



King's Research Portal

DOI:

[10.1111/eje.12286](https://doi.org/10.1111/eje.12286)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Escudier, M. P., Woolford, M. J., & Tricio, J. A. (2018). Assessing the application of knowledge in clinical problem-solving: The structured professional reasoning exercise. *European Journal of Dental Education*, 22(2), e269-e277. <https://doi.org/10.1111/eje.12286>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Assessing the Application of Knowledge in Clinical Problem Solving: The Structured Professional Reasoning Exercise

Running Head: Assessing Professional Reasoning

Michael P. Escudier¹, Mark J. Woolford¹, Jorge A. Tricio^{1,2}

¹ King's College London Dental Institute, London, United Kingdom.

² Faculty of Dentistry, University of the Andes, Santiago, Chile.

Corresponding author:

Profesor Michael P. Escudier

18th Floor Tower Wing, Guy's Hospital

London SE1 9RT, United Kingdom

Telephone: +44 (20) 7188 4399

e-mail: michael.escudier@kcl.ac.uk

Disclosure Statement: The authors report no conflicts of interests related to this study.

Keywords: Clinical reasoning, structured assessment, simulation.

Abstract

Introduction: Clinical reasoning is a fundamental and core clinical competence of healthcare professionals. The study aimed to investigate the utility of the Structured Professional Reasoning Exercise (SPRE), a new competence assessment method designed to measure dental students' clinical reasoning in simulated scenarios, covering the clinical areas of Oral Disease, Primary Dental Care and Restorative Dentistry, Child Dental Health, and Dental Practice and Clinical Governance.

Materials and Methods A total of 313 year-5 students sat for the assessment. Students spent 45 minutes assimilating the scenarios, before rotating through four pairs of 39 trained examiners who each independently assessed a single scenario over a ten-minute period, using a structured marking sheet. After the assessment, all students and examiners were invited to complete an anonymous perception questionnaire of the exercise. These questionnaires and the examination scores were statistically analysed.

Results and Discussion: Oral Disease showed the lowest scores; Dental Practice and Governance the highest. The overall Intraclass Correlation Coefficient (ICC) was 0.770, while examiner training helped to increase the ICC from 0.716 in 2013 to 0.835 in 2014. Exploratory factor analysis revealed one major factor with an eigenvalue of 2.75 (68.8% of total variance). The Generalizability coefficient was consistent at 0.806. A total of 295 students and 32 examiners completed the perception questionnaire. Students' lowest examination perceptions were an "Unpleasant" and "Unenjoyable" experience, while the highest were "Interesting", "Valuable" and "Important". The majority of students and examiners reported the assessment as acceptable, fair and valid.

Conclusion: The SPRE offers a reliable, valid and acceptable assessment method, provided it comprises at least four scenarios with two independently marking and trained assessors.

Introduction

Clinical Reasoning is a broad concept as well as a fundamental and core clinical competence of healthcare professionals (1). It is the process by which a clinician or a clinical student utilise the foundations of critical thinking to efficiently and effectively interact with a patient, assess his/her condition, collect clinical and scientific investigations, generate and test hypotheses, and weigh and prioritise demands to determine an optimal diagnosis and the best action or treatment for the particular patient's case (2).

As an essential component of clinical practice, clinical reasoning is increasingly being used in higher education and professional development to improve patient care (3, 4). Aligned with this, several countries' dental regulatory bodies require the teaching and assessment of students' use of critical thinking, clinical judgment and problem solving skills (5-7).

Many dental students with good theoretical knowledge, struggle to apply it in clinical contexts (8). This supports the need for reflective based teaching which promotes a deep approach to learning (9). Notwithstanding, the implementation of reflective teaching to promote clinical reasoning in healthcare education is challenging (3, 10). Despite this, several methods have been described both to promote (formative assessment) and evaluate (summative assessment) clinical reasoning skills.

Kramer et al. (11) recommended the use of structured essays as the preferred assessment technique for dental students' critical thinking and problem solving skills. However, writing and speaking critically are two very different processes (12). The latter face-to-face type of clinical assessment allows the student to be asked to suggest alternative explanations for the patient's condition, diagnosis hypotheses as well as different treatment options (12). Hence, Albino et al. (13) suggested the use of oral examinations to assess students' ability to synthesize information within a given context and apply their knowledge in clinical problem solving. "Orals" have been criticised, among other reasons, for issues related to examiners' bias and stringency variation and subjective standards for passing, resulting in poor reliability (14). However, the need to validly assess students' overall clinical performance in the workplace through face-to-face interaction with a trainer, gave rise to several reliable structured workplace-based assessment tools to assess clinical judgement (15). The reliable assessment of clinical reasoning and ill-defined problem solving skills at the "Shows How" level of Miller's pyramid (16) has received less attention.

One example of this is the use of Objective Structured Clinical Examinations (OSCE) for the assessment of a wide range of clinical skills (17). Despite the OSCE's positive results in dental education (18, 19), its classic design to assess clinical competence has been reported to poorly correlate with students' higher cognitive reflective skills (20). The OSCE, as a competency examination, properly assesses students' understanding and automatic decision making but restricted to the relevant features and aspects of the presented situation; probably due to the time constrain and student anxiety (21).

Consequently, some OSCEs have incorporated reflective scenarios (22-26) though the validity of the assessment has been questioned by some authors when only one OSCE station addresses clinical reasoning, especially in high stakes examinations (22, 26). Criticisms of this approach include the "special" reflective scenario station being distinct from the other OSCE stations (24), and that clinical reasoning is largely invisible and can only be inferred when certain students question a simulated patient (27). Moreover, clinical reasoning requires wider sampling to get reliable judgements (28).

A significant step in the assessment of clinical decision-making skills was the development (29) and future adaptation (30, 31) of the Key-Feature Questions. These consist of a brief clinical stem focused on a difficult aspect in the diagnosis and management at which candidates are most likely to make mistakes, followed by few written questions. Properly designed, Key-Feature Questions can generate reliable and valid examinations (1).

Another method to assess clinical reasoning is the Script Concordance Tests (initially known as Diagnostic Script Questionnaires) (32) in which candidates are presented with ill-defined clinical scenarios followed by a number of items and successive new information to test the proposed hypothesis. Responses are then compared with those from experts and credited accordingly (33). Despite the strong evidence for Script Concordance Tests' validity (34), they have been criticised for the complex aggregate scoring method, improper reliability assessment and problems with anchors at the extreme of the scale (35). The Clinical Reasoning Problem is another method to assess diagnostics skills that though time consuming, can produce reliable results (36). It uses the stimulus of a clinical scenario to ask candidates the two most likely diagnoses for the case as well as a list of clinical features they considered important in formulating the two diagnoses (37).

Taking into account the presented previous experiences to assess clinical reasoning, the aim of this cross-sectional study was to investigate the utility of a new competence assessment method specifically designed to measure the clinical reasoning construct while students apply theoretical knowledge to solve simulated clinical dental scenarios, named the Structured Professional Reasoning Exercise (SPRE). The study was developed based on two hypotheses. Firstly, all SPRE scenarios assessed the same construct; that is, clinical problem solving. Secondly, the SPRE was acceptable and fair to both students and staff.

Materials and Methods

Ethical approval

The study received full approval from the King's College London Biomedical Sciences, Dentistry, Medicine and Natural & Mathematical Sciences Ethical Committee (reference number BDM/11/12-21).

The instrument and the examination process

The Structured Professional Reasoning Exercise developed for this study was based on a similar assessment method used in Part 2 of the Membership of the Joint Dental Faculties (MJDF) of The Royal College of Surgeons of England (38). It was introduced at King's College London Dental Institute (KCLDI) for the Bachelor of Dental Surgery (BDS) Year 5 Finals examination to replace the "single unseen long case" in the assessment of undergraduate students' clinical reasoning. This study reports the results and analysis of the SPRE implementation in the 2013 and 2014 cohorts.

The design of the SPRE aimed to assess students' ability to evaluate, understand and apply theoretical knowledge orally intended to assess the clinical reasoning construct. For this purpose, structured clinical scenarios in the four key clinical areas of the KCLDI BDS Year 5 programme, as well as for the UK General Dental Council (GDC) learning outcomes (5), were developed. Each scenario contained carefully designed, tested and redesigned questions to assess students' problem solving competence on the four key areas. These areas were: Oral Disease, Primary Dental Care and Restorative Dentistry, Child Dental Health, and Dental Practice and Clinical Governance. Whilst the four clinical areas were the same for both cohorts, care was taken to develop different scenarios, with a similar level of difficulty, to eliminate "leakage" of material. In parallel, formative scenarios were uploaded to the College e-learning platform to enhance students' understanding of the new assessment and facilitate their learning.

In line with the Criteria for Assessment in Clinical Examinations from the Undergraduate Board of Examiners in the Dental Institute, examiners were thoroughly trained and calibrated, each year, on the Grade and Band of the structured marking scheme (Distinction, Merit, Pass, Fail) and the structured marking sheet which contained a written description of the same four criteria as in the marking scheme. Examiners also received explanations on the objectives of the examination which were:

“To assess the candidate’s interpretation and assessment of the presented evidence, appreciation of the key points presented by the scenario, implications of the findings to short and long term management of the scenario by articulating not only their theoretical knowledge and understanding, but also their ability to apply it to the specific clinical problems (problem solving), approach to reaching a diagnosis or conclusion, and ability to outline key points of management”. A sample scenario on Child Dental Health is shown in Figure 1.

Examiners also received a written guide specifically focused on the process of the SPRE. This included further reminders highlighting the assessment procedure:

- Candidates received 45 minutes to study the four different clinical cases, individually. As in the formative scenarios, students were advised to take approximate 10 minutes to read, understand and make notes of each case. Candidates then rotated through four pairs of examiners each of which addressed one of the four scenarios. Each examiner asked a series of structured questions to test the students’ knowledge and understanding of the scenario. The total number of questions per scenario ranged from 6 to 14 and these were consistently distributed between the two examiners.
- Each structured oral examination, with the pair of examiners, lasted for 10 minutes (5 minutes each examiner, always in the same sequence to standardise questions) with 1 minute for transfer of candidates. Hence, the full examination time corresponded to the study 45 minutes candidates were given in order to ensure a smooth flow of students.
- Supplementary questions were allowed to clarify a specific point.
- Examiners marked independently according to the marking scheme and the Answer Guide for each of the questions.
- Invigilators ensured accurate time keeping.
- Examiners were instructed to pass those candidates who could reasonably be termed a “safe beginner” (General Dental Practitioner) rather than an expert in the speciality.

Participants and data collection

All 313 BDS Year 5 students from the 2013 (n. 151) and 2014 (n. 162) cohorts sat for the SPRE and therefore took part in the study. Of these, 202 were females, 111 were males, and their mean age was 24.4 (sd=2.9).

Students were randomly assigned to one of three parallel circuits each consisting of Oral Disease, Primary Dental Care and Restorative Dentistry, Child Dental Health, and Dental Practice and Clinical Governance scenarios. The total number of examiners involved was 24 in 2013 and 39 in 2014.

At the end of their involvement in the examination all students and examiners were invited to voluntarily complete an anonymous questionnaire regarding their perceptions of the SPRE in terms of acceptability, fairness and validity, presented in a Likert scale.

Data analysis

SPRE examination scores and perception questionnaire responses from both cohorts were uploaded into a spreadsheet and imported into IBM Statistical Package for Social Sciences Windows® version 21 (SPSS Inc. IBM, Chicago, IL, USA), to be analysed as a single sample

Descriptive statistics were employed to express students’ gender and age characteristics. After assessing the normal distribution of the SPRE scores and questionnaire responses through visual inspection of a histogram

and by exploring the central tendency (mean, median, mode), students' SPRE marks were descriptively analysed as a whole and by each clinical scenario.

A one-way analysis of variance (ANOVA) test was performed to compare the SPRE scores of all four scenarios. Where the ANOVA showed significant results, a post-hoc analysis was carried out using Tukey's test.

The inter-examiner reliability was examined using Intraclass Correlation Coefficients (ICC) (95% confidence intervals) of each pair of examiners' independently marked SPRE scores. In order to identify and quantify underlying dimensions within the SPRE scenarios, an exploratory factor analysis with varimax rotation, was performed. Although each of the four different SPRE scenarios were designed having in mind the assessment of clinical reasoning as a single construct, we performed an exploratory factor analysis (using Varimax rotation) because at that time we were unsure about how many latent dimensions the SPRE would factor into. As not all datasets are suitable for factor analysis (39), the SPRE adequacy was explored beforehand using the Kaiser-Meyer-Olkin (KMO) Test and the Bartlett Test of Sphericity. Further, the internal consistency of the SPRE was studied using Cronbach's Alpha coefficients.

The final analysis regarding the SPRE instrument itself, was a G-study (Generalizability study) used to determine the contribution that all relevant factors (Scenario/Trainee/Assessor) made to the SPRE results and hence its reliability. Additionally, a D-study (Decision study) was used to estimate the effect of the number of scenarios and assessors on the reliability of the SPRE. Thus, a G-study with a nested three-facet (Scenario x Trainee x Assessor) random-effects using Trainee as a fixed facet was used (assessors were nested) as there were three parallel circuits meaning that not all students confronted the same pair of examiners.

The students' and examiners' questionnaire responses were statistically studied relative to their acceptability, fairness and validity.

Results

Participants and data collection

The overall SPRE average score from all 313 students was 66.5 (sd=12.9). According to the marking criteria, there were 86 (27%) distinctions, 150 (48%) merits, 72 (23%) pass, and 5 (2%) fails. As shown in Figure 2, the lowest mean score was for the Oral Disease (60.1, sd=11.7) scenario while the highest was for Dental Practice and Clinical Governance (70.3 sd=12.6).

The ANOVA test showed significant differences among the marks of the four scenarios ($F = 38.021$, $p < 0.0001$), with the post-hoc analysis using Tukey's test revealing that the higher score from Dental Practice and Clinical Governance was significantly ($p < 0.0001$) different from all other scenarios.

The overall Intraclass Correlation Coefficients (ICC) was 0.770 ($p < 0.0001$). The effects of training and calibration on inter-examiner reliability can be observed in Table 1. All of the 24 examiners that participated in the 2013 SPRE also did so in 2014 and so received training and calibration twice. All 15 new assessors in 2014 had previous experiences in other KCLDI clinical assessments e.g. Objective Structured Clinical Examination (OSCE). Notwithstanding this, new SPRE assessors were strategically paired with those who demonstrated high reliability in the 2013 SPRE. Consequently, the overall mean inter-examiner reliability (ICC) increased from 0.716 in 2013 to 0.835 in 2014 and was even greater in Primary Dental care (Table 1).

The SPRE datasets were found suitable for factor analysis (KMO=0.819, Bartlett's Test of Sphericity $\chi^2=505.656$ $p<0.0001$), which its four scenarios including 48 questions, revealed one major factor with an eigenvalue of 2.75 which accounted for 68.8% of total variance. Subsequently, a large drop in eigenvalues was observed in the other three components; 0.46, 0.42 and 0.36. Because one factor was identified

indicating a single construct, it was considered appropriate to calculate the internal consistency of the four SPRE scenarios as a whole (40). Hence, the Alpha coefficient of 0.848.

The Generalizability coefficient was 0.806. The contribution of all involved factors to the SPRE results are shown in Table 2 and the effect of the number of scenarios and assessors on the SPRE reliability in Table 3.

Students' and Examiners' perceptions

A total of 295 (94%) students and 32 (82%) examiners completed the anonymous questionnaires. Students rated the SPRE (1 to 7 scale) with the lowest mean score as an "Unpleasant" (3.71) and "Unenjoyable" (3.62) experience. However, they also rated the assessment with the highest scores as "Interesting" (5.48), "Valuable" (5.49) and "Important" (5.68) (Figure 3). Though there were no differences between cohorts, the "Unpleasant" and "Unenjoyable" scores were significantly different to all other question results ($F_{65,869}$, $p < 0.0001$).

When comparing the student's and examiners' perceptions there were quite large differences in their scores, although the trend was similar (Figure 4). Acceptability was high with around 80% of students rating the SPRE as "highly" or "moderately" acceptable. In contrast, almost all examiners thought this was the case (97%). Fairness for students was not perceived as positively as Acceptability and 18% of the students thought it was "not fair" (not at all + quite not answers). Again the vast majority of examiners (95%) thought the examination was "Fair". Perception of Validity showed a similar pattern to Fairness: although most students (65%) thought the assessment was Valid, there were again 19% of students that perceived it as "not valid" (not at all + quite not answers). There were also 6% of examiners that thought the assessment was "Quite Not" Valid

Discussion

The present cross-sectional study describes the utility of a new problem solving competence assessment method to assess the clinical reasoning construct, the Structured Professional Reasoning Exercise.

Four scenarios containing 48 questions in total were developed to assess students' ability to evaluate, understand and apply theoretical knowledge to solve clinical problems. Although scenarios included different key clinical areas from the KCLDI BDS programme, all questions were designed to assess the same construct. That is, dental clinical reasoning.

This assessment procedure was conceived and implemented as the culmination of a progressive, repeated and scaffolded teaching and learning process of knowledge integration that enabled students to develop an advanced reasoning ability, supplementing traditional didactic and skills teaching (41); something that higher education is supposed to accomplish (42). The success of this teaching and learning experience is evidenced by the low number of students failing the SPRE (2%).

The Oral Disease scenario showed the lowest score (average score 60.1) probably because of the greater amount of basic science prior knowledge required to solve the clinical problem (43, 44). In contrast and whilst Dental Practice and Clinical Governance (average score 70.3) requires significant knowledge in the areas of Principles of Management and Responsibility, Raising Concerns, Dental Team Working, Scope of Practice, and Complaints Handling (5), it is also conceivable that students used common sense to solve these scenarios resulting in higher scores (Figure 2).

The overall inter-examiner reliability assessed through Intraclass Correlation Coefficients (ICC) was statistically significant (0.770, $p < 0.0001$) (Table 1). It is remarkable to see the increase in inter-examiner reliability from 2013 to 2014. This was not only true for Primary Dental Care and Restorative Dentistry (ICC from 0.579 in 2013 to 0.796 in 2014) that got the lowest ICC score in 2013, but for all domains (Table 1). This effect might be explained by the thorough examiners' training and calibration which 62% of examiners underwent twice, as they participated in both SPREs. This also aligns with the suggestion that examiner

training achieves substantial gains in reliability (45). The use of structured questions and marking sheets might also have contributed (46) as may the fact that all new assessors had previous experience in OSCEs and were also paired with colleagues who demonstrated high reliability in the 2013 SPRE version. This is not only important psychometrically but also in maintaining an equitable assessment process (47).

Exploratory factor analysis identified only one factor with an eigenvalue greater than 1 (2.75), accounting for 68.8% of the variance. Loading with such a value has previously been considered “excellent” (40). Furthermore, there was a clear flattening between the second (0.46), third (0.42) and fourth (0.36) factors. As mentioned above, all four SPRE scenarios had in common the assessment of problem solving skills as a single construct, though based on different key clinical areas of dental education. The strong value of the first loading factor indicates the SPRE’s one-dimensionality. This together with the fact that all 48 questions were measuring problem solving skills, seems to suggest the SPRE was assessing clinical reasoning.

The high Generalizability coefficient (0.806) changed dramatically when the number of four scenarios with two examiners each were reduced through a D-Study (Table 3). This positive G coefficient might be explained by the fact that each students’ knowledge and understanding was deeply scrutinise through around 48 structured questions during an added 40 minutes’ examination time across four standardised cases marked independently by each examiner.

Regarding SPRE variance, the highest component was Trainee’s ability (37.7%) (Table 2). This was similar to that reported by Wilkinson et al. (46) in two long case examinations per medical candidate (33% in 2005 and 38% in 2006), In a more SPRE alike assessment, the same study also reported the results of four 15 minutes short case examinations per candidate. In this circumstance, the highest variance was Trainee x Case with 56% in 2005 and 58% in 2006. In our study, the SPRE Trainee x Case variance was much lower and reached 27.3%. This could be explained by the fact that our four cases were structured and all candidates were assessed based on scenarios of similar complexity. The same authors reported a D study reliability of ≥ 0.8 when two long cases (60 minutes each patient) were combined with four short cases (15 minutes each patient), to produce a three hours’ examination per student, questioning its feasibility. Similarly, Wass et al. (48) identified 10 as the minimum number of unstandardized long cases required to produce a reliable, high-stakes, examination.

As mentioned above, the SPRE was introduced to replace the “single” unseen long case which limited the ability to sample the curriculum widely resulting in poor reliability (49). However, the ability of the long case to assess the candidate’s overall critical approach to it with the possibility of asking the student to suggest different possible explanations, diagnosis hypotheses as well as treatment options (12) in a face-to-face clinical assessment, provided the basis for the SPRE. The advantage of the latter is a structured format and marking scheme, and higher examiner numbers and cases (49) increasing both reliability and validity (17). Further, in a previous study the SPRE showed a significant correlation ($p < 0.005$) with higher order thinking (20). Similar reflection correlations have been reported in medical students with “knowledge of actions to take” (50), which is one of the assessment criteria of our SPRE rubric.

Overall, 65% of students found the SPRE highly and moderately “fair” (Figure 4). Further, 79% declared it was highly and moderately “acceptable”. From the examiners viewpoint, “acceptability” and “fairness” perception was even higher (97% and 95%, respectively) (Figure 4). Whilst students evaluated the SPRE with the lowest scores as an “Unpleasant” and “Unenjoyable” experience (3.71 and 3.62, respectively) (Figure 3), in contrast with the highest scores in “Interesting”, “Valuable” and “Important” (5.48, 5.49 and 5.68, respectively). This could be interpreted as although students did not really like the examination, they considered it necessary and beneficial. Similar perceptions on the learning effect were reported by medical students after completing a reflective OSCE station, despite being critical about the scenario (24).

Despite these positive results, a number of 53 (18%) and 56 (19%) students thought the SPRE was “not fair” and “not valid” (not at all + quite not answers), respectively. Further, there were 4 examiners (13%) that did not express a positive view of the SPRE validity (“Quite Not” + “Neutral”). This outcome encouraged the assessment team to increase the number of formative scenarios uploaded into the College e-learning platform and to reiterate those competences to be assessed to the students. Furthermore, to increase validity for subsequent SPREs an increased number of examiners contributed to the design and implementation of the scenarios, with particular reference to course programmes learning outcomes to be assessed.

The development, organisation and implementation of a SPRE has some limitations. Despite 75% of the students being in the SPRE distinction and merit category, it would be interesting to know where those students with low marks are failing or having difficulty. Is it because there is a knowledge base problem? Is there a problem of comprehension? Or is it due to a weakness in knowledge application to problem solving? (51). The SPRE examination requires high logistics, such as organising time and space especially for big groups. Similarly, training and calibrating examiners is time consuming and our study was conducted in a single institution with relatively homogeneous student cohorts.

Conclusions

The Structured Professional Reasoning Exercise provides an acceptable and fair method of assessing final year dental students’ ability to evaluate, understand and apply meaningful knowledge to simulated clinical problem solving scenarios. Besides the structured clinical scenarios described in this study, other applications of the SPRE format might include the encouraging of reasoning in basic science courses and to further integrate them to practical cases. To ensure reliability, it is suggested that the SPRE comprises at least four scenarios eachone with two independently marking assessors.

References

- 1 Hrynchak P, Glover Takahashi S and Nayer M. Key-feature questions for assessment of clinical reasoning: A literature review. *Med Educ* 2014; 48 (9): 870-883.
- 2 Fish D and Coles C. *Medical education: Developing a curriculum for practice*. Maidenhead, Berkshire, UK: Open University Press, 2005.
- 3 Mann K, Gordon J and MacLeod A. Reflection and reflective practice in health professions education: A systematic review. *Adv in Health Sci Educ* 2009; 14 (4): 595-621.
- 4 Gonzalez M, Abu Kasim N and Naimie Z. Soft skills and dental education. *Eur J Dent Educ* 2013; 17 (2): 73-82.
- 5 General Dental Council. *Preparing for practice: Dental team learning outcomes for registration* London, UK: GDC, 2012: 1-104.
- 6 American Commission on Dental Accreditation CODA. *Accreditation standards for dental education programs*, 2010.
- 7 Australian Dental Council (ADC). *Professional attributes and competencies of the newly qualified dentist*. In: Council AD, ed. Melbourne, Victoria, Australia, 2010.
- 8 Maart R and Bitzer E. Aligning clinical assessment with course elements in prosthodontic dentistry: A south african case. *J Dent Educ* 2013; 77 (9): 1129-1139.
- 9 Sandars J. The use of reflection in medical education: Amee guide no. 44. *Med Teach* 2009; 31 (8): 685-695.
- 10 Ashley F, Gibson B, Daly B, Lygo Baker S and Newton J. Undergraduate and postgraduate dental students' reflection on learning: A qualitative study. *Eur J Dent Educ* 2006; 10 (1): 10-19.
- 11 Kramer G, Albino J, Andrieu S, et al. Dental student assessment toolbox. *J Dent Educ* 2009; 73 (1): 12-35.
- 12 Brookfield SD. *Teaching for critical thinking: Tools and techniques to help students question their assumptions*. Wiley, 2011.
- 13 Albino J, Young S, Neumann L, et al. Assessing dental students' competence: Best practice recommendations in the performance assessment literature and investigation of current practices in predoctoral dental education. *J Dent Educ* 2008; 72 (12): 1405-1435.
- 14 Boursicot K, Etheridge L, Setna Z, et al. Performance in assessment: Consensus statement and recommendations from the ottawa conference. *Med Teach* 2011; 33 (5): 370-383.
- 15 Norcini J and Burch V. Workplace-based assessment as an educational tool: Amee guide no. 31. *Med Teach* 2007; 29 (9): 855-871.
- 16 Miller G. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65 (9Suppl.): 63-67.
- 17 Wass V and Van Der Vleuten C. The long case. *Med Educ* 2004; 38 (11): 1176-1180.
- 18 Schoonheim-Klein M, Muijtjens A, Habets L, et al. On the reliability of a dental osce, using sem: Effect of different days. *Eur J Dent Educ* 2008; 12 (3): 131-137.

- 19 Schoonheim-Klein M, Habets L, Aartman I, van der Vleuten C, Hoogstraten J and van der Velden U. Implementing an objective structured clinical examination (osce) in dental education: Effects on students' learning strategies. *Eur J Dent Educ* 2006; 10 (4): 226-235.
- 20 Tricio J, Woolford M and Escudier M. Dental students' reflective habits: Is there a relation with their academic achievements? *Eur J Dent Educ* 2015; 19 (2): 113-121.
- 21 Brand H and Schoonheim-Klein M. Is the osce more stressful? Examination anxiety and its consequences in different assessment methods in dental education. *Eur J Dent Educ* 2009; 13 (3): 147-153.
- 22 Bogo M, Regehr C, Katz E, Logie C, Tufford L and Litvack A. Evaluating an objective structured clinical examination (osce) adapted for social work. *Res Soc Work Pract* 2012; 22 (4): 428-436.
- 23 Hudson J, Rienits H, Corrin L and Olmos M. An innovative osce clinical log station: A quantitative study of its influence on log use by medical students. *BMC Medical Education* 2012; 12: 8.
- 24 Miller E and Green AR. Student reflections on learning cross-cultural skills through a 'cultural competence' osce. *Med Teach* 2007; 29 (4): E76-E84.
- 25 Dory V, Charlin B, Vanpee D and Gagnon R. Multifaceted assessment in a family medicine clerkship: A pilot study. *Fam Med* 2014; 46 (10): 755-760.
- 26 Durning S, Artino A, Boulet J, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. *Med Teach* 2012; 34 (1): 30-37.
- 27 Kreiter CD and Bergus G. The validity of performance-based measures of clinical reasoning and alternative approaches. *Med Educ* 2009; 43 (4): 320-325.
- 28 Norman G. Research in clinical reasoning: Past history and current trends. *Med Educ* 2005; 39 (4): 418-427.
- 29 Page G and Bordage G. The medical council of canada's key features project: A more valid written examination of clinical decision-making skills. *Acad Med* 1995; 70 (2): 104-110.
- 30 Hatala R and Norman GR. Adapting the key features examination for a clinical clerkship. *Med Educ* 2002; 36 (2): 160-165.
- 31 Raupach T, Andresen JC, Meyer K, et al. Test-enhanced learning of clinical reasoning: A crossover randomised trial. *Med Educ* 2016; 50 (7): 711-720.
- 32 Charlin B. Script questionnaires: Their use for assessment of diagnostic knowledge in radiology. *Med Teach* 1998; 20 (6): 567-571.
- 33 Dory V, Gagnon R, Vanpee D and Charlin B. How to construct and implement script concordance tests: Insights from a systematic review. *Med Educ* 2012; 46 (6): 552-563.
- 34 Lubarsky S, Charlin B, Cook D, Chalk C and van der Vleuten C. Script concordance testing: A review of published validity evidence. *Med Educ* 2011; 45 (4): 329-338.
- 35 Lineberry M, Kreiter CD and Bordage G. Threats to validity in the use and interpretation of script concordance test scores. *Med Educ* 2013; 47 (12): 1175-1183.
- 36 Higgs J, Jones M, Loftus S and Christensen N. Clinical reasoning in the health professions. London, UK: Elsevier (Butterworth Heinemann), 2008.

- 37 Groves M, Scott I and Alexander H. Assessing clinical reasoning: A method to monitor its development in a pbl curriculum. *Med Teach* 2002; 24 (5): 507-515.
- 38 The Royal College of Surgeons of England. Membership of the joint dental faculties Structured Clinical Reasoning assessment. 2015.
- 39 Norman GR and Streiner DL. *Biostatistics: The bare essentials*. B.C. Decker, 2014.
- 40 Tavakol M and Dennick R. Post-examination interpretation of objective test data: Monitoring and improving the quality of high-stakes examinations: Amee guide no. 66. *Med Teach* 2012; 34 (3): e161-e175.
- 41 Chambers DW. Lessons from students in a critical thinking course: A case for the third pedagogy. *J Dent Educ* 2009; 73 (1): 65-82.
- 42 Bertolami CN. The role and importance of research and scholarship in dental education and practice. *J Dent Educ* 2002; 66 (8): 918-924.
- 43 Postma T and White J. Developing clinical reasoning in the classroom – analysis of the 4c/id-model. *Eur J Dent Educ* 2015; 19 (2): 74-80.
- 44 Woods N, Brooks L and Norman G. The value of basic science in clinical diagnosis: Creating coherence among signs and symptoms. *Med Educ* 2005; 39 (1): 107-112.
- 45 Norcini J. The death of the long case? *Br Med J* 2002; 324 (16): 408-409.
- 46 Wilkinson TJ, Campbell PJ and Judd SJ. Reliability of the long case. *Med Educ* 2008; 42 (9): 887-893.
- 47 Schuwirth L. Assessing medical competence: Finding the right answers. *Clin Teach* 2004; 1 (1): 14-18.
- 48 Wass V, van der Vleuten C, Shatzer J and Jones R. Assessment of clinical competence. *The Lancet* 2001; 357 (9260): 945-949.
- 49 Ponnampereuma G, Karunathilake I, McAleer S and Davis M. The long case and its modifications: A literature review. *Med Educ* 2009; 43 (10): 936-941.
- 50 Ambrose L and Ker J. Levels of reflective thinking and patient safety: An investigation of the mechanisms that impact on student learning in a single cohort over a 5 year curriculum. *Adv in Health Sci Educ* 2014; 19 (3): 297-310.
- 51 Askew K, Manthey D and Mahler S. Clinical reasoning: Are we testing what we are teaching? *Med Educ* 2012; 46 (6): 540-542.

Figure 1. Child Dental Health Reasoning Scenario.

This 7 year old girl comes to your dental practice as a new patient, as the family has recently moved to the area. She attends the appointment with her mother. The girl's mother lets you know that her daughter has had some discomfort/sensitivity from a tooth in the maxillary right quadrant and that she is nervous about coming to the dentist. The girl is medically fit and well.

1. What further information would you like from the patient/mother?
2. What are your clinical and radiographic findings?
3. What is your diagnosis?
4. What caries risk status would you place this patient?
5. With regards to prevention how would you manage this patient?
6. How would you manage this case/treatment options?



Figure 2. Structured Professional Reasoning Exercise scores per scenario.

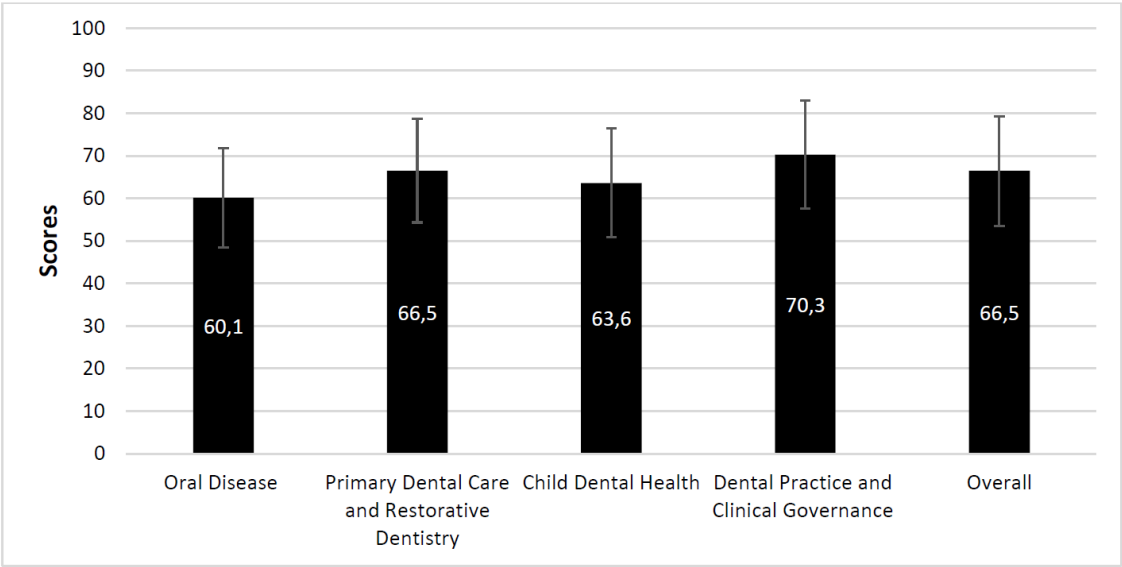


Figure 3. Students' perceptions on the Structured Professional Reasoning Exercise (n.295).

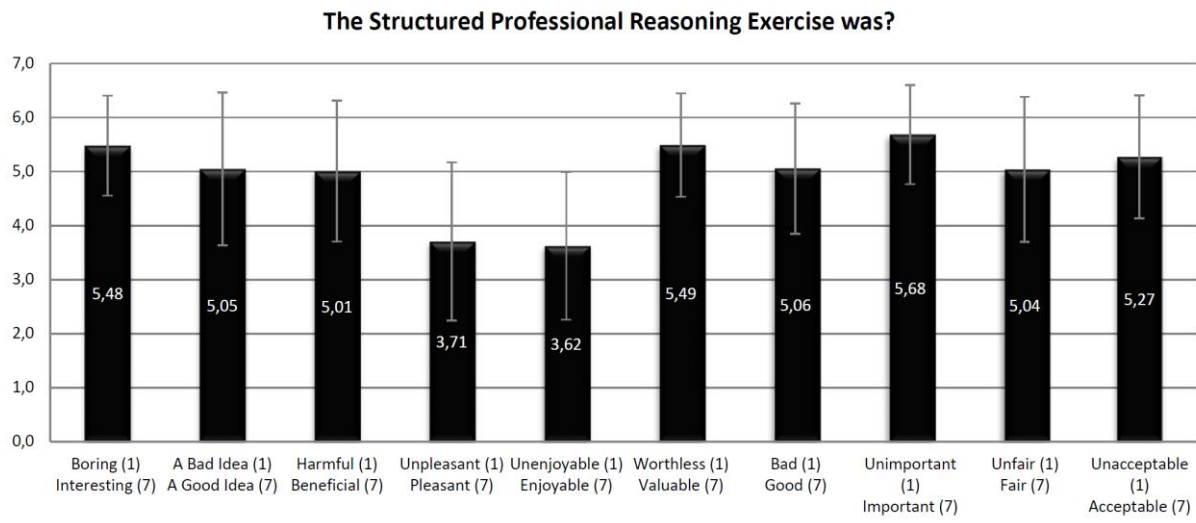


Figure 4. Students' (n.295) and examiners' (n.32) perceptions on the Structured Professional Reasoning Exercise (SPRE) Acceptability, Fairness and Validity.

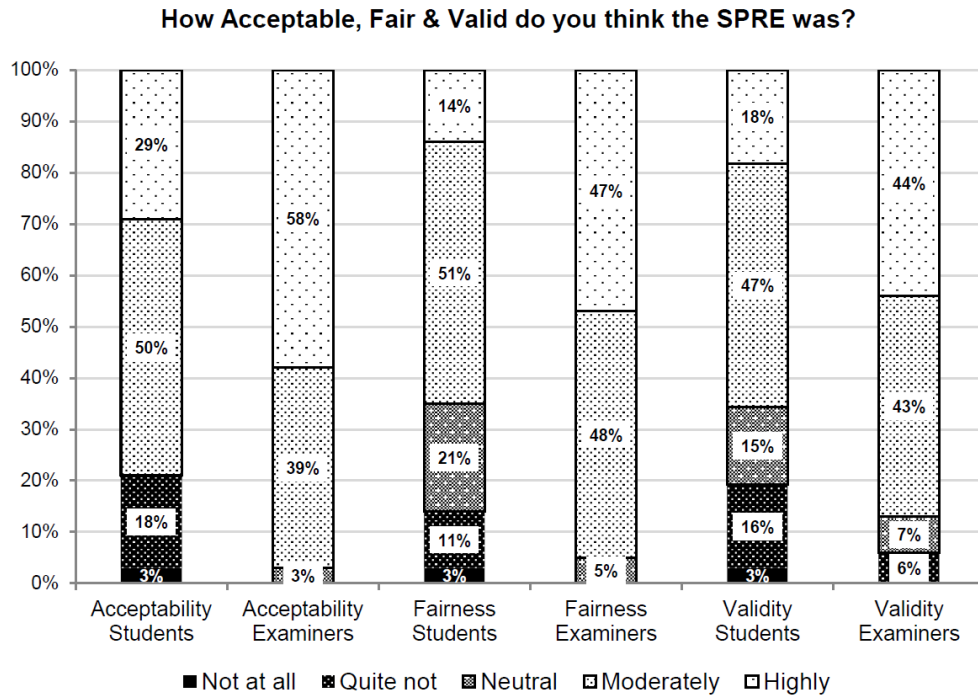


Table 1. Intraclass correlation coefficients (ICC) between the two independently marking assessors for each of the four Structured Professional Reasoning Exercise scenarios.

Scenario	Overall	2013		2014	
	ICC	ICC	Confidence Interval	ICC	Confidence Interval
Oral Disease (OD)	0.820*	0.805†	0.740-0.854	0.851†	0.802-0.888
Primary Dental Care / Restorative (PDC)	0.658*	0.579‡	0.462-0.679	0.796‡	0.731-0.846
Child Dental Health (CDH)	0.902*	0.828§	0.771-0.873	0.971§	0.961-0.979
Dental Practice / Governance (DPG)	0.675*	0.656	0.554-0.738	0.684	0.592-0.758
Overall	0.770*	0.716¶	0.675-0.753	0.835¶	0.809-0.857

* $p < 0.0001$

† Difference between OD 2013 and 2014 $p < 0.038$

‡ Difference between PDC 2013 and 2014 $p < 0.017$

§ Difference between CDH 2013 and 2014 $p < 0.008$

|| Difference between CDH 2013 and 2014 $p < 0.010$

¶ Difference between Overall 2013 and 2014 $p < 0.002$

Table 2. Variance components for the Structured Professional Reasoning Exercise.

Components	Proportion of overall variance (%)
Trainee	37.7
Scenario	14.6
Scenario/Trainee	27.3
Scenario/Trainee/Assessor	20.4

Table 3. Effect of the number of scenarios and assessors on the Structured Professional Reasoning Examination reliability.

Condition	Generalizability
4 scenarios with 2 examiners (as performed)	0.806
4 scenarios with 1 examiner	0.677
3 scenarios with 2 examiners	0.598