A Randomized-Controlled Study of Encounter Cards to Improve Oral Case Presentation Skills of Medical Students

Sarang Kim, MD,¹ Jennifer R. Kogan, MD,² Lisa M. Bellini, MD,² Judy A. Shea, PhD²

¹Division of General Internal Medicine, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, New Brunswick, NJ, USA; ²Department of Medicine, University of Pennsylvania, Philadelphia, Pa, USA.

OBJECTIVE: To determine the feasibility of oral case presentation (OCP) encounter cards as a tool for formative evaluation, to estimate the reliability and validity of the ratings when used in a medicine clerkship, and to examine whether the use of OCP encounter cards improves students' OCP skills.

DESIGN: Randomized controlled study.

SETTING: Medicine core clerkship at a U.S. medical school.

PARTICIPANTS/INTERVENTION: Students enrolled in the medicine core clerkship ($n=164$) from January to December of 2003 were randomly assigned to receive weekly feedback using OCP encounter cards rating nine presentation competencies or receive usual feedback. Mean OCP ratings were correlated with multiple summative assessments. Performance on an end-of-clerkship OCP was compared between intervention and control groups.

MAIN RESULTS: Eighty percent of cards were completed. The mean OCP rating averaged over 9 competencies was 7.7 (SD=0.8) on a 9-point scale. Standard error of ratings was 0.3. OCP ratings were correlated with inpatient evaluations ($r=.58$), inpatient ratings of presentation skills ($r=.43$), and final grades ($r=.40$). Final OCP performance was similar for the intervention and control groups (7.0 vs 7.2, $P=.09$).

CONCLUSION: OCP encounter cards are a novel and feasible tool to assess clerkship students' oral case presentation skills. OCP card ratings are reproducible, and validity is suggested by their correlation with multiple markers of performance. However, encounter cards did not improve performance on summative oral presentations.

KEY WORDS: medical education; medical students oral case presentations; clinical competence.

DOI: 10.1111/j.1525-1497.2005.0140.x

J GEN INTERN MED 2005; 20:743-747.

The Association of American Medical Colleges recognizes communication skills as a core learning objective for medical student education.¹ While much research has focused on physician-patient communication, few studies have been directed at communication between physicians. The oral case presentation (OCP) is an essential communication tool for physicians.^{1,2} In practice, it facilitates efficient transfer of clinical information between health care providers; during training, it also allows medical students to convey their clinical skills to resident or faculty evaluators.³ Despite its importance, little research has focused on evaluating OCP skills. Prior studies have examined the content of OCPs, how students learn OCP skills, and curricula to teach OCP skills.⁴⁻⁷ However, no study has examined the use of an assessment instrument to evaluate and improve OCP skills of medical students on a core clerkship.

Students generally learn OCP skills by trial and error.⁷ By providing formative feedback that is explicit and contextual, students may be able to learn OCP skills more efficiently. Encounter cards are a feasible method to provide case-specific, timely feedback about clinical skills.⁸⁻¹⁰ It is plausible that such encounter cards may also be useful in learning and evaluating oral case presentation skills. Therefore, we developed OCP encounter cards that enabled faculty and residents to rate students' OCP skills and provide timely, contextual feedback.

The objectives of this study were to: (1) determine the feasibility of OCP encounter cards as a tool for formative evaluation; (2) estimate the reliability and validity of the ratings when used in a medicine clerkship; and (3) examine whether the use of OCP encounter cards improved students' OCP skills. We hypothesized that OCP ratings would be moderately correlated with assessments of knowledge, clinical performance, and documentation skills. Our rationale was that these performance assessments rely, to some extent, on having a sound knowledge base and at least a rudimentary set of skills involving clinical synthesis and judgment. Correlations would not be too high, however, because each performance assessment requires certain unique skills. We also hypothesized that students using the OCP encounter cards would score higher on an end of clerkship presentation assessment than those not using the cards.

METHODS

OCP Encounter Cards

OCP encounter cards were developed using the medical education literature and were modeled on the mini-clinical evaluation exercises (mini-CEX).^{6,10-12} They had nine presentation competencies: history of present illness, past medical history, social/family history, physical exam, studies, assessment/differential diagnosis, plan, organization and coherence of the presentation, and general speaking ability (Appendix, available online). Each competency was graded on a nine-point scale where 1 to 3 was unsatisfactory, 4 to 6 was satisfactory, and 7 to 9 was superior. Evaluators could indicate whether a particular competency could not be assessed.

Participants

All students ($N=164$) enrolled in the medicine core clerkship from January to December of 2003 participated in the study. The medicine core clerkship is organized into four 9-week blocks. Each block includes a 6-week inpatient rotation and a three-week outpatient rotation. Within each block, half of the students were randomly assigned to begin with inpatient med-

The authors have no conflicts of interest, financial or otherwise, to report for this article or this research.

Address correspondence and requests for reprints to Dr. Kim: Division of General Internal Medicine, Clinical Academic Building, 2nd Floor, 125 Paterson Street, New Brunswick, NJ 08901 (e-mail: kimsa@umdnj.edu).

Received for publication October 28, 2004

and in revised form February 25, 2005

Accepted for publication March 1, 2005

icine. The other half began with outpatient medicine. Students worked with an average of three faculty and two residents during the inpatient rotation, and an average of three faculty during the outpatient rotation.

Within each block, we used an alternating design to assign students to the intervention or control group: for blocks 1 and 3, students starting the clerkship with inpatient medicine were assigned to the intervention group, whereas for blocks 2 and 4, students starting with outpatient medicine were assigned to the intervention group.

The control group ($N=83$) received usual feedback on their oral case presentations without encounter cards. The nature of feedback received by students in the control group was solely dependent on the faculty or residents with whom the students worked. Although there was no standard regarding the content of feedback, there was an expectation that feedback was given after a case presentation. The Department of Medicine runs annual feedback training sessions for all residents. Additionally, each quarter, faculty scheduled to attend on the inpatient service meet to discuss the goals and objectives of the clerkship. Providing feedback is a part of the agenda.

Students in the intervention group ($N=81$) each received a booklet of encounter cards on the first day of the clerkship and were instructed to have one OCP encounter card completed by an attending or resident during each of the 9 clerkship weeks (3 each from inpatient attendings, residents, and outpatient attendings). When students presented a case to a resident or faculty during rounds or office visits, they took out an encounter card and asked the evaluator to rate their presentation using the card. Evaluators completed the encounter card, provided the student with verbal feedback, and documented the time spent listening and providing feedback. Evaluators and students rated their satisfaction with the OCP card on a nine-point scale (1=low, 9=high). OCP encounter cards were then returned to the students for their review. At the end of the clerkship, students turned in the completed cards. The ratings on the OCP cards did not affect course grades.

Before implementing the OCP cards, faculty and residents were oriented to the rationale and use of the cards including a suggested approach for providing feedback. Inpatient attendings were oriented during quarterly mandatory meetings for faculty attending on inpatient services. Residents were oriented during morning report. Outpatient faculty were mailed orientation packets. Additionally, detailed instructions were (e)mailed to all evaluators. The Institutional Review Board approved this study.

Outcomes

To assess rating validity, outcome measures included: summative inpatient and outpatient clinical evaluations from faculty and residents using a 15-item form in which students are rated on a seven-point scale using behaviorally anchored items; the single item from the summative inpatient and outpatient evaluations that addresses oral presentation skills; National Board of Medical Examiners (NBME) medicine subject examination scores; case write-up grades; and final clerkship grades that are a weighted mean of inpatient and outpatient clinical evaluations, the subject examination score, and the case write-up grade.

The main outcome measure was an end of clerkship timed presentation of a student-selected, prepared case. Because

students' performance may vary depending on the type of presentation (inpatient or outpatient case), all students were instructed to prepare one inpatient and one outpatient case. When they arrived for the end of clerkship presentation, they were then randomly assigned to present either the inpatient or outpatient case. The end of clerkship presentation was evaluated using the OCP encounter cards. Evaluators were one of 6 non-attending faculty blinded to students' assignment to the intervention or control group. Presentations were audio-taped for independent, blinded review by 1 or 2 additional faculty.

Statistical Analysis

To determine feasibility, we examined the percentage of completed forms, average evaluation and feedback times, satisfaction ratings, and variability in students' OCP ratings. Differences in ratings by inpatient faculty, outpatient faculty, and residents were compared with the Kruskal-Wallis one-way analysis of variance. Internal consistency was assessed with Cronbach's α . Principal components analysis assessed dimensionality among the items. A generalizability study using a rater: student $\times$ competency design looked at rating precision. Generalizability analyses allow one to look at multiple sources of variation in a data set simultaneously, and estimate the relative magnitude of error that comes from each source. In this setting, we would expect large variability among students (i.e., there are true differences among students), moderate variability due to competencies (i.e., students do better on some competencies than others), and low rater-within-student variability (i.e., there would not be much variability among the multiple raters for a particular student). Following the original analyses in which we estimated a standard error of measurement in ratings for a student over all items and all raters, we examined how the standard error would change in scenarios involving fewer raters or items. Inter-rater reliability was estimated using κ coefficients. Relationships among OCP ratings and other student performance assessments were carried out with Spearman correlations. Disattenuated correlations (correlations corrected for error due to low reliability in one or both scores) were calculated assuming reliabilities of 0.70 among both the OCP ratings and the other assessment in each pair, with the exception of the medicine subject examination, which is a multi-item objective test with an estimated reliability of approximately 0.80 (A. Butler, National Board of Medical Examiners, personal communication, October 26, 2004). Intervention and control group means on the end of clerkship presentation were compared with the Kruskal-Wallis one-way analysis of variance. We present the means averaged over all raters, as well as those for the first, in-person rater. Parametric statistics were also used to compare group means since rating scales are often treated as interval data, to allow post hoc comparisons with Duncan's test, and to be consistent with other analyses (i.e., factor analyses, estimation of α), which rely on parametric assumptions. There were no differences in results, and only nonparametric results are presented. Because approximately 35 comparisons were made, we chose an α equal to 0.001 to note statistical significance.

RESULTS

Feasibility, reproducibility, and validity analyses are based on 79 of 81 students in the intervention group who completed at

least one OCP encounter card. 568 cards were completed (80% of target). The mean number of encounter cards per student was 7.2 (SD=2.4). As shown in Table 1, residents gave significantly higher mean ratings on all but two competencies than did either inpatient or outpatient faculty. All median competency ratings were 7 or 8, and ranges were typically 4 to 9 or 5 to 9 across resident and faculty evaluators. Overall, mean OCP presentation and feedback times were 16.8 (SD=12.4) and 7.1 (SD=5.1) minutes, respectively, and median times were 15 and 5, respectively. Students' mean presentation times differed significantly for outpatient and inpatient faculty and resident evaluators (11.9 vs 19.5 vs 16.5 minutes, respectively); and the median presentation time was 10 minutes for outpatient faculty compared with 15 minutes for inpatient residents and faculty. The mean evaluator and student satisfaction ratings were 7.3 (SD=1.7) and 6.8 (SD=2.0), respectively. Faculty and students liked the OCP cards less in the outpatient setting versus inpatient setting.

Students' individual mean OCP ratings averaged over all nine competencies and all encounters ranged from 5.6 to 8.8 (mean=7.7, SD=0.51). Mean ratings for the 9 individual competencies ranged from 7.6 (SD=1.0) for assessment/differential diagnosis and plan to 8.0 (SD=0.9) for general speaking ability (Table 1). The overall mean was 7.7 (SD=0.8). Medians for all scores were 8, and the ranges for all competencies were 4 to 9 except for physical exam and general speaking ability, which ranged from 5 to 9.

Principal components analysis revealed one primary factor that accounted for 75% of the variance. All 9 components were strongly associated with the first factor (factor loadings ranged from 0.83 to 0.90). Cronbach's α for the nine competencies was 0.96. Reproducibility of the OCP ratings estimated through generalizability analyses showed that the standard error of measurement around the ratings for a student, given all items and all raters, was 0.3. Reducing the number of raters to 6, 4, and 2 would change the standard error to 0.3, 0.4, and 0.6, respectively. Reducing the number of competencies from 9 to 3 would have little impact on the standard error.

Within the intervention group, mean OCP card ratings were significantly correlated with other summative assessments, including inpatient clinical evaluations, inpatient oral

presentation ratings, and the final clerkship grade (Table 2). Estimates of disattenuated correlations suggest the OCP skills are moderately to strongly related to other clinical assessments.

All 164 students performed an end of clerkship oral case presentation; 153 audio-taped presentations were reviewed by additional evaluators who were blinded to student assignments. Forty-nine percent were reviewed by one additional evaluator and 44% by two additional evaluators. The median kappa between pairs of raters over the 9 competencies was 0.27 and 0.35 for the average of the 9 competencies.

For the single in-person raters, mean ratings on end of clerkship OCP were 7.0 (SD=1.0) for the intervention group and 7.2 (SD=1.0) for the control group ($P=.09$). No significant differences existed in each of the nine competencies between the intervention and control group, reflecting the relatively large standard deviations (as well as the earlier generalizability results suggesting that more raters produce more stable scores). Multiple differences were observed when ratings were averaged over both the in-person rater and those who listened to the taped presentation. On 6 of the 9 competencies, students in the control group performed better than the students in the intervention group, but none reached statistical significance using an adjusted P value of .001 (Table 3). There were no significant differences between the intervention and control groups in presentation time ($P=.41$), inpatient summative clinical evaluations ($P=.72$), outpatient summative clinical evaluations ($P=.32$), NBME examination scores ($P=.75$), or final clerkship grades ($P=.91$).

DISCUSSION

We developed and implemented an encounter card system to evaluate medical students' OCP skills. The use of OCP encounter cards was feasible in a medicine clerkship for formative assessment. Students were assessed during multiple encounters in real-time clinical settings. The encounter cards were favorably perceived by students and evaluators, although the satisfaction ratings were lower in the outpatient than the inpatient setting. This may be due to time constraints inherent in busy outpatient practices that can limit the quantity and quality of feedback.¹³ As seen in other studies of clinical assessments,

Table 1. Oral Case Presentation Encounter Card Results

Competency	Overall Mean (SD)	Evaluator			P Value [†]
		Outpatient Faculty Mean (SD) (n=165 cards)	Inpatient Faculty Mean (SD) (n=213 cards)	Residents Mean (SD) (n=169 cards)	
1. History of present illness	7.8 (0.9)	7.6 (0.9)	7.7 (0.9)	8.0 (1.0)*	<.0001
2. Past medical history	7.7 (0.9)	7.5 (0.9)	7.6 (0.9)	8.1 (0.9)*	<.0001
3. Social/family history	7.8 (1.0)	7.6 (0.9)	7.6 (1.0)	8.1 (0.9)*	<.0001
4. Physical exam	7.7 (0.9)	7.6 (0.9)	7.6 (0.9)	7.9 (1.0)*	.0007
5. Studies	7.7 (1.0)	7.5 (1.0)	7.6 (1.0)	7.9 (0.9)	.0015
6. Assessment/differential	7.6 (1.0)	7.4 (1.0)	7.6 (1.0)	7.9 (1.0)*	<.0001
7. Plan	7.6 (1.0)	7.4 (1.0)	7.5 (1.0)	7.9 (1.0)*	<.0001
8. Organization/coherence	7.8 (1.0)	7.7 (0.9)	7.7 (1.0)	8.1 (1.0)*	<.0001
9. General speaking ability	8.0 (0.9)	7.9 (0.9)	7.9 (0.9)	8.2 (1.0)	.0014
Overall mean	7.7 (0.8)	7.6 (0.8)	7.6 (0.8)	8.0 (0.8)*	<.0001
Presentation time (minutes)	16.8 (12.4)	11.9 (6.5)*	19.5 (14.8)*	16.5 (9.7)*	<.0001
Feedback time (minutes)	7.1 (5.1)	6.6 (4.4)	8.0 (5.3)*	6.4 (5.1)	.0007
Evaluator satisfaction	7.3 (1.7)	6.5 (2.1)*	7.8 (1.2)*	7.4 (1.7)*	<.0001
Student satisfaction	6.8 (2.0)	6.2 (1.8)*	7.0 (2.2)	7.1 (1.8)	.0004

*Means different from one another based on post hoc comparisons with Duncan's test.

[†]P-values based on Kruskal-Wallis comparisons.

Table 2. Spearman Correlations of Mean OCP Ratings with Summative Assessments

Summative Assessments	Observed Correlation	Disattenuated Correlation*
Inpatient clinical evaluation [†]	0.58	0.83
Inpatient oral presentation rating [†]	0.43	0.61
Outpatient clinical evaluation	0.33	0.47
Outpatient oral presentation rating	0.21	0.30
Case write-up grade	0.19	0.27
Medicine subject exam score	0.24	0.32
Final clerkship grade [†]	0.40	0.57

*Observed correlations corrected for error due to unreliability in either score. We assumed that the reliability for all scores based on ratings and summative evaluations was 0.70, and the reliability of the medicine subject examination score was approximately 0.80.

[†]P < .001.

OCP, oral case presentation.

evaluators rarely used the full range of the rating scale, which resulted in limited variability in mean OCP ratings. In addition, as demonstrated in previous studies of in-training evaluation, residents consistently gave higher ratings than faculty.^{14,15}

With an average of 7.2 encounters per student, the OCP encounter card ratings were reproducible. In fact, the number of competencies and the number of encounters per student could easily be reduced to 3 and 4, respectively, with little change in the standard error. Our finding of low inter-rater reliability on the single end of clerkship presentation may be due to skewed ratings, and underscores the importance of multiple observations to obtain reliable ratings. Our data also show that evaluators failed to differentiate among the different competencies of oral case presentation skills. The competencies assessed on the OCP cards were highly correlated, and our

factor analysis indicates that evaluators were making global assessments rather than assessing each individual competency. This “halo” effect has been described in other studies that measure clinical competence and appears to be a characteristic of rating scales.¹⁶

Construct validity of OCP ratings is suggested given the correlations of encounter card ratings with other summative assessment measures such as the final clerkship grade. The high correlations with inpatient clinical evaluations are expected, given that many of the same faculty and residents were involved in OCP and clinical ratings. However, the imperfect disattenuated correlations and the low magnitude of the correlations with other skills and knowledge suggest that OCP encounter cards may be measuring distinct clinical skills.

A surprising finding was that the use of OCP encounter cards did not lead to superior performance on the end of clerkship oral case presentation. Several factors may account for this unexpected result. First, we measured the effectiveness of OCP encounter cards based on a single case presentation at the end of the clerkship, an inadequate measure of true OCP skills. Second, without an established “gold standard” for a successful oral case presentation, students may receive conflicting feedback about their presentation skills regardless of whether OCP encounter cards are used or not. Indeed, faculty assessment of student performance has been shown to be highly variable.^{17–22} Even our blinded raters, who held key positions in medical student education, did not agree on the quality of the summative case presentations. Third, although our faculty and residents were oriented to the use of OCP encounter cards, they were not specifically trained to assess oral case presentation skills. Faculty development aimed at improving evaluator rating skills and feedback skills specific to oral case presentations may have led to different results. A recent study suggests that a multifaceted approach to faculty development in direct observation of clinical skills improves faculty comfort and skill in direct observation and leads to more stringent rating behavior.²³

Our study has several limitations. First, this is a single institution study that limits generalizability. Second, while we tried to minimize potential contamination of the control group, it is likely that exposure to OCP cards by residents and faculty affected subsequent evaluation of students in the control group. Finally, we do not actually know the quality of feedback and/or messages received by students in the intervention group or control group, thus limiting our understanding of the results.

In summary, OCP encounter cards are a novel and feasible tool to evaluate oral case presentation skills of medical students in a medicine clerkship. The OCP ratings are reproducible, and validity is suggested by their correlation with multiple markers of performance. However, the use of OCP encounter cards did not significantly improve performance on a summative assessment of OCP skill. Most likely, the problem is with the unreliable nature of the single-rater summative OCP exam, rather than with the multiple OCP encounter cards. Future research focused on establishing the standards for successful oral case presentations, providing faculty development aimed at improving evaluators' direct observation skills of oral case presentations, as well as measuring the content and quality of feedback received by students may help clarify the role of OCP encounter cards in formative or summative evaluation.

Table 3. End of Clerkship Oral Case Presentation Performance*

Competency	Intervention Mean (SD) (n=83)	Control Mean (SD) (n=81)
Based on single in-person rater		
1. History of present illness	7.1 (1.1)	7.4 (1.2)
2. Past medical history	7.2 (1.0)	7.5 (1.1)
3. Social history/family history	7.2 (1.2)	7.2 (1.5)
4. Physical exam	6.9 (1.2)	7.1 (1.4)
5. Studies	6.8 (1.6)	7.2 (1.2)
6. Assessment/differential diagnosis	6.8 (1.5)	7.0 (1.5)
7. Plan	6.6 (1.6)	7.0 (1.3)
8. Organization/coherence	7.1 (1.3)	7.2 (1.6)
9. General speaking ability	7.2 (1.3)	7.3 (1.3)
Overall mean	7.0 (1.0)	7.2 (1.0)
Based on multiple (2 or 3) raters		
1. History of present illness	6.9 (0.9)	7.3 (1.1)
2. Past medical history	7.0 (0.8)	7.4 (1.0)
3. Social history/family history	6.8 (1.2)	7.2 (1.2)
4. Physical exam	6.8 (0.9)	7.2 (1.1)
5. Studies	6.8 (1.2)	7.1 (1.2)
6. Assessment/differential diagnosis	6.7 (1.3)	6.9 (1.4)
7. Plan	6.5 (1.5)	7.0 (1.1)
8. Organization/coherence	6.7 (1.2)	7.1 (1.3)
9. General speaking ability	6.9 (1.0)	7.3 (1.0)
Overall mean	6.8 (0.9)	7.2 (0.9)
Presentation time	7.1 (2.2)	7.5 (2.1)

*Mean scores did not differ significantly for any of the comparisons using a P value < .001 based on Kruskal-Wallis one-way analysis of variance.

We thank Laure Veet, MD, Christine Stoltz, MD, and Robin Garr-ell, MD, for their help as evaluators, and Ann Marie Hunt, for her administrative assistance.

REFERENCES

1. **AAMC Medical School Objectives Project.** American Association of Medical Colleges Web Site. Available at: <http://www.aamc.org/meded/msop>. Accessed February 10, 2004.
2. **ACGME Outcome Project.** Accreditation Council for Graduate Medical Education Web Site. Available at: <http://www.acgme.org>. Accessed February 10, 2004.
3. **Kasselbaum DG, Eaglen RH.** Shortcomings in the evaluation of students' clinical skills and behaviors in medical school. *Acad Med.* 1999;74:842-9.
4. **Klos M, Reuler JB, Nardone DA, Girard DE.** An evaluation of trainee performance in the case presentation. *J Med Educ.* 1983;58:432-4.
5. **Kihm JT, Brown JT, Divine GW, Linzer M.** Quantitative analysis of the outpatient oral case presentation. *J Gen Intern Med.* 1991;6:233-6.
6. **Wiese J, Varosy P, Tierney L.** Improving oral presentation skills with a clinical reasoning curriculum: a prospective controlled study. *Am J Med.* 2002;112:212-8.
7. **Haber RJ, Lingard LA.** Learning oral presentation skills, a rhetorical analysis with pedagogical and professional implications. *J Gen Intern Med.* 2001;16:308-14.
8. **Lye P, Bragg D, Simpson D.** Improving feedback with a clinical encounter form. *Acad Med.* 1997;72:444-5.
9. **Paukert JL, Richards ML, Olney C.** An encounter card system for increasing feedback to students. *Am J Surg.* 2002;183:300-4.
10. **Schum TR, Krippendorf RL, and the Advanced Education Faculty Development Group.** Feedback notes: a system for feedback to residents and students. *Acad Med.* 2000;75:556-7.
11. **Yurchak PM.** A guide to medical case presentations. *Resident Staff Physician.* 1981;109-15.
12. **Norcini JJ, Blank LL, Duffy FD, Fortna GS.** The Mini-CEX: a Method for Assessing Clinical Skills. *Ann Intern Med.* 2003;138:476-81.
13. **Irby DM.** Teaching and learning in ambulatory care settings: a thematic review of the literature. *Acad Med.* 1995;70:898-931.
14. **Kogan JR, Bellini LM, Shea JA.** Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. *Acad Med.* 2003;78(suppl):S33-5.
15. **Hatala R, Norman GR.** In-training evaluation during an internal medicine clerkship. *Acad Med.* 1999;74(suppl):S118-20.
16. **Haber RJ, Avins AL.** Do ratings on the American Board of Internal Medicine Resident Evaluation Form detect differences in clinical competence? *J Gen Intern Med.* 1994;9:140-5.
17. **McLeod PJ.** Faculty assessments of case reports of medical students. *J Med Educ.* 1987;62:673-7.
18. **Noel GL, Herbers JE Jr., Caplow MP, Cooper GS, Pangaro LN, Harvey J.** How well do internal medicine faculty members evaluate the clinical skills of residents? *Ann Intern Med.* 1992;117:757-65.
19. **Herbers JE Jr., Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ.** How accurate are faculty evaluations of clinical competence? *J Gen Intern Med.* 1989;4:202-8.
20. **Kalet A, Earp JA, Kowlowitz V.** How well do faculty evaluate the interviewing skills of medical students? *J Gen Intern Med.* 1992;7:499-505.
21. **Kroboth FJ, Hanusa BH, Parker S, et al.** The inter-rater reliability and internal consistency of a clinical evaluation exercise. *J Gen Intern Med.* 1992;7:174-9.
22. **Elliot DL, Hickam DH.** Evaluation of physical examination skills. Reliability of faculty observers and patient instructors. *JAMA.* 1987;258:3405-8.
23. **Holmboe ES, Hawkins RE, Huot SJ.** Effects of training in direct observation of medical residents' clinical competence. *Ann Intern Med.* 2004;140:874-81.

Supplementary Material

The following supplementary material is available for this article online:

Appendix. Oral Case Presentation Encounter Card.