

Continuous Workplace-based Assessment as an Indication of Clinical Competence in Paediatric Dentistry

Nadia Mohamed and Liezl Smit

Abstract

An authentic workplace setting provides the ideal opportunity for assessment of students' clinical competence at the 'does' level of performance. Final-year dental students in the Department of Paediatric Dentistry at the University of the Western Cape are evaluated in the clinical environment on a daily basis through multiple clinical evaluations which assess clinical and diagnostic skills over a year. An additional end-of-module clinical assessment in the form of a single blinded patient case (BPC) determines if students have reached the expected level of clinical competence in terms of patient evaluation and diagnosis. However, the reliability and feasibility of this single end-of-module clinical case have been questioned in this setting. This study aimed to determine if the current continuous workplace-based assessment (WPBA) results could be used as an indication of final-year students' clinical competence at the end of the module. A retrospective, quantitative, cross-sectional study was conducted of all complete assessment records. The correlation between the continuous WPBA components was analysed together with an evaluation of the reliability and validity of the assessment results. The continuous formative WPBA practices were found to be both valid and reliable when using Kane's (2013) and Royal's (2017) frameworks for analysis. However, the BPC should be reconsidered due to feasibility and reliability concerns.

Key words: Dental education, Paediatric Dentistry, clinical skills, workplace, summative, formative, continuous assessment

Un milieu de travail authentique offre l'occasion idéale d'évaluer les compétences cliniques des étudiants. Les étudiants de dernière année en médecine

ABOUT THE AUTHORS: NADIA MOHAMED, University of the Western Cape, South Africa. Email: namohamed@uwc.ac.za and LIEZL SMIT, University of Stellenbosch, South Africa

dentaire du Département de dentisterie pédiatrique de l'Université du Cap-Occidental sont évalués quotidiennement dans l'environnement clinique grâce à de multiples évaluations cliniques qui évaluent les compétences cliniques et diagnostiques sur une année. Une évaluation clinique supplémentaire en fin de module sous la forme d'un seul cas de patient à l'insu (BPC) détermine si les étudiants ont atteint le niveau attendu de compétence clinique. Cependant, la fiabilité et la faisabilité de ce cas clinique simple de fin-de-module ont été remises en question dans cet arrangement. Cette étude vise à déterminer si les résultats actuels de l'évaluation continue en milieu de travail (EBA) pouvaient être utilisés comme une indication des compétences cliniques des étudiants de dernière année à la fin du module. Une étude rétrospective, quantitative et transversale a été menée sur tous les dossiers d'évaluation complets. La corrélation entre les composantes continues de l'EBA a été analysée en même temps qu'une évaluation de la fiabilité et de la validité des résultats de l'évaluation. Les pratiques formatives continues de l'EBA se sont avérées à la fois valides et fiables lors de l'utilisation des cadres d'analyse de Kane (2013) et de Royal (2017). Toutefois, le BPC devrait être réexaminé en raison de préoccupations liées à la faisabilité et à la fiabilité.

Mots clés: éducation dentaire, dentisterie pédiatrique, compétences cliniques, milieu de travail, évaluation sommative, formative, continue

Introduction

Regardless of the discipline, the end result of training health professionals is to ensure that graduates are clinically competent and can treat patients safely. Workplace-based assessment (WPBA) is thought to be the most authentic way to assess clinical competence as students are evaluated in the work environment and have to solve real life clinical problems as they present. Workplace-based assessment practices are characterised by feedback and include reflection and skills development to encourage lifelong learning (Harris et al., 2017). Students are assessed at the 'does' level of Miller's pyramid (Beard, 2011; Harris et al., 2017). This uppermost tier of Miller's pyramid is synonymous with integration and higher-order thinking (Downing and Yudkowsky, 2009; Miller, 1990) which requires that students are also able to incorporate the lower tiers of the pyramid, which include knowledge and application. Such integration of theory and clinical practice is essential to be able to treat a patient holistically (Wimmers, 2006).

The Health Professions Council of South Africa (HPCSA, 2014) has incorporated an adapted version of the 2005 CanMEDS competency framework (The Royal College of Physicians and Surgeons of Canada, 2015) into its guidelines; where competencies are defined as observable and measur-

able abilities that, when actively integrated in practice, constitute health professional competence (Frank et al., 2010). Teaching, learning and assessment strategies should therefore enable students to develop clinical competence and ensure that competencies have been achieved and can be applied at the end of a course (Hays, Hamlin and Crane, 2015; South African Qualifications Authority, 2001).

The overall undergraduate dentistry curriculum at the University of the Western Cape in Cape Town, South Africa comprises six disciplines or modules of which Paediatric Dentistry is one. In order for students to sit the final examination, which will enable them to graduate from the dental programme, they are expected to pass each of the six modules individually during the fifth/final year of study. The final integrated examination is in a written, case-based format and includes contributions from all the disciplines within Dentistry. The need for sound assessment of the paediatric clinical competence of final-year students is thus critical during the course.

Assessment in the workplace can provide an ideal opportunity for assessment of clinical competencies in the context of professional practice (Beard, 2011; Epstein and Hundert, 2002) as students are evaluated on actual procedures that they will encounter in their working environment. In clinical disciplines, WPBA methods include Clinical Encounter Cards (CEC), Blinded Patient Encounters (BPE), Direct Observation of Procedural Skills (DOPS), Case-based Discussions (CbD), and the Mini-Clinical Evaluation Exercise (mini-CEX) (Norcini and Burch, 2007). All these methods have a feedback component based on specified criteria and are considered formative in nature (Norcini and Burch, 2007).

The mini-CEX has been widely used in the workplace-based environment (Pelgrim et al., 2011) and can be applied to a variety of settings (Norcini and Burch, 2007) including Dentistry. It is an abbreviated version of the traditional clinical examination (CEX) and consists of single patient encounters (Pelgrim et al., 2011) which are conducted over a period of time (Norcini, Blank, Duffy and Fortna, 2003).

Final year dental students in Paediatric Dentistry at the University of the Western Cape are assessed through continuous WPBA practices. These include daily direct observation assessments of student-patient interactions in the workplace over the course of a year (mini-CEX), completion of a minimum clinical quota of procedures, and two case presentations.

Due to the disruptions caused by student protests (#FeesMustFall) in October/ November 2015, final examinations for that year were cancelled. This forced the department to promote students based on their continuous assessment results. The question therefore arose as to whether these results were an accurate reflection of clinical competence.

There is a paucity of research on the application of WPBA tools in the

dentistry setting. This study therefore aimed to determine if the current continuous WPBA results in the Department of Paediatric Dentistry could be used as an indication of the clinical competence of final-year students at the end of the module and whether they could be used to make pass/fail decisions.

Methods

Paediatric Dentistry assessment

Assessment in the discipline of Paediatric Dentistry takes place in the clinical setting as well as in tutorials and traditional test and examination environments. In the final year, greater emphasis is placed on continuous assessment which includes a clinical component comprising of various WPBA tasks, as well as written tests on theoretical content. The clinical component carries a higher overall weighting of 60%, with 40% for the theoretical component (Figure 1). As this article focuses on clinical competence, the theoretical component will not be discussed further.

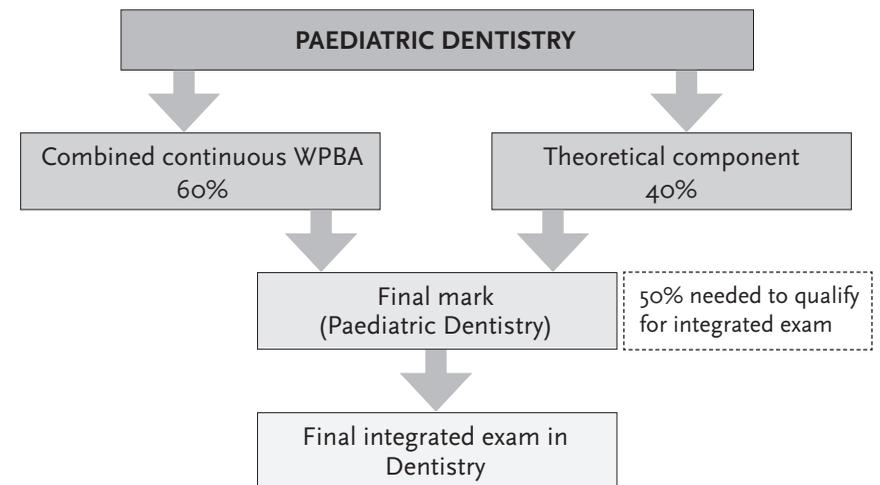


Figure 1: Overall assessment plan

A variety of assessment methods are used to determine whether a student can be considered clinically competent (Figure 2). Students are evaluated and receive feedback on all clinical aspects of Paediatric Dentistry, including their ability to formulate a diagnosis based on history taking, integration of knowledge, ability to address the patient's main complaint and actual treatment procedures. An end-of-module paediatric clinical assessment is also required to assist pass/fail decisions in the discipline. This assessment task

comprises a single blinded patient case (BPC)' however, the reliability of the single clinical case has been questioned (Harris et al., 2017).

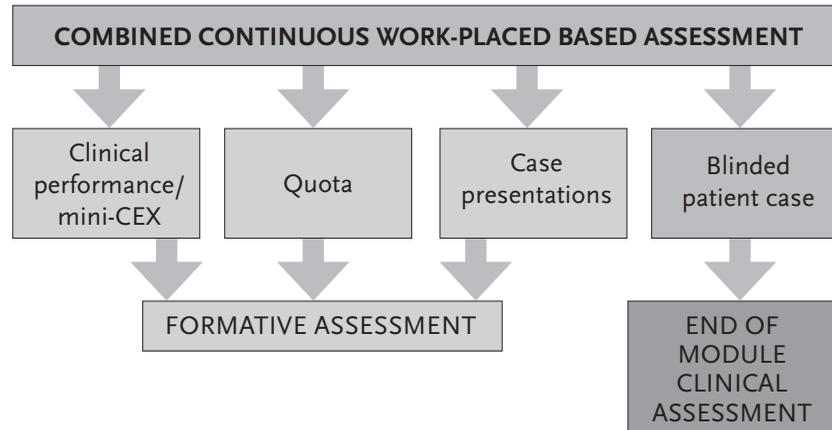


Figure 2: Clinical assessment components

Mini-CEX evaluations and logbook quotas

The department accommodates approximately 80 to 85 final-year students in the clinics every year over a period of roughly 30 weeks. Each student sees an average of ten patients during this time. These patients return for numerous visits so that all the treatment they require is completed. The initial visit is a diagnostic one where a treatment plan is drawn up. Subsequent appointments are dedicated to executing the treatment plan and completing all the treatment the patient requires. Students are supervised when treating these patients under local anaesthesia, general anaesthesia or sedation. This provides the ideal opportunity to assess students performing varied professional tasks in a range of workplace-based environments.

For the purposes of this article, the direct observation assessments of student-patient interactions are referred to as mini-CEX encounters although a rubric relevant to the Paediatric Dentistry context is used in place of the original mini-CEX scoring sheet published by Norcini and Burch (2007). This holistic clinical rubric with specific criteria (See appendix) is used to assess clinical ability, knowledge and application of knowledge as well as diagnostic and problem-solving skills.

Similar to the implementation of the mini-CEX in other clinical settings (Al Ansari, Kauser and Donnon, 2013; Norcini et al., 2003; Norcini and Burch, 2007), a student's performance is assessed by various staff members through direct observation whilst treating a range of patients.

All staff members are rotated through the clinics and students are thus exposed to a number of different staff during the year.

Based on staff and student feedback, the implementation of the clinical rubric in its current format has been shown to be feasible. The evaluation criteria are explicit and staff are required to provide feedback (in keeping with the principles of formative assessment) after each clinical session. Staff members discuss each student's performance with them. However, in order to be beneficial, students should understand why certain standards were not met and how to improve (Hays et al., 2015; Yorke, 2003). Specific verbal feedback relevant to the clinical session is provided. Students are required to reflect on their performance and highlight what they did well and where they feel they can improve. This approach is in line with recommendations in the literature that emphasise the need to develop an "action plan" which provides specific information on how students need to proceed to achieve the learning outcomes (Norcini and Burch, 2007, p. 866).

Staff are also required to record comments on the rubric to provide more detail regarding student performance. This is in line with other authors who encourage an additional narrative as opposed to merely recording a grade (Harris et al., 2017; Govaerts, Van de Wiel, Schuwirth, van der Vleuten and Muijtjens, 2013). These narratives allow scores to be interpreted more meaningfully (Govaerts et al., 2013). The grade allocated by the clinical supervisor is discussed with the student and both parties are required to sign the rubric as acknowledgement that this discussion has taken place. This improves accountability of the supervisor as well as the student and is the department's way of ensuring "sustainable assessment and feedback strategies" as recommended in the literature (Norcini and Burch, 2007, p. 869).

As most of the module credits are devoted to clinical time in the clinical workplace-based setting, the average of the mini-CEX encounters is calculated to reflect student performance throughout the year. Clinical practice is evaluated in conjunction with logbook experience (Beard, 2011) in order to calculate the final clinical mark. The mini-CEX episodes are therefore directly linked to procedures which form part of an expected clinical quota. Student protests have also forced management to question the value of logbook quotas and how they correlate with actual clinical competence.

Case presentations

Students are required to present patients seen in the workplace as part of a case for discussion. They motivate their treatment choices and receive feedback from their peers and supervisors. This is regarded as a continuation of WPBA as real patient cases are discussed (Norcini and Burch, 2007). The assessment focusses on clinical reasoning and the rationale for the clini-

cal choices (Norcini and Burch, 2007), thus honing critical thinking skills (Popil, 2011). Through presentations of their peers' patient cases, students are also exposed to a range of real-world patient cases that they would normally not have had access to (Popil, 2011).

Blinded Patient Case (BPC)

At the end of the module, students are assessed using a single BPC that is unknown to them. This is similar to the BPE reported on in the literature (Norcini and Burch, 2007). This assessment evaluates the student's ability to formulate a diagnosis based on history taking, integration of knowledge and ability to address the patient's main complaint.

The patients used for the BPC are not standardised, are unpredictable and have to be rescheduled if they fail to attend their appointments. In addition, as children younger than 12 are used in these assessments, expecting them to sit for more than one student examination is not feasible as they tire easily and have a limited attention span. Different patients are therefore booked, calling into question the fairness of this assessment. Students are also assessed by different examiners. These challenges raise questions regarding the reliability of this assessment.

Target and study population

The target population consisted of final-year students from the Paediatric Dentistry Department at the University of the Western Cape over a two-year period (2016-2017).

Data collection and management

The study participants were identified from routine university records. A retrospective, quantitative, cross-sectional study was conducted of all complete clinical assessment records. This included the completed mini-CEX rubrics (hard copies) as well as the routine annual departmental spreadsheets containing all clinical components of students' assessment marks. All final-year students were included, irrespective of final fail/pass/repeat assessment results. Incomplete records were excluded from the study. The theoretical component of the assessment was not included.

Data analysis

The data was analysed using STATA version 15. Descriptive statistics were used to assess the characteristics of the continuous WPBA components, i.e., means, proportions, standard deviation, median and maximum and minimum scores. Workplace-based assessment scores for each clinical encounter were plotted to track student progress. A Spearman's rank-order correlation was run to assess the relationship between the various

assessment components. This was illustrated using correlation matrices. The closer the correlation coefficient is to the value of 1, the better the correlation. One-sample t-tests were used to determine whether there was a statistically significant mean difference between various scores, i.e., when the p-value was less than or equal to 0.05.

In addition to statistical analysis, two qualitative approaches were used to assess the reliability and validity of the WPBA practices. Driessen et al.'s (2005) principles of credibility and dependability were applied to assess the reliability of the assessment, and validity was evaluated using a combined framework based on Kane (2013) and Royal (2017).

The characteristics of, and correlation between the continuous WPBA components were analysed together with an evaluation of the reliability and validity of the assessment results to ascertain whether inferences drawn from these results could assist the department to make decisions regarding the assessment practices currently employed.

Qualitative approach

Clinical supervisors evaluate and judge critical thinking and the quality of student responses when students motivate treatment choices and apply theory to the clinical context as part of the mini-CEX evaluations and case presentations. This is a subjective interpretation of the student's abilities, with assessment information presented as qualitative data.

An alternative qualitative approach to evaluate the reliability of assessment tasks was suggested where markers for credibility and dependability are included based on the supervisor's judgement (Driessen, van der Vleuten, Schuwirth, van Tartwijk, and Vermunt, 2005). Credibility or trustworthiness of an assessment must be supported by evidence and this can be achieved through three strategies, namely, triangulation (i.e., combining information from different sources), prolonged engagement over time and member checking which includes student discussion and feedback. Dependability refers to the quality assurance processes in place (Driessen et al., 2005). Using this qualitative analysis strategy (Driessen et al., 2005), the principles of credibility and dependability were applied to the assessment practices based on assessment records.

Ethical considerations

All data (including rubrics and spreadsheets) were handled and managed according to Good Clinical Practice (GCP) requirements and ethical standards. Students whose assessment scores were included in the study were assigned a unique student identifier number. This data was entered into an electronic database (an Excel spreadsheet). Student numbers and names were not entered into the electronic database, thereby ensuring confiden-

tiality. A request was made for a waiver of individual informed consent as this was a retrospective review of routinely collected data with minimal risk. Ethical clearance was obtained from the Health Research Ethics Committee at Stellenbosch University (X18/02/002) as well as the University of the Western Cape (UWCCRP070318NM).

Results

The study population consisted of 57 and 58 students in the 2016 and 2017 cohorts, respectively, i.e., a total of 115 complete records. Sixty students whose full complement of records was not available were excluded from the study. This included 36 records for 2016 and 24 for 2017.

A summary of the study cohort assessment data is presented in Table 1 with mean (%) and minimum and maximum scores for each assessment component (with standard deviations).

Table 1: Summary of assessment data for the cohort (n=115)

Assessment component	Average mini-CEX scores (%) [with SD]	Average case presentations (%) [with SD]	Logbook quota (%) [with SD]	Blinded patient case (%) [with SD]	Combined Continuous WPBA score (%) [with SD]
Mean	58.9 [4.3]	65.3 [6.4]	63.2 [11.6]	67.8 [8.2]	61.9 [4.9]
Minimum score	50	36.5	35	45	49.3
Maximum score	67.5	81.2	90	83	74.9

As depicted in Table 2, positive correlations were found between most of the assessment components.

Table 2: Spearman correlation coefficients between various assessment components (correlation row 1, p-value row 2)

	Average mini-clinical evaluation	Cases average	Logbook Quota	Blinded patient case	Combined continuous WPBA
Average mini-clinical evaluation	1.000				
Case presentation average	0.4114 0.0000	1.0000	0.3710 0.0000	0.4848 0.0000	
Logbook quota (procedures)	0.2550 0.0060		1.0000	0.0973 0.3011	
Combined continuous WPBA	0.7218 0.0000	0.7229 0.0000	0.7192 0.0000	0.5289 0.0000	1.0000
Blinded patient case	0.4704 0.0000			1.0000	

Moderate positive correlation was found between the average mini-CEX and the BPC ($r_s = 0.47, p < 0.0001$) which was statistically significant. The participants performed better in the scores for BPC (Table 1), 67.8% (8.2) than those for their average mini-clinical evaluation, 58.9% (4.3), a statistically significant difference of -8.82% (95% CI, -10.16 to -7.48), $p < 0.0001$.

The combined continuous WPBA moderately correlated with the BPC ($r_s = 0.52, p < 0.001$) which was statistically significant (Table 2). The participants performed better in the single clinical assessment, 67.8 (8.2) (Table 1) than in their combined continuous WPBA, 61.9 (4.9). A statistically significant difference was noted, -5.74 (95% CI, -7.03 to -4.45), $t(114) = -8.79, p < 0.0001$.

The combined continuous WPBA scores were also strongly correlated with the average mini-CEX score ($r_s = 0.72, p < 0.001$) as well as quota and cases ($r_s = 0.72, p < 0.001$). All of these findings were statistically significant (Table 2).

The principles of credibility and dependability (Driessen et al., 2005) were applied to the assessment practices based on the assessment records. A summary of the evaluation is presented in Table 3.

Table 3: Evaluation of reliability of the assessment approach using qualitative strategies (Driessen et al., 2005)

	Strategies	Application to Paediatric Dentistry
Credibility of assessment Dependability	Triangulation (combining information from different sources)	✓ Varied assessment used (mini-CEX, logbook quota, cases for discussion, blinded patient case)
	Prolonged engagement over time	✓ Multiple assessments over time ✓ Multiple examiners ✓ Multiple and varied patient cases
	Member checking	✓ Student feedback ✓ Student involvement in assessment process
	Quality assurance	✓ Rubrics with explicit criteria ✓ All staff attended training ✓ Rubrics are completed with detailed notes ✓ Regular evaluation and adaptation of assessment practices based on staff and student feedback

This qualitative analysis suggests that the continuous formative WPBA results are reliable.

Validity of WPBA

In order to structure an argument for validity, a combination of two frameworks was used that most accurately represented the department's WPBA practices, i.e., Kane's four inferences (2013) and Royal's four tenets (2017). The findings of the application of these frameworks are summarised in Table 4.

Table 4: Validity argument for WPBA

Tenets/ inferences	Application to Paediatric Dentistry
Scoring	Use of a detailed clinical rubric Staff trained to use rubric Evaluates both clinical and theoretical aspects of each patient case
Generalisation/ validity evidence	Multiple assessments Multiple examiners Varied patient cases
Implications	Correlation of scores between WPBA components WPBA as an indicator of clinical competence Value of end-of-module clinical assessment

** Adapted from Royal (2017) and Kane (2013)

This analysis suggests that the continuous formative WPBA results are valid.

Discussion

No previous studies could be directly compared with the combination of assessment practices employed by the Department of Paediatric Dentistry. Where applicable, the individual components are thus discussed in relation to the literature.

Formative results as predictors of summative performance

The mean scores obtained for the end-of-module BPC were statistically significantly higher than the average scores obtained in all the other assessment components during the module (Table 1). This is similar to the findings reported in a 2009 study where success in formative assessment tests was associated with better summative marks (Carrillo-de-la-Peña, Baillès, Caseras, Martínez, Ortet and Pérez, 2009). Two other studies support formative assessment's positive contribution to summative results (Anziani, Durham and Moore, 2008; Riaz, Yasmin S and Yasmin R, 2015).

It should be noted that the end-of-module assessment comprises of a single patient case. These clinical cases are not standardised and vary in terms of presentation of problems and complexity. The reliability of these results therefore has to be questioned. This score may thus not be a true reflection of competence when compared to the other assessments that take place over a period of time. However, the fact that the scores obtained for the end-of-module BPC were better even though a more structured, detailed assessment process was used, can also be regarded as evidence of improved performance as students were more thoroughly interrogated.

The continuous assessment results could be a better indication of student performance and could be considered when making a judgement call on whether or not a student should pass in this setting. Harris et al. (2017) concur that multiple assessments are a more accurate indicator of whether or not a student is fit to progress to the next level.

The value of multiple and varied assessments

Clinical practice requires a range of characteristics, as no single method of assessment is likely to provide sufficient data to make a valid and reliable judgement of competency (Norcini and Burch, 2007). Assessment methods should therefore be fit for purpose; with an understanding of the information they can provide. Individual assessment methods have strengths and weaknesses and issues of reliability, validity and feasibility should be considered. Furthermore, multiple and varied assessment tasks offer students the opportunity to showcase different competencies and strengths, making the assessment fairer (Downing and Yudkowsky, 2009; UWC Assessment Policy, 2018).

While most WPBA methods can be used for formative assessment on their own, multimodal assessment across a time period is recommended

for summative decisions (Schuwirth and van der Vleuten, 2010). Clinical competence assessment in the Paediatric Dentistry programme is pitched at the 'does' level of Miller's pyramid (Miller, 1990; Wass, van der Vleuten, Shatzer and Jones, 2001) and various WPBA methods are used over time to assess performance.

Because clinical skills are essential for a dentist, it is important that competence in performing clinical procedures is assessed. A logbook is used to monitor the number and type of clinical procedures students have completed. Whilst Dahllöf, Tsilingaridis and Hindbeck (2004) reported on the use of logbooks for reflection purposes and Anziani et al. (2008) used logbooks to compare formative and summative assessment scores, no studies have examined the correlation between clinical performance and the number of procedures completed. The average mini-CEX score in this study showed only a weak positive correlation with the logbook quota (Table 2). The results show that the number of procedures does not equate to clinical competence and that a more holistic view of student performance should be considered.

Presentation of patient cases seen in the clinical area enables students to appraise their chosen treatment options and defend their decisions with appropriate motivation from the literature. This helps to hone critical thinking and problem-solving skills as well as improve understanding of complex issues through integration of concepts (Popil, 2011; McDade, 1995).

The findings of this study suggest that the combined WPBA components provide a more accurate reflection of student competence than a single assessment task. This is in line with the principles of programmatic assessment which state that more than one component should be evaluated to obtain a more holistic view of student competence (van der Vleuten, Schuwirth, Driessen, Govaerts, and Heeneman, 2015). The combined continuous WPBA score includes all clinical work and a quota of procedures. It is therefore a true representation of the workplace as integration of theoretical knowledge and application thereof to the clinical context is assessed. This integrated assessment is important to determine clinical competence.

The case for reliability

Jonsson and Svingby (2007) suggest that multiple examiners and rubrics with explicit criteria strengthen reliability, especially if the criteria are strictly adhered to. In Paediatric Dentistry, all departmental staff have been trained to use the rubrics and emphasis was placed on identifying specific criteria to substantiate the mark that is allocated. Regular reinforcement is nonetheless essential to ensure that staff remain focussed on the purpose of the assessment. This helps to improve accuracy when completing the rubric.

The learning outcomes and the purpose of the assessment were used in

designing the rubric. As recommended by Gibson and Shaw (2010), specific aspects/characteristics that are linked to the learning outcomes and that need to be evaluated were identified and incorporated. Rubrics that are specifically designed for a particular clinical context also enhance reliability (Jonsson and Svingby, 2007). In this case, the clinical rubric was specifically designed for the Paediatric Dentistry context.

With the once-off clinical assessment, students can perform very well or very poorly depending on the type of patient case on that particular day. This could either be attributed to the complexity of the patient case or the student's inability to handle high-pressure situations. This was confirmed by Wass et al. (2001), who highlighted that these conditions could influence the reliability of the assessment.

The criteria delineated in the clinical rubric portray more than one description for each level (see appendix). It has been shown to be reproducible and the fact that staff and students often agree on the same score, attests to the clarity of the assessment criteria stipulated on the rubric. Overt performance where actual skills are assessed is easier to measure (Andreatta and Gruppen, 2009). However, the theoretical component that underpins the clinical practice is more open to subjective interpretation because covert performance like clinical reasoning is assessed (Andreatta and Gruppen, 2009). Examiner subjectivity comes into play with the mini-CEX evaluation, making it vulnerable to bias. On the other hand, as posited by Pelgrim et al. (2011), multiple assessors counter the effect of assessor subjectivity.

It should be borne in mind that factors like inherent personality traits, and staff's beliefs and opinions can also influence how students are scored (Tziner, Murphy and Cleveland, 2005). Scoring can therefore never really be "objective" as there is "no single true score"; however, all perspectives are valid (Govaerts and van der Vleuten, 2013, p. 1170). The fact that the average mini-CEX marks are recorded levels the playing field to some extent as any discrepancies resulting from subjective interpretation on the part of the clinical supervisor are somewhat mitigated.

Strong correlation between the combined continuous WPBA score and the individual WPBA components indicates good reliability of the assessment. Students are assessed by multiple supervisors who rotate through the clinic at different times and treat a broad spectrum of patients with varied needs. Reliability and validity are thus improved as students are assessed by more than one supervisor on multiple occasions over a period of time. This is in line with arguments presented in the literature (van der Vleuten and Schuwirth, 2005; Downing, 2004). Pelgrim et al. (2011) found that an acceptable level of reliability can be achieved using a minimum of ten encounters. The number of encounters included in the present study exceeds this as students have two encounters per week over a roughly 30-week period. As

noted by van der Vleuten and Schuwirth (2005), a larger sample accounts for variance and minimises errors, thereby improving reliability. The outcome of the assessment can therefore be considered reliable.

Based on Driessen et al.'s (2005) qualitative approach to reliability, the continuous formative WPBA practices in Paediatric Dentistry can be regarded as credible and dependable, and therefore reliable (Table 3).

The case for validity

According to Downing (2004), the reliability of an assessment is the main indication of its validity. The proven reliability of the WPBA in Paediatric Dentistry is therefore the first source of evidence in favour of validity of the departmental assessment practices. Content validity suggests that the entire course content should be covered (Schuwirth and van der Vleuten, 2010) and that competencies should be aligned with the learning objectives (Coderre, Woloschuk and McLaughlin, 2009). The importance of constructive alignment was also highlighted by Biggs (1999). In the module being evaluated, constructive alignment exists between the assessment methods employed and the learning outcomes as evidenced in the module descriptors and study guides. The latter sources were used when compiling the clinical rubric. Content is specific to the discipline and content validity is therefore enhanced. This is supported by Durning, Cation, Markert and Pangaro (2002).

Construct validity implies that an assessment tool should be able to discriminate between the high and low scorers (Schuwirth and van der Vleuten, 2010). It would seem that the clinical rubric does not discriminate between very good and very poor students. The tendency for departmental staff (part-time and permanent) to cluster their scores around a '4' does, however, suggest that supervisors find students to be competent. This score correlates well with the average mini-CEX scores over the entire year (Table 1). The latter all lie in the vicinity of 60%.

In terms of the BPC, each student is examined on a different patient case and it is therefore not standardised. This negatively affects the fairness and validity of the BPC component and is the motivation to consider removing the once-off BPC from the assessment.

According to Andreatta and Gruppen (2009), validity refers to whether decisions based on assessment data within a particular context can be trusted. Based on the combined framework of Kane (2013) and Royal (2017), evidence of validity in the departmental WPBA has been highlighted (Table 4). All individual assessment components are taken into consideration when making a decision on whether or not a student is competent, i.e., programmatic assessment practices are followed. This information richness adds to the credibility of pass/fail decisions (van der Vleuten et al.,

2015). Validity is thus enhanced as a more complete representation of a student's abilities can be obtained (Schuwirth and van der Vleuten, 2010).

The construct assessed, skills, knowledge and application of knowledge to a particular patient case, are clearly defined in Paediatric Dentistry with the measurement tool, i.e., the detailed clinical rubric, measuring what it is supposed to measure in an actual clinical setting. The overall WPBA practices can therefore be considered valid within this context.

Multiple mini-CEX assessments conducted over a period of time improve reliability and validity as student performance across a range of patients can be observed by different examiners (Al Ansari et al., 2013; Norcini et al., 2003). Interaction with different patients and assessors provides multiple opportunities for feedback (Norcini et al., 2003) and is therefore a valuable formative assessment exercise. Different supervisors are able to observe a range of skills over a period of time, thereby gaining a reasonable idea of the student's abilities (Norcini et al., 2003). This is essential when evaluating the level of clinical competence (Wass et al., 2001) as inferences made from multiple observations over time give a more accurate picture of competence to the point where a summative decision can perhaps be supported (Harris et al., 2017).

Limitations

The study is a single-centre review and its findings may not be transferable to other settings, although other resource-poor settings may share similar challenges. A further limitation was that roughly a third of the available records were excluded due to the fact that they were incomplete. However, the remaining records are sufficient to draw reasonable conclusions regarding the WPBA scores as an indication of clinical competence.

Conclusion

The continuous formative WPBA practices were found to be both valid and reliable when using Kane's (2013) and Royal's (2017) frameworks for analysis. They were shown to be feasible as they are integrated in the daily routine patient care provided in the paediatric dental clinics. The findings of this study suggest that the continuous formative WPBA scores are an indication of the clinical competence of final-year dental students and that they could be used to decide if students have reached the expected level of clinical competence in this module. However, the value of the single BPC should be reassessed.

Where the workplace provides an opportunity to assess students on all aspects of patient management in an authentic clinical environment, the complexity of the workplace-based setting poses a challenge where assessment practices are concerned. This is especially true when a judgement

call has to be made regarding a student's clinical competence. This study suggests that holistic evaluation of student performance is essential when making such a judgement call. A balance must be struck between retaining good clinical practice and ensuring that assessment practices are feasible, fair and more importantly, reliable and valid. The final decision on a student's competence should be made on the basis of professional judgements that are supported by evidence that is both defensible and credible.

References

- Al Ansari, A., Kauser, S., and Donnon T. (2013). The construct and criterion validity of the Mini-CEX: A meta-analysis of the published research. *Academic Medicine* 88 (3), 1-8.
- Andreatta, P.B., and Gruppen, L.D. (2009). Conceptualising and classifying validity evidence for simulation. *Medical Education* 43, 1028-1035.
- Anziani, H., Durham, J., and Moore, U. (2008). The relationship between formative and summative assessment of undergraduates in oral surgery. *European Journal of Dental Education* 12, 233-238.
- Beard, J. (2011). Workplace-based assessment: The need for continued evaluation and refinement. *The Surgeon* 9, S12-S13.
- Biggs, J. (1999). What the student does: teaching for enhanced learning. *Higher Education Research and Development* 18(1), 57-75.
- Carrillo-de-la-Peña, M.T., Baillès, E., Caseras, X., Martínez, A., Ortet, G., and Pérez, J. (2009). Formative assessment and academic achievement in pre-graduate students of health sciences. *Advances in Health Sciences Education* 14, 61-67.
- Coderre, S., Woloschuk, W., and McLaughlin, K. (2009). Twelve tips for blueprinting. *Medical Teacher* 31, 359-361.
- Cook, D.A., Brydges, R., Ginsburg, S., and Hatala, R. (2015). A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education* 49, 560-575.
- Dahlöf, G., Tsilingaridis, G., and Hindbeck, H. (2004). A logbook for continuous self-assessment during 1 year in paediatric dentistry. *European Journal of Paediatric Dentistry* 3, 163-169.
- Downing, S.M. (2003). Validity: on the meaningful interpretation of assessment data. *Medical Education* 37, 830-837.
- Downing, S.M. (2004). Reliability: on the reproducibility of assessment data. *Medical Education* 38, 1006-1012.
- Downing, S.M., and Yudkowsky, R. (2009). Introduction to assessment in Health Professions. In: *Assessment in Health Professions Education*, pp. 1-20. New York: Routledge.
- Driessen, E., van der Vleuten, C., Schuwirth, L., van Tartwijk, J., and Vermunt, J. (2005). The use of qualitative research for portfolio assessment as an alternative to reliability evaluation: a case study. *Medical Education* 39, 214-220.
- Durning, S.J., Cation, L.J., Markert, R.J., and Pangaro, L.N. (2002). Assessing the reliability and validity of the Mini-Clinical Evaluation Exercise for internal medicine residency training. *Academic Medicine* 77(9), 900-904.
- Epstein, R.M., and Hundert, E.M. (2002). Defining and assessing clinical competence. *JAMA* 287(2), 226-235.
- Frank, J.R., Snell, L.S., Cate, O.T., Holmboe, E.S., Carraccio, C., Swing, S.R., Harris, P., Glasgow, N.J., Campbell, C., Dath, D., Harden, R.M., Iobst, W., Long, D.W., Mungroo, R., Richardson, D.L., Sherbino, J., Silver, I., Taber, S., Talbot, M., and Harris, K.A. (2010). Competency-based medical education: theory to practice. *Medical Teacher* 32(8), 638-645.
- Gibson, K., and Shaw, C.M. (2010). Assessment of active learning. *The International Studies Encyclopaedia*. Denmark: Robert A. Blackwell Publishing, 2010. Available from: <https://oxfordre.com/internationalstudies/internationalstudies/view/10.1093/acrefore/9780190846626.001.0001/acrefore-9780190846626-e-120> [Accessed 3 July 2020].
- Govaerts, M., and van der Vleuten, C.P.M. (2013). Validity in work-based assessment: expanding our horizons. *Medical Education* 47, 1164-1174.
- Govaerts, M.J.B., Van de Wiel, M.W.J., Schuwirth, L.W.T., van der Vleuten, C.P.M., and Muijtjens, A.M.M. (2013). Workplace-based assessment: Raters' performance theories and constructs. *Advances in Health Sciences Education* 18, 375-396.
- Harris, P., Bhanji, F., Topps, M., Ross, S., Lieberman, S., Frank, J.R., Snell, L., and Sherbino, J. (2017). Evolving concepts of assessment in a competency-based world. *Medical Teacher* 39(6), 603-608.
- Hays, R.B., Hamlin, G., and Crane, L. (2015). Twelve tips for increasing the defensibility of assessment decisions. *Medical Teacher* 37, 433-436.
- Health Professions Council of South Africa (HPCSA). (2014). *Core competencies for undergraduate students in clinical associate, dentistry and medical teaching and learning programmes in South Africa*. Available from: https://chs.ukzn.ac.za/Libraries/News_Events/MDB_Core_Competencies_-_ENGLISH_-_FINAL_2014.pdf [Accessed 3 July 2020].
- Jonsson, A., and Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review* 2, 130-144.
- Kane, M.T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement* 50(1), 1-73.
- McDade, S.A. (1995). Case study pedagogy to advance critical thinking. *Teaching of Psychology* 22(1), 9-10.

- Miller, G.E. (1990). The assessment of clinical skills/ competence/ performance. *Academic Medicine* 65, S63-67.
- Norcini, J.J., Blank, L.L., Duffy, F.D., and Fortna, G.S. (2003). The Mini-CEX: A method for assessing clinical skills. *Annals of Internal Medicine* 138, 476-481.
- Norcini, J., and Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher* 29, 9-10, 855-871.
- Pelgrim, E.A.M., Kramer, A.W.M., Mokkink, H.G.A., van den Elsen, L., Grol, R.P.T.M., and van der Vleuten, C.P.M. (2011). In-training assessment during direct observation of single-patient encounters: a literature review. *Advances in Health Sciences Education* 16, 131-142.
- Popil, I. (2011). Promotion of critical thinking by using case studies as teaching method. *Nurse Education Today* 31, 204-207.
- Riaz, F., Yasmin, S., and Yasmin, R. (2015). Introducing regular formative assessment to enhance learning among dental students at Islamic International Dental College. *JPMA* 65, 1277-1282.
- Royal, K.D. (2017). Four tenets of modern validity theory for medical education assessment and evaluation. *Advances in Medical Education and Practice* 8, 567-570.
- South African Qualifications Authority (SAQA). (2001). *National policy and criteria for designing and implementing assessment for NQF qualifications and part-qualifications and professional designations in South Africa*. Available from: <http://www.saqa.org.za/docs/pol/2015/National%20Policy%20for%20Assessment.pdf> [Accessed 3 July 2020].
- SAQA. (2005). *Guidelines for integrated assessment*. Available from: <http://www.saqa.org.za/docs/guide/2005/intassessment.pdf> [Accessed 3 July 2020].
- Schuwirth, L.W.T., and van der Vleuten, C.P.M. (2010). How to design a useful test: the principles of assessment. In: Stanwick, T. (ed.) *Understanding Medical Education: Evidence, Theory and Practice*, pp 241-254. Wiley-Blackwell, John Wiley and Sons, Ltd.
- The Royal College of Physicians and Surgeons of Canada. *CanMEDS Physician competency framework*. (2015). Available from: http://canmeds.royalcollege.ca/uploads/en/framework/CanMEDS%202015%20Framework_EN_Reduced.pdf [Accessed 3 July 2020].
- Tziner, A., Murphy, K.R., and Cleveland J.N. (2005). Contextual and rater factors affecting rating behaviour. *Group and Organization Management* 30(1), 89-98.
- University of the Western Cape (UWC). (2018). *Assessment Policy*. General Calendar, pp. 49-54. Available from: [file:///C:/Users/Dell%20User/Downloads/2018%20GENERAL%20CALENDAR%20\(1\).pdf](file:///C:/Users/Dell%20User/Downloads/2018%20GENERAL%20CALENDAR%20(1).pdf) [Accessed 31 March 2020].
- Van der Vleuten, C.P.M., and Schuwirth L.W.T. (2005). Assessing professional competence: From methods to programmes. *Medical Education* 39, 309-317.
- van der Vleuten, C.P.M., Schuwirth L.W.T., Driessen, E.W., Govaerts, M.J.B., and Heeneman, S. (2015). Twelve tips for programmatic assessment. *Medical Teacher* 37, 641-646.
- Wass, V., van der Vleuten, C., Shatzer, J., and Jones, R. (2001). Assessment of clinical competence. *The Lancet* 357, 945-949.
- Wimmers, P. (2006). *Developing clinical competence*. Netherlands: Printpartners Ipskamp, B.V. Available from: https://www.researchgate.net/publication/254804646_Developing_Clinical_Competence [Accessed 3 July 2020].
- Yorke, M. (2003). Formative assessment in higher education: moves towards theory and the enhancement of pedagogic practice. *Higher Education* 45, 477-501.

Appendix: paedo/ortho clinical evaluation

Student name		Year:	Student code				
Staff name		Staff code					
Date:							

Rating	%
1	20
2	40
3	50
4	60
5	75+

	Pt codes
1	Child (Prim. Tooth)
2	Child (Perm Tooth)
3	Adult
Type codes	
1	Observed/ Grp discussion
2	Assisted
3	Treated

Patient Folder Number										Tooth No.	Pt	Type	Procedure Code						

Final mark (rubric)	Pt 1	Pt 2	Pt 3	Additional comments:	Procedures assisted
Clinical					
Theory					
Supervisor's signature: Student's signature:					

Mark	Clinical (includes professionalism)	Theory
1	<p><i>Patient bookings/professionalism</i></p> <ul style="list-style-type: none"> • Student absent/ Cancels patients without permission • No patient booked by student at least a week in advance <p><i>Clinical</i></p> <ul style="list-style-type: none"> • Detrimental to patient—mistake not rectifiable • Missed important information e.g. History of pain • Missed obvious pathology • Treatment plan unacceptable • Wrong procedure performed • Starting procedure without permission or signed treatment plan • Restorations clinically unacceptable 	<p>Complete lack of core knowledge—could not answer any of the questions posed</p> <ul style="list-style-type: none"> • Not able to justify material selection despite prompting • Cannot integrate theory with clinical practice • Complete lack of critical thinking • Cannot motivate treatment chosen

Mark	Clinical (includes professionalism)	Theory
2	<p><i>Patient bookings/professionalism</i></p> <ul style="list-style-type: none"> • Double-booking patients without permission • Arriving/starting late • Lack of time management/poor planning • Leaving without permission (if patient doesn't arrive) • Dismissing patient without supervisor's permission • Non-compliance with dress code <p><i>Clinical</i></p> <ul style="list-style-type: none"> • Instrument seal broken before patient arrives • Poor infection control, untidy cubicle, gloves not in biohazard container • Could not correlate history and clinical picture • Took radiographs before doing a clinical charting • Starting tooth preparation without radiographs where indicated • Missed no pathology clinically but treatment plan unacceptable • Quality of procedure unacceptable but mistake rectifiable • Restorations need major adjustment 	<ul style="list-style-type: none"> • Needs prompting to justify material selection/procedures <p>Could answer less than half of questions posed</p> <ul style="list-style-type: none"> • Has some idea of rationale for treatment • Some core knowledge • Critical thinking skills/reasoning not sound
3	<ul style="list-style-type: none"> • Incomplete History/Diagnosis • OH and Fluoride only • Reasonable history taking-- missed some things but could not answer when prompted • Treatment plan acceptable but needs major revision • Clinical work acceptable but guidance required with procedure • Clinically acceptable restorations but needs minor adjustment 	<p>Core knowledge acceptable/average</p> <ul style="list-style-type: none"> • Reasonable insight regarding procedures/materials • Fair idea for choosing treatment option • Theoretical knowledge and critical thinking skills sound
4	<ul style="list-style-type: none"> • Missed some things regarding history but could answer when prompted • Treatment plan acceptable. Needs only minor revision. • Very limited guidance needed with procedure • Good quality restoration (good contacts/ marginal adaptation)-- minor adjustment needed 	<p>Sound knowledge (better than average)</p> <ul style="list-style-type: none"> • Good motivation but doesn't cover all the possible treatment options • Competent • Able to justify material selection • Good insight demonstrated but can improve • Theoretical knowledge good
5	<ul style="list-style-type: none"> • Missed nothing. Covered all the bases without prompting • Diagnosis and treatment plan spot-on • No guidance needed with procedure • Exceptional patient management • Perfect restoration (good contacts/ marginal adaptation/ no overhangs)—no adjustment needed • All restorations polished—no rough spots 	<p>Substantial knowledge (more than expected)</p> <ul style="list-style-type: none"> • Good motivation— aware of all the treatment possibilities/ options • Displays in-depth understanding • Able to justify material selection/ choice of procedures critically • Excellent insight demonstrated • Theoretical knowledge excellent

NB! Lowest mark achieved gets recorded