Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training

Alberto Alves de Lima, Carlos Barrero, Sergio Baratta, Yanina Castillo Costa, Guillermo Bortman, Justo Carabajales, Diego Conde, Amanda Galli, Graciela Degrange, Cees Van DER Vleuten

Argentine Society of Cardiology Buenos Aires, Argentina
Instituto Cardiovascular de Buenos Aires, Education, Argentina
Educational Development and Research, Maastricht University, The Nederlands

First Published on: 21 September 2007
To cite this Article: de Lima, Alberto Alves, Barrero, Carlos, Baratta, Sergio, Costa, Yanina Castillo, Bortman, Guillermo, Carabajales, Justo, Conde, Diego, Galli, Amanda, Degrange, Graciela and Van DER Vleuten, Cees (2007) 'Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training', Medical Teacher, 29:8, 785 - 790
To link to this article: DOI: 10.1080/01421590701352261
URL: http://dx.doi.org/10.1080/01421590701352261

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training

ALBERTO ALVES DE LIMA¹, CARLOS BARRERO¹, SERGIO BARATTA¹, YANINA CASTILLO COSTA¹, GUILLERMO BORTMAN¹, JUSTO CARABAJALES¹, DIEGO CONDE², AMANDA GALLI¹, GRACIELA DEGRANGE¹ & CEEES VAN DER VLEUTEN³

¹Argentine Society of Cardiology Buenos Aires, Argentina, ²Instituto Cardiovascular de Buenos Aires, Education, Argentina, ³Educational Development and Research, Maastricht University, The Nederlands

Abstract

Aims: The purpose of the study was to determine the validity, reliability, feasibility and satisfaction of the Mini-CEX.

Methods and Results: From May 2003 to December 2004, 108 residents from 17 cardiology residency programs in Buenos Aires were monitored by the educational board of the Argentine Society of Cardiology. Validity was evaluated by the instrument’s capability to discriminate between pre-existing levels of clinical seniority. For reliability, generalisability theory was used. Feasibility was defined by a minimum number of completed observations: 50% of the residents obtaining at least four Mini-CEX’s. Satisfaction was evaluated through a one to nine rating scale from the evaluators, and residents’ perspectives. The total number of encounters was 253. Regarding validity, Mini-CEX was able to discriminate significantly between residents of different seniority. Reliability analysis indicated that a minimum of ten evaluations are necessary to produce a minimally reliable inference, but more are preferable. Feasibility was poor: 15% of the residents were evaluated four or more times during the study period. High satisfaction ratings from evaluators’ and residents’ were achieved.

Conclusion: Mini-CEX discriminates between pre-existing levels of seniority, requires considerable sampling to achieve sufficient reliability, and was not feasible within the current circumstances, but it was considered a valuable assessment tool as indicated by the evaluators’ and residents’ satisfaction ratings.

Introduction

The Mini-CEX has been designed to incorporate the skills that residents require in both actual patient encounters and in the educational interactions that they routinely encounter with attending physicians during teaching rounds (Norcini et al. 1995; Norcini et al. 1997; Holmboe et al. 1998; Norcini et al. 2003). A single faculty member observes and evaluates a resident while that resident conducts a focused history and physical examination on an inpatient or outpatient, or in the emergency department setting. After asking the resident for a diagnostic and treatment plan, the faculty member completes a short evaluation form and gives the resident feedback (Holmboe 2004). It is a performance-based evaluation method that is used to assess selected clinical competencies (e.g. patient charts and physical examination, also communication and interpersonal skills) in the medical training context. It is a performance-based assessment tool that intended to evaluate candidates at the ‘does’ level, that is, in real life settings and not in simulated situations (Miller 1990). As the interaction is relatively brief and occurs as a natural part of the process in the training environment, each individual can be evaluated on several occasions and by various faculty members.

Practice points

- The Mini-CEX has been designed to incorporate the skills that residents require in both actual patient encounters and in the educational interactions that they routinely encounter with attending physicians during teaching rounds.
- Regarding validity, Mini-CEX clearly was able to discriminate between pre-existing levels of global competency between residents.
- Mini-CEX has insufficient reproducibility when only a few evaluations are sampled.
- Mini-CEX has satisfaction rates according to both evaluators and residents satisfaction rates.
- Regarding feasibility, it appeared not easy to achieve the number of encounters required.

Assessment constitutes the most vital factor influencing student learning behavior (Newble et al. 1983; Newble et al. 1990; Van der Vleuten 1996). When students see that the recall of factual information is a predominant requirement in the examination system, they tend to adopt a rote-learning or
surface approach; however, if examiners wish to assess students at the ‘does’ level, they must evaluate the student’s habitual performance in daily practice (van der Vleuten 2000).

The major purpose of this study is to document if Mini-CEX is applied in a broad range of clinical settings and in a big number of cardiology residency programs leads to achieving adequate levels of validity, reliability, feasibility and satisfaction rates from residents and teachers.

**Methods**

For each Mini-CEX, a single faculty member observed and evaluated the resident while the latter conducts a history and physical examination on an in-patient or an outpatient or on a patient in the emergency department. After asking the resident for a diagnosis and treatment plan, the faculty member completes a short evaluation form and gives direct feedback. All formal mini-CEX evaluation data was collected on a one-page form that was the same in all of the different study sites where the study was carried out (Appendix 1). The form was previously translated and transculturally adapted into Spanish.

Research subjects were cardiology residents from 17 cardiology training programs from Buenos Aires, Argentina. All the programs are affiliated to the University of Buenos Aires and consist of a four-year training period. The total number of residents of the entire program was 118. All the program directors were invited to participate. It was a completely new assessment strategy for all of them. Participation was voluntary, no incentives were provided. Written instructions about the application of the format were distributed and required at least four encounters during the 19 months study period. There were no sanctions for failing to participate, but all of them accepted to participate. The assessment was used as a maximum-performance but formative evaluation. Results were not used in evaluating residents for promotion.

**Validity analysis**

Validity was evaluated by the ability of Mini-CEX to discriminate between pre-existing levels of expertise. In this case it was expected that significant mean resident performance differences are to be found between different years of training. The descriptive data were expressed as its means and standard deviations. For testing significance across expertise groups the non-parametric Mann Whitney test was used. A value of $p < 0.05$ was considered statistically significant.

To evaluate reliability, generalizability theory was used (Brennan 2001). An ANOVA was carried out by identifying Year-of-training with (Y), Residents-within-Year with (P:Y) and Evaluations-within-Residents-within-Year with (E:P:Y), and variance components were estimated using the URGenova program. Since there might be significant growth in mean ratings throughout the years, variance associated with year was estimated separately (Y) to arrive at a more unbiased estimate of the variance of trainees. Separate evaluations of a single trainee could either be done by the same examiner or by a different one. This might have led to an underestimate of the variance across evaluations (intra-rater variability is probably smaller than inter-rater variability). Two indices of reliability were estimated using the variance components: Dependability Coefficients (D) and Standard Errors of Measurement (SEM), both as a function of the number of evaluations. The D-coefficient can be interpreted as a reliability coefficient, i.e., the expected correlation between other random evaluations of the number indicated using other evaluators and patients at random. The SEM is an estimate of the standard error and can be used to estimate confidence intervals around the score of an individual resident on the original scoring scale. For a 95% confidence interval the SEM is multiplied by 1.96 (z-score under which 95% of normal distribution lies). The SEM should be below 0.26 (0.5/1.96) in order to produce a reliable inference on the scoring scale of at least one unit.

Feasibility was defined, according to the American Board of Internal Medicine’s guidelines for Mini-CEX’s implementations, on average a minimum of four Mini-CEX per resident (American Board of Internal Medicine 2005).

Satisfaction was evaluated through the examination of the Mini-CEX from the perspective of evaluators, emphasizing on the ratings of the residents and on their satisfaction of the format. Ratings were carried out on a nine-point rating scale.

**Results**

From May 2003 to December 2004, 253 Mini-CEX encounters were carried out. 108 residents and 53 evaluators from 17 cardiology residency programs participated in the study. Each resident had gone through 1 to 7 evaluations (mean: 2.34): 13.7% of the residents were in their first year, 34.8% in the second, 41.2% in the third and 10.3% in their fourth year of residency. Each evaluator conducted between 1 and 21 (mean 4.77) evaluations. The total numbers of encounters were 253 and constituted the basis of the analysis. Of the 253 encounters, 52% of the encounters occurred in the coronary care unit setting, 30% were carried out in step-down care unit, 6% in the emergency room, 6% in the ambulatory care unit and 6% in the cardiovascular intensive care unit. The overall competence ratings were similar in all settings. Forty-one percent of the encounters represented the first visit of a patient to a particular resident and 59% were return visits to that resident. The means total Mini-CEX time was 42.77 minutes (SD 19.97). If we divide the Mini-CEXs total time between assessment and feedback period, each individual assessment period takes 25.80 (SD 11.95, range 5–65) minutes and feedback period 17.31 (SD 11.28, range 5–65). The patient’s problems or diagnoses were specified by the evaluator and covered a broad range of problems in cardiology such as AMI, cardiac failure, unstable angina, atrial fibrillation, valvular disease and post-CABG. The mean ratings given by the 53 evaluators are reported in Table 1.

**Validity analysis**

Validity was evaluated by examining if the instrument was capable of discriminating between pre-existing levels of clinical seniority. Mini-CEX discriminates between pre-existing levels of global competency between residents; first year residents 7.19 (SD 0.74), second 7.51 (SD 0.82), third 7.76
Validity, reliability, feasibility and satisfaction of the Mini-CEX for cardiology residency training

Table 1. Mean Ratings given by all evaluators by year of training.

<table>
<thead>
<tr>
<th>Domain</th>
<th>1 year</th>
<th>2 year</th>
<th>3 year</th>
<th>4 year</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>7.16 (±0.64)</td>
<td>7.57 (±0.81)</td>
<td>7.57 (±0.92)</td>
<td>8.00 (±0.88)</td>
<td>0.002</td>
</tr>
<tr>
<td>Physical exam</td>
<td>7.12 (±0.84)</td>
<td>7.48 (±0.93)</td>
<td>7.59 (±0.96)</td>
<td>8.16 (±0.91)</td>
<td>0.0006</td>
</tr>
<tr>
<td>Professionalism</td>
<td>7.64 (±0.75)</td>
<td>7.82 (±0.84)</td>
<td>7.83 (±1.00)</td>
<td>8.20 (±0.93)</td>
<td>0.079</td>
</tr>
<tr>
<td>Clinical judgement</td>
<td>7.43 (±0.71)</td>
<td>7.56 (±0.86)</td>
<td>7.88 (±0.90)</td>
<td>8.20 (±0.93)</td>
<td>0.0004</td>
</tr>
<tr>
<td>Counselling</td>
<td>7.43 (±0.77)</td>
<td>7.44 (±0.84)</td>
<td>7.59 (±1.11)</td>
<td>8.12 (±0.90)</td>
<td>0.01</td>
</tr>
<tr>
<td>Organisation</td>
<td>7.32 (±0.94)</td>
<td>7.54 (±0.88)</td>
<td>7.68 (±1.01)</td>
<td>8.12 (±0.90)</td>
<td>0.008</td>
</tr>
<tr>
<td>Global Competency</td>
<td>7.19 (±0.74)</td>
<td>7.51 (±0.82)</td>
<td>7.76 (±0.86)</td>
<td>8.16 (±0.91)</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

(SD 0.86) and fourth year residents 8.16 (SD 0.91), this difference reaches statistical significance. P = 0.0008 (Table 1).

Reliability analysis

The generalizability study theory yielded variance components for Y, P:Y and E:P:Y which were respectively 0.16 (19.88% of total variance), 0.0482 (5.83%) and 0.614 (74.29%). Using the SEM benchmark of 0.26, a minimum of 10 evaluations were necessary to produce a minimal reliable inference (Table 2). This corresponds to a D-coefficient of 0.44.

Feasibility analysis

For feasibility analysis the data showed that only 14.81% of all the cohort was evaluated four or more times during the study period.

Satisfaction analysis

The residents (108) were generally satisfied with the mini-CEX format; their ratings ranged from 5 to 9 (mean 8.06 ± 0.74).

Discussion

The purpose of the study was to report logistic and psychometric data for the Mini-CEX format. Regarding validity, Mini-CEX clearly was able to discriminate between pre-existing levels of global competency between residents, has insufficient reproducibility when only a few evaluations are sampled and had high satisfaction rates according to both evaluators and residents satisfaction rates. Regarding feasibility, it appeared not possible to achieve the number of encounters required.

Some issues may explain this difficulty. It was a new assessment tool never applied earlier in this environment. We only developed written instructions, and perhaps in vivo faculties training programs for Mini-CEX would be preferable. Regarding reliability, the variance components for Y, P:Y and E:P:Y one-fifth of the variance can be attributed to growth in competence throughout the years. Provided that the Mini-CEX should offer information on growth towards a final level of competence and should be able to discriminate throughout years of training. Only approximately 6% of total variance is related to resident (or person) variance. Since the instrument is designed to discriminate among residents, this constitutes desirable (true score or universe score) variance and the larger it is the better. As usual in many competence and performance assessments, however, this component is relatively small (van der Vleuten 2005).

About three-quarters of variance is associated with differences between examiners/evaluation occasions and residual error. Using the SEM benchmark of 0.26, a minimum of ten evaluations are necessary to produce a minimal reliable inference. The D coefficient for ten observations was, however, rather low (0.44). Reliability coefficients of 0.80 or higher are generally accepted as a threshold for high-stakes judgements, such as the registration of a doctor for licensure (Crossley et al. 2002). In our dataset this was achieved with 50 observations.

The required number of encounters as derived from this study is more demanding than reported by other studies in the literature. Norcini et al. (2003) concluded that ten or more encounters produced relatively tight confidential intervals and that an increase in the number of encounters beyond that produced only small gains in consistency. Carline et al. (1992) concluded that seven independent ratings would be necessary to judge overall clinical performance. Similar results have been
reported by Kreiter and colleagues (more than eight ratings needed), Kwolek (7–8 ratings), Ramsey (11 ratings), Violato (10 ratings) and Kroboth (6–10 ratings) (Kroboth et al. 1992; Ramsey et al. 1993; Kwolek et al. 1997; Violato et al. 1997; Kreiter et al. 1998; Durning et al. 2002). All reports agree that somehow between seven and 11 ratings are necessary to achieve a generalizable global estimate of competence when ratings are based on a non-systematic sample of observations (Williams et al. 2003). Although there has been only limited research on the component skills included under the broad category of clinical competence, it is reasonable to expect that these abilities will develop at different rates and may differ in their stability in different situations. Results of performed work suggest that different numbers of observations will be required to establish stable estimates of clinical competence in various clinical competence areas (Williams et al. 2003).

We have not analysed inter-rater reliability since we took a time sampling perspective in which each observation is but one observation in a longer time-framework. We analysed reliability across these observations. Rater variance as well as performance variability variance across observations will be part of our reliability estimation (although we can’t partition out the two sources of variance). If only one observation is used for a Mini-CEX assessment it would be important to judge the inter-rater reliability. Our data, unfortunately, do not provide this information.

Consistent with previous works, examiners were satisfied with the format (Williams et al. 2003).

Limitations of the study

The number of residents participating in our study was relatively low and this group may perhaps not be fully representative of broader populations.

We are aware that improvement of performance across years of training is a weak form of construct validity. However, it is a fundamental one: absence of performance differences across different expertise groups would be detrimental to construct validity of the instrument used. This not being the case, we conclude to a first indication of validity. Further studies into construct validity should be the next step. Further looking at the incremental validity over existing more standardized performance instruments would provide compelling construct validity evidence.

The assessment used was a maximum-performance but formative evaluation. If we take into account that this assessment does not assign grades or certifications, this could seriously have affected the perception of the fellows and influence their satisfaction rates.

The reliability analysis carried out here used the usual assumption of local independence between the repeated measurement moments (one measurement is not influenced by another measurement). This assumption is clearly violated in Mini-CEX studies including ours. Every evaluation is actually meant to provide feedback and to change the performance of the person being assessed. The Mini-CEX evaluations are therefore not independent of each other. This is a general problem in the literature of performance measure which are dispersed in time and used for formative purposes (Alves de Lima et al. 2005).

Conclusion

The direct observation of residents’ behaviour is essential to assess clinical skills (Holmboe et al. 2004a; Holmboe et al. 2004b; Schuwirth & Van der Vleuten 2006). For decades, clinical supervisors have taken at face value the veracity of the history and physical examination presented on inpatient and outpatient rounds without ever directly observing how the trainee actually performed them (Holmboe 2004).

Medical educators’ major challenge lying ahead is to find the way to ensure that they themselves have not only strong clinical skills, but also the necessary skills to effectively observe, evaluate, and provide constructive feedback to trainees regarding clinical skills. In this direction Mini-CEX ensures direct observation and feedback of residents from different faculties in a broad range of patients’ problems in various settings. Furthermore, as this demonstrated feasibility is an issue. Application strategies should be reinforced. We don’t think that Mini-CEX requires any modification in itself but it should never be used as a unique assessment tool. Direct observation of trainees in clinical setting can be connected to other exercises that trainees may perform after their encounters with patients, such as oral case presentation, written exercises that assess the clinical reasoning, and literature searches. In addition, review videos of encounters with patients offer a powerful means of evaluating and providing feedback on trainees’ skills in clinical interaction (Epstein 2007). It can only be approved if this evaluation method becomes part of the clinical routine of the clinician and resident.

Notes on contributors

ALBERTO ALVES DE LIMA, MD, MHPE is a cardiologist, Head of the Education Department at the Instituto Cardiovascular de Buenos Aires, Buenos Aires, Argentina and a member of the educational department of the Argentine Society of cardiology.
CARLOS BARRERO, MD is a cardiologist and a member of the educational department at the Instituto Cardiovascular de Buenos Aires, Buenos Aires, Argentina.
AMANDA GALLI is educational Psychologist and a member of the Education Department of the Argentine Society of cardiology.
CEES VAN DER VLEUTEN, Professor, is Psychologist and Chairperson of the Department of Educational Development and Research, Maastricht University, The Netherlands.
Validity, reliability, feasibility and satisfaction of the Mini-CEX for cardiology residency training

References


# APPENDIX 1: The Mini-CEX form

| Evaluator: __________________________ | Date: __________ |
| Fellow: ____________________________ | F-1 | F-2 | F-3 |
| Patient Problem/Dx: | | | |
| Setting: | O Ambulatory | O In-patient | O ED | O Other |
| Patient: | Age: ____ | Sex: ____ | New: ____ | Follow-up: ____ |
| Complexity: | Low | Moderate | High | |
| Focus: | Data gathering | Diagnosis | Therapy | Counselling | |

1. Medical interviewing skills: (O not observed)
   - 1 2 3 4 5 6 | 7 8 9
   - Unsatisfactory | Satisfactory | Superior |

2. Physical examination skills: (O not observed)
   - 1 2 3 4 5 6 | 7 8 9
   - Unsatisfactory | Satisfactory | Superior |

3. Humanistic qualities / professionalism: (O not observed)
   - 1 2 3 4 5 6 | 7 8 9
   - Unsatisfactory | Satisfactory | Superior |

4. Clinical judgment: (O not observed)
   - 1 2 3 4 5 6 | 7 8 9
   - Unsatisfactory | Satisfactory | Superior |

5. Counselling skills: (O not observed)
   - 1 2 3 4 5 6 | 7 8 9
   - Unsatisfactory | Satisfactory | Superior |

6. Organization / efficiency: (O not observed)
   - 1 2 3 4 5 6 | 7 8 9
   - Unsatisfactory | Satisfactory | Superior |

7. Overall clinical competence: (O not observed)
   - 1 2 3 4 5 6 | 7 8 9
   - Unsatisfactory | Satisfactory | Superior |

Mini-CEX time: _____  Observing: Min _____  Providing feedback: Min _____

Evaluator satisfaction with mini-CEX:
- Low 1 2 3 4 5 6 7 8 9 High

Fellow’s satisfaction with mini-CEX:
- Low 1 2 3 4 5 6 7 8 9 High

Comments:

Fellow signature  Evaluator signature

Dx = diagnosis; ED = emergency department; min = minutes; F1 = first year fellow; F2 = second year fellow; F3 = third year fellow