

## Does clinical experience affect the reproducibility of cervical vertebrae maturation method?

Roberto Rongo<sup>a</sup>; Rosa Valleta<sup>b</sup>; Rosaria Bucci<sup>c</sup>; Giulio Alessandri Bonetti<sup>d</sup>; Ambrosina Michelotti<sup>e</sup>; Vincenzo D'Antò<sup>f</sup>

### ABSTRACT

**Objective:** To assess interobserver and intraobserver reproducibility of the cervical vertebrae maturation method (CVMM) among three panels of judges with different levels of orthodontic experience (OE).

**Materials and Methods:** Fifty individual lateral cephalograms of good quality with complete visualization of cervical vertebrae 1 to 4 were selected. Thirty clinicians, divided according to their OE into three groups (junior group, JU, OE ≤ 1 year; postgraduate group, PG, 2 ≤ OE ≤ 4 years; specialist group, SP, OE ≥ 7 years), evaluated the cephalograms in two sessions (T<sub>1</sub> and T<sub>2</sub>) at 3 weeks apart. Kendall's W and weighted Cohen's kappa (κ) coefficients were performed to assess interobserver and intraobserver agreement. The level of significance was set as  $P < .05$ . For both the interobserver and the intraobserver datasets, the percentage of perfect agreement (PPA) and the number of stages apart for each disagreement were calculated.

**Results:** Kendall's W at T<sub>1</sub> was SP = 0.61, PG = 0.70, and JU = 0.87; at T<sub>2</sub> it was SP = 0.78, PG = 0.85, and JU = 0.86. The percentage of total interobserver perfect agreement (Inter-PPA) was 42.3% at T<sub>1</sub> and 46.3% at T<sub>2</sub>. The JU group had the highest Cohen's κ coefficient at 0.78, while the PG and SP had coefficients of 0.64 each. The percentage of total intraobserver perfect agreement (Intra-PPA) was 54.2%.

**Conclusions:** The reproducibility of the method was not improved by the level of orthodontic experience. The group with the lowest level of orthodontic experience had the best performance. (*Angle Orthod.* 2015;85:841–847.)

**KEY WORDS:** Cervical vertebrae maturation method; Reproducibility; Skeletal maturation; Reliability

---

<sup>a</sup> Postdoctoral student, School of Orthodontics, Department of Neurosciences, Reproductive Sciences and Oral Sciences, University of Naples Federico II, Naples, Italy.

<sup>b</sup> Associate Professor, School of Orthodontics, Department of Neurosciences, Reproductive Sciences and Oral Sciences, University of Naples Federico II, Naples, Italy.

<sup>c</sup> PhD student, School of Orthodontics, Department of Neurosciences, Reproductive Sciences and Oral Sciences, University of Naples Federico II, Naples, Italy.

<sup>d</sup> Assistant Professor, Department of Orthodontics, University of Bologna, Bologna, Italy.

<sup>e</sup> Professor and Chair, School of Orthodontics, Department of Neurosciences, Reproductive Sciences and Oral Sciences, University of Naples Federico II, Naples, Italy.

<sup>f</sup> Research Fellow, School of Orthodontics, Department of Neurosciences, Reproductive Sciences and Oral Sciences, University of Naples Federico II, Naples, Italy; Division of Dentistry, Department of Pediatric Surgery, Bambino Gesù Children's Hospital, Rome, Italy.

Corresponding author: Dr Vincenzo D'Antò, Department of Neurosciences, Reproductive Sciences and Oral Sciences, University of Naples Federico II, Via Pansini 5, 80131 Naples, Italy (e-mail: vincenzo.danto@unina.it)

### INTRODUCTION

Craniofacial growth may play an important role in the success of orthodontic treatment.<sup>1,2</sup> The reliable prediction of patient mandibular and maxillary development could help in understanding the best therapeutic decision regarding treatment timing, appliance choice, and the possible need for surgery.<sup>3</sup> As most orthodontic patients are growing individuals, orthodontists have to consider their craniofacial growth path for successful treatment planning.<sup>4</sup> However, individuals with the same chronologic age may have different growth patterns regarding onset, duration, speed, direction, and amount of residual growth, as shown in several studies.<sup>5–9</sup>

---

Accepted: October 2014. Submitted: August 2014.

Published Online: January 26, 2015

© 2015 by The EH Angle Education and Research Foundation, Inc.

Many indicators have been suggested to evaluate the timing of mandibular growth peak and skeletal maturation,<sup>10-12</sup> and the most used methods for this assessment are based on radiographic analysis. Specifically, numerous authors investigated the relationship between mandibular growth and skeletal maturation estimated by means of hand-wrist bone analysis (HWBA) or the cervical vertebrae maturation method (CVMM).<sup>3,13-19</sup> Since the CVMM is advocated as a timing tool for orthopedic treatment, its validity and reproducibility should be assessed by procedures without any methodologic shortcomings.<sup>20</sup> Different investigations have studied the association between the CVMM and skeletal maturity and the CVMM and mandibular growth spurt, testing the validity.<sup>21-31</sup> Several studies found a good correlation between CVMM and HWBA, suggesting the possibility to use the CVMM instead of the HWBA to reduce the radiation dose.<sup>21-26</sup> However, a recent systematic review of the CVMM was not able to establish the validity of this method due to the lack of moderate/high quality papers on this topic.<sup>32</sup> Several studies criticized the validity of the CVMM, showing that it is unable to predict the start of the peak in mandibular growth.<sup>27-29</sup> Furthermore, it has been shown that the effective radiation dose for a lateral cephalogram without a thyroid shield is 1.5 fold higher than the effective dose for a lateral cephalogram with a thyroid shield plus a hand-wrist radiograph.<sup>33</sup>

On the other hand, the reproducibility of the CVMM, which can affect its clinical usefulness, is strongly debated. Very high levels of interobserver and intraobserver reproducibility (over 90%) were reported by some examinations.<sup>18,22,24</sup> However, these studies have some limitations, such as the analysis of traced vertebrae on the lateral cephalograms instead of actual radiographs, the presence of the authors among the group of judges performing the analysis, the use of small sample sizes, and the use of an improper statistical method.<sup>20</sup>

Some authors have succeeded in the investigation of the reproducibility overcoming the previous methodologic weakness but without exploring the influence of clinical experience on the reproducibility of the CVMM.<sup>20,30,31,34</sup> Indeed, there is little information about the impact of judges' clinical experience on the CVMM, even though it should not be underestimated.<sup>20,35</sup> Therefore, the aim of this study was to evaluate the interobserver and intraobserver reproducibility of the CVMM among three judge panels with different levels of orthodontic experience (OE). The null hypothesis was that the orthodontic clinical experience did not have any influence on the reproducibility of the CVMM.

## MATERIALS AND METHODS

The study was approved by the Local Ethical Committee of the University of Naples Federico II. Fifty good quality individual lateral cephalograms of patients attending the School of Orthodontics of the University of Naples Federico II, with complete visualization of cervical vertebrae 1 through 4, were selected for our sample. The cephalograms were randomly chosen from the school's electronic database, by means of the random function of a scientific calculator (EL-506VB, Sharp Corp, Osaka, Japan) and were equally divided by sex (mean age 12.4 years  $\pm$  3.2; 25 female and 25 male). Afterwards, the original lateral cephalograms were searched and scanned at 600 dpi (Perfection V750 Pro, Seiko Epson Corp, Suwa, Japan) for presentation as high-resolution images in TIFF format to maintain the radiographic quality.

To avoid any additional information that might influence the observer during the evaluation of the CVMM (as stage of dentition), the lateral cephalograms were cut to include only cervical vertebrae from C1 to C4.

The judges were divided into three groups according to the level of clinical experience: the junior group (JU) formed by 10 recent graduates in dentistry, with less than 1 year of orthodontic experience (mean age 25.1 years  $\pm$  1.2; OE 0.4 years  $\pm$  0.5; 6 female and 4 male); the postgraduate group (PG), comprising 10 postgraduate students in orthodontics with clinical experience ranging between 2 and 4 years (mean age 29.7 years  $\pm$  1.4; OE 2.6 years  $\pm$  0.5; 6 female and 4 male); and the specialist group (SP) including 10 specialists in orthodontics with more than 7 years of orthodontic experience (mean age 41.8 years  $\pm$  10.3; OE 19.1 years  $\pm$  10.4; 3 female and 7 male). None of the clinicians recruited for the assessment participated in the study design. All of the judges belonged to the School of Orthodontics of the University of Naples Federico II as tutor/professor (SP), postgraduate student (PG), or voluntary frequenter (JU).

Each observer was invited to perform two sessions of evaluation of cervical stage on the lateral cephalograms according to the method suggested by Baccetti et al.<sup>3</sup> Between the two sessions, a 3-week time interval was respected ( $T_1$  = initial;  $T_2$  = 3 weeks). The cephalograms were presented in a high-resolution file, randomly ordered for the two sessions. Before the first session, all participants attended a lecture on the CVMM by one investigator (Dr D'Antò). Moreover, at the start of each session the observers also received a copy of the paper by Baccetti et al.,<sup>3</sup> and beside each cephalogram a schematic representation of the CVMM was shown as well.

## Statistical Analysis

The statistical analysis was conducted to calculate Kendall's  $W$  coefficient for the interobserver agreement and weighted Cohen's kappa ( $\kappa$ ) coefficient for the intraobserver agreement. The Kendall's  $W$ , for the whole sample (Total) and for each group independently (JU, PG, SP) was calculated by means of the Statistical Package for the Social Sciences (SPSS, IBM, Armonk, NY) for the two sessions ( $T_1$  and  $T_2$ ). The weighted Cohen's  $\kappa$  was calculated by means of Statistical Analysis Software ver. 9.2 (SAS Inc, Cary, NC), with linear weights, comparing the two sessions for each observer and for each group (JU, PG, SP). The level of significance was set as  $P < .05$ . The range of variation of Kendall's  $W$  coefficient of concordance and of the weighted  $\kappa$  statistic is between 0 for no agreement and 1 for perfect agreement with five intermediate levels: *slight agreement* (0.01–0.20), *fair agreement* (0.21–0.40), *moderate agreement* (0.41–0.60), *substantial agreement* (0.61–0.80), and *almost perfect agreement* (0.81–0.99).<sup>36</sup> Furthermore, for both interobserver and intraobserver datasets, the percentage of perfect agreement (PPA) and the number of stages apart for each disagreement were calculated.

For the interobserver dataset, the percentage of perfect agreement (Inter-PPA) was the number of the same staging among each couple of judges, of each observation, independently for  $T_1$  and  $T_2$ . They were evaluated for all observers (Total) and for each group (JU, PG, SP).

For the intraobserver dataset, the percentage of perfect agreement (Intra-PPA) was the number of the same staging of each observation between  $T_1$  and  $T_2$  for each judge. Moreover, they were evaluated for all judges (Total) and for each group (JU, PG, SP).

Finally, the group top 10 was composed of the best 10 observers evaluated according to the results of the weighted Cohen's  $\kappa$ ; in addition, all of the variables were assessed for this group.

## RESULTS

The 30 participants performed a total of 3000 evaluations at two different time intervals. The Kendall's  $W$  coefficient for each group varied from 0.61–0.87. The interobserver agreement was the highest for the JU group in both time intervals and showed an *almost perfect agreement* ( $T_1$   $W = 0.87$ ;  $T_2$   $W = 0.86$ ). On the other hand, the SP group achieved the lowest Kendall's  $W$  values presenting a *substantial agreement* ( $T_1$   $W = 0.61$ ;  $T_2$   $W = 0.78$ ). The total interobserver agreement for the 30 participants at the two time intervals varied from  $W = 0.70$  at  $T_1$  (*substantial agreement*) to  $W = 0.81$  at  $T_2$  (*almost perfect agreement*) (Table 1).

**Table 1.** Kendall's  $W$  Coefficient for Interobserver Agreement at Two Time Intervals ( $T_1$  and  $T_2$ )

| Session | OE <sup>a</sup> | W    | P     | Interpretation <sup>b</sup> |
|---------|-----------------|------|-------|-----------------------------|
| $T_1$   | JU              | 0.87 | <.001 | Almost perfect              |
|         | PG              | 0.70 | <.001 | Substantial                 |
|         | SP              | 0.61 | <.001 | Substantial                 |
|         | Total           | 0.70 | <.001 | Substantial                 |
|         | Top 10          | 0.87 | <.001 | Almost perfect              |
| $T_2$   | JU              | 0.86 | <.001 | Almost perfect              |
|         | PG              | 0.85 | <.001 | Almost perfect              |
|         | SP              | 0.78 | <.001 | Substantial                 |
|         | Total           | 0.81 | <.001 | Almost perfect              |
|         | Top 10          | 0.89 | <.001 | Almost perfect              |

<sup>a</sup> OE indicates orthodontic experience: JU, junior; PG, postgraduate; and SP, specialist.

<sup>b</sup> Interpretation of the coefficient according Viera et al.<sup>36</sup>

A total of 21,750 comparisons between the evaluations for each couple of judges, for each session, was analyzed, to assess the Inter-PPA. The Inter-PPA was 42.3% ( $N = 9204/21,750$ ) at  $T_1$  and 46.3% ( $N = 10,082/21,750$ ) at  $T_2$ . The percentage of one stage apart disagreement was 40.1% ( $N = 8721/21,750$ ) at  $T_1$  and 41.4% ( $N = 9011/21,750$ ) at  $T_2$  (Table 2).

The intraobserver agreement for each observer ranged from  $\kappa = 0.24$  *fair agreement* to  $\kappa = 0.81$  *almost perfect agreement*. Among the 30 participants, however, only three showed a *fair agreement*, while 18 showed a *substantial agreement*, two showed an *almost perfect agreement*, and seven a *moderate agreement* (Table 3).

The group showing the best result was the JU, where the  $\kappa$  coefficient ranged from *moderate agreement* to *almost perfect agreement*. Moreover, the weighted Cohen's  $\kappa$  coefficient for the whole group was the highest ( $\kappa = 0.78$  *substantial agreement*). In the PG results there was one *fair agreement* but also one *almost perfect agreement*, while the SP had the worst performance. The latter two groups, PG and SP, showed a similar Cohen's  $\kappa$  coefficient for the whole group ( $\kappa = 0.64$  *substantial agreement*; Tables 3 and 4). For each observer the comparison between the evaluation at  $T_1$  and  $T_2$  of the 50 assessments was analyzed to evaluate the Intra-PPA. The total Intra-PPA was 54.7% ( $N = 821/1500$ ), while the percentage of intraobserver one stage apart disagreement was 34.3% ( $N = 515/1500$ ). The JU showed the highest Intra-PPA of 57.8% ( $N = 289/500$ ), and 34% (170/500) of their disagreements were of just one cervical stage (Table 5).

The Top 10 group was formed by six observers of the JU, only one of the PG, and three of the SP (Table 3). This group achieved the best results in the Inter-PPA, Intra-PPA, and Kendall's  $W$  (Tables 1, 2, and 5).

**Table 2.** Percentage of Interobserver Perfect Agreement (Inter-PPA) With Cervical Stage Differences

| Session        | OE <sup>a</sup> | N <sup>b</sup> | Inter-PPA (%)        | One Stage Apart (%) |
|----------------|-----------------|----------------|----------------------|---------------------|
| T <sub>1</sub> | Total           | 21,750         | 9204/21,750 (42.3)   | 8721/21,750 (40.1)  |
| T <sub>2</sub> | Total           | 21,750         | 10,082/21,750 (46.3) | 9011/21,750 (41.4)  |
| T <sub>1</sub> | JU              | 2250           | 1167/2250 (51.9)     | 916/2250 (40.7)     |
|                | PG              | 2250           | 912/2250 (40.5)      | 924/2250 (41.1)     |
|                | SP              | 2250           | 813/2250 (36.1)      | 886/2250 (39.4)     |
|                | Top 10          | 2250           | 1212/2250 (53.9)     | 862/2250 (38.3)     |
| T <sub>2</sub> | JU              | 2250           | 1063/2250 (47.2)     | 935/2250 (41.5)     |
|                | PG              | 2250           | 1124/2250 (50.0)     | 898/2250 (39.9)     |
|                | SP              | 2250           | 982/2250 (43.6)      | 919/2250 (40.8)     |
|                | Top 10          | 2250           | 1263/2250 (56.1)     | 857/2250 (38.1)     |

<sup>a</sup> OE indicates orthodontic experience: JU, junior; PG, postgraduate; and SP, specialist.

<sup>b</sup> N indicates number of comparisons.

## DISCUSSION

The aim of this study was to analyze the influence of the OE on the reproducibility of the CVMM using three judge panels with different levels of clinical experience. The SP was the group with the highest clinical experience (OE = 19.1), but this group achieved the

lowest values of W,  $\kappa$ , Inter-PPA, and Intra-PPA. On the contrary, the group with less than 1 year of clinical experience (JU), showed the highest values for all of the parameters investigated. These results suggest that the OE does not improve the reproducibility of the CVMM, and they could be explained by the different

**Table 3.** Weighted Cohen's  $\kappa$  Coefficient for Intraobserver Agreement and Percentage of Agreement by Observers<sup>a</sup>

| Observer | OE <sup>b</sup> | Cohen's $\kappa$ | SE          | 95% CI           | P               | PA <sup>c</sup> | Intra-PPA, <sup>c</sup> % | Interpretation <sup>d</sup> |
|----------|-----------------|------------------|-------------|------------------|-----------------|-----------------|---------------------------|-----------------------------|
| 1        | <b>JU</b>       | <b>0.77</b>      | <b>0.05</b> | <b>0.67–0.87</b> | <b>&lt;.001</b> | <b>33</b>       | <b>66</b>                 | <b>Substantial</b>          |
| 2        | <b>JU</b>       | <b>0.73</b>      | <b>0.06</b> | <b>0.62–0.84</b> | <b>&lt;.001</b> | <b>30</b>       | <b>60</b>                 | <b>Substantial</b>          |
| 3        | JU              | 0.65             | 0.06        | 0.53–0.76        | <.001           | 25              | 50                        | Substantial                 |
| 4        | <b>JU</b>       | <b>0.72</b>      | <b>0.06</b> | <b>0.61–0.83</b> | <b>&lt;.001</b> | <b>30</b>       | <b>60</b>                 | <b>Substantial</b>          |
| 5        | JU              | 0.63             | 0.05        | 0.52–0.73        | <.001           | 23              | 46                        | Substantial                 |
| 6        | JU              | 0.59             | 0.06        | 0.48–0.69        | .001            | 18              | 36                        | Moderate                    |
| 7        | JU              | 0.71             | 0.07        | 0.58–0.84        | <.001           | 31              | 62                        | Substantial                 |
| 8        | <b>JU</b>       | <b>0.81</b>      | <b>0.05</b> | <b>0.72–0.89</b> | <b>&lt;.001</b> | <b>34</b>       | <b>68</b>                 | <b>Almost perfect</b>       |
| 9        | <b>JU</b>       | <b>0.76</b>      | <b>0.06</b> | <b>0.65–0.87</b> | <b>&lt;.001</b> | <b>35</b>       | <b>70</b>                 | <b>Substantial</b>          |
| 10       | <b>JU</b>       | <b>0.75</b>      | <b>0.05</b> | <b>0.65–0.84</b> | <b>&lt;.001</b> | <b>30</b>       | <b>60</b>                 | <b>Substantial</b>          |
| 11       | PG              | 0.71             | 0.06        | 0.59–0.83        | <.001           | 29              | 58                        | Substantial                 |
| 12       | PG              | 0.63             | 0.06        | 0.52–0.73        | <.001           | 24              | 48                        | Substantial                 |
| 13       | PG              | 0.66             | 0.07        | 0.52–0.79        | <.001           | 28              | 56                        | Substantial                 |
| 14       | PG              | 0.68             | 0.06        | 0.58–0.79        | <.001           | 28              | 56                        | Substantial                 |
| 15       | PG              | 0.69             | 0.05        | 0.59–0.80        | <.001           | 26              | 52                        | Substantial                 |
| 16       | PG              | 0.57             | 0.05        | 0.47–0.68        | <.001           | 19              | 38                        | Moderate                    |
| 17       | PG              | 0.60             | 0.09        | 0.42–0.77        | <.001           | 32              | 64                        | Moderate                    |
| 18       | <b>PG</b>       | <b>0.81</b>      | <b>0.06</b> | <b>0.70–0.92</b> | <b>&lt;.001</b> | <b>39</b>       | <b>78</b>                 | <b>Almost perfect</b>       |
| 19       | PG              | 0.50             | 0.08        | 0.34–0.66        | <.001           | 24              | 48                        | Moderate                    |
| 20       | PG              | 0.30             | 0.09        | 0.12–0.47        | .01             | 20              | 40                        | Fair                        |
| 21       | SP              | 0.38             | 0.10        | 0.17–0.58        | .004            | 22              | 44                        | Fair                        |
| 22       | SP              | 0.64             | 0.06        | 0.52–0.76        | <.001           | 27              | 54                        | Substantial                 |
| 23       | SP              | 0.69             | 0.06        | 0.57–0.81        | <.001           | 30              | 60                        | Substantial                 |
| 24       | <b>SP</b>       | <b>0.74</b>      | <b>0.05</b> | <b>0.63–0.84</b> | <b>&lt;.001</b> | <b>32</b>       | <b>64</b>                 | <b>Substantial</b>          |
| 25       | <b>SP</b>       | <b>0.72</b>      | <b>0.07</b> | <b>0.59–0.85</b> | <b>&lt;.001</b> | <b>33</b>       | <b>66</b>                 | <b>Substantial</b>          |
| 26       | SP              | 0.49             | 0.09        | 0.31–0.67        | <.001           | 27              | 54                        | Moderate                    |
| 27       | SP              | 0.52             | 0.08        | 0.37–0.67        | <.001           | 25              | 50                        | Moderate                    |
| 28       | SP              | 0.24             | 0.09        | 0.07–0.41        | .034            | 13              | 26                        | Fair                        |
| 29       | <b>SP</b>       | <b>0.79</b>      | <b>0.05</b> | <b>0.70–0.88</b> | <b>&lt;.001</b> | <b>34</b>       | <b>68</b>                 | <b>Substantial</b>          |
| 30       | SP              | 0.41             | 0.10        | 0.22–0.60        | <.001           | 20              | 40                        | Moderate                    |

<sup>a</sup> Bold text indicates the observers in the Top 10 groups.

<sup>b</sup> OE indicates orthodontic experience: JU, junior; PG, postgraduate; and SP, specialist.

<sup>c</sup> PA indicates number of perfect agreements; Intra-PPA, intraoperator percentage of perfect agreement.

<sup>d</sup> Interpretation of the coefficient according Viera et al.<sup>36</sup>

**Table 2.** Extended

| Two Stages Apart (%) | Three Stages Apart (%) | Four Stages Apart (%) | Five Stages Apart (%) |
|----------------------|------------------------|-----------------------|-----------------------|
| 2758/21,750 (12.7)   | 805/21,750 (3.7)       | 241/21,750 (1.1)      | 21/21,750 (0.1)       |
| 2231/21,750 (10.3)   | 374/21,750 (1.7)       | 50/21,750 (0.2)       | 2/21,750 (approx. 0)  |
| 131/2250 (5.8)       | 34/2250 (1.5)          | 2/2250 (0.1)          | 0/2250 (0)            |
| 290/2250 (12.8)      | 93/2250 (4.1)          | 29/2250 (1.3)         | 2/2250 (0.1)          |
| 409/2250 (18.2)      | 105/2250 (4.7)         | 32/2250 (1.4)         | 5/2250 (0.2)          |
| 139/2250 (6.2)       | 31/2250 (1.4)          | 5/2250 (0.2)          | 1/2250 (approx. 0)    |
| 212/2250 (9.4)       | 34/2250 (1.5)          | 5/2250 (0.2)          | 1/2250 (approx. 0)    |
| 193/2250 (8.6)       | 30/2250 (1.3)          | 5/2250 (0.2)          | 0/2250 (0)            |
| 288/2250 (12.8)      | 58/2250 (2.6)          | 3/2250 (0.1)          | 0/2250 (0)            |
| 114/2250 (5.1)       | 16/2250 (0.7)          | 0/2250 (0)            | 0/2250 (0)            |

familiarity with the CVMM. Indeed, during their last year of undergraduate courses at the University of Naples Federico II, the JU were educated on the CVMM and they were extensively trained on this method. On the other hand, all of the components of the SP group did not use the CVMM in their daily clinical practice. This might lead to confirmation that to correctly use the CVMM, a more in-depth training may be critical.

The interobserver agreement for the overall group of observers was 0.81 at T<sub>2</sub>, higher than that reported by Gabriel et al.<sup>20</sup> (0.74). The main differences between the two works are the number of participants, the sample size, and the division of judge panels according to the level of clinical experience. Although the results are similar for the weighted Cohen's  $\kappa$  and for the Intra-PPA for all observers, in this study the JU and the PG groups reached a level of intraobserver and interobserver agreement higher than the results showed by Gabriel et al.<sup>20</sup>

These findings highlight that, probably, the level of practice and knowledge of the CVMM might be an important factor for its reproducibility, and that a simple use of a handout is likely not sufficient to obtain a good level of knowledge and familiarity with the CVMM.

In Zhao et al.,<sup>31</sup> the CVMM was explained to the observers by means of training sessions, and it showed values of W and  $\kappa$  coefficients similar to our

best performing sample (JU). Hence, to correctly use the CVMM there might be a need for multiple training sessions to understand how to assess the cervical stage and to acquire a consistent method of evaluation. Interestingly, the SP and PG groups showed an increase of the Kendall's W between T<sub>1</sub> and T<sub>2</sub> probably due to a training effect.

In a very recent study,<sup>34</sup> the CVMM was tested for accuracy and reproducibility, comparing the cephalometric evaluations of the concavities and shape of the cervical vertebrae with the visual evaluation of CVMM. The authors found the visual method accurate and reproducible; moreover, they stated that in their group the reproducibility was high, independently of the level of the experience of the observers, which is similar to our results.

In the current investigation, albeit the CVMM showed a good reproducibility, the Inter-PPA and Intra-PPA were low. In fact, in the JU group, the Intra-PPA was 57.8% as opposed to a high  $\kappa$  coefficient of 0.78. This means that in almost 4 of 10 cases a clinician could have evaluated a cervical stage differently and might have changed his treatment plan. A similar situation was observed for the other two groups as well. Moreover, the Inter-PPA was similar to the that found in the studies by Gabriel et al.<sup>20</sup> and Zhao et al.,<sup>31</sup> and it exceeded the 40% with a peak of 51.9% for JU at T<sub>1</sub>. This means that a clinician has almost a 50% of chance to disagree on cervical staging assessed by others. It has to be taken into account that the value of the weighted Cohen's  $\kappa$  coefficient increases with the number of categories of the assessed method; therefore, this can be a possible explanation of the difference between the high  $\kappa$  coefficient and the low perfect agreement.<sup>37</sup> Finally, it is interesting to note that in almost all three judge panels, on average, the 42% of the cervical staging were judged with only one stage apart of difference for

**Table 4.** Weighted Cohen's  $\kappa$  Coefficient for Intraobserver Agreement by Orthodontic Experience Groups

| OE <sup>a</sup> | $\kappa$ | SE   | 95% CI    | P     | Interpretation <sup>b</sup> |
|-----------------|----------|------|-----------|-------|-----------------------------|
| JU              | 0.78     | 0.02 | 0.75–0.81 | <.001 | Substantial                 |
| PG              | 0.64     | 0.02 | 0.61–0.68 | <.001 | Substantial                 |
| SP              | 0.64     | 0.02 | 0.60–0.68 | <.001 | Substantial                 |
| Top 10          | 0.76     | 0.02 | 0.73–0.79 | <.001 | Substantial                 |

<sup>a</sup> OE indicates orthodontic experience: JU, junior; PG, postgraduate; and SP, specialist.

<sup>b</sup> Interpretation of the coefficient according Viera et al.<sup>36</sup>

**Table 5.** Percentage of Intraobserver Perfect Agreement (Intra-PPA) With Cervical Stage Differences

| OE <sup>a</sup> | N    | Intra-PPA (%)   | One Stage Apart (%) | Two Stages Apart (%) | Three Stages Apart (%) | Four Stages Apart (%) | Five Stages Apart (%) |
|-----------------|------|-----------------|---------------------|----------------------|------------------------|-----------------------|-----------------------|
| JU              | 500  | 289/500 (57.8)  | 170/500 (34.0)      | 37/500 (7.4)         | 3/500 (0.6)            | 1/500 (0.2)           | 0/500 (0)             |
| PG              | 500  | 269/500 (53.8)  | 177/500 (35.4)      | 38/500 (7.6)         | 14/500 (2.8)           | 2/500 (0.4)           | 0/500 (0)             |
| SP              | 500  | 263/500 (52.6)  | 168/500 (33.6)      | 50/500 (10.0)        | 13/500 (2.6)           | 3/500 (0.6)           | 3/500 (0.6)           |
| Total           | 1500 | 821/1500 (54.7) | 515/1500 (34.3)     | 125/1500 (8.3)       | 30/1500 (2.0)          | 6/1500 (0.4)          | 3/1500 (0.2)          |
| Top 10          | 500  | 330/500 (66.0)  | 145/500 (29.0)      | 20/500 (4.0)         | 4/500 (0.8)            | 1/500 (0.2)           | 0/500 (0)             |

<sup>a</sup> OE indicates orthodontic experience: JU, junior; PG, postgraduate; and SP, specialist.

both interobserver and intraobserver analysis. This might be another factor that affects the  $\kappa$  coefficient; in fact, having most disagreements in just one category increases the weighted Cohen's  $\kappa$ . Therefore, even if the weighted Cohen's  $\kappa$  and the Kendall's W coefficients were sufficiently high, the level of perfect agreement in the intraobserver and interobserver analysis seems to be not adequate enough to support the reproducibility of the CVMM as a method to evaluate skeletal maturation. Also analyzing the top 10 group, even if the achieved results were better than the results of the other groups, there was still a low level of Inter-PPA and Intra-PPA that could affect the clinical decision, and so the usability of the CVMM.

Having a powerful means to predict the residual potential growth in young patients, without any added biological cost, is clinically helpful, and the CVMM could assume an important role in orthodontics, especially during clinical diagnoses and treatment planning decisions.<sup>3</sup> However, the level of reproducibility of this method may affect the clinical decision in the orthopedic orthodontic treatment.

One limitation of this work is that it cannot provide information on the validity of the CVMM due to the lack of a longitudinal sample analysis. Moreover, without using a gold standard observation there was not the possibility to assess the accuracy of the CVMM.

## CONCLUSIONS

- The main finding of this research was the lack of influence of orthodontic clinical experience on the interobserver and intraobserver reproducibility of the CVMM. The group with the lowest level of OE had the best performance. Hence, high level of orthodontic experience does not increase the reproducibility of the CVMM.
- The Inter-PPA (36.1%–56.1%) and the Intra-PPA (52.6%–66.0%) of the CVMM were too low to suggest the exclusive use of the CVMM in the assessment of skeletal growth.

## REFERENCES

1. Petrovic A. Auxologic categorization and chronobiologic specification for the choice of appropriate orthodontic

- treatment. *Am J Orthod Dentofacial Orthop.* 1994;105:192–205.
2. Baccetti T, Franchi L, Toth LR, McNamara JA Jr. Treatment timing for Twin-block therapy. *Am J Orthod Dentofacial Orthop.* 2000;118:159–170.
3. Baccetti T, Franchi L, McNamara JA Jr. The cervical vertebral maturation method for the assessment of optimal treatment timing in dentofacial orthopedics. *Semin Orthod.* 2005;11:119–129.
4. Verma D, Peltomäki T, Jäger A. Reliability of growth prediction with hand-wrist radiographs. *Eur J Orthod.* 2009;31:438–442.
5. Björk A. The significance of growth changes in facial pattern and their relationship to changes in occlusion. *Dent Rec (London).* 1951;71:197–205.
6. Bjork A. Variation in the growth pattern of the human mandible: longitudinal radiographic study by the implant method. *J Dent Res.* 1963;42:400–411.
7. Bambha JK, Van Natta P. Longitudinal study of facial growth in relation to skeletal maturation during adolescence. *Am J Orthod.* 1963;49:481–492.
8. Bishara SE, Peterson L, Bishara EC. Changes in facial dimensions and relationships between the ages of 5 and 25 years. *Am J Orthod.* 1984;85:238–252.
9. Bishara SE, Jakobsen JR. Longitudinal changes in three facial types. *Am J Orthod.* 1985;88:466–502.
10. Lewis AB, Garn SM. The relationship between tooth formation and other maturation factors. *Angle Orthod.* 1960;30:70–77.
11. Tanner JM. *Growth and Adolescence.* 2nd ed. Oxford, UK:Blackwell Scientific Publications; 1962.
12. Mitani H, Sato K. Comparison of mandibular growth with other variables during puberty. *Am J Orthod Dentofacial Orthop.* 1992;62:217–222.
13. Chapman SM. Ossification of the adductor sesamoid and the adolescent growth spurt. *Angle Orthod.* 1972;42:236–245.
14. Lamparski DG. *Skeletal Age Assessment Utilizing Cervical Vertebrae* [dissertation]. Pittsburgh:University of Pittsburgh; 1972.
15. Grave KC, Brown T. Skeletal ossification and the adolescent growth spurt. *Am J Orthod.* 1976;69:611–619.
16. Houston WJ, Miller JC, Tanner JM. Prediction of the timing of the adolescent growth spurt from ossification events in hand-wrist films. *Br J Orthod.* 1979;6:145–152.
17. Fishman LS. Radiographic evaluation of skeletal maturation. A clinically oriented method based on hand-wrist films. *Angle Orthod.* 1982;52:88–112.
18. Hassel B, Farman A. Skeletal maturation evaluation using cervical vertebrae. *Am J Orthod Dentofacial Orthop.* 1995;107:58–66.
19. 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.

20. Gabriel DB, Southard KA, Qian F, et al. Cervical vertebrae maturation method: poor reproducibility. *Am J Orthod Dentofacial Orthop.* 2009;136:478.e1–7.
21. Gandini P, Mancini M, Andreani F. A comparison of handwrist bone and cervical vertebral analysis in measuring skeletal maturation. *Angle Orthod.* 2006;76:984–989.
22. 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.
23. San Roman P, Palma JC, Oteo MD, Nevado E. Skeletal maturation determined by cervical vertebrae development. *Eur J Orthod.* 2002;24:303–311.
24. 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;130:622–628.
25. Stiehl J, Müller B, Dibbets J. The development of the cervical vertebrae as an indicator of skeletal maturity: comparison with the classic method of hand-wrist radiograph. *J Orofac Orthop.* 2009;70:327–335.
26. Wong RW, Alkhal HA, Rabie BM. Use of cervical vertebral maturation to determine skeletal age. *Am J Orthod Dentofacial Orthop.* 2009;136:484.e1–6.
27. Ball G, Woodside D, Tompson B, Hunter WS, Posluns J. Relationship between cervical vertebral maturation and mandibular growth. *Am J Orthod Dentofacial Orthop.* 2011;139:e455–461.
28. Beit P, Peltomäki T, Schätzle M, Signorelli L, Patcas R. Evaluating the agreement of skeletal age assessment based on hand-wrist and cervical vertebrae radiography. *Am J Orthod Dentofacial Orthop.* 2013;144:838–847.
29. Mellion ZJ, Behrents RG, Johnston LE Jr. The pattern of facial skeletal growth and its relationship to various common indexes of maturation. *Am J Orthod Dentofacial Orthop.* 2013;143:845–854.
30. 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.
31. Zhao XG, Lin J, Jiang JH, Wang Q, Ng SH. Validity and reliability of a method for assessment of cervical vertebral maturation. *Angle Orthod.* 2012;82:229–234.
32. 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;82:1123–1131.
33. Patcas R, Signorelli L, Peltomäki T, Schätzle M. Is the use of the cervical vertebrae maturation method justified to determine skeletal age? A comparison of radiation dose of two strategies for skeletal age estimation. *Eur J Orthod.* 2013;35:604–609.
34. Perinetti G, Caprioglio A, Contardo L. Visual assessment of the cervical vertebral maturation stages: a study of diagnostic accuracy and repeatability. *Angle Orthod.* 2014;84:951–956.
35. Baccetti T, Franchi L, McNamara JA Jr. Reproducibility of the CVM method: a reply. *Am J Orthod Dentofacial Orthop.* 2010;137:446–447.
36. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005;37:360–363.
37. Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology.* 1996;7:199–202.