

## Visual assessment of the cervical vertebral maturation stages *A study of diagnostic accuracy and repeatability*

Giuseppe Perinetti<sup>a</sup>; Alberto Caprioglio<sup>b</sup>; Luca Contardo<sup>c</sup>

### ABSTRACT

**Objective:** To evaluate the diagnostic accuracy and repeatability of the visual assessment of the cervical vertebral maturation (CVM) stages.

**Materials and Methods:** Ten operators underwent training sessions in visual assessment of CVM staging. Subsequently, they were asked to stage 72 cases equally divided into the six stages. Such assessment was repeated twice in two sessions (T1 and T2) 4 weeks apart. A reference standard for each case was created according to a cephalometric analysis of both the concavities and shapes of the cervical vertebrae.

**Results:** The overall agreement with the reference standard was about 68% for both sessions and 76.9% for intrarater repeatability. The overall kappa coefficients with the reference standard were up to 0.86 for both sessions, and 0.88 for intrarater repeatability. Overall, disagreements one stage and two stage apart were 23.5% (T1) and 5.1% (T2), respectively. Sensitivity ranged from 53.3% for CS5 (T1) to 99.9% for CS1 (T2), positive predictive values ranged from 52.4% for CS5 (T2) to 94.3% for CS6 (T1), and accuracy ranged from 83.6% for CS4 (T2) to 94.9% for CS1 (T1).

**Conclusions:** Visual assessment of the CVM stages is accurate and repeatable to a satisfactory level. About one in three cases remain misclassified; disagreement is generally limited to one stage and is mostly seen in stages 4 and 5. (*Angle Orthod.* 2014;84:951–956.)

**KEY WORDS:** Orthodontics; Cervical vertebrae; Diagnosis; Accuracy; Validation study

### INTRODUCTION

When dealing with skeletal disharmonies, the precise identification of skeletal maturity, that is, the growth phase, with particular regard to the onset of the pubertal growth spurt, has major clinical implications in terms of treatment efficacy and efficiency.<sup>1,2</sup> Several indices have been proposed to identify the skeletal maturation phases<sup>1,3–6</sup> and among them is the cervical vertebral maturation (CVM) method initially proposed

by Lamparski.<sup>7</sup> The CVM method has been correlated with both the statural and the mandibular growth spurt,<sup>8,9</sup> and even with levels of biomarkers of growth.<sup>10,11</sup>

However, two previous investigations<sup>12,13</sup> reported that this methodology is not sufficiently repeatable to be used alone in assessing skeletal maturation. Moreover, even though raters participating in those studies were experienced orthodontists, the repeatability was reported to be only from low to good. In this regard, the impact of dedicated sessions of training on the repeatability of the CVM method has still to be evaluated. To the best of our knowledge, a study on the diagnostic accuracy, that is, agreement with a reference standard, of the visual assessment of the CVM stage is still lacking.

The present study was thus designed to address the following questions: (1) Is the visual assessment of the CVM stage accurate and repeatable? and (2) if disagreement is seen, how is this structured among the different stages? This study ultimately aims to verify whether the CVM method may be proposed as a reliable indicator of growth phase and used in routine clinical practice.

<sup>a</sup> Research Fellow, Department of Medical, Surgical and Health Sciences, School of Dentistry, University of Trieste, Trieste, Italy.

<sup>b</sup> Associate Professor, Department of Orthodontics, Università dell'Insubria, Varese, Italy.

<sup>c</sup> Assistant Professor, Department of Medical, Surgical and Health Sciences, School of Dentistry, University of Trieste, Trieste, Italy.

Corresponding author: Dr Giuseppe Perinetti, Piazza Ospitale 1, Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, Friuli Venezia Giulia 34129, Italy (e-mail: G.Perinetti@fmc.units.it)

Accepted: February 2014. Submitted: December 2013.

Published Online: March 25, 2014

© 2014 by The EH Angle Education and Research Foundation, Inc.

## MATERIALS AND METHODS

### Study Design

A total of 10 operators underwent training sessions in visual assessment of CVM staging using a series of cases analyzed cephalometrically (reference standard). Subsequently, they were asked to assign stage in a different set of cases at baseline (T1) and again after 4 weeks (T2). The outcomes of these sessions were compared with the reference standard (objective analysis) and for diagnostic accuracy.

This study included subjects who were seeking orthodontic treatment and who had never been treated. Signed informed consent was obtained from the parents of the subjects before they were included in the study, and the protocol was reviewed and approved by the local Ethical Committee of the University of Trieste. In particular, a lateral cephalogram was taken at the first clinical examination as part of the initial clinical recording. The following inclusion criteria were applied: (1) age between 7 and 18 years, (2) absence of anomalies of the vertebrae, (3) good general health and an absence of any nutritional problems, (4) no history of trauma at the cervical region or right hand, and (5) Caucasian ethnicity. Patients with lateral head films of low quality were excluded from the study. An initial sample of more than 200 subjects was screened to obtain a total of 132 subjects (73 girls and 59 boys) for the study. These 132 subjects were equally divided into the six CVM stages according to the cephalometric analysis. Within each CVM stage, 10 cases (total, 60 cases) were used for the training sessions; the remaining 12 cases per CVM stage (total, 72 cases) were used for the diagnostic accuracy and reproducibility part of the study. In particular, these 132 cases were selected according to the outcome of the cephalometric analysis (reference standard, see the section that follows). All the lateral cephalograms were cropped to include cervical vertebrae C2 to C4 and to eliminate any additional information.

### Raters and Training Sessions

A total of 10 raters were equally derived from two different universities involved in the study. Among these raters were six postgraduate students, two postdoctoral students, one assistant professor, and one undergraduate student. None of the authors were among the raters. An expert researcher in the CVM method instructed the raters. Each rater, independently of his or her experience in orthodontics underwent at least two training sessions in the CVM staging. In particular, all the raters participated together at the first session, where the series of 60 cases was used. After a detailed explanation of the rules to be followed for

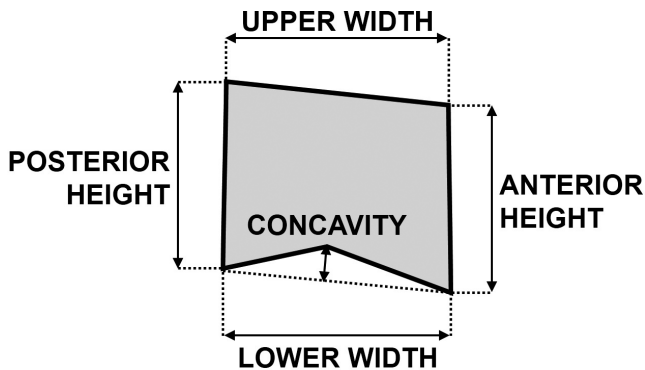
assigning CVM stages, each rater had to assign a stage to each of the traced cases visually and blinded. The followed sequence was based on identification of concavities at C2, C3, and C4 and, subsequently, assignment of shapes of C3 and C4. After a first assignment, the raters could see the superimposed cervical tracings and the numeric data of the cephalometric analysis. Any conflict was discussed immediately. At the end of this session, each rater had the files with images and numeric data. After 4–6 weeks the raters underwent a second training session within either university under the supervision of two different instructors, and training was considered successful only if at least 75% of the cases were correctly identified. Those raters failing to reach this result underwent a further session of training 2 weeks later.

### CVM Method

The most recent CVM method, as proposed by Baccetti et al.,<sup>1</sup> including six stages: two prepubertal (stages 1 and 2), two pubertal (stages 3 and 4), and two postpubertal (stages 5 and 6) was used. When apparent disagreement was seen between the presence of concavities at the lower borders and the shapes of the cervical vertebrae, the most mature stage of either the concavities or shape was assigned.

### Reference Standard (Objective Analysis)

A customized digitization regimen and analysis with cephalometric software (Viewbox, version 3.0, dHAL Software, Kifissia, Greece) was used for all cephalograms examined in this study. This analysis included measurements generated from 21 landmark markers, with 11 derived linear variables (Figure 1). Among these derived variables, three were for concavities of lower borders of C2-C4 and eight related to the anterior and posterior heights and upper and lower widths for the C3 and C4. Herein the threshold of 1.0 mm<sup>1</sup> was used to assess the presence of concavity. Similarly, assessment of the shape of the vertebral bodies of C3 and C4 was defined according to the height-width difference calculated with the following formula:  $[(\text{posterior height} + \text{anterior height})/2] - [(\text{upper width} + \text{lower width})/2]$ . Clustering was as follows: values below  $-1.0$  mm assessed a trapezoid or rectangular horizontal shape, values between  $-1.0$  mm and  $+1.0$  mm assessed a square shape, and values above  $+1.0$  mm assessed a rectangular vertical shape. All cephalograms were traced by an operator and checked for accuracy by a second investigator. Full method error was quantified through the method of moments variance estimator<sup>13</sup> calculated on 20 pairs of recordings randomly selected and expressed as mean (95% confidence interval [CI]).



**Figure 1.** Diagram of the cephalometric measurements with derived linear variables. Only a vertebral body is shown for clarity. In cervical vertebra 2, only concavity was measured.

**Statistical Analysis**

The percentages of agreement between each rater and the reference standard (within each session) were calculated, along with the percentages of intrarater repeatability between the two sessions. To determine the degree of agreement between each rater and the reference standard (within each session), along with the degree of repeatability between the two sessions and within each rater, a weighted kappa coefficient was used and presented as mean and 95% CI. This kappa coefficient was weighted linearly as this makes the coefficient less sensitive to the number of categories or stages.<sup>14</sup> The kappa coefficient ranges from zero for no agreement to 1 for perfect agreement, and the following standards for strength of agreement for the kappa coefficient have proposed: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and >0.80 almost perfect.<sup>15</sup>

Within each session, interrater agreement was assessed by the Kendall’s W coefficient of concordance. Moreover, the mean percentages of disagreement one stage, two stage and three stage apart were calculated according to each CVM stage for both sessions. Within each CVM stage, a more comprehensive diagnostic performance analysis was performed, including sensitivity, specificity, positive and negative predictive values, and accuracy,<sup>15</sup> and presented as mean and 95% CI. Finally, when calculating the overall means for these parameters, including the weighted kappa values, the paired nature of that data was taken into account.

SPSS software 13.0 (SPSS Inc, Chicago, Ill), MedCalc software 12.3.3.0 (MedCalc Software, Mariakerke, Belgium), and Comprehensive Meta-Analysis, version 2 (Biostat, Englewood, NJ) were used to perform the statistical analyses. A P value <.05 was used for rejection of the null hypothesis.

**Table 1.** Percentages of Agreement with the Reference Standard According to Recording Session (T1 and T2) and for Intrarater Agreement Between the Recording Sessions for Each Rater

Rater	With Reference Standard		Intrarater T1–T2 % (n)
	T1 % (n)	T2 % (n)	
1	77.8% (56)	75.0% (54)	83.3% (60)
2	81.9% (59)	79.2% (57)	84.7% (61)
3	72.9% (57)	73.6% (53)	86.1% (62)
4	77.8% (56)	81.9% (59)	95.8% (69)
5	80.6% (58)	76.4% (55)	83.3% (60)
6	48.6% (35)	61.1% (44)	80.6% (58)
7	44.4% (32)	38.9% (28)	52.8% (38)
8	58.3% (42)	62.5% (45)	62.5% (45)
9	62.5% (45)	68.1% (49)	65.3% (47)
10	70.8% (51)	66.7% (48)	75.0% (54)
Overall <sup>a</sup>	68.2% (49.1)	68.3% (49.2)	76.9% (55.4)

<sup>a</sup> Overall results presented as means of percentages (n) among raters; n = 12 for each cervical vertebral maturation stage.

**RESULTS**

The final group of 72 subjects used for the diagnostic accuracy and reproducibility analyses was composed of 37 girls and 35 boys (mean age, 12.9 ± 2.6 years; range, 7.3–17.9 years). The method error for the cephalometric parameters as mean (95% CI) were between 0.21 mm (0.10–0.36) and 0.26 mm (0.12–0.43) for the concavities of C3 and C2, respectively, and between 0.22 mm (0.11–0.37) and 0.49 mm (0.23–0.81) for the lower widths of C3 and C4, respectively.

The percentage of full agreement between each rater and the reference standard and within each rater is summarized in Table 1. The full agreement with the reference standard for each rater was generally similar between the two recording sessions, with values ranging from 38.9% to 81.9%. Intrarater agreements ranged from 52.8% to 95.8%. The overall agreement with the reference standard was about 68% for both sessions and 76.9% for intrarater comparisons.

The weighted kappa coefficients between each rater and the reference standard and within each rater are summarized in Table 2. The kappa coefficients as mean (95% CI) with the reference standard for each rater was generally similar between the two recording sessions, with values ranging from 0.52 (0.41–0.62) to 0.90 (0.84–0.95). The overall kappa coefficients of the whole group of raters with the reference standard were 0.82 (0.77–0.88) and 0.81 (0.77–0.87) for the T1 and T2 session, respectively, and 0.88 (0.84–0.93) for intrarater repeatability.

The Kendall’s W coefficients of concordance for interrater agreement were 0.90 and 0.91 (P < .001) for the T1 and T2 sessions, respectively. The percentages of disagreement with the reference standard and for the whole group of raters, according to each CVM

**Table 2.** Weighted Kappa Coefficients for Interrater Agreement with the Reference Standard According to Recording Session (T1 and T2) and for Intrarater Agreement Between the Recording Sessions for Each Rater

Rater	With Reference Standard		Intrarater T1–T2
	T1	T2	
	Mean (95% CI)	Mean (95% CI)	
1	0.87 (0.80–0.93)	0.85 (0.78–0.92)	0.90 (0.84–0.96)
2	0.88 (0.82–0.95)	0.87 (0.80–0.93)	0.92 (0.87–0.97)
3	0.88 (0.82–0.94)	0.83 (0.76–0.91)	0.91 (0.86–0.97)
4	0.88 (0.82–0.94)	0.90 (0.84–0.95)	0.98 (0.95–0.99)
5	0.89 (0.83–0.94)	0.86 (0.80–0.93)	0.91 (0.85–0.96)
6	0.70 (0.63–0.78)	0.78 (0.71–0.85)	0.89 (0.84–0.95)
7	0.64 (0.54–0.74)	0.52 (0.41–0.62)	0.61 (0.50–0.71)
8	0.74 (0.65–0.82)	0.74 (0.64–0.84)	0.79 (0.73–0.86)
9	0.73 (0.63–0.82)	0.81 (0.74–0.88)	0.76 (0.68–0.85)
10	0.83 (0.76–0.90)	0.79 (0.71–0.87)	0.86 (0.79–0.92)
Overall <sup>a</sup>	0.82 (0.77–0.88)	0.81 (0.77–0.87)	0.88 (0.84–0.93)

<sup>a</sup> Overall results presented as means (95% CI) among raters; n = 12 for each cervical vertebral maturation stage.

stage, are summarized in Table 3. These disagreements were generally similar between the two recording sessions, with the CS1 and CS5 showing the lowest and greater disagreements, respectively. Moreover, CS1 showed disagreements only one stage apart (up to 1.5%), CS2 and CS3 showed disagreements up to two stages apart (up to 3.8%, CS3, T1), and CS4 to CS5 showed disagreements up to three stages apart (up to 3.8%, CS4, T2).

The full diagnostic accuracy analysis for the whole group of raters is summarized in Table 4. All the diagnostic parameters were generally similar between the two recording sessions. In particular, sensitivity ranged from 50.0% for CS4 (T2) to 99.9% for CS1 (T2), specificity values ranged from 90.3% for CS4 (T2) to 98.9% for CS6 (T1), positive predictive values ranged from 52.4% for CS4 (T2) to 94.3% for CS6 (T1), negative predictive values ranged from 99.9% for CS1 (T2) to 90.0% for CS4 (T2), and accuracy ranged from 83.6% for CS4 (T2) to 94.9% for CS1 (T1).

**DISCUSSION**

The present study analyzed the diagnostic accuracy and repeatability of the CVM method in a heterogeneous group of 10 raters previously subjected to at least two sessions of training.

In previous investigations<sup>12,13</sup> each observer was trained in the CVM method using only exact figures and legends reported in literature.<sup>1</sup> However, descriptive pictures were usually simplified versions of the full range of possibilities, and this is not a substitute for having expert instructors teach all the skills needed to become confident with any procedure. Moreover, these investigations did not assess whether the operators' training was successful.

**Table 3.** Percentages of Cervical Stage Disagreement with the Reference Standard According to Recording Session (T1 and T2) for the Whole Group of Raters<sup>a</sup>

CVM Stage	Disagreement, No. of Stages Apart	With Reference Standard	
		T1 Mean (Min-Max)	T2 Mean (Min-Max)
1	1	1.5% (0–8.3)	—
	2	—	—
	3	—	—
2	1	28.0% (0–66.7)	26.5% (0–58.2)
	2	0.8% (0–8.3)	1.5% (0–8.3)
	3	—	—
3	1	21.2% (0–58.3)	19.7% (0–50.0)
	2	3.8% (0–16.7)	3.0% (–8.3)
	3	—	—
4	1	28.8% (0–50.0)	26.5% (0–50.0)
	2	14.4% (0–33.3)	15.2% (0–25.0)
	3	0.8% (0–8.3)	3.8% (0–16.7)
5	1	34.1% (0–66.7)	33.3% (0–66.7)
	2	6.8% (0–16.7)	5.3% (0–25.0)
	3	1.5% (0–8.3)	0.8% (0–8.3)
6	1	27.3% (0–75.0)	29.5% (0–66.7)
	2	3.8% (0–16.7)	5.3% (0–41.7)
	3	0.8% (0–8.3)	1.5% (0–16.7)
Overall <sup>b</sup>	1	23.5%	22.6%
	2	4.9%	5.1%
	3	0.5%	1.0%

<sup>a</sup> CVM indicates cervical vertebral maturation stage; Min, minimum; Max, maximum; —, no cases.

<sup>b</sup> Overall results presented as means of percentages among CVM stages; n = 12 for each CVM stage.

Because this is the first study evaluating the diagnostic accuracy of the CVM method by using a standard reference, comparisons with previous data are not possible. An overall satisfactorily diagnostic accuracy was seen in the present study, even though difference in terms of percentages of agreement and kappa values were seen between the raters from the two centers at both clinical sessions. This may be explained by the different degree of training in the visual assignment of the CVM stage. Generally, the raters yielding greater weighted kappa values were from the university (raters 1 to 5, Tables 1 and 2) in which relevant training in CVM staging is part of the curriculum and where the research team has been using this method for several years. A similar behavior was seen for the intrarater repeatability. This reinforces the concept that regular training is necessary to obtain high diagnostic accuracy and intrarater repeatability in the visual assignment of the CVM stages.

Because CVM staging is based on an ordinal scale rather than being a dichotomous assessment, the number of stages apart in case of disagreement is also important. In the present study, most of the disagreements were one stage apart (Table 3). Apart from the percentages of agreement and kappa values, satisfactory diagnostic accuracy was seen; accuracy values

**Table 4.** Diagnostic Accuracy Parameters for Each Cervical Vertebral Maturation (CVM) Stage According to Recording session (T1 and T2) for the Whole Group of Raters<sup>a</sup>

CVM Stage <sup>b</sup>	Parameter	With Reference Standard	
		T1 Mean (95% CI)	T2 Mean (95% CI)
1	Sensitivity	98.3% (91.3–100)	99.9% (99.8–100)
	Specificity	94.2% (88.3–100)	93.7% (87.6–99.7)
	PPV	78.3% (58.5–98.1)	79.1% (61.3–96.6)
	NPV	99.6% (98.1–100)	99.9% (99.8–100)
	Accuracy	94.9% (89.8–99.9)	94.7% (89.6–99.8)
2	Sensitivity	68.3% (43.4–93.3)	96.2% (45.0–93.4)
	Specificity	93.3% (87.1–99.5)	94.2% (88.3–100)
	PPV	70.0% (47.8–92.1)	73.9% (52.1–95.7)
	NPV	93.6% (87.5–99.8)	93.9% (88.0–99.9)
	Accuracy	89.2% (82.1–96.2)	90.0% (83.2–96.8)
3	Sensitivity	72.5% (50.4–94.6)	75.0% (52.6–97.4)
	Specificity	92.7% (86.2–99.1)	93.5% (87.4–99.6)
	PPV	67.4% (44.4–90.5)	71.7% (48.5–94.9)
	NPV	94.5% (88.8–100)	95.0% (89.6–100)
	Accuracy	89.3% (82.3–96.3)	90.4% (83.7–97.1)
4	Sensitivity	51.7% (24.2–79.1)	50.0% (22.4–77.6)
	Specificity	90.7% (83.4–97.9)	90.3% (82.9–97.8)
	PPV	54.4% (26.9–81.9)	52.4% (24.8–80.0)
	NPV	90.4% (82.9–97.8)	90.0% (82.5–97.6)
	Accuracy	84.2% (75.8–92.5)	83.6% (75.1–92.1)
5	Sensitivity	53.3% (28.6–78.1)	55.8% (30.1–81.5)
	Specificity	92.0% (85.2–98.8)	91.5% (84.5–98.5)
	PPV	56.4% (30.3–82.6)	56.5% (28.7–84.3)
	NPV	90.9% (83.7–98.0)	91.3% (84.3–98.3)
	Accuracy	85.6% (77.6–93.5)	85.6% (77.5–93.6)
6	Sensitivity	65.0% (40.8–89.2)	60.0% (35.7–84.3)
	Specificity	98.9% (96.4–100)	98.8% (96.1–100)
	PPV	94.3% (81.7–100)	92.1% (74.4–100)
	NPV	93.6% (87.6–99.5)	92.7% (86.5–99.0)
	Accuracy	93.3% (87.6–99.0)	92.3% (86.3–98.4)

<sup>a</sup> CVM indicates cervical vertebral maturation; PPV, positive predictive value; NPV, negative predictive value.

<sup>b</sup> For each CVM stage, n = 12.

were above 83.6% for each CVM stage (Table 4). However, when dealing with several possible clusters (ie, six CVM stages), important diagnostic parameters are the sensitivity (when CVM stages are equally distributed) and positive predictive value; these indicate a given rater's capability in identifying any CVM stage, irrespective of the number of true negative cases belonging to the other stages. The least satisfactory diagnostic accuracy was seen for the CVM stages 4 and 5, where the greatest sensitivity was low as 55.8% (stage 5, T2) and positive predictive values were as low as 52.4% (stage 4, T2). The fine morphologic transitions between two following stages may be behind such evidence. This is particularly true of the visual assessment of the shape of a vertebral body. Interestingly, CVM stage 4 is based on both initial concavity on C4 and initial appearance of a square shape on either C3 or C4. Previous evidence

showed that intrarater repeatability was much lower in assessing the shape of the bodies of C3 and C4 compared with that obtained in assessing the concavities of C2–C4.<sup>13</sup>

When considering intrarater repeatability between the two sessions, the overall percentage of agreement was 76.9% (Table 1). Previous studies using CVM methods, although not focused on repeatability, reported very high percentages of agreement<sup>8,16</sup>. On the contrary, in the previous study of CVM method repeatability,<sup>12</sup> an average of 62% was reported. The reason for the high agreements seen in most studies has been explained by the use of third-person tracings to stage the lateral cephalograms, or because the authors of these studies were also raters themselves.<sup>12</sup>

### Clinical Implications

According to the present results, the visual assessment of the CVM stages appears to be accurate and repeatable as long as training is followed. This accuracy appears independent of the rater's own experience in orthodontics. However, assessment of CVM stages 4 and 5 requires more careful evaluation to avoid unreliable diagnosis and treatment plans. Any case in which CVM stage 4 is misclassified as stage 5, or vice versa, may thus miss the opportunity for proper treatment. Therefore, when the visual assessment of the CVM staging is uncertain (especially for stages 4 and 5), a cephalometric analysis or the use of a further morphologic indicator may be indicated.

### CONCLUSIONS

- When specific training is provided along with precise guidelines in assessing visually each stage, the CVM method proves to be accurate and repeatable to a satisfactory level
- About one of three patients remain misclassified, though disagreement is generally only one stage apart and is mostly seen in stages 4 and 5.

### ACKNOWLEDGMENT

The authors are deeply grateful to all the raters involved in the study.

### REFERENCES

1. Baccetti T, Franchi L, McNamara JA Jr. The cervical vertebral maturation (CVM) method for the assessment of optimal treatment timing in dentofacial orthopedics. *Semin Orthod.* 2005;11:119–129.
2. Petrovic A, Stutzmann J, Lavergne J. Mechanism of craniofacial growth and modus operandi of functional appliances: a cell-level and cybernetic approach to orthodontic decision making. In: Carlson DS, ed. *Craniofacial Growth Theory and Orthodontic Treatment*. Monograph 23. Cra-

- niofacial Growth Series. Ann Arbor: Center for Human Growth and Development, University of Michigan; 1990: 13–74.
3. Fishman LS. Radiographic evaluation of skeletal maturation. A clinically oriented method based on hand-wrist films. *Angle Orthod.* 1982;52:88–112.
  4. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. 2nd ed. Stanford, CA: Stanford University Press; 1959.
  5. Hassel B, Farman AG. Skeletal maturation evaluation using cervical vertebrae. *Am J Orthod Dentofacial Orthop.* 1995; 107:58–66.
  6. Perinetti G, Baccetti T, Contardo L, Di Lenarda R. Gingival crevicular fluid alkaline phosphatase activity as a non-invasive biomarker of skeletal maturation. *Orthod Craniofac Res.* 2011;14:44–50.
  7. Lamparski DG. *Skeletal Age Assessment Utilizing Cervical Vertebrae* [dissertation]. Pittsburgh, PA: University of Pittsburgh; 1972.
  8. Franchi L, Baccetti T, McNamara JA Jr. Mandibular growth as related to cervical vertebral maturation and body height. *Am J Orthod Dentofacial Orthop.* 2000;118:335–340.
  9. 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;134:227–237.
  10. Masoud M, Masoud I, Kent RL Jr, Gowharji N, Cohen LE. Assessing skeletal maturity by using blood spot insulin-like growth factor I (IGF-I) testing. *Am J Orthod Dentofacial Orthop.* 2008;134:209–216.
  11. Perinetti G, Franchi L, Castaldo A, Contardo L. Gingival crevicular fluid protein content and alkaline phosphatase activity in relation to pubertal growth phase. *Angle Orthod.* 2012;82:1047–1052.
  12. 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–e7; discussion 78–80.
  13. Nestman TS, Marshall SD, Qian F, Holton N, Franciscus RG, Southard TE. Cervical vertebrae maturation method morphologic criteria: poor reproducibility. *Am J Orthod Dentofacial Orthop.* 2011;140:182–188.
  14. Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology.* 1996;7:199–202.
  15. Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. *BMJ.* 1997;315:540–543.
  16. Ozer T, Kama JD, Ozer SY. A practical method for determining pubertal growth spurt. *Am J Orthod Dentofacial Orthop.* 2006;130:131.e1–131.e6.