Published in final edited form as:

J Hosp Med. 2013 April ; 8(4): 191–200. doi:10.1002/jhm.2023.

Development of a handoff evaluation tool for shift-to-shift physician handoffs: The handoff CEX

Leora I. Horwitz, MD, MHS^{1,2}, David Rand, DO, MPH³, Paul Staisiunas⁴, Peter H. Van Ness, PhD, MPH⁵, Katy L.B. Araujo, MPH⁵, Jeanne M. Farnan, MD, MHPE⁶, and Vineet M. Arora, MD, MAPP⁶

¹Section of General Internal Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT

²Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT

³Yale School of Public Health, New Haven CT. Dr. Rand is now with the Department of Medicine, University of Vermont College of Medicine, Burlington, VT

⁴Department of Medicine, University of Chicago Medicine, Chicago, IL. Mr. Staisiunas is now with the Law School, Marquette University, Milwaukee, WI

⁵Section of Geriatrics, Department of Internal Medicine, Yale School of Medicine, New Haven, CT

⁶Department of Medicine, University of Chicago Medicine, Chicago, IL

Abstract

Background—Increasing frequency of shift-to-shift handoffs coupled with regulatory requirements to evaluate handoff quality make a handoff evaluation tool necessary.

Objective—To develop a handoff evaluation tool

Design—Tool development

Setting—Two academic medical centers

Subjects—Nurse practitioners, medicine house staff and hospitalist attendings

Intervention—Concurrent peer and external evaluations of shift-to-shift handoffs.

Measurements—The Handoff CEX consists of 6 subdomains and one overall assessment, each scored from 1–9, where 1–3 is unsatisfactory and 7–9 is superior. We assessed range of scores, performance among subgroups, internal consistency, and agreement among types of raters.

Results—We conducted 675 evaluations of 97 unique individuals during 149 handoff sessions. Scores ranged from unsatisfactory to superior in each domain. The highest rated domain for handoff providers was professionalism (median 8, interquartile range [IQR] 7–9); the lowest was content (median 7, IQR 6–8). Scores at the two institutions were similar, and scores did not differ significantly by training level. Spearman correlation coefficients among the CEX sub-domains for

Access to data

Leora Horwitz had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior presentations

An earlier version of this work was presented as a poster presentation at the Society of General Internal medicine Annual Meeting in Orlando, FL on May 9, 2012

Conflicts of interest

The authors declare they have no conflicts of interest.

provider scores ranged from 0.71–0.86, except for setting (0.39–0.40). Third-party external evaluators consistently gave lower marks for the same handoff than peer evaluators did. Weighted kappa scores for provider evaluations comparing external evaluators to peers ranged from 0.28 (95% CI, 0.01–0.56) for setting to 0.59 (0.38–0.80) for organization.

Conclusions—This handoff evaluation tool was easily used by trainees and attendings, had high internal consistency and performed similarly across institutions. Because peers consistently provided higher scores than external evaluators, this tool may be most appropriate for external evaluation.

Keywords

handoff; evaluation; house staff; hospitalist; transfer of care

Introduction

Transfers among trainee physicians within the hospital typically occur at least twice a day and have been increasing among trainees as work hours have declined.¹ The 2011 Accreditation Council for Graduate Medical Education (ACGME) guidelines,² which restrict intern working hours to 16 hours from a previous maximum of 30, have likely increased the frequency of physician trainee handoffs even further. Similarly, transfers among hospitalist attendings occur at least twice a day, given typical shifts of 8–12 hours.

Given the frequency of transfers, and the potential for harm generated by failed transitions,^{3–6} the end of shift written and verbal handoff have assumed increasingly greater importance in hospital care among both trainees and hospitalist attendings.

The ACGME now requires that programs assess the competency of trainees in handoff communication.² Yet, there are few tools for assessing the quality of sign-out communication. Those that exist primarily focus on the written sign-out, and are rarely validated.^{7–12} Furthermore, it is uncertain whether such assessments must be done by supervisors, or whether peers can participate in evaluation. In this prospective, multi-institutional study, we assess the performance characteristics of a verbal sign-out evaluation tool for internal medicine house staff and hospitalist attendings, and examine whether it can be used by peers as well as by external evaluators. This tool has previously been found to effectively discriminate between experienced and inexperienced nurses conducting nursing handoffs.¹³

Methods

Tool design and measures

The Handoff CEX is a structured assessment based on the format of the mini-CEX, an instrument used to assess the quality of history and physical examination by trainees for which validation studies have previously been conducted.^{14–17} We developed the tool based on themes we identified from our own expertise,^{1,5,6,8,18–29} the ACGME core competencies for trainees² and the literature in order to maximize content validity. First, standardization has numerous demonstrable benefits for safety in general, and handoffs in particular.^{30–32} Consequently we created a domain for organization in which standardization was a characteristic of high performance. Second, there is evidence that people engaged in conversation routinely overestimate peer comprehension,²⁷ and that explicit strategies to combat this overestimation such as confirming understanding, explicitly assigning tasks rather than using open-ended language, and using concrete language are effective.³³ Accordingly we created a domain for communication skills, which is also an ACGME competency. Third, while there were no formal guidelines for sign-out content when we

developed this tool, our own research had demonstrated that the content elements most often missing and felt to be important by stakeholders were related to clinical condition and explicating thinking processes,^{5,6} so we created a domain for content that highlighted these areas and met the ACGME competency of medical knowledge. In accordance with standards for evaluation of learners, we incorporated a domain for judgment to identify where trainees were in the RIME spectrum of reporter, interpreter, master and educator. Next we added a section for professionalism in accordance with the ACGME core competencies of professionalism and patient care.³⁴ To avoid the disinclination of peers to label each other unprofessional, we labeled the professionalism domain as “patient-focused” on the tool. Finally, we included a domain for setting because of an extensive literature demonstrating increased handoff failures in noisy or interruptive settings.^{35–37} We then revised the tool slightly based on our experiences among nurses and students.^{13,38} The final tool included the six domains described above and an assessment of overall competency. Each domain was scored on a 9-point scale and included descriptive anchors at high and low ends of performance. We further divided the scale into three main sections: unsatisfactory (score 1–3), satisfactory (4–6) and superior (7–9). We designed two tools, one to assess the person providing the handoff and one to assess the handoff recipient, each with its own descriptive anchors. The recipient tool did not include a content domain. (Appendix A)

Setting and subjects

We tested the tool in two different urban, academic medical centers: the University of Chicago Medicine (UCM) and Yale-New Haven Hospital (Yale). At UCM, we tested the tool among hospitalists, nurse practitioners and physician assistants during the Monday and Tuesday morning and Friday evening sign-out sessions. At Yale, we tested the tool among house staff during the evening sign-out session from the primary team to the on-call covering team.

The UCM is a 550-bed urban, academic medical center in which the non-teaching hospitalist service cares for patients with liver disease, or end stage renal or lung disease awaiting transplant, and a small fraction of general medicine and oncology patients when the housestaff service exceeds its cap. No formal training on sign-out is provided to attending or mid-level providers. The non-teaching hospitalist service operates as a separate service from the house staff service and consists of 38 hospitalist clinicians (hospitalist attendings, nurse practitioners and physicians assistants). There are two handoffs each day. In the morning the departing night hospitalist hands off to the incoming daytime hospitalist or mid-level provider. These handoffs occur at 7:30am in a dedicated room. In the evening the daytime hospitalist or mid-level provider hands off to an incoming night hospitalist. This handoff occurs at 5:30pm or 7:30pm in a dedicated location. The written sign-out is maintained on a Microsoft Word document on a password-protected server and updated daily.

Yale is a 946-bed urban, academic medical center with a large internal medicine training program. Formal sign-out education which covers the main domains of the tool is provided to new interns during the first three months of the year,¹⁹ and a templated, EMR-based electronic written handoff report is produced by the house staff for all patients.³⁹ Approximately half of inpatient medicine patients are cared for by house staff teams, which are entirely separate from the hospitalist service. House staff sign out occurs between 4pm and 7pm every night. At a minimum, the departing intern signs out to the incoming intern; this handoff is typically supervised by at least one second- or third-year resident. All patients are signed out verbally; in addition the written handoff report is provided to the incoming team. Most handoffs occur in a quiet charting room.

Data collection

Data collection at UCM occurred between March-December, 2010 on three days of each week: Mondays, Tuesdays and Fridays. On Mondays and Tuesdays the am handoffs were observed; on Fridays the pm handoffs were observed. Data collection at Yale occurred between March-May, 2011. Only evening handoffs from the primary team to the overnight coverage were observed. At both sites, participants provided verbal informed consent prior to data collection. At the time of an eligible sign-out session, a research assistant (D.R. at Yale, P.S. at UC) provided the evaluation tools to all members of the incoming and outgoing teams, and observed the sign-out session himself. Each person providing a handoff was asked to evaluate the recipient of the handoff; each person receiving a handoff was asked to evaluate the provider of the handoff. In addition, the trained third-party observer (D.R., P.S.) evaluated both the provider and recipient of the handoff. The external evaluators were trained in principles of effective communication and the use of the tool, with specific review of anchors at each end of each domain. One evaluator had an MD and was completing an MPH. The second evaluator was an experienced clinical research assistant whose training consisted of supervised observation of 10 handoffs by a physician-investigator. At Yale, if a resident was present, s/he was also asked to evaluate both the provider and recipient of the handoff. Consequently every sign-out session included at least two evaluations of each participant, one by a peer evaluator and one by a consistent external evaluator who did not know the patients. At Yale many sign-outs also included a third evaluation by a resident supervisor.

The study was approved by the institutional review boards at both University of Chicago and Yale School of Medicine.

Statistical analysis

We obtained mean, median and interquartile range of scores for each subdomain of the tool as well as the overall assessment of handoff quality. We assessed convergent construct validity by assessing performance of the tool in different contexts. To do so, we determined whether scores differed by type of participant (provider or recipient), by site, by training level of evaluatee, or by type of evaluator (external, resident supervisor or peer) by using Wilcoxon rank sum tests and Kruskal-Wallis tests. For the assessment of differences in ratings by training level, we used evaluations of sign-out providers only, because the two sites differed in scores for recipients. We also assessed construct validity by using Spearman's rank correlation coefficients to describe the internal consistency of the tool in terms of the correlation between domains of the tool, and we conducted an exploratory factor analysis to gain insight into whether the subdomains of the tool were measuring the same construct. In conducting this analysis, we restricted the dataset to evaluations of sign-out providers only, and used a principal components estimation method, a promax rotation, and squared multiple correlation communality priors. Finally, we conducted some preliminary studies of reliability by testing whether different types of evaluators provided similar assessments. We calculated a weighted kappa using Fleiss-Cohen weights for external versus peer scores and again for supervising resident versus peer scores (Yale only). We were not able to assess test-retest reliability by nature of the sign-out process. Statistical significance was defined by a p value ≤ 0.05 and analyses were performed using SAS 9.2 (SAS Institute, Cary, NC).

Results

A total of 149 handoff sessions were observed: 89 at UC and 60 at Yale. Each site conducted a similar total number of evaluations: 336 at UC, 337 at Yale. These sessions

involved 97 unique individuals, 34 at UC and 63 at Yale. Overall scores were high at both sites, but a wide range of scores was applied (Table 1).

Handoff providers

A total of 343 evaluations of handoff providers were completed regarding 67 unique individuals. For each domain, scores spanned the full range from unsatisfactory to superior. The highest rated domain on the handoff provider evaluation tool was professionalism (median 8, interquartile range [IQR] 7–9). The lowest rated domain was content (median 7, IQR 6–8). (Table 1)

Handoff recipients

A total of 330 evaluations of handoff recipients were completed regarding 58 unique individuals. For each domain, scores spanned the full range from unsatisfactory to superior. The highest rated domain on the handoff provider evaluation tool was professionalism, with a median of 8 (IQR 7–9). The lowest rated domain was setting, with a median score of 7 (IQR 6–9). (Table 1)

Validity testing

Comparing provider scores to recipient scores, recipients received significantly higher scores for overall assessment (Table 1). Scores at UC and Yale were similar in all domains for providers but were slightly lower at UC in several domains for recipients (Appendix B). Scores did not differ significantly by training level (Table 2). Third-party external evaluators consistently gave lower marks for the same handoff than peer evaluators did (Table 3).

Spearman rank correlation coefficients among the CEX sub-domains for provider scores ranged from 0.71–0.86, except for setting (Table 4). Setting was less well correlated with the other subdomains, with correlation coefficients ranging from 0.39–0.41. Correlations between individual domains and the overall rating ranged from 0.80–0.86, except setting, which had a correlation of 0.55. Every correlation was significant at $p < 0.001$. Correlation coefficients for recipient scores were very similar to those for provider scores (Appendix C).

We analyzed 343 provider evaluations in the factor analysis; there were six missing values. The scree plot of eigenvalues did not support more than one factor; however, the rotated factor pattern for standardized regression coefficients for the first factor and the final communality estimates showed the setting component yielding smaller values than did other scale components (Appendix D).

Reliability testing

Weighted kappa scores for provider evaluations ranged from 0.28 (95% CI, 0.01–0.56) for setting to 0.59 (0.38–0.80) for organization, and were generally higher for resident versus peer comparisons than for external versus peer comparisons. Weighted kappa scores for recipient evaluation were slightly lower for external versus peer evaluations but agreement was no better than chance for resident versus peer evaluations (Table 5).

Discussion

In this study we found that an evaluation tool for direct observation of house staff and hospitalists generates a range of scores and was well-validated in the sense of performing similarly across two different institutions and among both trainees and attendings, and having high internal consistency. However, external evaluators gave consistently lower marks than peer evaluators at both sites, resulting in low reliability when comparing these two groups of raters.

It has traditionally been difficult to conduct direct evaluations of handoffs, since they may occur at haphazard times, in variable locations and without very much advance notice. For this reason several attempts have been made to incorporate peers in evaluations of handoff practices.^{5,40,41} Using peers to conduct evaluations also has the advantage that peers are more likely to be familiar with the patients being handed off and might recognize handoff flaws that external evaluators would miss. Nonetheless, peer evaluations have some important liabilities. Peers may be unwilling or unable to provide honest critiques of their colleagues given that they must work closely together for years. Trainee peers may also lack sufficient clinical expertise or experience to accurately assess competence. In our study we found that peers gave consistently higher marks to their colleagues than did external evaluators, suggesting they may have found it difficult to criticize their colleagues. We conclude that peer evaluation alone is likely an insufficient means of evaluating handoff quality.

Supervising residents gave very similar marks as intern peers, suggesting that they also are unwilling to criticize, are insufficiently experienced to evaluate, or, alternatively, that the peer evaluations were reasonable. We suspect the latter is unlikely given that external evaluator scores were consistently lower than peers. One would expect the external evaluators to be biased towards higher scores given that they are not familiar with the patients and are not able to comment on inaccuracies or omissions in the sign-out.

The tool appeared to perform less well in most cases for recipients than for providers, with a narrower range of scores and low weighted kappa scores. Although recipients play a key role in ensuring a high quality sign-out by paying close attention, ensuring it is a bidirectional conversation, asking appropriate questions and reading back key information, it may be that evaluators were unable to place these activities within the same domains that were used for the provider evaluation. An altogether different recipient evaluation approach may be necessary.⁴²

In general, scores were clustered at the top of the score range, as is typical for evaluations. One strategy to spread out scores further would be to refine the tool by adding anchors for satisfactory performance, not just the extremes. A second approach might be to reduce the grading scale to only 3 points (unsatisfactory, satisfactory, superior) to force more scores to the middle. However, this approach might limit the discrimination ability of the tool.

We have previously studied the use of this tool among nurses. In that study, we also found consistently higher scores by peers than by external evaluators. We did, however, find a positive effect of experience, in which more experienced nurses received higher scores on average. We did not observe a similar training effect in this study. There are several possible explanations for the lack of a training effect. It is possible that the types of handoffs assessed played a role: at UC some assessed handoffs were night staff to day staff, which might be lower quality than day staff to night staff handoffs, whereas at Yale all handoffs were day to night teams. Thus average scores at UC (primarily hospitalists) might have been lowered by the type of handoff provided. Given that hospitalist evaluations were conducted exclusively at UC and house staff evaluations exclusively at Yale, lack of difference between hospitalists and house staff may also have been related to differences in evaluation practice or handoff practice at the two sites, not necessarily related to training level. Third, in our experience attending physicians provide briefer, less comprehensive sign-outs than trainees, particularly when communicating with equally-experienced attendings; these sign-outs may appropriately be scored lower on the tool. Fourth, the great majority of the hospitalists at UCM were within 5 years of residency and therefore not very much more experienced than the trainees. Finally, it is possible that skills do not improve over time given widespread lack of observation and feedback during training years for this important skill.

The high internal consistency of most of the subdomains and the loading of all subdomains except setting onto one factor are evidence of convergent construct validity, but also suggest that evaluators have difficulty distinguishing among components of sign-out quality. Internal consistency may also reflect a halo effect, in which scores on different domains are all influenced by a common overall judgment.⁴³ We are currently testing a shorter version of the tool including domains only for content, professionalism and setting in addition to overall score. The fact that setting did not correlate as well with the other domains suggests that sign-out practitioners may not have or exercise control over their surroundings. Consequently, it may ultimately be reasonable to drop this domain from the tool, or alternatively to refocus on the need to ensure a quiet setting during sign-out skills training.

There are several limitations to this study. External evaluations were conducted by personnel who were not familiar with the patients, and they may therefore have overestimated the quality of sign-out. Studying different types of physicians at different sites might have limited our ability to identify differences by training level. As is commonly seen in evaluation studies, scores were skewed to the high end, although we did observe some use of the full range of the tool. Finally, we were limited in our ability to test inter-rater reliability because of the multiple sources of variability in the data (numerous different raters with different backgrounds at different settings, rating different individuals).

In summary, we developed a hand-off evaluation tool that was easily completed by house staff and attendings without training, that performed similarly in a variety of different settings at two institutions, and that can in principle be used either for peer evaluations or for external evaluations, although peer evaluations may be positively biased. Further work will be done to refine and simplify the tool.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Development and evaluation of the sign-out CEX was supported by a grant from the Agency for Healthcare Research and Quality (1R03HS018278-01). Dr. Arora is supported by a National Institutes on Aging (K23 AG033763). Dr. Horwitz is supported by the National Institute on Aging (K08 AG038336) and by the American Federation for Aging Research through the Paul B. Beeson Career Development Award Program. Dr. Horwitz is also a Pepper Scholar with support from the Claude D. Pepper Older Americans Independence Center at Yale University School of Medicine (P30AG021342 NIH/NIA). No funding source had any role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality, the National Institute on Aging, the National Institutes of Health, or the American Federation for Aging Research.

References

1. Horwitz LI, Krumholz HM, Green ML, Huot SJ. Transfers of patient care between house staff on internal medicine wards: a national survey. *Arch Intern Med.* 2006; 166(11):1173–1177. [PubMed: 16772243]
2. Accreditation Council for Graduate Medical Education. [Accessed Aug 23, 2011.] Common Program Requirements. 2011. http://www.acgme-2010standards.org/pdf/Common_Program_Requirements_07012011.pdf
3. Petersen LA, Brennan TA, O'Neil AC, Cook EF, Lee TH. Does housestaff discontinuity of care increase the risk for preventable adverse events? *Ann Intern Med.* Dec 1; 1994 121(11):866–872. [PubMed: 7978700]
4. Sutcliffe KM, Lewton E, Rosenthal MM. Communication failures: an insidious contributor to medical mishaps. *Acad Med.* Feb; 2004 79(2):186–194. [PubMed: 14744724]

5. Arora V, Johnson J, Lovinger D, Humphrey HJ, Meltzer DO. Communication failures in patient sign-out and suggestions for improvement: a critical incident analysis. *Qual Saf Health Care*. Dec; 2005 14(6):401–407. [PubMed: 16326783]
6. Horwitz LI, Moin T, Krumholz HM, Wang L, Bradley EH. Consequences of inadequate sign-out for patient care. *Arch Intern Med*. Sep 8; 2008 168(16):1755–1760. [PubMed: 18779462]
7. Borowitz SM, Waggoner-Fountain LA, Bass EJ, Sledd RM. Adequacy of information transferred at resident sign-out (in-hospital handover of care): a prospective survey. *Qual Saf Health Care*. Feb; 2008 17(1):6–10. [PubMed: 18245212]
8. Horwitz LI, Moin T, Krumholz HM, Wang L, Bradley EH. What are covering doctors told about their patients? Analysis of sign-out among internal medicine house staff. *Qual Saf Health Care*. Aug; 2009 18(4):248–255. [PubMed: 19651926]
9. Gakhar B, Spencer AL. Using direct observation, formal evaluation, and an interactive curriculum to improve the sign-out practices of internal medicine interns. *Acad Med*. Jul; 2010 85(7):1182–1188. [PubMed: 20375830]
10. Raduma-Tomas MA, Flin R, Yule S, Williams D. Doctors' handovers in hospitals: a literature review. *Qual Saf Health Care*. Jan 5.2011
11. Bump GM, Jovin F, Destefano L, et al. Resident sign-out and patient hand-offs: opportunities for improvement. *Teach Learn Med*. Apr; 2011 23(2):105–111. [PubMed: 21516595]
12. Helms AS, Perez TE, Baltz J, et al. Use of an appreciative inquiry approach to improve resident sign-out in an era of multiple shift changes. *J Gen Intern Med*. Mar; 2012 27(3):287–291. [PubMed: 21997480]
13. Horwitz LI, Dombroski J, Murphy TE, Farnan JM, Johnson JK, Arora VM. Validation of a handoff assessment tool: the Handoff CEX. *J Clin Nurs*. Jun 7.2012
14. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med*. Nov 15; 1995 123(10):795–799. [PubMed: 7574198]
15. Norcini JJ, Blank LL, Arnold GK, Kimball HR. Examiner differences in the mini-CEX. *Adv Health Sci Educ Theory Pract*. 1997; 2(1):27–33. [PubMed: 16180056]
16. Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med*. Sep; 2002 77(9):900–904. [PubMed: 12228088]
17. Holmboe ES, Huot S, Chung J, Norcini J, Hawkins RE. Construct validity of the miniclinical evaluation exercise (miniCEX). *Acad Med*. Aug; 2003 78(8):826–830. [PubMed: 12915378]
18. Horwitz LI, Meredith T, Schuur JD, Shah NR, Kulkarni RG, Jenq GY. Dropping the Baton: A Qualitative Analysis of Failures During the Transition From Emergency Department to Inpatient Care. *Ann Emerg Med*. Jun 13.2008
19. Horwitz LI, Moin T, Green ML. Development and implementation of an oral sign-out skills curriculum. *J Gen Intern Med*. Oct; 2007 22(10):1470–1474. [PubMed: 17674110]
20. Horwitz LI, Moin T, Wang L, Bradley EH. Mixed methods evaluation of oral sign-out practices. *J Gen Intern Med*. Apr.2007 22(S1):S114.
21. Horwitz LI, Parwani V, Shah NR, et al. Evaluation of an asynchronous physician voicemail sign-out for emergency department admissions. *Ann Emerg Med*. Sep; 2009 54(3):368–378. [PubMed: 19282064]
22. Horwitz LI, Schuster KM, Thung SF, et al. An institution-wide handoff task force to standardise and improve physician handoffs. *BMJ Qual Saf*. May 23.2012
23. Arora V, Johnson J. A model for building a standardized hand-off protocol. *Jt Comm J Qual Patient Saf*. Nov; 2006 32(11):646–655. [PubMed: 17120925]
24. Arora V, Kao J, Lovinger D, Seiden SC, Meltzer D. Medication discrepancies in resident sign-outs and their potential to harm. *J Gen Intern Med*. Dec; 2007 22(12):1751–1755. [PubMed: 17963009]
25. Arora VM, Johnson JK, Meltzer DO, Humphrey HJ. A theoretical framework and competency-based approach to improving handoffs. *QualSaf Health Care*. Feb; 2008 17(1):11–14.
26. Arora VM, Manjarrez E, Dressler DD, Basaviah P, Halasyamani L, Kripalani S. Hospitalist handoffs: A systematic review and task force recommendations. *Journal of Hospital Medicine*. 2009; 4(7):433–440. [PubMed: 19753573]

27. Chang VY, Arora VM, Lev-Ari S, D'Arcy M, Keysar B. Interns overestimate the effectiveness of their hand-off communication. *Pediatrics*. 2010; 125(3):491–496. [PubMed: 20142285]
28. Johnson JK, Arora VM. Improving clinical handovers: Creating local solutions for a global problem. *Quality and Safety in Health Care*. 2009; 18(4):244–245. [PubMed: 19651924]
29. Vidyarthi AR, Arora V, Schnipper JL, Wall SD, Wachter RM. Managing discontinuity in academic medical centers: strategies for a safe and effective resident sign-out. *J Hosp Med*. Jul; 2006 1(4): 257–266. [PubMed: 17219508]
30. Salerno SM, Arnett MV, Domanski JP. Standardized sign-out reduces intern perception of medical errors on the general internal medicine ward. *Teach Learn Med*. Apr-Jun; 2009 21(2):121–126. [PubMed: 19330690]
31. Haig KM, Sutton S, Whittington J. SBAR: a shared mental model for improving communication between clinicians. *Jt Comm J Qual Patient Saf*. Mar; 2006 32(3):167–175. [PubMed: 16617948]
32. Patterson ES. Structuring flexibility: the potential good, bad and ugly in standardisation of handovers. *Qual Saf Health Care*. Feb; 2008 17(1):4–5. [PubMed: 18245211]
33. Patterson ES, Roth EM, Woods DD, Chow R, Gomes JO. Handoff strategies in settings with high consequences for failure: lessons for health care operations. *Int J Qual Health Care*. Apr; 2004 16(2):125–132. [PubMed: 15051706]
34. Ratanawongsa N, Bolen S, Howell EE, Kern DE, Sisson SD, Larriviere D. Residents' perceptions of professionalism in training and practice: barriers, promoters, and duty hour requirements. *J Gen Intern Med*. Jul; 2006 21(7):758–763. [PubMed: 16808778]
35. Coiera E, Tombs V. Communication behaviours in a hospital setting: an observational study. *BMJ*. Feb 28; 1998 316(7132):673–676. [PubMed: 9522794]
36. Coiera EW, Jayasuriya RA, Hardy J, Bannan A, Thorpe ME. Communication loads on clinical staff in the emergency department. *Med J Aust*. May 6; 2002 176(9):415–418. [PubMed: 12056992]
37. Ong MS, Coiera E. A systematic review of failures in handoff communication during intrahospital transfers. *Jt Comm J Qual Patient Saf*. Jun; 2011 37(6):274–284. [PubMed: 21706987]
38. Farnan JM, Paro JA, Rodriguez RM, et al. Hand-off education and evaluation: piloting the observed simulated hand-off experience (OSHE). *J Gen Intern Med*. Feb; 2010 25(2):129–134. [PubMed: 19924489]
39. Horwitz LI, Schuster KM, Thung SF, et al. An institution-wide handoff task force to standardise and improve physician handoffs. *BMJ Qual Saf*. May 23. 2012
40. Kitch BT, Cooper JB, Zapol WM, et al. Handoffs causing patient harm: a survey of medical and surgical house staff. *Jt Comm J Qual Patient Saf*. Oct; 2008 34(10):563–570. [PubMed: 18947116]
41. Li P, Stelfox HT, Ghali WA. A Prospective Observational Study of Physician Handoff for Intensive-Care-Unit-to-Ward Patient Transfers. *Am J Med*. Sep; 2011 124(9):860–867. [PubMed: 21854894]
42. Greenstein E, Arora V, Banerjee S, Staisiunas P, Farnan J. Characterizing Physician Listening Behavior During Hospitalist Handoffs using the HEAR Checklist. *BMJ Qual Saf*. In press.
43. Thorndike EL. A constant error in psychological ratings. *J Appl Psychol*. 1920; 4(1):25.

Table 1
Median, mean and range of handoff CEX scores in each domain, providers and recipients

Domain	Provider N=343			Recipient N=330			P-value
	Median (IQR)	Mean (SD)	Range	Median (IQR)	Mean (SD)	Range	
Setting	7 (6–9)	7.0 (1.7)	2–9	7 (6–9)	7.3 (1.6)	2–9	0.05
Organization	7 (6–8)	7.2 (1.5)	2–9	8 (6–9)	7.4 (1.4)	2–9	0.07
Communication	7 (6–9)	7.2 (1.6)	1–9	8 (7–9)	7.4 (1.5)	2–9	0.22
Content	7 (6–8)	7.0 (1.6)	2–9	--	--	--	--
Judgment	8 (6–8)	7.3 (1.4)	3–9	8 (7–9)	7.5 (1.4)	3–9	0.06
Professionalism	8 (7–9)	7.4 (1.5)	2–9	8 (7–9)	7.6 (1.4)	3–9	0.23
Overall	7 (6–8)	7.1 (1.5)	2–9	7 (6–8)	7.4 (1.4)	2–9	0.02

IQR: interquartile range; SD: standard deviation

Table 2

Handoff CEX scores by training level, providers only

Domain	Median (Range)				P-value
	NP/PA N=33	Sub-internal or intern N=170	Resident N=44	Hospitalist N=95	
Setting	7 (2–9)	7 (3–9)	7 (4–9)	7 (2–9)	0.89
Organization	8 (4–9)	7 (2–9)	7 (4–9)	8 (3–9)	0.11
Communication	8 (4–9)	7 (2–9)	7 (4–9)	8 (1–9)	0.72
Content	7 (3–9)	7 (2–9)	7 (4–9)	7 (2–9)	0.92
Judgment	8 (5–9)	7 (3–9)	8 (4–9)	8 (4–9)	0.09
Professionalism	8 (4–9)	7 (2–9)	8 (3–9)	8 (4–9)	0.82
Overall	7 (3–9)	7 (2–9)	8 (4–9)	7 (2–9)	0.28

NP/PA: Nurse Practitioner/Physician Assistant

Table 3

Handoff CEX scores by peer versus external evaluators

Domain	Provider			P-value	Recipient			P-value
	Peer N=152	Resident Supervisor N=43	Median (Range)		Peer N=145	Resident Supervisor N=43	Median (Range)	
Setting	8 (3-9)	7 (3-9)	7 (2-9)	0.02	8 (2-9)	7 (3-9)	7 (2-9)	<0.001
Organization	8 (3-9)	8 (3-9)	7 (2-9)	0.18	8 (3-9)	8 (6-9)	7 (2-9)	<0.001
Communication	8 (3-9)	8 (3-9)	7 (1-9)	<0.001	8 (3-9)	8 (4-9)	7 (2-9)	<0.001
Content	8 (3-9)	8 (2-9)	7 (2-9)	<0.001	N/A	N/A	N/A	N/A
Judgment	8 (4-9)	8 (3-9)	7 (3-9)	<0.001	8 (3-9)	8 (4-9)	7 (3-9)	<0.001
Professionalism	8 (3-9)	8 (5-9)	7 (2-9)	0.02	8 (3-9)	8 (6-9)	7 (3-9)	<0.001
Overall	8 (3-9)	8 (3-9)	7 (2-9)	0.001	8 (2-9)	8 (4-9)	7 (2-9)	<0.001

Table 4

Spearman correlation coefficients, provider evaluations (N=342)

Spearman Correlation Coefficients						
	Setting	Organization	Communication	Content	Judgment	Professionalism
Setting	1.000	0.40	0.40	0.39	0.39	0.41
Organization	0.40	1.00	0.80	0.71	0.77	0.73
Communication	0.40	0.80	1.00	0.79	0.82	0.77
Content	0.39	0.71	0.79	1.00	0.80	0.74
Judgment	0.39	0.77	0.82	0.80	1.00	0.78
Professionalism	0.41	0.73	0.77	0.74	0.78	1.00
Overall	0.55	0.80	0.84	0.83	0.86	0.82

All p values <0.0001

Table 5

Weighted kappa scores

Domain	Provider		Recipient	
	External v. Peer N=144	Resident v. Peer N=42	External v. Peer N=134	Resident v. Peer N=43
Setting	0.39 (0.24,0.54)	0.28 (0.01,0.56)	0.34 (0.20,0.48)	0.48 (0.27, 0.69)
Organization	0.43 (0.29,0.58)	0.59 (0.39,0.80)	0.39 (0.22, 0.55)	0.03 (−0.23, 0.29)
Communication	0.34 (0.19, 0.49)	0.52 (0.37,0.68)	0.36 (0.22, 0.51)	0.02 (−0.18, 0.23)
Content	0.38 (0.25,0.51)	0.53 (0.27,0.80)	NA (NA)	NA (NA)
Judgment	0.36 (0.22, 0.49)	0.54 (0.25,0.83)	0.28 (0.15, 0.42)	−0.12 (−0.34, 0.09)
Professionalism	0.47 (0.32,0.63)	0.47 (0.23,0.72)	0.35 (0.18, 0.51)	−0.01 (−0.29, 0.26)
Overall	0.50 (0.36,0.64)	0.45 (0.24,0.67)	0.31 (0.16, 0.48)	0.07 (−0.20, 0.34)