Joint Diseases and Related Surgery

Jt Dis Relat Surg
2026;37(3):714-722

ORIGINAL ARTICLE

Professional experience and reliability of cervical vertebral maturation staging among unfamiliar raters

Yiğit Umur Cırdı, MD[1], Alp Er Tunga Bölükbaşı, MD[2], Mehmet Oğuz Çolak, MD[1], Enes Kanay, MD[2], Danyal Gümüş, MD[2], Mehmet Demirel, MD[3], Okan Özkunt, MD[4], Kerim Sarıyılmaz, MD[1]

[1]Department of Orthopedics and Traumatology, Acıbadem Mehmet Ali Aydınlar University, İstanbul, Türkiye
[2]Department of Orthopedics and Traumatology, Acıbadem Ataşehir Hospital, İstanbul, Türkiye
[3]Department of Orthopedics and Traumatology, İstanbul University, İstanbul Faculty of Medicine, İstanbul, Türkiye
[4]Department of Orthopedics and Traumatology, Biruni University, İstanbul, Türkiye

ABSTRACT

Objectives: This study aims to evaluate the reliability of cervical vertebral maturation (CVM) staging in determining skeletal maturity in patients with adolescent idiopathic scoliosis (AIS), focusing on the effect of professional experience among orthopedic surgeons.

Patients and methods: Between January 2023 and January 2025, this reliability and agreement study included a total of 72 patients aged between nine and 16 years who were diagnosed with AIS. Intra- and inter-rater reliability were analyzed using weighted kappa (κ) statistics with standard errors (SEs) and 95% confidence intervals (CIs). A total of 72 full-spine lateral radiographs were independently assessed by 10 raters with varying levels of orthopedic expertise, including medical students, residents, consultants, and spinal surgeons. None of the raters had prior experience with the CVM classification. Each rater performed CVM staging twice with a four-week interval between assessments.

Results: Of the patients, 33 were male and 39 were female with a mean age of 13.07 ± 1.97 (range, 9 to 16) years. Overall intra-rater reliability was substantial (κ = 0.64, SE = 0.032, 95% CI: 0.569-0.714), and overall inter-rater reliability was moderate (κ = 0.52). Although spinal surgeons showed slightly higher intra-rater reliability overall, inter-rater reliability declined between assessments (first assessment κ = 0.61 *vs*. second assessment κ = 0.34). Consultants exhibited the highest inter-rater agreement (κ = 0.62), whereas residents showed improvement between assessments (κ = 0.40 *vs*. 0.52); however, no significant difference observed between groups.

Conclusion: The reliability of CVM staging for assessing skeletal maturity in AIS patients does not significantly differ across levels of orthopedic expertise. While CVM staging has potential as a radiation-sparing method, the overall moderate agreement, particularly among physicians without prior experience, indicates that reliability is not significantly influenced by professional experience.

Keywords: Adolescent idiopathic scoliosis, cervical vertebral maturation, classification, diagnostics, skeletal maturity.

Adolescent idiopathic scoliosis (AIS) is defined as more than 10° curvature in spine in the coronal plane and particularly manifests during early adolescence.[1] The initial diagnosis is typically made between 10 and 16 years of age, with a reported prevalence of 1 to 3% in this population.[2] Due to dynamic nature of AIS, recognition of skeletal growth potential is crucial to predict which cases are likely to progress. Rapid osteochondral growth accelerates progression of deformities, eventually causing symptoms.[3,4]

The estimation of remaining growth velocity remains as the cornerstone for an individualized treatment plan. Despite conservative treatment achieving the success rates up to 75%, accurate estimation of growth potential is critical for both,

Received: December 17, 2025
Accepted: March 10, 2026
Published online: July 24, 2026

Correspondence: Enes Kanay, MD. Acıbadem Ataşehir Hastanesi, Ortopedi ve Travmatoloji Bölümü, 34758 Ataşehir, İstanbul, Türkiye.

E-mail: eneskanay@gmail.com

Doi: 10.52312/jdrs.2026.2784

Citation: Cırdı YU, Bölükbaşı AET, Çolak MO, Kanay E, Gümüş D, Demirel M, et al. Professional experience and reliability of cervical vertebral maturation staging among unfamiliar raters. Jt Dis Relat Surg 2026;37(3):714-722. doi: 10.52312/jdrs.2026.2784.

for opting between surgical and non-surgical management, and for optimizing the timing and treatment modality.[5] Chronological age alone is an insufficient indicator due to wide inter-individual variation in biological growth, highlighting the importance of reliable radiological parameters.[6,7]

Several methods have been established to assess the skeletal maturity in patients with AIS, including Greulich-Pyle and Tanner-Whitehouse methods utilize wrist radiographs evaluated by atlas comparison; the Risser sign references iliac apophysis growth, distal radius and ulna classification stages the ossification centers, and the epiphyseal to metaphyseal ratio of the proximal tibia assesses physeal development. All these approaches require additional imaging, thereby exposing patients to radiation.[8-10] Later, the cervical vertebral maturation (CVM) method introduced by Lamparski[11] and further refined by Bacetti et al.[12] demonstrated its importance in orthodontics for optimizing treatment timing.[13] The CVM method is based on predictable chronological changes of the cervical vertebrae morphology; wherein the flat inferior border of C2 in the prepubertal stage becomes progressively concave over time. Similarly, the vertebral bodies of C3 and C4 are typically trapezoidal in the prepubertal stage; with growth, they transform into a rectangular shape with concave inferior borders, gradually increases in height to adopt a square appearance, and finally evolves into a vertical rectangle. These developmental transformations can be staged into six subgroups, providing a structural framework for assessing skeletal maturity. In patients with AIS, CVM offers a practical advantage, wherein the skeletal maturity can be assessed using whole-spine lateral radiographs that are already obtained for diagnostic purposes, thereby avoiding further radiation exposure. Ronkers et al.[14] reported that a scoliosis patient was exposed to an average of 22.9 X-rays during treatment which led to a significantly higher rate of chest neoplasia. Thus, CVM may be a viable radiation-free alternative to other methods for assessing skeletal maturity.

On the other hand, the utility of CVM in practice depends on its reliability. A valid classification tool must not only reflect actual growth rate but also be highly reproducible. In clinical settings, intra- and inter-rater agreement are crucial, as they guide and potentially alter clinical decisions. Recent literature has raised concerns regarding the limited reliability of CVM as a skeletal maturity index.[15-17] However, many studies suffer from methodological limitations, particularly in assessing reproducibility and diagnostic accuracy.[18] The reported reliability of CVM is variable, and its dependence on professional expertise has not been investigated.

In the present study, we hypothesized that professional experience could have minimal impact on reliability of the classification, particularly among the raters without prior exposure or routine use of the method. We, therefore, aimed to measure the reliability of the CVM method as a tool for evaluating skeletal maturity in patients with AIS, with particular emphasis on the influence of professional experience in the field.

PATIENTS AND METHODS

This reliability and agreement study was conducted at Acıbadem Mehmet Ali Aydınlar University, Department of Orthopedics and Traumatology between January 2023 and January 2025. Patients aged between nine and 16 years who were diagnosed with AIS were included. Inclusion criteria were as follows: diagnosis of AIS with cobb angle more than 10° and proper view of anteroposterior (AP) and lateral whole spine X-ray in standing position. Patients with a previous history of spinal and pelvic surgery, those with concomitant growth-related disorder, those with a history of infection, trauma or tumor, and improperly acquired cervical images were excluded from the study. Among 300 images, consecutive radiographs were screened sequentially until the predetermined sample size was reached. A total of 82 were identified as candidates. Of these, 10 were subsequently excluded based on predefined exclusion criteria (n = 2 previous spinal surgery, n = 7 inadequate quality of diagnostic imaging, and n = 1 receiving treatment due to precocious puberty) (Figure 1). In total, 72 blinded full-spine lateral radiographs were evaluated by 10 independent raters with varying levels of expertise, none of whom had prior experience with CVM staging. In accordance with Obuchowski[19] recommendation that at least 10 raters are required to ensure a reliable assessment, the specifications of raters in this study are summarized in Table I.

Radiological data were retrieved from the institutional Picture Archiving and Communication Systems (PACS) system. All personal data were anonymized while importing radiographs to protect the participants' privacy and confidentiality. No personal information was disclosed, and all digital data used in this study. A written informed consent was obtained from the parents and/or

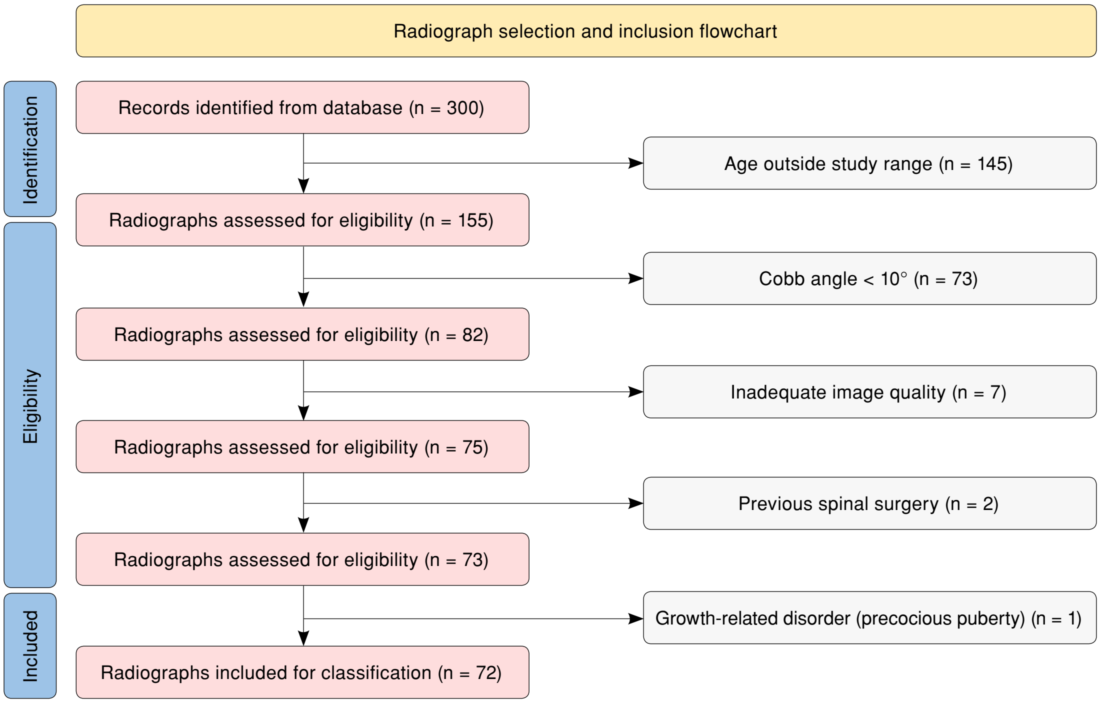

FIGURE 1. Study flowchart.

legal guardians of the patients for all diagnostic and therapeutic procedures. The study protocol was approved by the Acıbadem Mehmet Ali Aydınlar University Institutional Review Board (Date: 26.06.2025, No.: 2025-10/401). The study was conducted in accordance with the principles of the Declaration of Helsinki. The Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were followed.

A blinded researcher who conducted literature research and designed the study also prepared a PowerPoint (Microsoft® Redmond, WA, USA) slide consisting of whole spine X-rays and cropped lateral cervical images of the patients. To avoid bias and reliability inflation, this researcher did not participate in evaluations. Prior to classification, all raters were given the CVM method user guide by McNamara from key reference article and were given sufficient time to independently review the distinguishing morphological features of each cervical stage, until they felt confident in differentiating between stages.[20,21] No additional workshops or calibration sessions were conducted to improve classification skills, as the study was intentionally designed to simulate realistic clinical settings in which clinicians encounter a method in literature, review the related article, and apply it independently in practice. This eliminated the influence of prior familiarity with the method and isolated the effect of professional experience on reliability of the classification system. Subsequently,

TABLE I
Demographic and professional characteristics of the raters who participated in the study

Rater group	Rater ID	Experience years	Prior experience for CVM	Independent assessment
Medical students	R-1, R-2	0	None	Yes
Residents	R-3, R-4	2-4 in residency	None	Yes
Consultants	R-5, R-6	< 3	None	Yes
Senior consultants	R-7, R-8	5-7	None	Yes
Spinal surgeons	R-9, R-10	≥ 15	None	Yes

CVM, cervical vertebral maturation.

staging was conducted by examining the sequential morphological alterations of the C2, C3, and C4 vertebrae, including evaluation of inferior border concavity and the transformation of vertebral body shape, in accordance with the provided user guide and the assigned CVM stages were recorded (Figure 2). After a four-week interval, slides were shuffled using the RANDOM function in Excel (Microsoft®, Redmond, WA, USA), and the grading process was repeated.

Statistical analysis

Statistical analysis was performed using the R software version 4.5.1 software (R Foundation, Vienna, Austria). Descriptive data were expressed in mean ± standard deviation (SD), median (min-max) or number and frequency, where applicable.

Sample size determination was based on reliability specific methodology rather than conventional hypothesis testing power analysis. According to the simulation-based recommendations of Cicchetti,[22] for an ordinal classification system with k categories, the minimum required sample size for valid weighted kappa (κ) analysis is approximately $2k^2$. Therefore, for a six-category classification system, a minimum of 72 cases is required, and accordingly 72 radiographs were included in the present study. Weighted Cohen's κ values (linear) were calculated to assess intra-rater reliability. Inter-rater reliability was assessed via weighted κ values (linear), and average values were calculated by rater pairs (Light's κ approach). Kappa values less than 0.20 indicated slight agreement; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; and greater than 0.80, very good agreement.[23] Kappa coefficients were compared across professional experience groups using the Kruskal-Wallis test. A p value of < 0.05 was considered statistically significant with 95% confidence interval (CI).

RESULTS

Of the patients, 33 were male and 39 were female with a mean age of 13.07 ± 1.97 (range, 9 to 16) years.

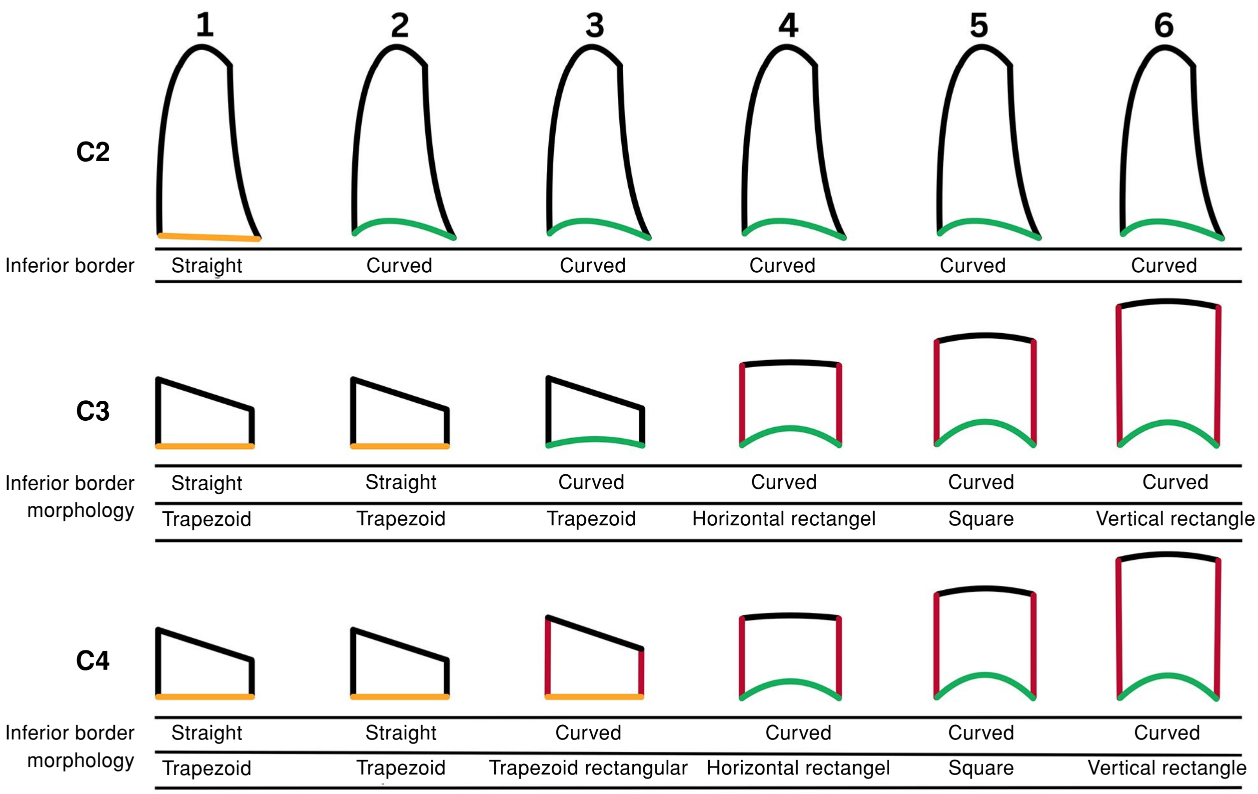

FIGURE 2. A visual diagram of the chronological morphological changes of the C2, C3, and C4 vertebrae used for cervical vertebral maturation staging. Inferior border morphology (straight versus concave) and vertebral body shape transformation (trapezoid, horizontal rectangle, square, and vertical rectangle) are demonstrated. Different colors are employed to represent different structural characteristics.

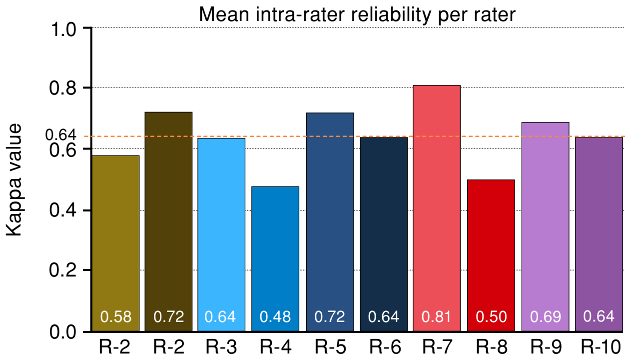

FIGURE 3. Intra-rater reliability per observer, expressed as linear weighted kappa (κ) values. Each bar represents an individual rater (Obs1-Obs10), with the dotted line indicating the overall mean intra-rater reliability (κ = 0.64).

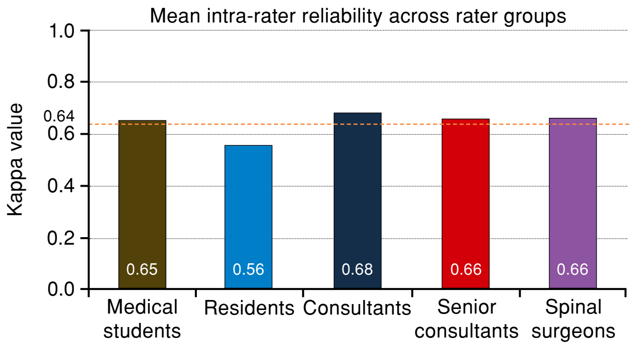

FIGURE 4. Mean intra-rater reliability across observer groups, expressed as linear weighted kappa (κ) values. Bars represent group averages; the dotted line indicates the overall mean intra-rater reliability (κ = 0.64).

Intra-rater reliability

Raters demonstrated overall substantial agreement with a weighted linear κ of 0.64 (standard error [SE] = 0.03, 95% CI: 0.57-0.71). Individual intra-rater κ values were as follows: R-1 (medical student) = 0.58, R-2 (medical student) = 0.72, R-3 (resident) = 0.64, R-4 (resident) = 0.48, R-5 (consultant) = 0.72, R-6 (consultant) = 0.64, R-7 (senior consultant) = 0.81, R-8 (senior consultant) = 0.50, R-9 (spinal surgeon) = 0.69, and R-10 (spinal surgeon) = 0.64 (Figure 3). When averaged within each experience group, mean intra-rater κ values were 0.65 for medical students, 0.56 for residents, 0.68 for consultants, 0.66 for senior consultants, and 0.66 for spinal surgeons, indicating that consultants achieved the highest intra-rater κ values, whereas residents showed the lowest between their two assessments (Figure 4). The Kruskal-Wallis test comparing the distribution of κ coefficients across the five professional experience groups revealed no statistically significant differences across levels of professional expertise ($p = 0.60$).

Inter-rater reliability

The overall inter-rater reliability was moderate (κ = 0.52), with a mean κ value of 0.51 for the first assessment and 0.52 for the second. At the group level, consultants (R-5 *vs.* R-6) showed the highest agreement (κ = 0.63 at the first assessment and κ = 0.59 at the second), followed by medical students (R-1 *vs.* R-2: κ = 0.58 and 0.60). Residents (R-3 *vs.* R-4) showed the lowest agreement at the

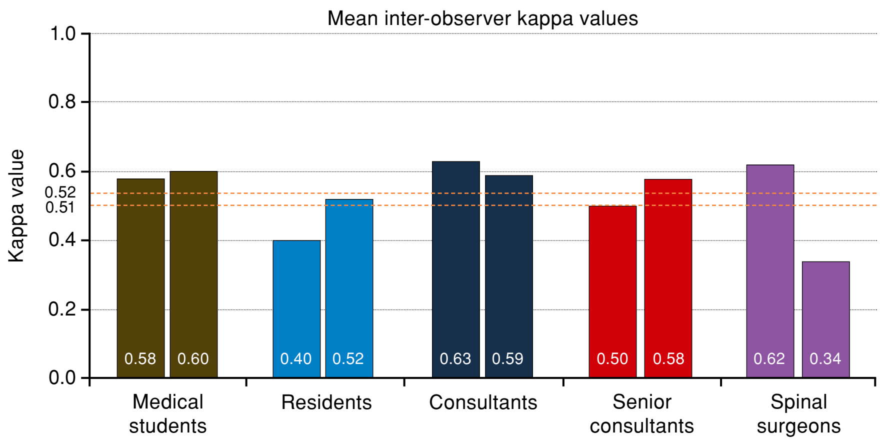

FIGURE 5. Demonstration of inter-rater reliability by groups. Mean inter-rater reliability across observer groups is expressed as linear weighted kappa (κ) values for the first (dashed line, κ = 0.51) and second (solid line, κ = 0.52) assessments. Bars indicate group averages, with the first bar representing the first assessment and the second bar representing the second assessment.

first assessment but improved markedly by the second round (κ = 0.40 → 0.52). Spinal surgeons (R-9 *vs.* R-10) had the second-highest agreement among all groups at the first assessment (κ = 0.62) but displayed a marked drop by the second (κ = 0.34). The Kruskal-Wallis test showed no statistically significant difference across groups (p = 0.49) (Figure 5).

DISCUSSION

In the present study, we evaluated the reliability of CVM staging in determining skeletal maturity in patients with AIS. The main finding of the study was that the overall intra-rater agreement was substantial (κ = 0.64, SE = 0.03, 95% CI: 0.57-0.71), while overall inter-rater reliability was moderate (κ = 0.52) for CVM staging with no statistically significant difference observed across levels of professional experience. Consultants demonstrated the highest overall intra- and inter-rater agreement; however, no difference were observed between groups. Performance among spinal surgeons significantly dropped between assessments (first assessment κ = 0.61 *vs.* second assessment κ = 0.34).

Although the CVM method has been widely used by orthodontists to assess skeletal maturation, its adoption by orthopedists remains limited, despite recent meta-analyses demonstrating strong correlation with established skeletal maturity indicators such as the Risser staging system, Sanders simplified maturity system, and hand wrist maturation methods.[24-30] Additionally, a recent study has reported that CVM staging can successfully predict curve progression, with Stage 3 representing peak growth velocity and Stage 6 indicating the end of growth.[31] Taken together, CVM has been shown to reflect skeletal maturity comparably to these commonly used methods and offers a radiation-free alternative. Nevertheless, the clinical value of a staging system strongly dependent on the reliability of the staging process. A maturity indicator can only be meaningful, if it can be applied consistently and reproducibly. From an orthopedic perspective, dosage and rotational issues in radiographic imaging are frequently encountered in clinical practice. McNamara and Franchi[20] reported that up to 5% of lateral cephalograms were unsuitable for CVM staging due to both biological variability and radiological errors. Furthermore, skeletal growth is inherently dynamic and not always predictable, indicating that maturation does not always follow expected timelines.

Whether professional experience can mitigate these limitations and enhance the reliability of CVM staging is unclear. In general, experience is expected to have a positive effect on the reliability of classification systems.[32-34] McNamara and Franchi[20] reported that reliability of the CVM method improved with clinical experience, but no supporting data were provided. In the current study, spinal surgeons showed slightly better κ values; however, these differences did not reach statistical significance across experience levels. This finding should be interpreted with caution and does not denote that professional experience has no effect on CVM reliability. Rather, it demonstrates that within the current methodological framework, sample size, and rater composition, no measurable difference attributable to professional experience was detected. Despite the possible experience-related effects, professional seniority alone appears insufficient to mitigate the inherent subjectivity of CVM staging. Additionally, the prominent drop in inter-rater reliability between the first and second assessments, particularly among spinal surgeons, suggests that professional seniority does not necessarily translate into greater reproducibility.

Previous studies on CVM staging have shown very high intra- and inter-rater agreement, reaching as high as 98%, depending on the statistical method applied.[25,27,35-37] However, many relied solely on the intraclass correlation coefficient (ICC) or raw percentage agreement without accounting for chance agreement, which reflects consistency based on the similarity rather than categorical agreement.[16,25,30,38,39] Accordingly, recent meta-analyses have highlighted this limitation and advocated the use of chance-corrected agreement coefficients (Cohen's κ, Fleiss' κ, weighted κ) for categorical data, as they can account for chance agreement and specifically assess agreement rather than association.[40]

Despite this previously reported high agreement, recent meta-analyses have emphasized that most studies suffer from methodological flaws and inappropriate statistical approaches.[17,18] Predko-Engel et al.[16] highlighted that studies reporting high reliability often recruit raters with a high level of expertise on the subject, which may inflate κ values. Likewise, Gabriel et al.,[15] reported that the lead authors also served as raters, indicating a research-level familiarity of CVM staging, which likely inflated reliability scores. To mitigate bias in our study, the individual who prepared the presentation slides and the lead author were not involved in the scoring process. This methodological separation strengthens the objectivity of the evaluation and supports that the reliability outcomes reflect the accurate reproducibility of the classification in a realistic clinical setting. Notably, our study demonstrated lower κ values compared to literature.

Gabriel et al.,[15] reported moderate to substantial agreement (weighted κ = 0.40-0.80) and a mean intra-rater agreement among 10 raters of just 62%, even after comprehensive training. Approximately one-quarter of raters' assessments differed by two or more CVM stages, indicating a key practical issue: the difficulty of sharply distinguishing adjacent stages, which may explain the absence of improvement in reliability with increasing professional experience among raters unfamiliar with the classification. Rainey et al.,[41] conducted a study involving 20 orthodontists who were unfamiliar with the classification but trained pre-staging, whose intra-rater reliability improved from 0.70 to 0.82 when standardized images were used. In our study, cervical vertebrae were assessed using full-spine radiographs used in routine clinical practice, which may have contributed to the relatively low agreement. However, Rainey et al.'s[41] raters consisted of both residents and specialists, and no information was provided regarding the potential relationship between professional experience and reliability.

Similarly, previous studies that recruited two spine surgeons in one and a radiologist and three orthodontists in another reported excellent and substantial agreement, respectively. Interestingly, the least experienced raters showed better κ values, though investigating the effect of professional experience was not the primary objective of either study.[25,42] Furthermore, Rongo et al.,[43] classified orthodontist raters as junior, postgraduate, and specialist and reported the highest κ value in the least experienced group. In agreement with our results, increased professional seniority did not improve reliability; moreover, the inter-rater agreement of spinal surgeons significantly decreased. We attribute this difference to the fact that less experienced raters tend to adhere more strictly to the classification guidelines without introducing clinical biases derived from routine practice.

Consequently, most studies involved raters with varying levels of expertise primarily to increase rater count rather than to evaluate the effect of experience on reliability. Predko-Engel et al.,[16] defined "experienced" raters by routine use of the method rather than years in practice. They reported moderate intra-rater (κ = 0.44) and fair inter-rater reliability (κ = 0.28), and although experienced raters had significantly greater ICC values, their self-agreement rates were nearly similar to that of the inexperienced group (54.7% *vs.* 55.4%), highlighting the limited reproducibility of CVM staging for clinical use. Given that all observers in our study were unfamiliar with the CVM method, our cohort is most comparable to the inexperienced group in the aforementioned study. Taken together, neither clinical seniority nor familiarity with the classification alone promised sufficient reproducibility for CVM staging to be a reliable clinical tool, which is consistent with our findings and suggests that the CVM method should not be used as a clinical decision-making tool solely on the basis of professional experience. From the clinical point of view, moderate inter-rater reliability indicates that reliance on CVM staging alone may introduce variability in growth-dependent clinical decisions.

Nonetheless, the present study has several limitations that should be acknowledged. First, excessive deterioration in coronal balance can cause the lower margin of the cervical vertebra to appear distorted in clinical settings. The reliability of this classification in X-rays in advanced scoliosis cases should be confirmed. Second, skeletal maturation indices are measured mostly with the appendicular skeleton, but CVM staging evaluates the vertebra with the axial skeleton, which seems to be an advantage but is more subjective and depends on other variables such as the quality of the image. This may also explain the limited κ value in the present study.

In conclusion, our study results showed that the reliability of CVM staging did not differ across different levels of orthopedic expertise and demonstrated substantial intra-rater and moderate inter-rater reliability among raters unfamiliar with method. As a promising radiation-sparing classification tool, it can become more clinically applicable via structured training and calibration session; however, combining it with other skeletal maturation assessment tools is recommended. Future studies should investigate the impact of different medical specialties and structured training programs on reliability and their influence on clinical decision-making.

Data Sharing Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions: Y.U.C.: Idea/concept; Y.U.C., M.D.: Design, writing the article; K.S., O.Ö.: Control/supervision; A.E.T.B., D.G., E.K., M.O.Ç., M.D., K.S., O.Ö.: Data collection and/or processing; M.D., O.Ö.: Analysis and/or interpretation; Y.U.C., M.D., K.S., M.O.Ç.: Literature review; O.Ö., K.S.: Critical review; Y.U.C., M.O.Ç.: References and fundings; M.O.Ç.: Materials. All authors approve of the content of the manuscript and agree to be held accountable for the work.

Conflict of Interest: The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding: The authors received no financial support for the research and/or authorship of this article.

AI Disclosure: The authors declare that artificial intelligence (AI) tools were not used, or were used solely for language editing, and had no role in data analysis, interpretation, or the formulation of conclusions. All scientific content, data interpretation, and conclusions are the sole responsibility of the authors. The authors further confirm that AI tools were not used to generate, fabricate, or 'hallucinate' references, and that all references have been carefully verified for accuracy.

REFERENCES

1. Lenz M, Oikonomidis S, Harland A, Fürnstahl P, Farshad M, Bredow J, et al. Scoliosis and prognosis-a systematic review regarding patient-specific and radiological predictive factors for curve progression. Eur Spine J 2021;30:1813-22. doi: 10.1007/s00586-021-06817-0.
2. Weinstein SL, Dolan LA, Cheng JC, Danielsson A, Morcuende JA. Adolescent idiopathic scoliosis. Lancet 2008;371:1527-37. doi: 10.1016/S0140-6736(08)60658-3.
3. Charles YP, Daures JP, de Rosa V, Diméglio A. Progression risk of idiopathic juvenile scoliosis during pubertal growth. Spine (Phila Pa 1976) 2006;31:1933-42. doi: 10.1097/01.brs.0000229230.68870.97.
4. Tan KJ, Moe MM, Vaithinathan R, Wong HK. Curve progression in idiopathic scoliosis: Follow-up study to skeletal maturity. Spine (Phila Pa 1976) 2009;34:697-700. doi: 10.1097/BRS.0b013e31819c9431.
5. Weinstein SL, Dolan LA, Wright JG, Dobbs MB. Effects of bracing in adolescents with idiopathic scoliosis. N Engl J Med 2013;369:1512-21. doi: 10.1056/NEJMoa1307337.
6. Manzetti M, Ruffilli A, Barile F, Viroli G, Traversari M, Vita F, et al. Is there a skeletal age index that can predict accurate curve progression in adolescent idiopathic scoliosis? A systematic review. Pediatr Radiol 2024;54:299-315. doi: 10.1007/s00247-023-05834-z.
7. Cheung JPY, Cheung PWH, Samartzis D, Luk KD. Curve progression in adolescent idiopathic scoliosis does not match skeletal growth. Clin Orthop Relat Res 2018;476:429-36. doi: 10.1007/s11999.0000000000000027.
8. Gerdhem P. CORR Insights®: Curve progression in adolescent idiopathic scoliosis does not match skeletal growth. Clin Orthop Relat Res 2018;476:437-8. doi: 10.1007/s11999.0000000000000128.
9. Huang LF, Furdock RJ, Uli N, Liu RW. Estimating skeletal maturity using wrist radiographs during preadolescence: The epiphyseal: Metaphyseal ratio. J Pediatr Orthop 2022;42:e801-5. doi: 10.1097/BPO.0000000000002174.
10. Cheung JP, Cheung PW, Samartzis D, Cheung KM, Luk KD. The use of the distal radius and ulna classification for the prediction of growth: peak growth spurt and growth cessation. Bone Joint J 2016;98-B:1689-96. doi: 10.1302/0301-620X.98B12.BJJ-2016-0158.R1.
11. Lamparski DG. Skeletal age assessment utilizing cervical vertebrae. Am J Orthod 1975;67:458-9.
12. Baccetti T, Franchi L, McNamara JA. The cervical vertebral maturation (CVM) method for the assessment of optimal treatment timing in dentofacial orthopedics. Semin Orthod 2005;11:119-29.
13. Kahan JB, Li DT, Schneble CA, Elabd A, Attia E, Esparza R, et al. Measuring proximal tibial metaphyseal width and lateral epiphyseal height improves determination of remaining growth in children. J Pediatr Orthop 2021;41:e739-44. doi: 10.1097/BPO.0000000000001853.
14. Ronckers CM, Land CE, Miller JS, Stovall M, Lonstein JE, Doody MM. Cancer mortality among women frequently exposed to radiographic examinations for spinal disorders. Radiat Res 2010;174:83-90. doi: 10.1667/RR2022.1.
15. Gabriel DB, Southard KA, Qian F, Marshall SD, Franciscus RG, Southard TE. Cervical vertebrae maturation method: Poor reproducibility. Am J Orthod Dentofacial Orthop 2009;136:478.e1-7. doi: 10.1016/j.ajodo.2007.08.028.
16. Predko-Engel A, Kaminek M, Langova K, Kowalski P, Fudalej PS. Reliability of the Cervical Vertebrae Maturation (CVM) method. Bratisl Lek Listy 2015;116:222-6. doi: 10.4149/bll_2015_043.
17. Sohrabi A, Babay Ahari S, Moslemzadeh H, Rafighi A, Aghazadeh Z. The reliability of clinical decisions based on the cervical vertebrae maturation staging method. Eur J Orthod 2016;38:8-12. doi: 10.1093/ejo/cjv030.
18. Santiago RC, de Miranda Costa LF, Vitral RW, Fraga MR, Bolognese AM, Maia LC. Cervical vertebral maturation as a biologic indicator of skeletal maturity. Angle Orthod 2012;82:1123-31. doi: 10.2319/103111-673.1.
19. Obuchowski NA. How many observers are needed in clinical studies of medical imaging? AJR Am J Roentgenol 2004;182:867-9. doi: 10.2214/ajr.182.4.1820867.
20. McNamara JA Jr, Franchi L. The cervical vertebral maturation method: A user's guide. Angle Orthod 2018;88:133-43. doi: 10.2319/111517-787.1.
21. Baccetti T, Franchi L, McNamara JA Jr. An improved version of the Cervical Vertebral Maturation (CVM) method for the assessment of mandibular growth. Angle Orthod 2002;72:316-23. doi: 10.1043/0003-3219(2002)072<0316:AIVOTC>2.0.CO;2.
22. Cicchetti DV. Testing the normal approximation and minimal sample size requirements of weighted kappa when the number of categories is large. Appl Psychol Meas 1981;5:101-4.
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
24. Ferrillo M, Curci C, Roccuzzo A, Migliario M, Invernizzi M, de Sire A. Reliability of cervical vertebral maturation compared to hand-wrist for skeletal maturation assessment in growing subjects: A systematic review. J Back Musculoskelet Rehabil 2021;34:925-36. doi: 10.3233/BMR-210003.

25. Qi H, Tian X, Zu F, Zhao Z, Wang C, Wang C, et al. Evaluating the feasibility of cervical vertebral maturation stages as an alternative to risser sign in assessing skeletal maturity in adolescent idiopathic scoliosis patients. World Neurosurg 2024;192:e523-32. doi: 10.1016/j.wneu.2024.10.014.
26. Soegiharto BM, Moles DR, Cunningham SJ. Discriminatory ability of the skeletal maturation index and the cervical vertebrae maturation index in detecting peak pubertal growth in Indonesian and white subjects with receiver operating characteristics analysis. Am J Orthod Dentofacial Orthop 2008;134:227-37. doi: 10.1016/j.ajodo.2006.09.062.
27. Uysal T, Ramoglu SI, Basciftci FA, Sari Z. Chronologic age and skeletal maturation of the cervical vertebrae and hand-wrist: Is there a relationship? Am J Orthod Dentofacial Orthop 2006;130:622-8. doi: 10.1016/j.ajodo.2005.01.031.
28. Gandini P, Mancini M, Andreani F. A comparison of hand-wrist bone and cervical vertebral analyses in measuring skeletal maturation. Angle Orthod 2006;76:984-9. doi: 10.2319/070605-217.
29. Szemraj A, Wojtaszek-Słomińska A, Racka-Pilszak B. Is the Cervical Vertebral Maturation (CVM) method effective enough to replace the Hand-Wrist Maturation (HWM) method in determining skeletal maturation?-A systematic review. Eur J Radiol 2018;102:125-8. doi: 10.1016/j.ejrad.2018.03.012.
30. Yucekul A, Yilgor C, Demirci N, Gurel IE, Orhun O, Karaman MI, et al. A comparative analysis of axial and appendicular skeletal maturity staging systems through assessment of longitudinal growth and curve modulation after VBT surgery. Eur Spine J 2025;34:251-62. doi: 10.1007/s00586-024-08488-z.
31. Cheung STY, Cheung GCC, Cheung JPY, Cheung PWH. The utility of cervical vertebral maturation method for staging skeletal growth and curve progression in patients with adolescent idiopathic scoliosis. JB JS Open Access 2025;10:e25.00184. doi: 10.2106/JBJS.OA.25.00184.
32. Yıldırım C, Muratoğlu OG, Turan K, Ergün T, Mısır A, Aydın M. The intra- and interobserver reliability of five commonly used intertrochanteric femur fracture classification systems. Jt Dis Relat Surg 2022;33:187-92. doi: 10.52312/jdrs.2022.498.
33. Demirel M, Yenigün MY, Mert L, Kendirci AŞ, Yağci TF, Demir TB, et al. Intra- and inter-observer reliability of Dias-Tachdjian classification in pediatric ankle fractures: Do clinical experience and expertise matter? J Pediatr Orthop B 2024;33:240-5. doi: 10.1097/BPB.0000000000001097.
34. Prevezas N, Antypas G, Louverdis D, Konstas A, Papasotiriou A, Sbonias G. Proposed guidelines for increasing the reliability and validity of Letournel classification system. Injury 2009;40:1098-103. doi: 10.1016/j.injury.2009.06.005.
35. Alkhal HA, Wong RW, Rabie AB. Correlation between chronological age, cervical vertebral maturation and Fishman’s skeletal maturity indicators in southern Chinese. Angle Orthod 2008;78:591-6. doi: 10.2319/0003-3219(2008)078[0591:CBCACV]2.0.CO;2.
36. Flores-Mir C, Burgess CA, Champney M, Jensen RJ, Pitcher MR, Major PW. Correlation of skeletal maturation stages determined by cervical vertebrae and hand-wrist evaluations. Angle Orthod 2006;76:1-5. doi: 10.1043/0003-3219(2006)076[0001:COSMSD]2.0.CO;2.
37. Zhang Y, Shu S, Gu Q, Liu Z, Zhu Z, Qiu Y, et al. Cervical Vertebral Maturation (CVM) stage as a supplementary indicator for the assessment of Peak Height Velocity (PHV) in Adolescent Idiopathic Scoliosis (AIS). Quant Imaging Med Surg 2020;10:96-105. doi: 10.21037/qims.2019.11.07.
38. Cunha AC, Cevidanes LH, Sant'Anna EF, Guedes FR, Luiz RR, McNamara JA, et al. Staging hand-wrist and cervical vertebrae images: A comparison of reproducibility. Dentomaxillofac Radiol 2018;47:20170301. doi: 10.1259/dmfr.20170301.
39. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96-106. doi: 10.1016/j.jclinepi.2010.03.002.
40. Hussain U, Ul Hassan F, Kamran MA, Alnazeh AA, Ullah I, Alam S. Inter-observer and intra-observer agreement of cervical vertebral maturation staging: A systematic review and meta-analysis. Int Orthod 2024;22:100874. doi: 10.1016/j.ortho.2024.100874.
41. Rainey BJ, Burnside G, Harrison JE. Reliability of cervical vertebral maturation staging. Am J Orthod Dentofacial Orthop 2016;150:98-104. doi: 10.1016/j.ajodo.2015.12.013.
42. Jaqueira LM, Armond MC, Pereira LJ, Alcântara CE, Marques LS. Determining skeletal maturation stage using cervical vertebrae: Evaluation of three diagnostic methods. Braz Oral Res 2010;24:433-7. doi: 10.1590/s1806-83242010000400010.
43. Rongo R, Valleta R, Bucci R, Bonetti GA, Michelotti A, D'Antò V. Does clinical experience affect the reproducibility of cervical vertebrae maturation method? Angle Orthod 2015;85:841-7. doi: 10.2319/080414-544.1.