Comparison of two data capture methods and gender during clinical assessment in osteopathy: The impact on student/tutor satisfaction ratings

AUTHORS:

Philip Brighta PhD PGCHE PGCRM BSc
ORCID ID: https://orcid.org/0000-0002-5339-8281

Ollie Brightb

a. European School of Osteopathy,
b. Undergraduate, University of Bath.

KEY WORDS

Mini-CEX; Satisfaction; Computer-based format

ABBREVIATIONS

Mini-CEX - Mini Clinical Examination Exercise

What this paper adds:

Influences on student and examiner satisfaction with clinical assessment may include the process of capture. This study investigated osteopathy students’ and tutors’ satisfaction using the mini-CEX via paper-based and online versions. Findings indicate that satisfaction is independent of capture medium, but aligned between student and examiner, further modified by gender.

Abstract

Background: Direct observation of students with patients is important for assessing clinical skills prior to professional registration. The mini-clinical evaluation exercise (mini-CEX) is established as part of a broad assessment profile. Differences between student and tutor satisfaction, when deploying this assessment through different methods, are not widely explored. This study explored gender bias in osteopathy students and tutors with satisfaction ratings, using the mini-CEX via online and paper-based versions.

Methods: An online mini-CEX was initially trialled as a post-hoc data entry administration tool. Android tablets were then used for online capture of observed clinical practice of students by tutors. Comparison with a paper counterpart over the course of three academic years was undertaken. Influence of gender and assessment capture was analysed using summary, correlation and regression statistics to explore the data in depth.

Results: 736 assessments of patient encounters were analysed (550 (75%) online). The influence of paper and online process on satisfaction scores was not significant (odds ratio 1, CI 0.86 – 1.15). Student satisfaction ratings for female students assessed by male tutors indicated lower scores compared to same-sex pairings (P<.007). Correlation between all student and tutor satisfaction ratings was moderate (r2=0.62, 95% CI 0.57 – 0.66, P<.00001).

Conclusions: The findings suggest that there is no difference between the two methods of delivery and satisfaction for either examiner or student, suggesting support for use of the online...
version. The relevance to the teaching environment within osteopathy has applicability to wider clinical healthcare. The role of gender as an influence in the satisfactory conduct of assessment warrants further investigation.

Introduction: Assessment of students’ clinical competence in healthcare education, typically involves the evaluation of their performance during a patient consultation (Mortaz Hejri et al., 2017). This provides focus upon the students’ knowledge, skills and attributes (Vaughan and Moore, 2016), and gives opportunity for valuable feedback for development and performance improvement (Torre et al., 2007). This process is crucial in clinical education, but current guidance (Lefroy et al., 2015) suggests it may not be fully utilised, with paper-based assessment initiating unqualified tick-box responses. A possible solution to this problem includes the use of specialist tools supporting clinical assessment, such as the mini-clinical evaluation exercise (mini-CEX), and direct observation of procedural skills (DOPS) (Lörwald et al., 2018).

The mini-CEX, initially developed by the American Board of Internal Medicine (Durning et al., 2002), focuses upon history-taking skills, examination skills, clinical judgement, professionalism, organisation and clinical competence during a patient consultation (Moore and Vaughan, 2016). Multiple examiners produce ratings for each of these areas and provide written/verbal feedback based on the students’ performance (Norcini and Burch, 2007). The additional formative feedback and rating system of the mini-CEX appears superior to more traditional methods (Habibi et al., 2013), with suggested increase in improvements to clinical skills. Consequently, the paper-based instrument is now fully established across a range of healthcare disciplines, such as medical education, nursing, social work, pharmacy and general practice as a core assessment tool (Alkureishi et al., 2018).

The scope for engagement is further developed with an electronic Personal Digital Assistant version, suggesting high-scale reliability (Cronbach’s alpha coefficient: 0.89) (Torre et al., 2007). Commensurate reports indicate that the mini-CEX also displays appropriate qualities as an educational feedback aid (Durning et al., 2002; Nair et al., 2008). This is supported, irrespective of the influence of examiners’ characteristics, such as seniority and gender or mode of assessment (Chang et al., 2017). Further exploration of influences is warranted given the contextualised, cross-sectional nature of the directly observed mini-CEX events (Rogausch et al., 2015). An exaggerated positive or negative bias, the ‘halo and horn’ effect, can be implicated in practical examination and this may be attributable to gender or racial bias (Minter et al., 2005; Guraya et al., 2016). Gender bias has also been observed in relation to acceptance of new technology in the workplace (Venkatesh, Morris and Ackerman, 2000; Venkatesh and Bala, 2008), but it is not clear how stereotypes may influence student expectations or satisfaction around online assessment (Maruping et al., 2017).

Such considerations may be offset given the mini-CEX’s reported cost-effectiveness and satisfying acceptability for both examiners and students, but qualitative, meaningful, experiential results are lacking around this area (Pelgrim et al., 2011). While positive satisfaction is mooted, there are doubts about the validity, potentially due to inadequate examiner training, which would detract from a satisfactory outcome (Vaughan, MacFarlane and Florentine, 2014). Indicators suggested that examiners felt the tool debased their core role of supervision (Moore and Vaughan, 2016), and being paper-based, added unwanted inconvenience (Torre et al., 2007), thus influencing satisfaction ratings. Such potential reductions in satisfaction scores could be limiting uptake of the paper-based instrument. Online replication of the mini-CEX has been established in an emergency healthcare setting with seemingly positive benefits over the paper equivalent (Chang et al., 2017). However, these findings around accessibility and acceptability may not be transferable to other settings in healthcare and education. It is not currently established if an online, form-based version of the instrument would influence satisfaction for students and assessors, in examination processes within an undergraduate manual therapy programme.

Aims and Objectives: The aim of this study was to explore how satisfaction with assessment is influenced by the process of a paper-based, compared to an online assessment method. The objectives were to create an online version of the mini-CEX, with a view to determine the examiner and student satisfaction, in comparison to the paper-based version, when used as part of an ongoing clinical assessment schedule. This was with a view to answer the following research question: Does gender and the method of assessment influence satisfaction scores in the use of the mini-CEX?

Methods: This was a cross-sectional, questionnaire study.

Participants: Third year students and clinic tutors, at an osteopathic educational teaching clinic, took part in organised formative and summative assessments from August 2016 to March 2019. These were standard, clinically-based examinations, required as part of the pre-registration process for entry into UK osteopathy practice. The clinical lead made it known, during induction activities
for each cohort, that the suitability of delivery of the clinical assessment was being determined as an action research process.

Ethics: The study was approved by the Research Ethics Committee of a UK-based institution, providing osteopathic pre-registration education outside of an NHS setting. The work was carried out in accordance with the Declaration of Helsinki, providing assurance that the anonymity of participants was upheld, following the informed consent of participants.

Procedure: An online mini-CEX questionnaire was developed using the cloud-based Google Forms (see Supplementary Material), and was initially piloted as a data entry tool. Administration staff used the form to upload details from completed paper assessment materials, following direct observation. The online form was directly based on these paper versions, as previously reported and validated in a similar educational environment (Vaughan and Moore, 2016). Subsequently, Android-based, 8-inch, tablets were prospectively phased-in to provide access to the online form, for direct reporting by tutors on observed clinical practice of students. These facilitated a retrospective comparison to the paper counterpart over the course of three academic years, as indicated by the schedule in Table 1.

Table 1: Assessment schedule and method of mini-CEX reporting

<table>
<thead>
<tr>
<th>Assessment Period</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>August-December 2016</td>
<td>Paper</td>
</tr>
<tr>
<td>February-April 2017</td>
<td>Online</td>
</tr>
<tr>
<td>May 2017</td>
<td>Paper/Online</td>
</tr>
<tr>
<td>June-July 2017</td>
<td>Online</td>
</tr>
<tr>
<td>August 2017</td>
<td>Paper/Online</td>
</tr>
<tr>
<td>September 2017</td>
<td>Online</td>
</tr>
<tr>
<td>October-November 2017</td>
<td>Paper</td>
</tr>
<tr>
<td>December 2017</td>
<td>Paper/Online</td>
</tr>
<tr>
<td>January 2018</td>
<td>Paper</td>
</tr>
<tr>
<td>February-March 2018</td>
<td>Paper/Online</td>
</tr>
<tr>
<td>April 2018</td>
<td>Online</td>
</tr>
<tr>
<td>May-June 2018</td>
<td>Paper/Online</td>
</tr>
<tr>
<td>July 2018</td>
<td>Online</td>
</tr>
<tr>
<td>August 2018</td>
<td>Paper</td>
</tr>
<tr>
<td>October-December 2018</td>
<td>Paper/Online</td>
</tr>
<tr>
<td>January-December 2019</td>
<td>Online</td>
</tr>
</tbody>
</table>

Students were assessed by different clinical tutors as standard, to afford a range of independent markers for each examinee; pragmatically, there may have been instances where a tutor had assessed a student on more than one occasion, or where lack of immediate access to tablet devices necessitated paper use. The satisfaction scores were captured at each assessment as integral items common to the mini-CEX format (Durning et al., 2002). These are formed of two six-point scales (one for examiner and student), where 1 indicates low satisfaction, and 6, high satisfaction. Both students and tutors were present at the time of rating the satisfaction and would have sight of the score reported.

Statistical Analysis: The assessment data collected from the forms were exported to a spreadsheet and downloaded into Microsoft Excel version 14 (Microsoft Corporation, Redmond, WA, USA) for generation of pivot table summaries, and measures of central tendency and dispersion. The influence of paper and online methods of assessment capture was explored with binary regression, with the dependent variables of student and tutor satisfaction. Student and tutors were assigned identified gender categories (Male (M) or Female (F)) to determine groups of same or mixed gender assessment pairings (MM/FM/MF/FF); the student was indicated by the lead character in each pair. The difference between the gender groups’ satisfaction scores was then explored using the Kruskal Wallis test, with a post-hoc, Steel-Dwass-Critchlow-Fligner pairwise, distribution-free, multiple comparison completed if statistical significance was demonstrated (Spurrier, 2006). The combination of gender pairs and potential influence with the capture process was also explored using the Chi\(^2\) test, Kruskal Wallis and aforementioned post-hoc test strategy. Correlation between student and tutor satisfaction scores were tested with Spearman’s Test to report \(r^2\) values and 95% confidence intervals (CI). The statistical tests were run using Analyse-it version 4.65.3 (Analyse-it Software, Ltd, Leeds, UK), with significance set to 5% and confidence intervals reported at a level of 95%.

Results: A total of 736 mini-CEX assessment of patient encounters were included in the analysis; 550 (75%) were completed online and the profile against the yearly assessment schedule can be seen in Table 2. Forty-four tutors (32% female) assessed 159 students (69% female) across this schedule.

The satisfaction scores from examiners were complete for all 736 records, but 15 records (2%) failed to record student satisfaction, and of these, two originated during the paper-based assessment. These were the consequence of tutors completing the submission of the process without the
student present, due to over-running of the process, or staffing issues; pairwise deletion was applied to these cases. The gender pairings of student and tutor equated to uneven groups: 391, FM; 99, FF; 51, MF; 180, MM (Chi², P<.0001). The central tendencies of the satisfaction scores can be seen in Table 3, with lower overall satisfaction reported by tutors. Median values indicated an overall ‘good’ level of satisfaction with the examination process.

### Table 3: Summary of satisfaction scores

<table>
<thead>
<tr>
<th>Method</th>
<th>Measure</th>
<th>Overall Median (IQR)</th>
<th>Online Median (IQR)</th>
<th>Paper Median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Student satisfaction</td>
<td>4 (2)</td>
<td>4 (2)</td>
<td>4 (2)</td>
</tr>
<tr>
<td></td>
<td>using mini-CEX</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tutor satisfaction</td>
<td>4 (3)</td>
<td>4 (3)</td>
<td>4 (3)</td>
</tr>
<tr>
<td></td>
<td>using mini-CEX</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There was no difference between satisfaction scores for the paper method of assessment, compared to the online process (odds ratio 1, CI 0.86 – 1.15). The groupings around paired gender satisfaction ratings indicated that male tutors assessing female students (FM group), demonstrated lower summary values when compared to female tutors and students (FF group) (P<.005)*, following pairwise comparisons (Table 4). Allied to this finding, student satisfaction ratings for female students assessed by male tutors, also indicated lower summary scores when compared to male students assessed by male tutors (MM group) (P<.007)**. The pairwise comparisons are also indicated in Table 4, subsequent to Kruskal Wallis outcomes (P<.006).

There was only one significant difference determined around examiner satisfaction scores and capture process, when explored with student-tutor gender pairings. With respect to online capture in the grouping of same gender female pairing, a difference was found when compared to paper-based capture of female students assessed by male tutors (3.9 vs 3.2 (mean), P<.045). The correlation between all student and tutor satisfaction ratings indicated an r² value of 0.62 (95% CI 0.57 – 0.66, P<.00001), or over 60% of variance was associated between scores, leaving 40% of the overall variance unaccounted for by the statistical model.

### Table 4: Steel-Dwass-Critchlow-Fligner all pairs comparisons of tutor and student satisfaction scores

<table>
<thead>
<tr>
<th>Group comparisons (mean score)</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF(4.0) - FM(3.9)</td>
<td>0.0 to 1.0</td>
<td>.7503</td>
</tr>
<tr>
<td>FF(4.0) - MF(4.3)</td>
<td>-1.0 to 0.0</td>
<td>.6245</td>
</tr>
<tr>
<td>FM(3.9) - MM(4.3)</td>
<td>-1.0 to 0.0</td>
<td>.4026</td>
</tr>
<tr>
<td>MF(4.3) - MM(4.3)</td>
<td>-1.0 to 1.0</td>
<td>.9998</td>
</tr>
</tbody>
</table>

Conclusions: The aim of this study was to explore how the satisfaction with assessment is influenced by the method of capture and gender. There was no inferred influence on satisfaction rating of the assessment, based on the online or paper-based capture of observed clinical practice. While differences were found across both parties’ satisfaction scores when female students were assessed by male tutors, this particular grouping was the most frequently occurring in the pairings. There was a moderate level of association between the satisfaction ratings of both students and tutors across the range of assessments.

The finding of no influence on satisfaction ratings from the mini-CEX capture method, conflicts with reported elements captured in an emergency setting; Chang et al. (Chang et al., 2017) indicated odds (OR 1.47) in favour of a computer-based format prompting the presence of positive feedback, developmental indicators and agreed action plans. This
finding was established from 1101 assessment events compared to 736 in the current study. Undertaking assessment within trauma medicine also has potential implications, not only in terms of the life and death scenario, but in the hierarchy of healthcare professionals involved in emergency scenarios. This ‘seniority’ was seen to have an impact that was not possible to explore in the monotechnic osteopathy teaching clinic. While the structure of tutors is hierarchically ‘flat’, a small number of management staff would be involved in assessment duties. While this influence may be equivocal, osteopathy has been characterised as a profession of divisive attitudes (Kasiri-Martino and Bright, 2016), split between the values of traditionalists and progressives, embodying professional artistry, technical rationalism or evidence-informed pragmatism (Thomson, Petty and Moore, 2014; Figg-Latham and Rajendran, 2017). These attitudes were not captured or identifiable in the current sample; there may be issues around students aligning to their seniors’ shared and voiced sensibilities and expectations, that then relate to mutual satisfaction of experience (Borghi, Mainardes and Silva, 2016). As the satisfaction scoring was unblinded, the scope for mutuality was present and could have informed cognitive bias. Chang et al. underline the role that professional standing has on feedback, but not the satisfaction rating of the experience; this may relate to a wider shared philosophical viewpoint, or cultural code of examiner and examinee, where deviation from rational judgement arises (Chang et al., 2017). The implications for the authenticity of the assessment experience may require triangulation with actual patients, rather than virtual ones, to further establish integrity (Forsberg et al., 2016; Perrella, 2016).

The influence of technology in this study was seemingly marginal in relation to the ‘good’ satisfaction level reported; further indication that clinical assessment can be facilitated through an online process as an alternative to paper capture (Froud et al., 2018). This also corroborates findings reported within a medical school setting where assessment satisfaction was apparent in over 90% of observers (Ferenchick et al., 2013). The application of the Technology Acceptance Model is supported in that perceived usefulness and perceived ease of use, can be inferred from the lack of discernible change between scores (Venkatesh and Bala, 2008). The inference is that the satisfaction rating is independent of the medium of capture. Barriers to the acceptance of technology in this discipline may have become moderated in line with attitudes reported in earlier stages of education (Ertemer et al., 2012). This may be facilitated by the context of the assessment remaining consistent between the paper and online capture phases; setting, expectation and dosage were in line with review findings (Lörwald et al., 2018). The fidelity and quality of assessment may be prone to inconsistency in application, outside of the medium of capture, given the measures of dispersion recorded around satisfaction in this study. These may be influenced by the factors reported in the wider literature that were not observed here, such as the variance in perceived complexity of the observed event, dependent on the patient presentation (Cook et al., 2009; Rogausch et al., 2015). Prior reports indicate that an extensive range of musculoskeletal problems and associated issues, in keeping with a primary healthcare discipline, are encountered in these pre-registration environments (Rajendran et al., 2015; Judkins, Vaughan and Mulcahy, 2017). The cross-sectional nature of the clinical assessment provides the challenge of exposure to these complexities.

Gender has been reported as a potential factor influencing assessment outcome, with female academic performance outpacing male across general education, seen partly attributable to a more considered feminine trait regarding strategy and planning (Carvalho, 2016). This may also be supported by increased self-efficacy informing expectations and promulgating satisfaction, but female trends are more implied in this instance, as these characteristics are seen as generally present in high academic achievement (Doménech-Betoret, Abellán-Roselló and Gómez-Artiga, 2017). The findings around gender influence on satisfaction in the current study are prone to bias. Uneven pairings indicated higher mean scores for male students assessed by male tutors, with lowest mean scores reported for female students assessed by male tutors. This may be indicative of linked satisfaction with test anxiety, emotionality and performance (Hill et al., 2009; Steinmayr et al., 2016; Nasir and Iqbal, 2019), but one proposed benefit of assessing observed practice is the reduction in exam stress due to familiarity of setting (Ansari, Ali and Donnon, 2013). While these findings have the potential to support misogynist tendencies, particularly given the female student majority (Morley, 2011), conflicting with the male assessor hegemony (Burke, 2017), the dynamics of the student/tutor ratio would indicate that female/male pairings are the majority group in this sample, and hence, conservative interpretation is warranted (median values are comparable across all pairs). The 40% of variance that is unaccounted for in the current findings may also be further elucidated by both known and unknown factors around participant characteristics, such as age, attitude, emotional intelligence, ethnicity and prior achievement. Further matched group assessment may be possible in the future, given the current equilibrium between male and female osteopaths that are state registered (GOsC, 2019). Aspirations of growing inclusivity and diversity within UK osteopathic education, following on from wider calls in Higher Education (Bracken and Novak, 2019), should become a reality and allow for unequivocal comparison.
On comparing all sets of satisfaction scores, there was indication of alignment between the students and tutors, with moderate correlation indicated ($r^2=0.62$). The suggestion from the overall sample is that students have a tendency to follow their seniors and satisfaction is mutually agreed, particularly as the students were not blinded to their assessors’ score. This may be indicative of the legacy of paternalism in this healthcare educational structure (Padua Filho, Padua and Fernandes, 2019). In similar clinical assessments, independent reports of satisfaction with examination processes between examiners and students, indicated no difference in ratings either (Amiri and Nickbakht, 2012; Dhinakaran, Mullai, Jugesh Chattwal, 2015). These reports do not emphasise strength or direction of relationship within the two parties’ scores. The Objective Structured Clinical Examination (OSCE) used in these studies is also a more fixed assessment, potentially avoiding the clinical uncertainty that directly observed practice may proffer (Spanke et al., 2019). The level of satisfaction reported across the use of mini-CEX in an Australian pre-registration programme, indicates a higher overall rating (median 5, mean 4.75 – 4.81) (Vaughan and Moore, 2016), but again the strength of relationship between these scores is unreported. The Level 6/7 programme content is comparable with that of the current UK study (GOsC, 2019), although Australia’s extended clinical course duration led to Year 4 and 5 students being assessed. This may account for the difference in satisfaction due to the additional clinical experience the students would have gained compared to Year 3 students included in this study.

Strengths and Limitations: The large range of the sample suggests there is potential for paper-based assessments to be revised as online tools within osteopathy education and other clinical settings. The option to expedite data capture and analysis can then provide contemporary feedback to students, whilst also ensuring the health of a course and its curriculum through monitoring. This can provide effective use of educational staff time, more accessible data and further support the student experience.

The limitations of the study are that the ordinal scores do not fully capture the nuances around the components that inform a satisfactory experience. The scope to conduct qualitative studies to explore the experience of the administration of the mini-CEX and any perceived influences in this type of assessment, is warranted. The phenomenon of alignment around satisfaction ratings between student and examiner, including the nuances of practitioner sensibilities, also warrants further investigation. The option to include a blinded approach to capture the satisfaction rating may provide more open reporting, but disentangling this from direct observation may be problematic. A technological solution in this area requires further innovation and development.

Implications of gender dynamics is a moot point given the imbalance in the groups of female and male participants. Further exploration with balanced groups in the profile of arranging assessments would be pertinent and could form the basis of a comparative study. The assessment process itself assumes a silent third party in that the patient experience is not captured which may authenticate the satisfaction levels. Adopting a patient voice in the process could add vital context to the assessment, providing triangulation around student/tutor relationship, using a patient specific online mini-CEX. The inclusion of such data may provide additional benefit to the student and assist with the development of communication and professionalism as demanded by regulated practice. Future research should look to employ mixed methods to explore the replication and qualification of these results. There is scope to adopt more inclusive designs, with models to account for greater nuances of personal student and tutor characteristics.

Summary: The aim of this study was to explore how the satisfaction with assessment is influenced by the process of capture. The findings suggest that there is no statistically significant difference between the two methods of delivery in terms of satisfaction of use for either examiner or student, potentially indicative of the suitability of the online version. While this has relevance to the teaching environment within osteopathy, there is applicability to other clinical healthcare areas. The role of gender as an influence in the satisfactory conduct of assessment warrants further investigation. In-depth qualitative investigation is warranted with students, examiners and patients in a range of clinical assessment settings to contextualise these findings.

References:


