BMC Medical Education

RESEARCH

Open Access

Factors influencing the agreement between teachers and students in the assessment of preclinical endodontics using a rubric

B. Baracco[1], N. Escribano[1*], D. Da Silva[1], V. Belliard[1], L. Ceballos[1] and V. Fuentes[1]

Abstract

Background Students' self-assessment and rubrics are pedagogical tools designed to enhance learning and evaluation processes. Achieving convergence between learners and teachers is paramount during the learning of complex treatments such as endodontic procedures, although it can be challenging due to uncontrollable factors that may affect task performance and perception. This study aimed to evaluate the influence of various factors (anatomical difficulty, type of instrumentation system, and training level) on the assessments of root canal treatments (RCT) by teachers and students using a rubric, and to determine the degree of agreement between them.

Methods 144 RCT were performed on extracted human molars by 36 dental students using two mechanized systems and subsequently evaluated using a rubric by both the students and four teachers. Rubric yielded a total score for the sum of 4 items analyzed individually: radiographic evaluation, access cavity, instrumentation, and obturation. The influence of the three following factors: anatomical difficulty (categories: minimal, moderate and high), instrumentation system (categories: Protaper Next and Reciproc Blue), and training level (categories: initial and advanced) on teachers' and students' RCT total scores and for each item were analyzed by three-way ANOVA. Agreements between teachers and students were measured by intraclass correlation coefficients and quadratic weighted Kappa. Statistical analyses were conducted at a pre-set alpha of 0.05 using Stata 16.

Results No significant influence of anatomical difficulty, instrumentation system, or training level was observed on RCT total scores given by teachers and students ($p > 0.05$). However, training level influenced the assessments by teachers in instrumentation as those by students in radiographic evaluation. Agreement was moderate for RCT total scores and substantial for teeth with minimal or high difficulty, after using Protaper Next, and at the initial training level. Agreement was substantial in obturation, and moderate in radiographic evaluation, access cavity, and instrumentation. Some categories in instrumentation and obturation items showed substantial agreement, while fair agreement was observed only in access cavity.

Conclusions Educational factors in Endodontics such as anatomical difficulty, mechanized instrumentation system, and level of training did not influence the total scores of preclinical RCT by teachers or students. Agreement for RCT total scores ranged from moderate to substantial. The strongest agreement was observed in obturation, while access cavity yielded the lowest. The rubric as an assessment tool in Endodontics teaching is recommended.

*Correspondence:
N. Escribano
nuria.escribano@urjc.es

Full list of author information is available at the end of the article

Keywords Agreement, Self-assessment, Endodontics training, Anatomical difficulty, Rubric, Root Canal treatment, Dental education

Background

Self-assessment is recognized as a continuous and reflective progress of self-knowledge and self-criticism for anyone embarking on a learning experience. This process is highly beneficial for students, as it increases their awareness of academic progression and facilitates the identification of areas needing improvement [1, 2]. Consequently, self-evaluation also serves as a motivating factor, enhancing students' autonomy and responsibility [3]. Additionally, it provides teachers with a valuable tool to deeply understand students' learning pace, enabling them to adapt pedagogical strategies and resources to meet students' needs [4, 5].

The incorporation of rubrics as an assessment method aims to align the perspectives of teachers and students as closely as possible [4]. Rubrics provide guidance and objectivity in evaluating students across various disciplines, enhancing consistency among multiple assessors [6]. However, the criteria of experienced evaluators, the teachers, and novice practitioners, the students, may differ despite the use of rubrics, especially in situations that are difficult to standardize, which is quite common in clinical subjects of Health Sciences [7, 8].

In Dentistry, the use of rubrics as an assessment tool has grown in recent years and in various contexts, such as preclinical training in Integrated Dentistry [9] and Prosthodontics [10] or clinical performance in Periodontics [11]. Endodontics teaching has also explored this approach and, although references are scarce, there are recent studies on the application of rubrics for assessing endodontic treatments [12, 13] or portfolios [14], as shown in a previous publication by the authors of the present study.

Performing an entire root canal treatment (RCT) requires an organized sequence of several clinical stages, in accordance with the European Society of Endodontology (ESE) quality guidelines [15]. The development of RCT can be influenced by various recognizable and, to a certain extent, modifiable preconditions, as well as by unexpected events occurring during the procedure [16]. The former includes a vast anatomical variability of pulpal chambers, roots, and canals, which can lead to highly defiant situations [17].

To address some of these challenges, mechanized endodontic instrumentation has become increasingly prevalent in clinical practice [18, 19]. However, all new technologies require a learning curve for the operator, who must acquire the necessary skills and competencies to use them to their clinical benefit [20, 21].

Training students to excel in a clinical environment necesitates confronting realistic scenarios, working on natural teeth, and mastering various instrumentation systems [18, 22]. Encountering these challenging cases, especially during initial experiences, can alter students' self-evaluation perceptions [2] and even influence teachers' criteria, affecting the alignment between them. Since agreement enhances learning outcomes [4, 23, 24], understanding how these factors may affect alignment is crucial for improving Endodontics education.

While the ESE indicates that training in Endodontics for undergraduate students involves acquiring skills in an appropriate preclinical environment, it acknowledges that defining a minimum level of knowledge or skills-based input is not possible. Nor is it feasible or desirable to provide definitive directions to schools on course planning and delivery. Consequently, despite the importance of learning endodontics, there is a significant lack of specific assessment procedures [25]. It is well known that when clinical skills are practiced without evaluation, errors are reinforced [26]. Therefore, it can be inferred that preclinical training not evaluated with structured criteria, will result in inadequate clinical practices.

Given these challenges, this study aimed to (1) evaluate whether anatomical complexity, instrumentation system, and training level influence the assessment of RCT by teachers and students using a rubric, and (2) determine the level of agreement between them. Accordingly, the null hypotheses to be tested were: (1) Factors such as anatomical difficulty, instrumentation system, or training level do not influence the RCT total scores or the individual item assessments by teachers and students; and (2) Teachers and students show a similar level of agreement in their assessments of total scores and individual items in preclinical RCT.

Materials and methods

Performance of preclinical RCT

This study was conducted at Rey Juan Carlos University (Madrid, Spain). The Ethics committee of this institution considered that explicit approval was not necessary (document with registration number 2002201904819), as the present study was carried out once the mandatory preclinical practice concluded, so it did not interfere with the students' grades.

Fourth-year undergraduate dental students performed the RCT during the preclinical endodontic training for the course Dental Pathology and Therapeutics II, being the second year using mechanized instrumentation as well as the evaluation rubric. The treatments were carried

out by a total of 36 students, comprising all the students enrolled in the course during that academic year. Each student completed four RCT on extracted human molars ($n = 144$), as part of the five-week preclinical block of the course.

The students provided the teeth, while the teachers, applying the exclusion criteria, selected which ones would be used. Molars exhibiting substantial loss of coronal structure, canals not radiographically visible, canal obliteration, extreme curvatures, incomplete root formation, extensive apical resorption, or internal resorption, were discarded. Consequently, no attempt was made to distribute the anatomical complexity of the molars or the instrumentation system among the students in any specific manner.

Afterwards, the approximate working length (WL) of each root canal was determined by superimposing a file over the preoperative radiograph and measuring the distance between a coronal reference point and each radiographic apex. This distance was then reduced by 0.5 mm to ensure that the working length would approximately reach the location of the apical constriction. During the conductometry radiograph, it was verified whether the initially established WL was correct or needed modification, so that the preparation and obturation would reach the area of the apical constriction. Access cavities were created using high-speed diamond burs under water-cooling, followed by canal location with an endodontic probe. Students scouted the root canals with a K-file diameter 10, achieving apical patency at WL + 0.5 mm. Irrigation with 5.25% sodium hypochlorite delivered by syringe was maintained throughout the entire shaping procedure.

Canal instrumentation was performed using Protaper Next (continuous rotary motion, $n = 74$) or Reciproc Blue (reciprocating motion, $n = 70$), both manufactured by Dentsply Sirona (Konstanz, Germany). Students were required to use both systems for the first two cases, allowing them to choose their preferred system for the remaining teeth. To calibrate the effect of training, the first and second RCT were labeled as the initial group ($n = 72$), while the third and fourth were classified as the advanced group ($n = 72$).

The obturation of all canals was performed using standardized gutta-percha points with a 0.02 taper and AH-Plus root canal sealer (Dentsply Sirona), with the cold lateral condensation technique. Students took various radiographic records of all teeth during the preclinical practice: preoperatively, before initiating endodontic treatment; during instrumentation to verify the working length in the root canals; prior to obturation to check the fit of the master gutta-percha cones; and a final radiograph upon completion of the canal obturation. The teeth were placed horizontally on their lingual or palatal surfaces, and the X-ray was directed perpendicularly to the long axis of the tooth. Periapical size 2 EF-speed X-ray films (Henry Schein, Melville, NY, USA) were used for radiographic documentation. The X-ray generator used was a Kodak 2200 Intraoral X-ray System (Carestream Dental, Atlanta, GA, USA) operated at 65 kV-DC and 7 mA. Films were processed manually using Carestream Dental X-ray processing chemicals (Carestream Dental).

Evaluation of preclinical RCT

Authors of the present study designed a previous rubric for the assessment of endodontic portfolios. After verifying the evaluators' calibration, the rubric demonstrated enhanced agreement among them [14]. The published version, with minor modifications (the deletion of two sections focused on the portfolio document), was used in the present study. This approach allowed teachers and students to concentrate exclusively on the RCT, without undermining the proven validity of the rubric, which considered four items: radiographic evaluation, access cavity, instrumentation procedure, and obturation (Appendix 1). Items were rated on a scale of 1 to 5, with 1 indicating the lowest quality of performance and 5 representing perfection. The rubric produced a RCT total score on a ten-point scale by summing the scores obtained from each item. This document was provided to the students before the practice to serve as a guide.

After completion of each RCT, students were asked to self-assess each procedure using the rubric, with no time limitations. Upon completion of this process, the sample was randomly divided into four groups (36 molars each), which were assigned to four blinded teachers experienced in Endodontics. These teachers individually evaluated the RCT using the rubric, with their analysis divided into six sessions over six consecutive days to avoid fatigue. Subsequently, the same evaluators collaboratively assessed the anatomical complexity of each molar based on the American Association of Endodontists (AAE) case difficulty form. This evaluation considered exclusively factors for extracted teeth, including crown and canal morphology, the radiographic appearance of the canals, and the presence of apical resorption. As a result, the molars difficulty was classified into the following categories: minimal ($n = 20$), moderate ($n = 88$), and high ($n = 36$) (Table 1).

Statistical analysis

Firstly, descriptive statistics were conducted on the marks given by teachers and students for RCT total scores and each rubric item. Secondly, we analyzed the influence of independent variables (anatomical difficulty, instrumentation system, and training level) on RCT total scores and rubric items for both teachers and students (three-way ANOVA). Data normality and homogeneity were tested

Table 1 Mean marks and standard deviations (sd) given by teachers and students for RCT total scores and for each item, regarding the different factors and their categories, as well as the level of agreement between them

Items of the rubric →			**RCT total scores**		**Radiographic evaluation**		**Access cavity**		**Instrumentation procedure**		**Obturation**	
		Evaluators ↓	**Mean (sd)**	**Agreement**	**Mean (sd)**	**Agreement**	**Mean (sd)**	**Agreement**	**Mean (sd)**	**Agreement**	**Mean (sd)**	**Agreement**
Marks (*n*=144)		Teachers	7.81 (1.17)	0.60 [95% CI 0.49-0.69]	4.33 (0.81)	0.48 [95% CI 0.33-0.64]	3.90 (1.00)	0.48 [95% CI 0.32-0.63]	4.08 (0.95)	0.57 [95% CI 0.42-0.71]	3.62 (0.96)	0.69 [95% CI 0.60-0.78]
		Students	7.93 (1.03)	Moderate	4.26 (0.81)	Moderate	4.15 (0.87)	Moderate	4.04 (0.79)	Moderate	3.60 (0.90)	Substantial
Factors ↓	*Categories ↓*											
Anatomical difficulty	Minimal (*n*=20)	Teachers	7.93 (1.20)	0.67 [95% CI 0.34-0.86]	4.55 (0.83)	0.53 [95% CI 0.06-0.61]	3.90 (0.97)	0.56 [95% CI 0.20-0.93]	4.25 (0.97)	0.62 [95% CI 0.35-0.88]	3.55 (1.00)	0.67 [95% CI 0.45-0.90]
		Students	7.86 (0.07)	Substantial	4.35 (0.93)	Moderate	4.10 (0.85)	Moderate	4.15 (0.67)	Substantial	3.40 (0.88)	Moderate
	Moderate (*n*=88)	Teachers	7.79 (1.19)	0.56 [95% CI 0.40-0.68]	4.35 (0.76)	0.47 [95% CI 0.26-0.67]	3.88 (1.02)	0.54 [95% CI 0.37-0.71]	4.05 (0.91)	0.44 [95% CI 0.24-0.65]	3.62 (0.93)	0.67 [95% CI 0.55-0.79]
		Students	7.96 (1.08)	Moderate	4.19 (0.76)	Moderate	4.08 (0.96)	Moderate	4.09 (0.74)	Moderate	3.70 (0.89)	Moderate
	High (*n*=36)	Teachers	7.81 (1.15)	0.69 [95% CI 0.48-0.83]	4.14 (0.90)	0.50 [95% CI 0.25-0.72]	3.94 (1.01)	0.21 [95% CI 0.09-0.52]	4.06 (1.04)	0.76 [95% CI 0.60-0.91]	3.64 (1.05)	0.74 [95% CI 0.58-0.90]
		Students	7.88 (0.99)	Substantial	4.36 (0.87)	Moderate	4.36 (0.59)	Fair	3.86 (0.96)	Substantial	3.47 (0.94)	Substantial
Instrumentation system	ProTaper Next (*n*=74)	Teachers	7.79 (1.17)	0.67 [95% CI 0.53-0.78]	4.26 (0.85)	0.41 [95% CI 0.18-0.63]	3.91 (0.98)	0.58 [95% CI 0.43-0.80]	3.97 (1.01)	0.64 [95% CI 0.45-0.83]	3.62 (0.93)	0.66 [95% CI 0.52-0.80]
		Students	7.88 (1.06)	Substantial	4.31 (0.79)	Moderate	4.14 (0.94)	Moderate	4.03 (0.83)	Substantial	3.55 (0.89)	Substantial
	Reciproc Blue (*n*=70)	Teachers	7.85 (1.18)	0.52 [95% CI 0.33-0.67]	4.40 (0.77)	0.58 [95% CI 0.38-0.78]	3.89 (1.04)	0.32 [95% CI 0.09-0.56]	4.19 (0.87)	0.47 [95% CI 0.29-0.66]	3.61 (1.00)	0.72 [95% CI 0.61-0.83]
		Students	7.98 (1.01)	Moderate	4.20 (0.83)	Moderate	4.17 (0.80)	Fair	4.06 (0.76)	Moderate	3.66 (0.92)	Substantial
Training level	Initial (*n*=72)	Teachers	7.67 (1.13)	0.61 [95% CI 0.45-0.74]	4.29 (0.88)	0.41 [95% CI 0.15-0.65]	3.88 (1.07)	0.52 [95% CI 0.32-0.71]	3.90 (1.01)	0.50 [95% CI 0.26-0.73]	3.57 (0.95)	0.64 [95% CI 0.49-0.79]
		Students	7.77 (1.09)	Substantial	4.15 (0.80)	Moderate	4.06 (0.95)	Moderate	3.99 (0.78)	Moderate	3.53 (0.90)	Substantial
	Advanced (*n*=72)	Teachers	7.96 (1.20)	0.58 [95% CI 0.41-0.72]	4.36 (0.84)	0.56 [95% CI 0.39-0.74]	3.92 (0.95)	0.42 [95% CI 0.18-0.68]	4.25 (0.85)	0.65 [95% CI 0.49-0.80]	3.67 (0.98)	0.73 [95% CI 0.63-0.84]
		Students	8.10 (0.96)	Moderate	4.36 (0.81)	Moderate	4.25 (0.78)	Moderate	4.10 (0.81)	Substantial	3.68 (0.90)	Substantial

using Shapiro-Wilk's and Levene's tests, respectively. Partial eta squared (η^2p) was calculated as a measure of effect size. According to Cohen's guidelines, η^2p values of 0.01, 0.06, and 0.14 indicate small, medium, and large effects, respectively. The absence of association between variables was verified to eliminate potential bias (Spearman correlation coefficient, $p > 0.20$). Finally, we assessed inter-rater agreement between teachers and students using intraclass correlation coefficients (ICC) for RCT total scores and quadratic weighted Kappa for rubric items, considering the three factors (anatomical difficulty, instrumentation system, and training level). Reliability results were categorized according to the Landis and Koch criteria: poor (0), slight (0.01–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect agreement (0.81-1.00). All statistical tests were performed at a pre-set alpha of 0.05 using Stata/IC 16.1 (Stata Corp LLC, College Station, TX, USA).

Results

Table 1 presents the mean marks and standard deviations (sd) for RCT total scores and for each item, as well as the level of agreement between teachers and students.

Three-way ANOVA revealed that RCT total scores from teachers and students were not significantly affected by anatomical difficulty ($p > 0.05$), the instrumentation system ($p > 0.05$), the training level ($p > 0.05$), or their interactions ($p > 0.05$). Regarding the influence of the three analyzed factors on the rubric items, students' ratings for radiographic evaluation were significantly affected by the interaction between training level and anatomical difficulty ($p = 0.013$; $\eta^2p = 0.058$). Students self-scored higher for radiographic evaluation in molars with moderate difficulty when training was advanced compared to the beginning ($p = 0.009$, Student t-test). Additionally, marks for the instrumentation procedure given by teachers were influenced by training level ($p = 0.002$; $\eta^2p = 0.073$) and the interaction between training level and anatomical difficulty ($p = 0.010$; $\eta^2p = 0.068$). Teachers' scores were higher in the advanced group than in the initial group ($p = 0.027$, Student t-test); even for low difficulty cases ($p = 0.003$, Student t-test).

The agreement between teachers and students was moderate for RCT total scores, with variation across different factors. Agreement was substantial for cases with minimal or high anatomical difficulty, and moderate for cases with moderate difficulty. Regarding instrumentation systems, Protaper Next yielded higher agreement than Reciproc Blue. Finally, agreement by training level was substantial for the initial group and decreased to moderate in the advanced group.

Regarding the items of the rubric, agreement was moderate for radiographic evaluation, access cavity, and instrumentation, and substantial for obturation. Taking into account different factors tested, agreement was moderate and substantial for all items except for cavity access in high difficulty molars and when Reciproc Blue system was used. Table 2 shows the distribution of ratings from teachers and students across the rubric for each item. For all items, ratings of 1 and 2 were the least frequent, whereas ratings of 4 and 5 collectively constituted more than 50% of the evaluations from both teachers and students. The highest rating, indicative of perfection, was selected more frequently by teachers than by students in the items of radiographic evaluation, instrumentation, and obturation. There is a general preference for higher ratings among both groups, with teachers tending to give the highest ratings more often than students in specific items.

Discussion

Like many other dental procedures, RCT encounters a wide range of situations, posing challenges to both the operator's and the evaluator's judgment. For instance, the potential impact of anatomical difficulty or the instrumentation system on daily assessments is a concern, as the former remains inherently invariable, and the latter continuously evolves. Nevertheless, our results showed that assessments by teachers and students were not influenced by any of the variables considered: anatomical complexity, type of instrumentation system and training level. Scores for the different items of the rubric were also unaffected, except for students' ratings in radiographic evaluation or teachers' scores in instrumentation procedure. Consequently, the first hypothesis must be partially accepted.

The training level contributed to these differences, with higher scores observed among teachers and students in the advanced group. While this may seem intuitive, as repeated practice is widely recognized to refine technique, enhance confidence, and improve skill acquisition [27, 28], a deeper analysis of our data reveals only the early stages of this effect. Specifically, teachers assigned higher ratings to the instrumentation item, a complex and advanced aspect of RCT, reflecting their strong critical capacity. In contrast, students rated themselves higher in the radiography section, the only category not reliant on presumed motor improvement. This raises the question of whether performing a greater number of RCT over an extended period would have led to more pronounced differences in evaluations, particularly among students at the final stages of the endodontic procedure. However, due to the retrospective nature of our study, students did not receive the teacher's correction after self-assessment. From a learning perspective, this represents a significant limitation of our study, as real-time feedback has been demonstrated to be an effective teaching method [29], even in the field of Health Sciences [30].

Table 2 Percentage distribution of teachers' and students' scores for each item of the rubric

Items →	Radiographic evaluation					Access cavity					Instrumentation procedure					Obturation				
Values →	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
Teachers	0	2.1	15.2	30.6	52.1	1.4	9.0	20.9	36.1	32.6	0.7	6.2	18.1	34.7	40.3	0	11.9	36.8	29.1	22.2
Students	0	2.8	14.6	36.8	45.8	0.7	4.9	12.5	42.3	39.6	0.7	2.8	16.7	51.3	28.5	0.7	9.7	34.7	38.2	16.7

The lack of immediate feedback for the students may have hindered their complete assimilation of the rubric. However, the intention in this study was to isolate the rubric's ability to bring students and teachers closer to consensus and to demonstrate the objectivity it provides. It should be noted that students were already familiar with this assessment tool. During their first year of preclinical endodontic treatments, they were evaluated using a rubric of similar structure and with the same evaluation items but intended for single-rooted teeth.

Our results revealed that the level of agreement between teachers and students was predominantly rated as moderate, with some categories and items showing even higher concordance. Thus, the second hypothesis was partially accepted.

Regarding anatomical difficulty, agreement was substantial for molars with high and minimal difficulty, and moderate for molars with moderate intricacy. Students might consider intermediate levels of difficulty as standard cases, assessing their performance favorably, and identifying this anatomy as the most common clinical occurrence. Interestingly, their self-score for the minimally difficult group, where third molars with a single wide canal were prevalent, was the lowest. Students might have perceived treatments in these teeth as much easier than those on molars with multiple or narrow canals, and consequently, less deserving of a higher mark, despite the rubric's supposed rigidity. This hypothetical statement suggests that students might not have adhered to the rubric as closely as teachers, as noted in other studies [31, 32]. The higher level of experience and training of the professors enables them to assess these types of cases more fairly, in which students tend to underestimate their performance.

In respect of the instrumentation systems, the level of agreement was higher for Protaper Next than for Reciproc Blue. Although students were allowed to choose the system for 2 out of 4 molars, the distribution of mechanized systems in the sample was uniform. Students had practiced with Protaper Next in the previous course, but this was their first exposure to the reciprocating system. Therefore, certain characteristics of the system itself might explain the higher self-assessments after using Reciproc Blue. The reduced number of files required and the perceived increased control with the reciprocating motion may have given them the impression that errors were less likely to occur during the instrumentation sequence, as previously reported for undergraduate students [33].

Regarding the training level, agreement was higher in the initial group compared to the advanced group. Although a clear trend of improvement was observed throughout the practice, the self-assigned scores of students in the advanced group were the highest recorded

in the entire study, which reduced the level of consensus. As previously mentioned, the self-empowering effect of repetitive exercises [9, 27, 31] may have been manifested in two ways. On the one hand, repetition genuinely improved students' skills and performance, as reflected in the fact that teachers' ratings for the advanced group were the highest they assigned. On the other hand, this ongoing practice may have led students to subjectively overestimate their performance during the final stage of their learning.

RCT involves the completion of various interdependent procedures, which were evaluated separately in our rubric, in accordance with previous studies [11, 34]. Agreement between teachers and students for different items was consistent, with radiographic evaluation, access cavity and instrumentation scoring moderate, while obturation achieved substantial agreement. This stronger agreement in obturation could be attributed to the quantitative nature of this item in the rubric, such as counting canals with radiographically unfilled spaces. In contrast, access cavity and instrumentation included subjective terms. Notably, the access cavity was the only item where no substantial agreement was recorded, with some fair agreements observed. Sections linked to intuitive appraisals may confuse students, thereby reducing both the objectivity intended by the rubric and the agreement between multiple assessors [4].

Limited research has focused on the agreement between teachers and students in Endodontics, yielding varied outcomes. Two studies from Saudi Arabia identified a clear divergence, although their methodologies were not entirely comparable to ours. AlRahabi et al. [35] considered only single RCT performed by novice students in Endodontics, which could be determinant for dissent [36, 37]. Almohaimede [12] reported disagreement in clinical RCT because students overestimated their efficiency. It is important to note the significant differences between their education system and that of a country like Spain [38, 39]. Cultural norms, collectivistic versus individualistic values, teacher authority, and gender segregation [39] influence the relationships between teachers and students [40, 41]. In a cultural context more similar to our study, in Taiwan, adequate concordance between teachers and students was observed when evaluating RCT on resin printed teeth [13]. Our results are consistent with theirs, as the strongest agreement was found in obturation and the weakest in access cavity. Although this first manual procedure may seem less clinically relevant, its correct execution prevents potentially iatrogenic mishaps at later stages of the RCT [18]. In our study, many of the access cavities were excessively small, which the rubric considered a penalty, specifically assigning it the second-lowest rating (Appendix 1). A plausible explanation for this frequent finding may lie in two factors. From an educational standpoint, as previously noted, students may not fully adhere to the rubric, reducing their attention to its content (notably, the mention of an access cavity preserving the pulp chamber roof appears at the end of the rubric cell). From a clinical perspective, prior experience with single-rooted teeth may lead students to replicate smaller access cavities, potentially misinterpreting the larger pulp chamber dimensions in molars.

The capacity of self-assessment may be influenced by inherent student characteristics, such as cultural factors, gender, and academic performance [37, 42, 43]. However, neither our study nor the previously mentioned studies considered these aspects. For instance, it has been shown that dental students in the United States of America tend to overrate their performance more frequently than students in Japan [44]. Additionally, male dental students generally overestimate their abilities more than their female counterparts [45], although some studies have not detected this difference, even in groups with a clear female predominance [46]. In our study, 75% of the students were women.

To the authors' knowledge, no studies in the Western context have analyzed the influence of characteristic factors on the evaluation of RCT by teachers and students, nor the quality of the agreement between them. Future research should consider a broader range of variables related to learners and the specific characteristics of RCT, with the aim of improving educational methodologies in Endodontics.

Conclusions

Factors potentially affecting the assessment of Endodontics in an educational preclinical environment, such as dental anatomical difficulty, type of mechanized instrumentation system and training level, did not influence the RCT total scores by either teachers or students.

A specifically designed rubric allowed teachers and students to achieve predominantly moderate agreement, supporting its use as an assessment tool in the practical teaching of Endodontics. Among the RCT components, the highest concordance was observed in obturation, while the lowest occurred in access cavity preparation. This suggests that the initial stages of RCT may be more challenging to standardize and achieve strong inter-rater agreement.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-025-07193-9.

Supplementary Material 1

Acknowledgements
None.

Author contributions
BB, LC, NE and VF participated in the conception and design of the study. BB, NE, DS and VB designed the rubric and evaluated all the endodontic treatments. DS and VB obtained rubrics from the students' self-assessment. Coordination of the evaluation process was carried out by BB and NE. Statistical analysis was done by VF. BB wrote the manuscript. BB and VF prepared the tables. LC, NE and VF performed a critical review. All authors read and approved the final manuscript.

Funding
None.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
The authors submitted the protocol to the Ethics Committee of Rey Juan Carlos University. This board issued a report in which it clearly stated that this study did not require express approval. The need for informed consent from the participants was waived by the Ethics Committee of Rey Juan Carlos University. This report is available upon request. All methods in this study were carried out in accordance with guidelines and regulations in the Declaration of Helsinki.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Faculty of Health Sciences, IDIBO Research Group, Universidad Rey Juan Carlos, Madrid, Spain

Received: 26 November 2024 / Accepted: 17 April 2025
Published online: 28 April 2025

References

1. Taras M. Student self-assessment: processes and consequences. Teach High Educ. 2010;15(2):199–209.
2. Andrade H, Valtcheva A. Promoting learning and achievement through self-assessment. Theory Pract. 2009;48(1):12–9.
3. Yan Z, Brown GTL. A cyclical self-assessment process: towards a model of how students engage in self-assessment. Assess Eval High Educ. 2017;42(8):1247–62.
4. Panadero E, Jonsson A. The use of scoring rubrics for formative assessment purposes revisited: A review. Educ Res Rev. 2013;9:129–44.
5. Ross JA. The reliability, validity, and utility of self-assessment. PARE. 2006;11(1):10.
6. Jonsson A, Svingby G. The use of scoring rubrics: reliability, validity and educational consequences. Educ Res Rev. 2007;2(2):130–44.
7. Piano M, Diemer K, Hall M, Hui F, Kefalianos E, Lawford BJ, et al. A rapid review of challenges and opportunities related to diversity and inclusion as experienced by early and mid-career academics in the medicine, dentistry and health sciences fields. BMC Med Educ. 2023;23(1):288.
8. Inayah AT, Anwer LA, Shareef MA, Nurhussen A, Alkabbani HM, Alzahrani AA, et al. Objectivity in subjectivity: do students' self and peer assessments correlate with examiners' subjective and objective assessment in clinical skills? A prospective study. BMJ Open. 2017;7(5):e0122897.
9. Tenkumo T, Fuji T, Ikawa M, Shoji S, Sasazaki H, Iwamatsu-Kobayashi Y, et al. Introduction of integrated dental training jaw models and rubric criteria. Eur J Dent Educ. 2019;23(1):e17–31.
10. Habib SR. Rubric system for evaluation of crown Preparation performed by dental students. Eur J Dent Educ. 2018;22(3):e506–13.
11. Satheesh KM, Brockmann LB, Liu Y, Gadbury-Amyot CC. Use of an analytical grading rubric for self-assessment: A pilot study for a periodontal oral competency examination in predoctoral dental education. J Dent Educ. 2015;79(12):1429–36.
12. Almohaimede AA. Comparison between students' self-evaluation and faculty members' evaluation in a clinical endodontic course at King Saud university. Eur J Dent Educ. 2022;26(3):569–76.
13. Tsai ST, Ho YC, Tsai CL, Yang SF, Lai YL, Lee SY. Evaluation of students' self-assessment performance in preclinical endodontic training by means of rubrics and a 3D printed model. J Formos Med Assoc. 2022;121(11):2203–10.
14. Escribano N, Belliard V, Baracco B, Da Silva D, Ceballos L, Fuentes MV. Rubric vs. numeric rating scale: agreement among evaluators on endodontic treatments performed by dental students. BMC Med Educ. 2023;23(1):197.
15. Löst C. Quality guidelines for endodontic treatment: consensus report of the European society of endodontology. Int Endod J. 2006;39(12):921–30.
16. Ng YL, Mann V, Rahbaran S, Lewsey J, Gulabivala K. Outcome of primary root Canal treatment: systematic review of the literature - Part 2. Influence of clinical factors. Int Endod J. 2008;41(1):6–31.
17. Vertucci FJ. Root Canal morphology and its relationship to endodontic procedures. Endod Top. 2005;10(1):3–29.
18. Arias A, Peters OA. Present status and future directions: Canal shaping. Int Endod J. 2022;55(Suppl 3):637–55.
19. Cheung MC, Peters OA, Parashos P. Global survey of endodontic practice and adoption of newer technologies. Int Endod J. 2023;56(12):1517–33.
20. Koch M, Eriksson HG, Axelsson S, Tegelberg Å. Effect of educational intervention on adoption of new endodontic technology by general dental practitioners: A questionnaire survey. Int Endod J. 2009;42(4):313–21.
21. Parashos P, Messer HH. The diffusion of innovation in dentistry: A review using rotary nickel-titanium technology as an example. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2006;101(3):395–401.
22. Luz D, dos Ourique S, Scarparo F, Vier-Pelisser RK, Morgental FV, Waltrick RD. Preparation time and perceptions of Brazilian specialists and dental students regarding simulated root canals for endodontic teaching: A preliminary study. J Dent Educ. 2015;79(1):56–63.
23. Meyer ML, McDonald SA, DellaPietra L, Wiechnik M, Dasch-Yee K. Do students overestimate their contribution to class? Congruence of student and professor ratings of class participation. JoSoTL. 2018;18(3):44–54.
24. Lai TL. Effects of student-teacher congruence on students' learning performance. Soc Sci Q. 2015;96(5):1424–35.
25. De Moor R, Hülsmann M, Kirkevang LL, Tanalp J, Whitworth J. Undergraduate curriculum guidelines for endodontology. Int Endod J. 2013;46(12):1105–14.
26. Dunnington GL, Angeles L, Wright K, Hoffman K. Pilot experience with competency-based clinical skills assessment in a surgical clerkship. Am J Surg. 1994;167(6):604–6.
27. Duvivier RJ, Van Dalen J, Muijtjens AM, Moulaert VR, Van Der Vleuten CP, Scherpbier AJ. The role of deliberate practice in the acquisition of clinical skills. BMC Med Educ. 2011;11(1):101.
28. Anders Ericsson K, Harwell KW. Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: why the original definition matters and recommendations for future research. Front Psychol. 2019;10:2396.
29. Maurer TW, Kropp JJ. The impact of the immediate feedback assessment technique on course evaluations. Teach Learn Inq. 2015;3(1):31–46.
30. Fatteh R, Phayo Y, Raza M, Naing TT, Phyo Z, Arja SB. A multi-institutional study regarding the perceptions of students and faculty members about constructive feedback for medical students in medical education. Adv Med Educ Pract. 2024;15:1361–71.
31. Shipman D, Roa M, Hooten J, Wang ZJ. Using the analytic rubric as an evaluation tool in nursing education: the positive and the negative. Nurse Educ Today. 2012;32(3):246–9.
32. Gan S, Bhattamisra SK, Mohd Z, Ho KL. Pharmacy students' and supervisors' perceptions of the effectiveness of assessment rubrics for undergraduate research performance review. Pharm Educ. 2023;23(1):193–204.
33. Kwak SW, Cheung GSP, Ha JH, Kim SK, Lee H, Kim HC. Preference of undergraduate students after first experience on nickel-titanium endodontic instruments. Restor Dent Endod. 2016;41(3):176.
34. Vantorre T, Bécavin T, Deveaux E, Marchandise P, Chai F, Robberecht L. Are the evaluation criteria used in preclinical endodontic training courses relevant? A preliminary study. Aust Endod J. 2020;46(3):374–80.
35. AlRahabi MK, AlKady AM, Ghabbani HM. Agreement between faculty member assessments and student self-assessments in a preclinical endodontic programme. Aust Endod J. 2019;45(3):346–51.

36. Yan Z, Panadero E, Wang X, Zhan Y. A systematic review on students' perceptions of self-assessment: usefulness and factors influencing implementation. Educ Psychol Rev. 2023;35(3):81.
37. Tuncer D, Arhun N, Yamanel K, Çelik Ç, Dayangaç B. Dental students' ability to assess their performance in a preclinical restorative course: comparison of students' and faculty members' assessments. J Dent Educ. 2015;79(6):658–64.
38. Alharbi EAR. Higher education in Saudi Arabia: challenges to achieving world-class recognition. IJCH. 2016;2(4):169–72.
39. Hakiem RAAD. I can't feel like an academic': gender inequality in Saudi Arabia's higher education system. High Educ. 2023;86(3):541–61.
40. Xu C, Huizinga M, De Luca G, Pollé S, Liang R, Sankalaite S, et al. Cultural universality and specificity of teacher-student relationship: a qualitative study in Belgian, Chinese, and Italian primary school teachers. Front Psychol. 2023;14:1287511.
41. Fabris MA, Lin S, Longobardi C. A cross-cultural comparison of teacher-student relationship quality in Chinese and Italian teachers and students. J Sch Psychol. 2023;99:101227.
42. Lee C, Asher SR, Chutinan S, Gallucci GO, Ohyama H. The relationship between dental students' assessment ability and preclinical and academic performance in operative dentistry. J Dent Educ. 2017;81(3):310–7.
43. Benson LM, Martin AJ, Williams PA. The role of demographic factors in university students' self-assessment skills. J Educ Psychol. 2021;113(4):763–78.
44. Tabassian LJ, Nagasawa M, Ba AK, Akiba N, Akiba Y, Uoshima K, et al. Comparing dental student preclinical self-assessment in the united States and Japan. J Dent Educ. 2022;86(1):21–8.
45. Kornmehl DL, Patel E, Agrawal R, Harris JR, Ba AK, Ohyama H. The effect of gender on student self-assessment skills in operative preclinical dentistry. J Dent Educ. 2021;85(9):1511–7.
46. Liang L, Nagasawa M, Ha V, Lin AJ, Akiba Y, Akiba N, et al. Association between gender and self-assessment skills amongst Japanese dental students. J Dent Sci. 2024;19(3):1533–9.

Publisher's note