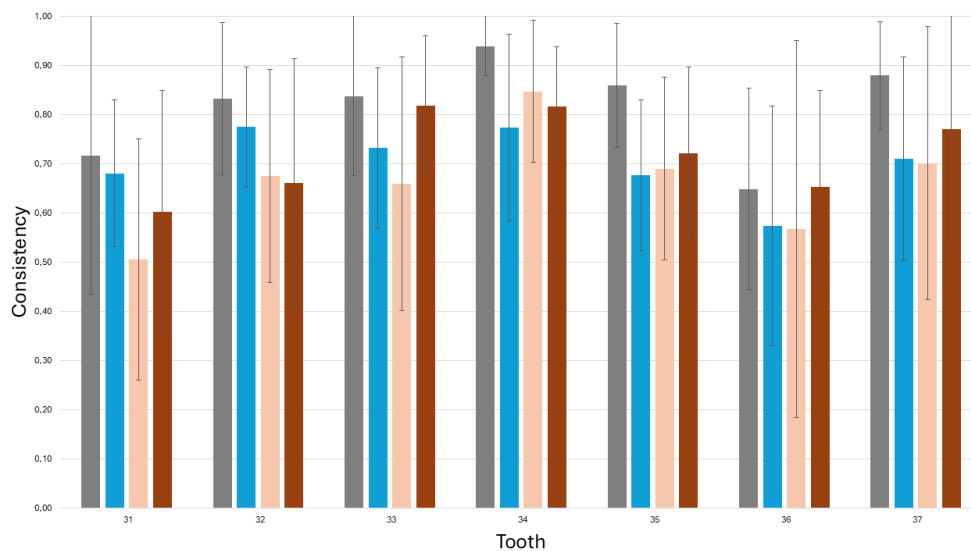


Figure 2. Overview of consistency rates for dental staging with Moorrees' system across the seven tooth positions and the four groups of participants allocated based on their academic level, being undergraduate (grey), specialization (blue), master's (pink) and doctoral (red) students.



4. DISCUSSION

Dental age assessment has experienced technological transformation led by artificial intelligence-driven research [18-21]. Studies have been developed to improve performance metrics, especially the accuracy of dental age estimates. Part of them has been conducted on specific procedures within the dental age assessment process, such as operator-dependent tasks. Among these tasks, radiographic tooth measurements and dental staging are the most common. However, while fully or semi-automated radiographic staging has been evolving, training the models still require ground truth information created by human observers. This makes studies on traditional dental staging and operator consistency relevant even in era of computer-based solutions. The present study compared the inter-observer consistency of Demirjian's [7] and Moorrees' [6] staging systems across four groups of observers with different levels of academic training.

The first aspect to be discussed is the training profile behind each sample group. Undergraduate students of the last year were addressed, right after the completion of the dental age assessment module into the Forensic Odontology discipline. The module

contained six hours of subadult dental age assessment theory, followed by two hours of on-campus practice and two hours of homework. The latter consisted of simulation cases using panoramic radiographs. The only staging systems though in the module were Demirjian's and Moorrees'. In contrast, specialization, master's and doctoral students, had 40 hours of subadult dental age assessment training distributed in on-campus lectures in the course of their activities, being 1.5, 2, and 3-year training programmes, respectively. During this time, they were exposed to approximately 20 dental age assessment methods 20 dental age assessment methods, including Demirjian's (original [7] and variations [22]) and Moorrees' [6].

With this in mind, interpretations of the obtained outcomes are pertinent: The overall results of the present study showed similar staging between the two systems across the different groups of observers. Moreover, the absolute difference of consistency metrics between Demirjian's and Moorrees' systems were below 0.10 in 92.9% of the analyses (26 out of 28 observations). Despite the discrete differences, the most evident outcome were observed for master's students assessing tooth 36, which reached absolute differences of 0.18, respectively. Interestingly, it must be noted that for tooth 36 the difference between systems was more pronounced, since only three stages of Demirjian's [7] system were used while five of Moorrees' [6] stages were used. In this example, the greater staging granularity of Moorrees' system (five stages versus three in Demirjian's system) resulted in a 23.7% reduction in consistency for the ten first molars – from a kappa value of 0.76 with Demirjian's to 0.58 with Moorrees' systems. Complementarily, Gwet's AC2 results were confirmatory for the four groups, corroborating the agreement patterns observed with weighted kappa statistics. Because Gwet's AC2 is less sensitive to prevalence and marginal distribution effects than kappa statistics, these findings provide an additional layer over the consistency between observers and systems. Naturally, these are only point-specific differences worth observing from the overall outcomes of the present study.

The findings of the present pilot study can also be considered in relation to a recent investigation by Angelakopoulos et al. [23], which evaluated the reliability of Demirjian's and Moorrees' systems among 18 experienced forensic odontologists and forensic experts using 250 panoramic radiographs. Despite differences in study design and observer profiles, the two investigations showed broadly consistent findings regarding the reproducibility of the two staging systems. The previous study [23] reported high inter-observer agreement for both systems, with disagreements predominantly involving adjacent developmental stages, and found slightly higher weighted reliability for Moorrees' system but higher exact agreement for Demirjian's system. In the present study, both systems likewise demonstrated broadly comparable patterns of inter-observer consistency across the four academic groups, without evidence of a consistent superiority of either system. These findings suggest that, although the greater granularity of the Moorrees system may increase the cognitive demands of stage allocation, particularly at specific developmental stages, it does not necessarily result in a substantial reduction in overall inter-observer consistency. Importantly, the present study extends the previous evidence by examining observers at different stages of formal academic training rather than focusing exclusively on experienced practitioners. This distinction is relevant because it suggests that the reproducibility of dental staging should be considered not only in relation to the characteristics of the staging system itself, but also in relation to the academic level and training context of the observer.

When comparisons were assessed per tooth position, important differences were concentrated in the permanent mandibular left central incisor. Merdietio Boedi et al. [13] (2025) already pointed to this finding in a study that evaluated Demirjian's staging system between eight forensic experts. A reasonable explanation for the low consistency in staging in central incisors is the fact that the anterior region of the mandible is normally affected by radiolucent images from the anterior-posterior superimposition with the cervical vertebrae [24]. In adolescents this region is also commonly affected by crowding [25], which can hamper visualization of the root contours. Another tooth that presented lower consistency values, was the permanent mandibular left first molar, especially for undergraduate and specialization students (for master's and doctoral students, tooth 36 showed low consistency only with Moorrees' staging system). Curiously, in the sampled age range (6.5 to 13.5 years, mean: 9.45 ± 2.45 years), both the teeth 31 and 36 are progressing into late stages of root formation. Hence, assessing these teeth in practice would depend on observer's decisions between a few numbers of stages – compared to the second molar, for instance. This can be especially challenging when the staging system has higher granularity in late root formation. A study conducted in 2020 [26] with third molars assessed with the staging system of Gunst and Hunt [9] (1955) modified by Köhler et al. [8] (1994) discussed the difficulties of allocating stages of apical root formation, namely full and $\frac{1}{2}$ apical closure, for instance. The same level of granularity is found in Moorrees' [6] last stages, which could pose difficulties to the staging of incisors and first molars, but not necessarily to the second molar given the sample's upper age limit of 13.5 years. It must be noted that the sample's age range was restricted between 6.5 and 13.5 years for two main reasons: the first was to address age categories that are commonly considered in dental age estimation scientific research (normally starting around the age of 5 or 6, except for atlas methods extending sample collection up to intrauterine life) and the second was to avoid a high number of fully developed teeth (after the age of 13.5 years for instance), which could artificially boost the agreement rates, given the higher frequency of stages for apical closure.

It should also be noted that the interpretation of consistency according to tooth position requires careful consideration. In the present study, all seven mandibular teeth were assessed simultaneously from the same panoramic radiograph. Consequently, when assigning a developmental stage to a given tooth, observers inevitably had access to the developmental status of the adjacent dentition and the overall maturation pattern of the individual. This methodological decision was intentional, as it reproduces routine forensic practice, in which teeth are evaluated within their anatomical and developmental context rather than as isolated, cropped images. Therefore, the reported consistency for each tooth should not be interpreted as reflecting exclusively the intrinsic difficulty of staging that tooth, but rather its assessment within the broader dentomaxillofacial context available to the observer. Nevertheless, future studies may investigate observer consistency using isolated cropped teeth under

controlled conditions, thereby removing contextual information from adjacent teeth and enabling the evaluation of stage allocation based solely on the morphological characteristics of the developing tooth. Comparing contextual and decontextualized assessment protocols may further clarify the influence of surrounding anatomical information on observer consistency.

Another finding of interest was the overall high consistency of undergraduate students in the present study. This phenomenon may be explained from different perspectives. This group of observers had dental age assessment theory and practice “fresh” in the memory, since their recruitment was accomplished right after the conclusion of the dental age assessment module. Another reason would be the fact that observers with more training time (even cumulative training since some of them followed a sequential academic track of specialization, master’s and doctoral courses in forensic odontology) did not necessarily work with dental age assessment cases in their routine. In Brazil, country of origin of the sample, practical dental age assessment is mainly performed by forensic odontologists working in public security services, mostly within the medicolegal police response. A recent survey with 28 forensic odontologists in the country, however, showed that almost half of the respondents had less than five dental age assessment cases per year [27]. Moreover, half of all the cases was related to postmortem identification [27], which normally involves adults, where dental staging is not the method of choice. On top of it, more experienced observers may be less intuitive during stage allocation, especially because they are trained in several dental age assessment systems, other than Demirjian’s and Moorrees’ thought to undergraduate students. The latter, on the other hand, may have a reduced tendency to overanalyze borderline stages and be less critical and detailed during stage allocation. Additionally, this outcome also should be interpreted from the perspective of study’s limitations. Since the 46 observers analyzed ten radiographs previously selected based on age, only part of the dental stages that compose the two systems (Demirjian’s and Moorrees’) were sampled. This recalls the preliminary potential of the present pilot study to support future research in the field.

From an educational perspective, these findings may have implications for the teaching of dental age assessment and the structure of forensic odontology curricula. The relatively high consistency observed among undergraduate students, who had recently completed a focused module on dental age assessment, suggests that recent, structured and method-specific training may support consistent stage allocation. Conversely, the greater heterogeneity of training among graduate students highlights the potential importance of maintaining focused training and regular practice when teaching observer-dependent staging procedures. These findings should not be interpreted as evidence that undergraduate training is superior to advanced academic training, but rather as preliminary evidence that the organization, recency and specificity of training may be relevant to observer consistency. Dental education curricula may therefore benefit from standardized practical exercises using representative radiographic cases and periodic calibration exercises to reinforce consistent stage allocation. Further studies with larger samples and longitudinal educational designs are needed to determine whether specific training strategies can improve inter- and intra-observer consistency.

Regarding future research, the present pilot study was designed to assess observer consistency rather than the accuracy of dental age estimation. Consequently, high consistency rates should not be interpreted as evidence of correct stage allocation, since reproducibility and accuracy represent distinct methodological constructs. No comparisons were planned against a gold-standard observer in the present methodological setup because the focus was limited to consistency, but future investigations of accuracy and consistency should consider the inclusion of a reference comparator [28]. Comparisons with chronological age were also not conducted because this was beyond the scope of the present study; such comparisons should be considered in future research examining systematic differences between estimated and chronological ages and evaluating observers’ performance through diagnostic accuracy metrics. Additionally, the radiographic sample was intentionally restricted to an age range representative of common age categories addressed in dental age estimation literature of methods applicable to children (using developing permanent teeth). Although this approach aids on the practical relevance of the findings, broader age ranges—including younger individuals and third molar development—would enable the evaluation of additional dental age estimation systems and a supporting to investigation of a wider spectrum of developmental stages. Future studies may also benefit from adopting sampling strategies based on the frequency of developmental stages rather than chronological age bands, thereby achieving a more balanced distribution of stage categories and enabling more realistic comparisons of observer consistency across the entire developmental process. Finally, expanding the radiographic sample and incorporating observers’ repeated assessments to evaluate intra-observer consistency, while maintaining a large sample of observers, would further strengthen the methodological robustness of future investigations and provide a more comprehensive understanding of the relationship between observer consistency, staging accuracy, and the forensic applicability of dental age estimation systems.

5. CONCLUSION

The present pilot study demonstrated that Demirjian’s and Moorrees’ dental developmental staging systems produced broadly comparable levels of inter-observer consistency across observers with different academic backgrounds. Although some variability was observed according to tooth position, particularly for teeth 31 and 36, the overall reproducibility patterns remained similar between the two staging methods. These findings should be interpreted within the scope of the study, which was designed to evaluate observer consistency rather than staging accuracy or chronological age estimation performance. As such, the results provide preliminary evidence supporting the reproducibility of both systems while highlighting the need for future investigations incorporating larger and more diverse radiographic samples, broader developmental stages, reference

standards, chronological age comparisons, and accuracy metrics to better establish the forensic applicability of dental developmental staging.

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This study was financed in part by the Coordination for the Improvement of Higher Education Personnel – Brazil (CAPES) – Finance Code 001. This study was financed in part by the National Council for Scientific and Technological Development (CNPq).

Authors Contribution

R.P., M.E.C., L.S.: conception of work- data acquisition- writing original draft- writing final version-approved the final version, N.A. : design of work-data analysis- writing final version-approved the final version D.M., A.F. : conception of work- data interpretation- writing original draft- writing final version-approved the final version

The available contribution categories include: Conceptualization, Methodology, Formal Analysis, Funding Acquisition, Data Curation, Original Draft Writing, Supervision and Editing, and Administration.

Informed Consent Statement: Informed consent was obtained from the participants who were recruited as observers and allocated dental developmental stages for the two systems. Given the retrospective nature of the study, which exclusively involved the analysis of previously acquired anonymized radiographic images, ethical approval was obtained to assess and collect images from a preexisting radiological database.

Data Availability Statement: The data supporting this study's findings are available from the corresponding author, upon reasonable request.

Acknowledgments: None

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Committee of Ethics in Human Research at the Faculdade São Leopoldo Mandic (Protocol No. 95980626.1.0000.5374).

References

- [1] Dursun D, Geçer RB. Dental age estimation from panoramic radiographs: A comparison of orthodontist and ChatGPT-4 evaluations using the London Atlas, Nolla, and Haavikko methods. *Diagnostics (Basel)* 2025; 15(18): 2389. <https://doi.org/10.3390/diagnostics15182389>
- [2] Mohamed EG, Díaz Redondo RP, Koura A, EL-Mofty MS, Kayed M. Dental age estimation using deep learning: a comparative survey. *Computation* 2023; 11(2): 18. <https://doi.org/10.3390/computation11020018>
- [3] Baumann P, Widek T, Merken H, Boldt J, Petrovic A, Urschler M, et al. Dental age estimation of living persons: Comparison of MRI with OPG. *Forensic Sci Int* 2015; 253: 76-80. <https://doi.org/10.1016/j.forsciint.2015.06.001>
- [4] Franco A, Murray J, Heng D, Lygate A, Moreira D, Ferreira J, et al. Binary decisions of artificial intelligence to classify third molar development around the legal age thresholds of 14, 16 and 18 years. *Sci Rep* 2024; 14(1): 4668. <https://doi.org/10.1038/s41598-024-55497-5>
- [5] Nolla CM. The development of permanent teeth. *J Dent Child* 1960; 27(4): 254-66.
- [6] Moorrees CF, Fanning EA, Hunt EE Jr. Age variation of formation stages for ten permanent teeth. *J Dent Res* 1963; 42: 490-502.
- [7] Demirjian A, Goldstein H, Tanner JM. A new system of dental age assessment. *Hum Biol* 1973; 45(2): 211-27.
- [8] Köhler S, Schmelzle R, Loitz C, Püschel K. Development of wisdom teeth as a criterion of age determination. *Ann Anat* 1994; 176(4): 339-45.
- [9] Gleiser I, Hunt EE Jr. The permanent mandibular first molar: its calcification, eruption and decay. *Am J Phys Anthropol* 1955; 13(2): 253-83. <https://doi.org/10.1002/ajpa.1330130206>
- [10] Nicodemo RA, Moraes LC, Médiçi Filho E. Tabela cronológica da mineralização dos dentes permanentes entre brasileiros. *Rev Fac Odontol São José dos Campos* 1974; 3(1): 55-6.

- [11] AlQahtani SJ, Hector MP, Liversidge HM. Brief communication: the London atlas of human tooth development and eruption. *Am J Phys Anthropol* 2010; 142(3): 481-90. <https://doi.org/10.1002/ajpa.21258>
- [12] Esan TA, Yengopal V, Schepartz LA. The Demirjian versus the Willems method for dental age estimation in different populations: A meta-analysis of published studies. *PLoS One* 2017; 12(11): e0186682. <https://doi.org/10.1371/journal.pone.0186682>
- [13] Boedi RM, Angelakopoulos N, Franco A, Polukhin N, Kumagai A, Galic I, et al. Evaluation of Demirjian staging method reproducibility and reliability for dental age estimation between eight forensic experts. *Int J Legal Med* 2025; 139(6): 2883-90. <https://doi.org/10.1007/s00414-025-03560-9>
- [14] Franco A, Moreira DD, Cidade R, Machado M, Bueno J, Malschitzky C, Boedi RM. The Brazilian (FRANCO) method for dental age estimation: Willems' model revisited. *Clin Oral Investig* 2024; 28(9): 495. <https://doi.org/10.1007/s00784-024-05869-y>
- [15] American Dental Association. ADA technical report no. 1077 for human age assessment by dental analysis; American Board of Forensic Odontology standards and guidelines for dental age assessment. Chicago: American Dental Association; 2020. Available from: <https://abfo.org>
- [16] Enhancing the Quality and Transparency of Health Research. EQUATOR network. Available at: <https://www.equator-network.org/>
- [17] World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. *JAMA*. 2025;333(1):71–74. doi: 10.1001/jama.2024.21972
- [18] Banar N, Bertels J, Laurent F, Boedi RM, De Tobel J, Thevissen P, et al. Towards fully automated third molar development staging in panoramic radiographs. *Int J Legal Med* 2020; 134(5): 1831-41. <https://doi.org/10.1007/s00414-020-02283-3>
- [19] Shi Y, Ye Z, Guo J, Tang Y, Dong W, Dai J, et al. Deep learning methods for fully automated dental age estimation on orthopantomograms. *Clin Oral Investig* 2024; 28(3): 198. <https://doi.org/10.1007/s00784-024-05598-2>
- [20] Mohammad N, Ahmad R, Gaus MHA, Kurniawan A, Yusof MYPM. Accuracy of automated forensic dental age estimation lab (F-DentEst Lab) on large Malaysian dataset. *Forensic Sci Int* 2024; 361: 112150. <https://doi.org/10.1016/j.forsciint.2024.112150>
- [21] Akshaya M, Vijayalakshmi KR, Kumar TNR, Rakshana M. Automated dental age estimation model for panoramic radiographs. *J Forensic Leg Med* 2026; 120: 103148. <https://doi.org/10.1016/j.jflm.2026.103148>
- [22] Demirjian A, Goldstein H. New systems for dental maturity based on seven and four teeth. *Ann Hum Biol* 1976; 3(5): 411-21. <https://doi.org/10.1080/03014467600001671>
- [23] Angelakopoulos N, Boedi RM, De Luca S, et al. Does increased staging complexity influence expert agreement in dental age estimation?. *Int J Legal Med* (2026). <https://doi.org/10.1007/s00414-026-03946-3>
- [24] Perschbacher S. Interpretation of panoramic radiographs. *Aust Dent J* 2012; 57(Suppl 1): 40-5. <https://doi.org/10.1111/j.1834-7819.2011.01655.x>
- [25] Patano A, Malcangi G, Inchingolo AD, Garofoli G, De Leonardis N, Azzollini D, et al. Mandibular crowding: Diagnosis and management—A scoping review. *J Pers Med* 2023; 13(5): 774. <https://doi.org/10.3390/jpm13050774>
- [26] Franco A, Vetter F, Coimbra EF, Fernandes Â, Thevissen P. Comparing third molar root development staging in panoramic radiography, extracted teeth, and cone beam computed tomography. *Int J Legal Med* 2020; 134(1): 347-53. <https://doi.org/10.1007/s00414-019-02206-x>
- [27] Valente RPA, Ribeiro JMC, Sakamoto SPS, Rodrigues R, Franco A, Paranhos LR. Survey on the dental age estimation practice among Brazilian experts. *Observ Latinoam Econ* 2024; 22(3): e3806. <https://doi.org/10.55905/oelv22n3-138>
- [28] Franceschetti L, Merelli VG, Corona S, Magli F, Maggioni L, Cummaudo M, et al. Analysis of interrater reliability in age assessment of minors: how does expertise influence the evaluation? *Int J Legal Med* 2022; 136(1): 279-285. <https://doi.org/10.1007/s00414-021-02707-8>