Article

Not peer-reviewed version

# Initial Evidence of Validity and Reliability of Rubrics for the Evaluation of Preclinical Practices in Orthodontics in Dentistry Students

Aguado-Gil JM *, Arena-Echeverry S, Suárez A, Gómez-Sánchez M, González-Cuevas G

Posted Date: 21 April 2026

doi: 10.20944/preprints202604.1394.v1

Keywords: rubrics; dental education; dentistry degree; preclinical practice; evaluation tools; psychometry

 doi:10.20944/preprints202604.1394.v1



*Article*

# Initial Evidence of Validity and Reliability of Rubrics for the Evaluation of Preclinical Practices in Orthodontics in Dentistry Students

**Aguado-Gil JM [1,*], Arena-Echeverry S [1], Suárez A [1], Gómez-Sánchez M [2] and González-Cuevas G [3]**

[1] Department of Preclinical Dentistry II, Faculty of Biomedical and Health Sciences, Universidad Europea de Madrid, C/ Tajo s/n, Villaviciosa de Odón, Madrid 28670, Spain

[2] Faculty of Dentistry. Alfonso X El Sabio University

[3] Central Arizona College, Arizona (United States), Boise State University, Idaho (United States)

* Correspondence: josemaria.aguado@universidadeuropea.es; Tel.: +34914268071

**Abstract**

The objective evaluation of preclinical practices in Dentistry is a challenge of great relevance. In recent years, tools have been developed aimed at improving the objectivity, standardization and transparency of the evaluation process. In this context, the present study analyses the initial evidence of reliability, internal structure and academic usefulness of evaluation rubrics applied to preclinical practices of Orthodontics at the European University of Madrid. Material and method: The research followed a quantitative, observational and correlational design, of a cross-sectional nature and with an instrumental component, with a sample of 175 third-year students of the Bachelor's Degree in Dentistry. The reliability and internal structure of the rubrics were analysed using Cronbach's alpha and exploratory factor analysis, as well as their relationship with academic performance variables using Pearson correlations and partial correlations. Results: The results showed moderate to adequate levels of internal consistency ($\alpha$ = 0.61 and 0.71) and results compatible with a one-dimensional structure in both instruments. Likewise, certain academic variables, such as performance in the cephalometry seminar and in the theoretical evaluation, showed significant associations with practical performance. Conclusion: The results support the use of rubrics not only as grading tools, but also as potentially useful training instruments to structure preclinical assessment in specific contexts.

**Keywords:** rubrics; dental education; dentistry degree; preclinical practice; evaluation tools; psychometry

## 1. Introduction

Assessment plays a central role in contemporary university education, especially in degrees with a high practical component, such as the Bachelor's Degree in Dentistry. In this context, assessment not only fulfils an accrediting function, but also acts as a structural element of the learning process, conditioning the progressive acquisition of practical competences. For this reason, the development of instruments that unify criteria, increase objectivity and provide effective feedback to the student has become a priority of higher health education. (Pasillera Díaz et al., 2010) (Consolation Menzala - Peralta & Ortega- Menzala, 2021; Jornet Meliá, 2017; Kanthan & Senger, 2011).

The concept of assessment has evolved from grading-focused approaches to continuous, systematic and formative models, based on the collection and interpretation of evidence to improve both learning and teaching practice. In this framework, informed decision-making acquires a key role, overcoming the reductionist vision of evaluation as a mere assignment of a score. (Hernán-Losada, 2009; López-Pastor, 2011)

 doi:10.20944/preprints202604.1394.v1



In Dentistry, this need is especially relevant in preclinical practices, where the student must demonstrate not only theoretical knowledge, but also psychomotor skills, technical precision and ability to apply standardized criteria before joining clinical practice. The evaluation of this type of activity poses specific difficulties, since it involves complex, multidimensional tasks with a potential component of subjectivity on the part of the evaluator. (Mays & Branch- Mays, 2016) (Chambers & Labarre, 2014; Sanderson et al., 2016; Tenkumo et al., 2019; Vahtivuori-Hänninen et al., 2007; Velasco-Martínez et al., 2017)

In response to these limitations, rubrics have been consolidated as instruments of interest in health and dental education, by making it possible to make explicit criteria, performance levels and learning expectations. Its use has been associated with greater transparency of the evaluation process, improved feedback and greater coherence between evaluators. However, in the preclinical dental field, the literature has focused mainly on its perceived usefulness or its application in specific contexts, and there are still few studies that systematically analyse its psychometric properties. (Allen & Tanner, 2006; Morrow et al., 2014; Oh et al., 2018; Satheesh et al., 2015; Saunders et al., 2018; Segura Egea & Jiménez Rubio-Manzanares, 2013) (Baracco et al., 2025)

This limitation is relevant, since the incorporation of rubrics should not be based only on their acceptance, but also on evidence that supports the interpretation of the scores. In this sense, the analysis of reliability, internal structure and their relationship with academic variables allows us to determine if these instruments generate consistent and coherent measures with the constructed evaluated. (Bharuthram & Patel, 2017; Bindayel, 2017; Habib, 2018)

Within preclinical practices in Orthodontics, wire bending and cephalometry are two contexts that are particularly suitable for this type of analysis. Both activities are part of the fundamental training of the student, require technical precision and application of observable criteria, and present a structure susceptible to evaluation by means of analytical instruments. In addition, they involve differentiated competencies, since wire bending is mainly related to manual dexterity, while cephalometry integrates analysis, anatomical localization and measurement skills. (Bhoopathi & Atchison, 2015; Gansky et al., 2004; Schwibbe et al., 2016) (Jauhar et al., 2016) (Fields et al., 2017)

Likewise, it is still relevant to analyse whether the scores obtained through these instruments are related to other indicators of academic performance, which would allow assessing their usefulness within a competency-based assessment approach. (Donaldson & Gray, 2012)

In this context, the objective of this study was to analyse the psychometric properties and academic usefulness of two evaluation rubrics applied to preclinical practices of Orthodontics of the Degree in Dentistry of the European University of Madrid. Specifically, the students' performance, reliability and construct validity of the instruments were evaluated, as well as their relationship with academic performance variables. (Fluit et al., 2010; Francis, 2018; Heathman et al., 2021).

## 2. Materials and Methods

### *2.1. Design*

The research was designed as an observational, correlational and quantitative study, of a cross-sectional nature with an instrumental component aimed at the analysis of the psychometric properties of two evaluation rubrics used in preclinical practices of Orthodontics. The study was carried out in the Bachelor's Degree in Dentistry at the European University of Madrid.

### *2.2. Participants*

The sample was composed of 175 third-year students enrolled in the subject of Orthodontics I during the academic years 2017–2018 and 2018–2019.

Only students evaluated by the same teacher were included, in order to reduce the variability associated with possible differences in the application of the evaluation criteria.

Groups with English language teaching were excluded due to possible difficulties in linguistic comprehension in perception surveys.

 doi:10.20944/preprints202604.1394.v1



*2.3. Development and Validation of Rubrics*

The design and development of the rubrics was structured in several phases. In an initial stage, a Community of Practice was constituted made up of 30 professors of the Bachelor's Degree in Dentistry, with the aim of agreeing on common evaluation criteria for preclinical practices. This collaborative process made it possible to identify relevant transversal competencies (responsibility, punctuality, attitude, appropriate use of material and respect for the environment), as well as to establish general guidelines aligned with the White Paper on Dentistry and the principles of the EHEA.

Based on this consensus, an initial regulation was drawn up that defined serious faults, weightings, rating scales and a limited number of evaluation criteria.

Subsequently, specific rubrics were developed for the preclinical practices of Orthodontics I, specifically for wire bending and cephalometry. These pilot rubrics were implemented during the 2015–2016 academic year and were disseminated to students through face-to-face explanatory sessions and through the institutional virtual campus.

After this initial phase, limitations related to the number of items, the evaluation load for the teaching staff and a tendency to increase average grades were identified, which led to a review and adjustment process.

In a later phase, the rubrics were subjected to external review by orthodontic experts from outside the institution, who evaluated the adequacy of the criteria, the clarity of the descriptors and the weighting of the items using a structured questionnaire.

The observations made by the experts were used to make minor adjustments to the weighting of some criteria and to the wording of certain descriptors, with the aim of improving their clarity and applicability in the teaching context. The revised and agreed versions of the rubrics were those used in this study (Figure 1).

| Concepto | % sobre la nota | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| **Diseño** | 40 | El alambre se desvía más de 3 o más veces del dibujo. | Se desvía 2 veces del dibujo. | Se desvía 1 veces del dibujo. | El alambre sigue al 100% el diseño del dibujo. |
| **Muescas** | 20 | 3 o más muescas. | 2 muescas | 1 muescas | No hay muescas. |
| **Plano** | 30 | Hay 3 desviaciones del alambre sobre el plano o la desviación es superior a 2 mm. | Hay 2 desviaciones del alambre sobre el plano. | Hay 1 desviación del alambre sobre el plano. | El alambre está en un solo plano. |

**Figure 1.** Ultimate Wire Bending Rubric.

*2.4. Preclinical Practices Evaluated*

The practices analysed corresponded to two thematic blocks: wire bending and cephalometry. Each block was developed over three training sessions and a final evaluation session, following a progressive model with prior training feedback.

Wire bending was evaluated based on three main criteria (design, notches, and plane), while cephalometry included five dimensions (tracing, location of points, lines, and planes, cephalometric measurements, and cleanliness).

All criteria were assessed using four-level ordinal scales (1-4), according to the original design of the rubrics. Given the ordinal nature of these scales, the scores were interpreted as performance level indicators within each criterion evaluated. (Al Amri et al., 2018; Habib, 2018)

The final scores were obtained by adding weighted percentages of the previously established criteria. 2 corresponded to the minimum required to pass the criterion evaluated (5/pass), 1 to an

 doi:10.20944/preprints202604.1394.v1



insufficient level (2.5/fail), 3 to a remarkable performance and 4 to the achievement of all the criteria set out for that level (outstanding).

### *2.5. Academic Variables and Student Perception*

Additional academic variables were collected, including the grade of the theoretical exam, the grade of the cephalometry seminar and the final grade of the internship. Likewise, satisfaction surveys were applied before and after each practical block, with five-point Likert-type scales, aimed at analysing the students' perception of the clarity, usefulness and adequacy of the rubrics as assessment instruments.

The surveys were anonymous and voluntary and were administered in paper format in person during the internship. The students completed them immediately and handed them in a few minutes later.

In case of incomplete records in the surveys, the analyses were carried out using the available cases for each variable.

### *2.6. Statistical Analysis*

The statistical analysis was performed with IBM SPSS Statistics 30.0. Analyses were carried out using descriptive statistics, mean comparison tests (Student's *t* and repeated measures ANOVA), reliability analysis using Cronbach's alpha coefficient and (from Vet et al., 2017) construct validity using exploratory factor analysis, after checking the assumptions of sample adequacy (KMO) and Bartlett's sphericity.

Finally, the associations between the scores obtained in the rubrics and other variables of academic performance were analysed using Pearson correlations and partial correlations.

In all analyses, a statistical significance level of $p < 0.05$ was considered.

## 3. Results

The sample analysed was composed of 175 students, with a balanced distribution by sex and a mean age of 23.14 years. In the analyses corresponding to the pre- and post-evaluation perception surveys, 156 students finally participated, due to the presence of incomplete questionnaires.

### *3.1. Academic Performance in Preclinical Internships*

In the practice of wire bending, 82.86% of the students passed the test, with an average grade of 6.80/10 (SD=1.86).

To contextualize these results, the grades obtained were compared with those recorded in immediately preceding academic cohorts of the same course, whose mean was 7.48/10 (SD = 1.26). The comparison using a Student's *t-test* for independent samples showed statistically significant differences between both groups, $t(355) = 4.04$, $p < 0.01$, d = 0.43, with higher mean scores in the previous cohorts.

When analyzing the different criteria evaluated in practice (Design: M = 2.54, SD = 1.86; Notches: M = 2.83, SD = 0.85; Flat: M = 2.73, SD = 0.85), a repeated-measures ANOVA was applied. The effect of the evaluated criterion was significant, $F(2, 348) = 7.81$, $p = 0.01$, $\eta^2p = 0.04$, indicating differences in the scores obtained according to the criterion considered.

Post-hoc comparisons with Bonferroni correction showed that the ratings were significantly lower in the Design criterion compared to Notches ($p < 0.01$) and Flat ($p < 0.05$). No significant differences were observed between Notches and Flat ($p = 0.53$). Nor were statistically significant differences observed according to sex.

In the cephalometry practice, 78.86% of the students achieved a pass, with an average grade of 6.57/10 (SD=2.13). Like the previous practice, these scores were compared with those corresponding to previous cohorts of the same academic year (M = 7.19, SD = 2.17): Student's t-test for independent

 doi:10.20944/preprints202604.1394.v1



samples showed statistically significant differences, $t(355) = 2.74$, $p < 0.01$, $d = 0.29$, with higher mean scores in the previous cohorts.

The analysis of the different criteria evaluated in cephalometry practice (Trace: M = 2.41, SD = 0.92; Points: M = 3.11, SD = 0.94; Lines and planes: M = 2.86, SD = 1.07; Measurements: M = 2.37, SD = 1.05; Cleanliness: M = 2.82, SD = 0.94), repeated measures ANOVA showed a significant effect of the evaluated criterion, $F(4, 696) = 26.90$, $p = 0.01$, $\eta^2p = 0.13$. Post-hoc comparisons with Bonferroni correction showed that the scores were significantly higher for Points compared to the rest of the sections ($p < 0.05$) and lower for plotting and measurements with the rest of the sections ($p < 0.01$). No significant differences were found between the Layout and Measurements sections ($p = 1$).

In relation to sex, significant differences were only observed in the Cleanliness criterion, with higher scores in women. No significant differences were identified when comparing the overall performance between the two practices.

As for the complementary academic tests, the average mark of the theoretical exam was 5.84 (SD=2.05), while the overall average grade of the internship was 7.00 (SD=1.31), with no statistically significant differences according to sex.

### *3.2. Students' Perception of the Use of Rubrics*

The results of the surveys showed a high initial assessment of the rubrics in both practices.

In the case of wire bending, a significant decrease in the assessment was observed after the evaluation, especially in the items related to theoretical knowledge and expectations, while manual skill remained the best valued aspect (Table 1).

**Table 1.** Changes in perception in the surveys on the wire bending rubric.

| Items | Medium (DT) Pre | Medium (DT) Post | Post-Pre Difference | *T-Test* | Sig. | *d* |
|---|---|---|---|---|---|---|
| **OVERALL SCORE** | **4.16 (0,61)** | **3.75 (0.77)** | **-0.40** | **5,47** | **< 0.01** | **0.44** |
| **Theoretical knowledge** | 3.52 (1,07) | 2.92 (1.38) | -0.60 | 4,41 | < 0.01 | 0.35 |
| **Manual skill** | 4.70 (0,61) | 4.33 (0.84) | -0.37 | 4,73 | < 0.01 | 0.38 |
| **Objectives** | 4.24 (0,82) | 3.79 (1.01) | -0.45 | 4,24 | < 0.01 | 0.34 |
| **Performance** | 4.20 (0,81) | 3.94 (0.95) | -0.26 | 2,95 | < 0.01 | 0.24 |
| **Expectations** | 4.13 (0,85) | 3.78 (0.99) | -0.35 | 4,35 | < 0.01 | 0.35 |

In cephalometry practice, global perception remained stable after the assessment, with only a slight decrease in the performance item (Table 2).

**Table 2.** Changes in perception in the surveys of the Cephalometry rubric.

| Items | Medium (DT) Pre | Medium (DT) Post | Post-Pre Difference | *T-Test* | Sig. | *d* |
|---|---|---|---|---|---|---|
| **OVERALL SCORE** | **4.16 (0.58)** | **3.75 (0.60)** | **-0.07** | **1.3** | **0.30** | **-** |
| **Theoretical knowledge** | 4.29 (0.83) | 4.24 (0.74) | -0.05 | 0.63 | 0.53 | - |
| **Manual skill** | 3.87 (1.00) | 3.87 (1.01) | 0.00 | 0.00 | 1.00 | - |
| **Objectives** | 4.31 (0.71) | 4.16 (0.83) | -0.15 | 1.91 | 0.06 | 0.15 |
| **Performance** | 4.17 (0.78) | 3.98 (0.93) | -0.19 | 2.10 | < 0.05 | 0.17 |
| **Expectations** | 4.17 (0.75) | 4.25 (0.80) | 0.04 | -0.51 | 0.61 | - |

When analysing the item-by-item results using repeated-measures ANOVA, significant differences were observed in both the pre-assessment ($F(4, 620) = 11.61$, $p = 0.01$, $\eta^2p = 0.07$) and the post-assessment ($F(4, 620) = 8.17$, $p = 0.01$, $\eta^2p = 0.05$). The highest mean was obtained by the item of

 doi:10.20944/preprints202604.1394.v1



objectives in the pre-evaluation survey and expectations in post-evaluation. The lowest score was observed in the item Manual Skill at both evaluation moments.

In both practical blocks, a positive association was observed between the grade obtained and the subsequent assessment of the evaluation system.

*3.3. Reliability and Psychometric Validity of the Instruments*

In relation to psychometric validation, the wire bending rubric showed moderate reliability ($\alpha$ = 0.61), while the cephalometry rubric reached an adequate level of internal consistency ($\alpha$ = 0.71). The analysis of the functioning of the items did not indicate the need to eliminate any of them. Satisfaction surveys showed Cronbach's alpha values above 0.70 in all versions. Table 3.

**Table 3.** Cronbach's alpha.

| Variable | Cronbach's alpha ($\alpha$) |
|---|---|
| **Wire bending rubric** | **0.61** |
| **Cephalometry rubric** | **0.71** |
| Pre-evaluation Survey Wire Bending | 0.77 |
| Post-evaluation Survey Wire Bending | 0.79 |
| Pre-assessment Cephalometry Survey | 0.76 |
| Post-evaluation Cephalometry Survey | 0.73 |

In the case of the wire bending rubric, when analysing the performance of the items, the elimination of any of them is not recommended since the total alpha of the scale decreases when they are eliminated and the item-total correlation is above the cut-off point 0.3 (Table 4). Similar results were observed in the cephalometry rubric, although the cleaning item showed a slightly lower contribution to the scale as a whole (Table 5).

**Table 4.** Analysis of the operation of items for the Wire Bending rubric.

| Item Total Statistics | | | |
|---|---|---|---|
| **Scale Average if the item has been deleted** | Scale variance if the element has been deleted | Total correlation of corrected elements | Cronbach's alpha if the element has been deleted |
| **5.57** | 1.89 | 0.45 | 0.47 |
| **5.27** | 2.10 | 0.39 | 0.55 |
| **5.38** | 2.04 | 0.42 | 0.52 |

**Table 5.** Analysis of the functioning of items for the cephalometry rubric.

| Item Total Statistics | | | | |
|---|---|---|---|---|
| | **Scale Average if the item has been deleted** | Scale variance if the element has been deleted | Total correlation of corrected elements | Cronbach's alpha if the element has been deleted |
| **Layout** | **11.16** | 7.87 | 0.48 | 0.66 |
| **Points** | **10.46** | 7.07 | 0.65 | 0.58 |
| **Lineas_planos** | **10.71** | 7.54 | 0.43 | 0.68 |
| **Measurements** | **11.21** | 7.62 | 0.43 | 0.68 |
| **Cleaning** | **10.76** | 8.40 | 0.36 | 0.70 |

The exploratory factor analysis confirmed the one-dimensionality of the rubrics and surveys, with adequate sample adequacy indices (KMO) and statistically significant Bartlett sphericity tests.

In the case of the wire bending rubric, the Kaiser-Meyer-Olkin sample adequacy index (KMO = 0.64) and the Bartlett sphericity test, $c\ 2(3) = 56.21$, $p < 0.01$, indicated that the data were adequate for

 doi:10.20944/preprints202604.1394.v1



factor analysis. A single factor with an eigenvalue greater than 1 was extracted, which explained 56.33% of the total variance. The factor loads of the items exceeded the value of 0.5 (Table 6).

**Table 6.** Factorial loads of the wire bending rubric.

| | Factor Loads |
|---|---|
| **Design** | 0.78 |
| **Notches** | 0.72 |
| **Map** | 0.75 |

In the cephalometry rubric, the KMO index was 0.74 and Bartlett's test was significant ($c$ 2(10) = 164.78, $p < 0.001$). The factor analysis identified a single factor that explained 47.07% of the total variance, with factor loads greater than 0.5 in all items (Table 7).

**Table 7.** Factor loads of the cephalometry rubric.

| | Factor Loads |
|---|---|
| **Layout** | 0.71 |
| **Points** | 0.84 |
| **Lines and plans** | 0.64 |
| **Measurements** | 0.66 |
| **Cleaning** | 0.55 |

In the satisfaction surveys corresponding to the practice of wire bending, both in the pre- and post-evaluation phase, the KMO indices (0.79) and the Bartlett tests were significant, extracting a single factor that explained 53.28% of the total variance.

Similar results were observed in the surveys corresponding to the cephalometry practice, with KMO indices of 0.77 in the pre-evaluation phase and 0.76 in the post-evaluation phase, and statistically significant Bartlett tests.

*3.4. Relationship with Academic Variables*

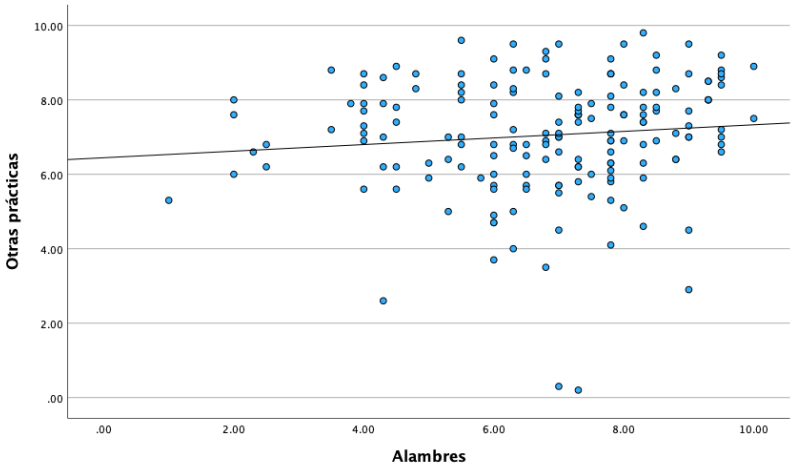
Performance in the wire bending practice did not show a significant association with performance in other practices ($r$ (175) = 0.11, $p$ = 0.17). See Graph 1.

Conversely, performance in the cephalometry seminar showed a positive association of moderate magnitude with performance in cephalometry practice ($r$ (175) = 0.36, $p < 0.01$). See graph 2.

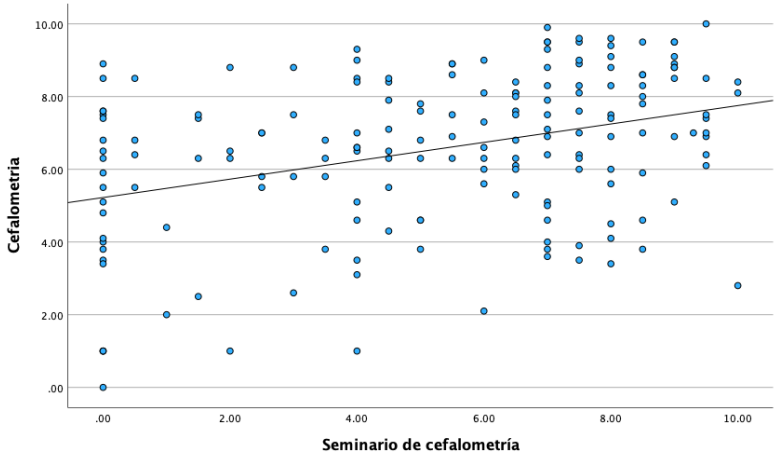
Likewise, performance in cephalometry practice was positively associated with theoretical performance in this block even after statistically controlling for the effect of the seminar (*partial* $r$ (172) = 0.31, $p < 0.01$).

 doi:10.20944/preprints202604.1394.v1






**Graph 1.** Dispersion of the linear relationship between the variables of wire practice and other practices.




**Graph 2.** Dispersion of the linear relationship between the variables of the cephalometry seminar and the practice of cephalometry.

## 4. Discussion

The evaluation of preclinical practices in Dentistry poses an important methodological challenge, since it involves simultaneously assessing cognitive, psychomotor and attitudinal dimensions in highly applied learning contexts.

This complexity has been repeatedly pointed out in the literature, especially due to the difficulties in guaranteeing adequate levels of objectivity, coherence and standardization in the assessment of student performance in clinical and preclinical settings. In this scenario, rubrics have been proposed as instruments capable of making the evaluation criteria explicit, structuring the teaching judgment and reinforcing the formative function of the evaluation process. (Smith et al., 2016; Velasco-Martínez et al., 2016)

The present study analysed two rubrics applied to preclinical practices of Orthodontics from three complementary perspectives: academic performance, student perception and initial evidence of reliability and internal structure of the instruments. In addition, the development of a digital platform aimed at facilitating its systematic application in the teaching context was described.

The design of the study, based on a Community of Practice formed by teachers of the Bachelor's Degree in Dentistry, responds to previous recommendations that underline the importance of teaching consensus in the definition of evaluation criteria. Although the initial review of the rubrics

 doi:10.20944/preprints202604.1394.v1



was carried out internally, the subsequent external review by external experts provided an additional layer of contrast in relation to the clarity of the descriptors and the relevance of the selected criteria, in line with what has been pointed out in other works on the design and refinement of assessment instruments. The performance levels of the rubrics created also coincide with several publications. However, it should be noted that this evidence of content validity was mainly qualitative and based on expert review, so future research should complement it with more structured procedures for quantifying expert agreement. (Alfakhry et al., 2023; Katser et al., 2025) (Abdulghani & Jellil, 2024; To the Moaleem et al., 2025; Slieman & Camarata, 2019).

The sample size used, higher than that of much of the previous research in Dentistry, in line with a large number of authors, allowed a preliminary analysis of the internal consistency and internal structure of the rubrics to be carried out with a reasonable degree of stability. The decision to exclude students from international groups responded to the need to minimize possible biases derived from the linguistic understanding of perception surveys. However, this decision reduces the generalizability of the results and suggests the desirability of future studies that include more heterogeneous and multicenter samples. Similarly, the inclusion of a single evaluator favored homogeneity in the application of the criteria, but prevents the assessment of interevaluator reliability, an aspect that is especially relevant in studies on evaluation by means of rubrics. (Alfakhry et al., 2023; Mittal et al., 2024; Tsai et al., 2022) (Al Ansari et al., 2017; Bujang et al., 2018; Horsten et al., 2013; Wollenschläger et al., 2016) (Ormond et al., 2019)

In relation to academic performance, the results showed lower average grades than those described in previous academic years. Although this difference could be compatible with a stricter or more explicit application of the evaluation criteria, the study design does not allow this finding to be causally attributed to the use of rubrics, since other potentially influential factors, such as teacher changes, cohort characteristics or differences in the academic context, were not controlled. Therefore, this comparison should be interpreted with caution and understood as a contextual element rather than as a demonstration of the effect of the evaluative intervention. Regarding differences by sex, the absence of significant differences in most indicators coincides with what has been described in the literature, where the results are generally inconsistent or of limited magnitude, but, in general, there are no statistically significant differences. (Daghrery et al., 2024; Liang et al., 2024) (Boar et al., 2025)

The students' perception of the use of rubrics was initially high in both practices. However, after the evaluation, a decrease was observed in some indicators, especially in the practice of wire bending. This pattern has been previously described and can be related to several factors, both a more realistic understanding of the demands of the task after practical experience, a greater critical capacity of the student, as well as the possible influence of the grade obtained on the perception of the evaluation system. The fact that the subsequent assessment was positively associated with the grade obtained suggests, precisely, that the perception of the instrument does not depend only on its formal characteristics, but also on the academic result experienced by the student. Despite this variability, the surveys used showed adequate internal consistency, which supports their usefulness as instruments to explore the perception of students in similar contexts. (Cockett & Jackson, 2018; Lau et al., 2021; McKenzie, 2013) (Saunders et al., 2018)

From the psychometric point of view, the analyzed rubrics showed favorable initial evidence, although of unequal magnitude. The cephalometry rubric reached an adequate level of internal consistency ($\alpha = 0.71$), while the wire bending rubric presented a more modest value ($\alpha = 0.61$). The latter result must be interpreted with caution. Although it is below the conventional threshold of 0.70, it can be considered reasonable in a brief instrument, composed of only three items, and aimed at assessing a task with an important psychomotor component, where a certain heterogeneity of performance is to be expected. In fact, there are authors who advocate a greater number of items to be evaluated. (Nemec et al., 2018) (Alfehaid et al., 2018; Katser et al., 2025) (Arunachalam et al., 2023)

Regarding the internal structure, the exploratory factor analysis showed results compatible with a one-dimensional organization both in the rubrics and in the satisfaction surveys. However, this interpretation should also be formulated with caution, especially in the case of the wire bending

 doi:10.20944/preprints202604.1394.v1



heading, given the small number of items. In this sense, rather than talking about a definitive confirmation of one-dimensionality, the results can be considered preliminary evidence of sufficient internal coherence to justify the use of a global score in this specific context. These findings are, broadly speaking, consistent with previous studies that have shown that rubrics can exhibit acceptable psychometric functioning when their criteria are clearly defined and aligned with the performance assessed. However, future research should incorporate more robust approaches to ordinal data, such as alpha ordinal, omega coefficients or factor analyses based on polychoric matrices, in order to more accurately estimate the metric properties of this type of instrument. (Baig et al., 2014; Brown et al., 2019; Oh et al., 2018)

In relation to the associations with academic variables, the results suggest that not all practices provide the same type of information on student performance. Performance in wire bending did not show a significant association with other practices, while the cephalometry seminar was positively related to performance in the corresponding practice, and this, in turn, was associated with theoretical performance even after controlling for the effect of the seminar. Rather than evidencing a strong predictive capacity of the rubrics, these results point to an academically significant relationship between certain training activities and performance in tasks that integrate practical and cognitive components. These findings are consistent with previous studies that question the use of single tests of manual dexterity as predictors of future performance but recognize their usefulness as partial indicators within a broader training process. Another major problem that has been pointed out is that students, even having the rubrics with which they are going to be evaluated, continue to score themselves above what the teaching staff would do. (Gadbury-Amyot et al., 2014; Normon et al., 2006; Othman et al., 2023) (Alghilan et al., 2025; Ellakany et al., 2023; Pérez-Higueras et al., 2025)

The present work has several limitations that must be considered when interpreting the results. These include the fact that the study was carried out in a single centre, the use of a single assessor, the absence of inter-rater reliability analysis, the exclusion of international groups and the cross-sectional nature of the design, which makes it impossible to assess the temporal stability of the results. Likewise, the ordinal nature of the scales used and the small number of items in one of the rubrics make it advisable to interpret the psychometric results as initial evidence, not as a definitive validation of the instrument.

Overall, the results support the use of rubrics as potentially useful instruments to structure preclinical evaluation in Dentistry and provide greater transparency to the evaluation process. However, their consolidation as more psychometrically robust tools will require additional studies with multicenter designs, multiple evaluators and more advanced analytical strategies.

## 5. Conclusions

The results of this study indicate that rubrics are a viable tool for the evaluation of preclinical practices in Dentistry, with a psychometric behavior consistent with their design and an observable relationship with academic performance in certain activities. These findings provide initial evidence that supports its usefulness in structuring preclinical assessment and guides future research on competency-based assessment in the dental field.

## References

Abdulghani, M. N., & Jellil, E. I. (2024). Effectiveness of Using Analytical Grading Method for Crown Preparation Performed by Dental Students—An In-Vitro Study. *Indian Journal of Dental Research*, *35*(4), 438–442. https://doi.org/10.4103/ijdr.ijdr_308_24

Al Amri, M. D., Sherfudhin, H. R., & Habib, S. R. (2018). Effects of Evaluator's Fatigue and Level of Expertise on the Global and Analytical Evaluation of Preclinical Tooth Preparation. *Journal of Prosthodontics*, *27*(7), 636–643. https://doi.org/10.1111/jopr.12558

 doi:10.20944/preprints202604.1394.v1



Al Ansari, A., Strachan, K., Hashim, S., & Otoom, S. (2017). Analysis of psychometric properties of the modified SETQ tool in undergraduate medical education. *BMC Medical Education*, *17*(1), 1–9. https://doi.org/10.1186/s12909-017-0893-4

Al Moaleem, M. M., Adawi, H. A., Alahmari, N. M., Sayed, M. E., Albar, N. H., Okshah, A., Meshni, A. A., Gadah, T. S., Alshehri, A. H., & AL-Makramani, B. M. A. (2025). Using an analytic rubric system for the evaluation of anterior ceramic crown preparation performed by preclinical dental students. *PLoS ONE*, *20*(4 April), 1–17. https://doi.org/10.1371/journal.pone.0318752

Alfakhry, G., Mustafa, K., Ybrode, K., Jazayerli, B., Milly, H., Abohajar, S., Hassan, H., Alhomsi, K., & Jamous, I. (2023). Evaluation of a workplace assessment method designed to improve self-assessment in operative dentistry: a quasi-experiment. *BMC Medical Education*, *23*(1), 1–11. https://doi.org/10.1186/s12909-023-04474-z

Alfehaid, L. S., Qotineh, A., Alsuhebany, N., Alharbi, S., & Almodaimegh, H. (2018). The Perceptions and Attitudes of Undergraduate Healthcare Sciences Students of Feedback: A Qualitative Study. *Health Professions Education*, *4*(3), 186–197. https://doi.org/10.1016/j.hpe.2018.03.002

Alghilan, M. A., Munaga, S., & Aldakhil, A. (2025). A Comparison of Dental Students' Self-Assessment and Instructors' Assessment in Competency Examinations in a Preclinical Operative Dentistry Course. *Advances in Medical Education and Practice*, *16*(February), 177–188. https://doi.org/10.2147/AMEP.S493719

Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. In *CBE Life Sciences Education*. https://doi.org/10.1187/cbe.06-06-0168

Arunachalam, S., Pau, A., Nadarajah, V. D., Babar, M. G., & Samarasekera, D. D. (2023). Entrustable professional activities in undergraduate dental education: A practical model for development and validation. *European Journal of Dental Education*, *27*(2), 332–342. https://doi.org/10.1111/eje.12809

Baig, M., Ali, S. K., Ali, S., & Huda, N. (2014). Evaluation of multiple choice and short essay question items in basic medical sciences. *Pakistan Journal of Medical Sciences*. https://doi.org/10.12669/pjms.301.4458

Baracco, B., Escribano, N., Da Silva, D., Belliard, V., Ceballos, L., & Fuentes, V. (2025). Factors influencing the agreement between teachers and students in the assessment of preclinical endodontics using a rubric. *BMC Medical Education*, *25*(1). https://doi.org/10.1186/s12909-025-07193-9

Bharuthram, S., & Patel, M. (2017). Co-constructing a rubric checklist with first year university students: A self-assessment tool. *Apples—Journal of Applied Language Studies*. https://doi.org/10.17011/apples/urn.201708073430

Bhoopathi, V., & Atchison, K. A. (2015). Pilot testing a predoctoral dental public health curriculum: Reporting process and outcomes. *Journal of Public Health Dentistry*, *75*(5), S25–S32. https://doi.org/10.1111/jphd.12125

Bindayel, N. A. (2017). Reliability of rubrics in the assessment of orthodontic oral presentation. *Saudi Dental Journal*, *29*(4), 135–139. https://doi.org/10.1016/j.sdentj.2017.07.001

Brown, A., Nidumolu, A., McConnell, M., Hecker, K., & Grierson, L. (2019). Development and psychometric evaluation of an instrument to measure knowledge, skills, and attitudes towards quality improvement in health professions education: The Beliefs, Attitudes, Skills, and Confidence in Quality Improvement (BASiC-QI) Scale. *Perspectives on Medical Education*, *8*(3), 167–176. https://doi.org/10.1007/s40037-019-0511-8

Bujang, M. A., Omar, E. D., & Baharum, N. A. (2018). A review on sample size determination for cronbach's alpha test: A simple guide for researchers. *Malaysian Journal of Medical Sciences*, *25*(6), 85–99.

Chambers, & Labarre. (2014). Why professional judgment is better than objective description in dental faculty evaluations of student performance. *Journal of Dental Education*, *78*(5), 681–693. https://doi.org/10.1111/j.1600-0579.1998.tb00029.x

Cockett, A., & Jackson, C. (2018). The use of assessment rubrics to enhance feedback in higher education: An integrative literature review. *Nurse Education Today*, *69*, 8–13. https://doi.org/10.1016/j.nedt.2018.06.022

Consuelo Menzala-Peralta, C. I., & Ortega-Menzala, E. I. (2021). Rubrics as an assessment instrument in higher education Rubric como instrumento de avaliação no ensino superior. *Educational Sciences Research Article*, *7*(ISSN: 2477-8818), 1020–1034. http://dominiodelasciencias.com/ojs/index.php/es/index

Daghrery, A., Alwadai, G. S., Alamoudi, N. A., Alqahtani, S. A., Alshehri, F. H., Al Wadei, M. H., Abogazalah, N. N., Pereira, G. K. R., & Al Moaleem, M. M. (2024). Students' performance in clinical class II composite

 doi:10.20944/preprints202604.1394.v1



restorations: a case study using analytic rubrics. *BMC Medical Education*, *24*(1). https://doi.org/10.1186/s12909-024-06261-w

de Vet, H. C. W., Mokkink, L. B., Mosmuller, D. G., & Terwee, C. B. (2017). Spearman–Brown prophecy formula and Cronbach's alpha: different faces of reliability and opportunities for new applications. *Journal of Clinical Epidemiology*, *85*, 45–49. https://doi.org/10.1016/j.jclinepi.2017.01.013

Donaldson, J. H., & Gray, M. (2012). Systematic review of grading practice: Is there evidence of grade inflation? *Nurse Education in Practice*, *12*(2), 101–114. https://doi.org/10.1016/j.nepr.2011.10.007

Ellakany, P., El Tantawi, M., Al-Eraky, D. M., Aly, N. M., & Al-Eraky, M. M. (2023). Self-assessment and peer-assessment of the psychomotor skills in preclinical prosthodontics: A prospective study. *European Journal of Dental Education*, *27*(3), 601–609. https://doi.org/10.1111/eje.12846

Fields, H. W., Kim, D. G., Jeon, M., Firestone, A. R., Sun, Z., Shanker, S., Mercado, A. M., Deguchi, T., & Vig, K. W. L. (2017). Evaluation of objective structured clinical examination for advanced orthodontic education 12 years after introduction. *American Journal of Orthodontics and Dentofacial Orthopedics*, *151*(5), 840–850. https://doi.org/10.1016/j.ajodo.2016.10.031

Fluit, C. R. M. G., Bolhuis, S., Grol, R., Laan, R., & Wensing, M. (2010). Assessing the quality of clinical teachers: A systematic review of content and quality of questionnaires for assessing clinical teachers. *Journal of General Internal Medicine*, *25*(12), 1337–1345. https://doi.org/10.1007/s11606-010-1458-y

Francis, J. E. (2018). Linking rubrics and academic performance: An engagement theory perspective. *Journal of University Teaching and Learning Practice*, *15*(1).

Gadbury-Amyot, C. C., McCracken, M. S., Woldt, J. L., Brennan, R. L., C.C., G.-A., M.S., M., J.L., W., & R.L., B. (2014). Validity and reliability of portfolio assessment of student competence in two dental school populations: a four-year study. *Journal of Dental Education*, *78*(5), 657–667. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L373340928%5Cnhttp://bj7rx7bn7b.search.serialssolutions.com?sid=EMBASE&issn=19307837&id=doi:&atitle=Validity+and+reliability+of+portfolio+assessment+of+student+competence+in+two+dent

Gansky, S. a, Pritchard, H., Kahl, E., Mendoza, D., Bird, W., Miller, A. J., & Graham, D. (2004). Reliability and validity of a manual dexterity test to predict preclinical grades. *Journal of Dental Education*, *68*(9), 985–994. http://www.ncbi.nlm.nih.gov/pubmed/15342660

Habib, S. R. (2018). Rubric system for evaluation of crown preparation performed by dental students. *European Journal of Dental Education*, *22*(3), e506–e513. https://doi.org/10.1111/eje.12333

Heathman, A. K., Mitchell, T. V., Van Ness, C. J., Godley, L. W., Ehrenberger, C. J., Wendland, M. E., & Gadbury-Amyot, C. C. (2021). Reliability of a grading rubric designed to evaluate the reflective ability of predoctoral dental students. *European Journal of Dental Education*, *25*(4), 664–669. https://doi.org/10.1111/eje.12644

Hernán-Losada, I. (2009). Conclusions on the application of Bloom's Taxonomy to the design of pedagogical tools. (March). http://eciencia.urjc.es/handle/10115/3525

Horsten, L. R., Beltràn, C. A., & Moreno, M. P. (2013). Validation of a work performance evaluation instrument. *Ibero-American Journal of Psychology: Science and Technology*.

Jabali, A., Alroomy, R., Mashyakhy, M., Bajawi, H., Al Malwi, A. A., & Karobari, M. I. (2025). Rubric system for the evaluation of root canal treatment of single canal tooth performed by preclinical dental students: a cross-sectional study. *Saudi Dental Journal*, *37*(1). https://doi.org/10.1007/s44445-025-00008-z

Jauhar, P., Mossey, P. A., Popat, H., Seehra, J., & Fleming, P. S. (2016). A survey of undergraduate orthodontic teaching and factors affecting pursuit of postgraduate training. *Bdj*, *221*, 487. http://dx.doi.org/10.1038/sj.bdj.2016.778

Jornet Meliá, J. (2017). Standardized evaluation. *Ibero-American Journal of Educational Evaluation*, *10*(1), 5–8.

Kanthan, R., & Senger, J. L. B. (2011). An appraisal of students' awareness of "self-reflection" in a first-year pathology course of undergraduate medical/dental education. *BMC Medical Education*, *11*(1). https://doi.org/10.1186/1472-6920-11-67

Katser, M. S., Veremis, B. M., Danciu, T. E., Ramaswamy, V., & Popov, V. (2025). Rallying for Reflection: Pilot Use of Rubric to Facilitate Self-Reflection in Dental Education. *European Journal of Dental Education*, 1–8. https://doi.org/10.1111/eje.13088

 doi:10.20944/preprints202604.1394.v1



Lau, M. N., Kamarudin, Y., Zakaria, N. N., Sivarajan, S., Tahir, N. N. Z. M., Bahar, A. D., Naimie, Z., Othman, S. A., & Wan Hassan, W. N. (2021). Comparing flipped classroom and conventional live demonstration for teaching orthodontic wire-bending skill. *PLoS ONE*, *16*(7 July), 1–12. https://doi.org/10.1371/journal.pone.0254478

Liang, L., Nagasawa, M., Ha, V., Lin, A. J., Akiba, Y., Akiba, N., Yamakami, S. A., Uoshima, K., & Ohyama, H. (2024). Association between gender and self-assessment skills amongst Japanese dental students. *Journal of Dental Sciences*, *19*(3), 1533–1539. https://doi.org/10.1016/j.jds.2023.12.025

López-Pastor, V. M. (2011). The role of formative assessment in competency-based assessment: contributions from the formative and shared assessment network in university teaching. *REDU. Journal of University Teaching*. https://doi.org/10.4995/redu.2011.6185

Mays, K. A., & Branch-Mays, G. L. (2016). A Systematic Review of the Use of Self-Assessment in Preclinical and Clinical Dental Education. *Journal of Dental Education* ■, *80*(8), 902–913. http://www.jdentaled.org/content/80/8/902.full.pdf

McKenzie, C. T. (2013). Dental student perceptions of case-based educational effectiveness. *Journal of Dental Education*, *77*(6), 688–694. http://www.ncbi.nlm.nih.gov/pubmed/23740905

Mittal, P., Jadhav, G. R., Pawar, M., Banerjee, S., Wangaskar, S., Di Blasio, M., Matterhorn, G., & Minervini, G. (2024). Impact of self-assessment on dental student's performance in pre-clinical conservative dentistry course. *BMC Oral Health*, *24*(1), 1–7. https://doi.org/10.1186/s12903-024-04140-w

Morrow, J. A., Pulido, M. T., Smith, P. B., McDaniel, T. F., & Willcox, A. B. (2014). Effective use of e-grading in the dental simulation clinic. *Journal of Dental Education*, *78*(6), 829–837. http://www.ncbi.nlm.nih.gov/pubmed/24882768

Nemec, E. C., Baker, D. M., Zhang, D., & Dintzner, M. (2018). Development of valid and reliable tools for student evaluation of teaching. *Currents in Pharmacy Teaching and Learning*, *10*(5), 549–557. https://doi.org/10.1016/j.cptl.2018.02.009

Normon, G., Eva, K., Brooks, L., & Hamstra, S. (2006). Expertise in Medicine and Surgery. In *Cambridge Handbook of Expertise and Expert Performance*. https://doi.org/10.1017/CBO9780511816796

Oh, S. L., Liberman, L., & Mishler, O. (2018). Faculty calibration and students' self-assessments using an instructional rubric in preparation for a practical examination. *European Journal of Dental Education*, *22*(3), e400–e407. https://doi.org/10.1111/eje.12318

Ormond, K. E., Hallquist, M. L. G., Buchanan, A. H., Dondanville, D., Cho, M. K., Smith, M., Roche, M., Brothers, K. B., Coughlin, C. R., Hercher, L., Hudgins, L., Jamal, S., Levy, H. P., Raskin, M., Stosic, M., Uhlmann, W., Wain, K. E., Currey, E., & Faucett, W. A. (2019). Developing a conceptual, reproducible, rubric-based approach to consent and result disclosure for genetic testing by clinicians with minimal genetics background. *Genetics in Medicine*. https://doi.org/10.1038/s41436-018-0093-6

Othman, S. A., Kamarudin, Y., Sivarajan, S., Soh, E. X., Lau, M. N., Zakaria, N. N., Wey, M. C., Wan Hassan, W. N., Bahar, A. D., Mohd Tahir, N. @. N. Z., Razi, R. M., & Naimie, Z. (2023). Students' perception on flipped classroom with formative assessment: A focus group study. *European Journal of Dental Education*, *27*(3), 419–427. https://doi.org/10.1111/eje.12823

Pasillera Díaz, M., Fullana Noell, J., Planas Lladó, A., & del Valle Gómez, A. (2010). Adaptation to the European Higher Education Area in Spain: the changes/challenges involved in competency-based teaching and guidelines to respond to them. *Ibero-American Journal of Education*, *52*(4), 1–13.

Pérez-Higueras, J. J., Hidalgo Arroquia, J. J., & Gancedo-Caravia, L. (2025). Rubric for peer evaluation of oral presentations: Use and perceptions among experienced and non-experienced students. *Journal of Dental Education*, (July 2024), 1–10. https://doi.org/10.1002/jdd.13831

Sanderson, T. R., Kearney, R. C., Kissell, D., & Salisbury, J. (2016). Search Evaluating Student Self-Assessment through Video- Recorded Patient Simulations.

Satheesh, K. M., Brockmann, L. B., Liu, Y., & Gadbury-Amyot, C. C. (2015). Use of an Analytical Grading Rubric for Self-Assessment: A Pilot Study for a Periodontal Oral Competency Examination in Predoctoral Dental Education. *Journal of Dental Education*, *79*(12), 1429–1436.

 doi:10.20944/preprints202604.1394.v1



Saunders, C., Palesy, D., & Lewis, J. (2018). Systematic review and conceptual framework for health literacy training in health professions education. *Health Professions Education*. https://doi.org/10.1016/j.hpe.2018.03.003

Schwibbe, A., Kothe, C., Hampe, W., & Konradt, U. (2016). Acquisition of dental skills in preclinical technique courses: influence of spatial and manual abilities. *Advances in Health Sciences Education*, *21*(4), 841–857. https://doi.org/10.1007/s10459-016-9670-0

Segura Egea, J. J., & Jiménez Rubio-Manzanares, A. (2013). Methodological proposal for the evaluation of clinical practices in dentistry (pp. 1–34).

Slieman, T. A., & Camarata, T. (2019). Case-Based Group Learning Using Concept Maps to Achieve Multiple Educational Objectives and Behavioral Outcomes. *Journal of Medical Education and Curricular Development*, *6*, 238212051987251. https://doi.org/10.1177/2382120519872510

Smith, K. M., Geletta, S., & McArdle, A. (2016). The Use of Rubrics in the Clinical Evaluation of Podiatric Medical Students Objectification of the Subjective Experience. *Journal of the American Podiatric Medical Association*, *106*(1), 60–67. https://doi.org/10.7547/14-038

Tenkumo, T., Fuji, T., Ikawa, M., Shoji, S., Sasazaki, H., Iwamatsu-Kobayashi, Y., Okuyama, Y., Matsui, H., Shiraishi, N., Furuuchi, T., Higuchi, K., Hashimoto, W., Takeuchi, Y., Simauchi, H., & Sasaki, K. (2019). Introduction of integrated dental training, jaw models and rubric criteria. *European Journal of Dental Education*, *23*(1), e17–e31. https://doi.org/10.1111/eje.12395

Tsai, S. T., Ho, Y. C., Tsai, C. L., Yang, S. F., Lai, Y. L., & Lee, S. Y. (2022). Evaluation of students' self-assessment performance in preclinical endodontic training by means of rubrics and a 3D printed model. *Journal of the Formosan Medical Association*, *121*(11), 2203–2210. https://doi.org/10.1016/j.jfma.2022.03.021

Vahtivuori-Hänninen, S., Karaharju-Suvanto, T., & Suomalainen, K. (2007). Characteristics of pedagogical models in the mobile teaching-studying-learning (TSL) environments: Preliminary findings of the I-trace project. *1st International Workshop on Pen-Based Learning Technologies, PLT 2007*. https://doi.org/10.1109/PLT.2007.5

Velasco-Martínez, L. C., Díaz-Barriga, F., & Tójar-Hurtado, J.-C. (2017). Acquisition and Evaluation of Competencies by the Use of Rubrics. Qualitative Study on University Faculty Perceptions in Mexico. *Procedia—Social and Behavioral Sciences*, *237*(June 2016), 869–874. https://doi.org/10.1016/j.sbspro.2017.02.185

Velasco-Martínez, L. C., Díaz-Barriga, F., Tójar-Hurtado, J.-C., Trace, J., Meier, V., Janssen, G., Smith, K. M., Geletta, S., McArdle, A., Sadid-Zadeh, R., Nasehi, A., Davis, E., Katsavochristou, A., Ali, S., Shoaib, H., Rehman, R., Panadero, E., Romero, M., Strijbos, J. W., ... Jonsson, A. (2016). Acquisition and Evaluation of Competencies by the Use of Rubrics. Qualitative Study on University Faculty Perceptions in Mexico. *Nurse Education Today*, *26*(4), 94–99. https://doi.org/10.1016/j.sbspro.2017.02.185

Wollenschläger, M., Hattie, J., Machts, N., Möller, J., & Harms, U. (2016). What makes rubrics effective in teacher-feedback? Transparency of learning goals is not enough. *Contemporary Educational Psychology*. https://doi.org/10.1016/j.cedpsych.2015.11.003

Zawawi, K., Afify, A. R., Othman, H. I., & Al-Dharrab, A. A. (2013). Correlation of psychomotor skills and didactic performance among dental students in Saudi Arabia. *Advances in Medical Education and Practice*, 223. https://doi.org/10.2147/AMEP.S53319