Published in final edited form as:

Acad Pediatr. 2016 August ; 16(6): 524–531. doi:10.1016/j.acap.2016.04.002.

Reliability of Verbal Handoff Assessment and Handoff Quality Before and After Implementation of a Resident Handoff Bundle

Angela M. Feraco, MD^{a,b}, Amy J. Starmer, MD, MPH^a, Theodore C. Sectish, MD^a, Nancy D. Spector, MD^c, Daniel C. West, MD^d, and Christopher P. Landrigan, MD, MPH^{a,e}

^aDepartment of Medicine, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA 02115 and Department of Pediatrics, Harvard Medical School, 25 Shattuck Street, Boston MA 02215

^bDepartment of Pediatric Hematology/Oncology, Dana-Farber/Boston Children's Cancer and Blood Disorders Center, 450 Brookline Ave, Dana 3, Boston, MA 02215

^cSt. Christopher's Hospital for Children, 160 East Erie Avenue, Philadelphia, PA 19134, and Department of Pediatrics, Drexel University College of Medicine, 2990 W. Queen Lane, Philadelphia, PA 19129

^dUCSF Benioff Children Hospital and Department of Pediatrics, University of California, San Francisco, 505 Parnassus Avenue, San Francisco, CA 94143

^eSleep and Patient Safety Program, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115

Abstract

Objective—1) To develop validity evidence for the use of the Verbal Handoff Assessment Tool (VHAT) and examine the reliability of VHAT scores, and 2) to determine whether implementation of a resident handoff bundle (RHB) was associated with improved verbal patient handoffs among pediatric resident physicians.

Corresponding Author: Angela M. Feraco, MD, 450 Brookline Avenue, Dana 3, Boston, MA, 02215, office phone 617-632-5991, fax 617-632-4410, Angela_Feraco@dfci.harvard.edu.

Potential Conflict of Interest Disclosures: Drs. Starmer, Sectish, Spector, West, and Landrigan are recipients of a grant from the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation; Grant no. 1R18AE000029. Drs. Starmer, Sectish, Spector, West, and Landrigan report personal fees from Massachusetts General Hospital (MGH), New York Presbyterian Hospital (NYPH), and MD Anderson through service agreements with Boston Children's Hospital outside the submitted work. Dr. Landrigan is partially supported by the Child Health Corporation of America for his work as a member of the PRIS Network Executive Council, and reports personal fees from Virgin Pulse for consultation related to a sleep health development program. Dr. Spector is the recipient of an unrestricted medical education grant from Pfizer. Sponsors had no involvement in study design, data collection, data analysis, data interpretation, manuscript drafting, revision, or the decision to submit the manuscript.

Other disclosures: Drs. Starmer, Sectish, Spector, West, and Landrigan are the recipients of a grant from the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation; Grant no. 1R18AE000029; Drs. Starmer, Sectish, Spector, West, and Landrigan report personal fees from Massachusetts General Hospital (MGH), New York Presbyterian Hospital (NYPH), and MD Anderson through service agreements with Boston Children's Hospital outside the submitted work; Dr. Landrigan reports personal fees from Virgin Pulse for consultation related to a sleep health development program; Dr. Spector is the recipient of an unrestricted medical education grant from Pfizer; Dr. Feraco is currently supported by NHLBI training grant 5T32HL007574-33.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Method—In a pre-post design, prospectively audio recorded verbal patient handoffs conducted at Boston Children’s Hospital before and after implementation of the RHB were rated using the VHAT, which was developed for this study (primary outcome). Using Generalizability theory, we evaluated the reliability of VHAT scores.

Results—Overall, VHAT scores increased following RHB implementation (mean 142 vs. 191, possible score 0–500, $p<0.0001$). When accounting for clustering by resident physician, hospital unit, unit census, and patient complexity, implementation of the RHB was associated with a 63-point increase in VHAT score. Using Generalizability theory, we determined that a resident’s mean VHAT score based on handoff of 15 patients assessed by a single observer was sufficiently reliable for relative ranking decisions (i.e. norm-based; G-coefficient 0.81), whereas a VHAT score based on handoff of 21 patients would be sufficiently reliable for high-stakes, standard-based decisions (Phi 0.80).

Conclusions—Verbal handoffs improved following implementation of a resident handoff bundle, though gains were variable across the two clinical units. The VHAT shows promise as an assessment tool for resident handoff skills. If used for competency or entrustment decisions, a resident’s mean VHAT score should be based on observation of verbal handoff of 21 patients.

Keywords

Graduate medical education; patient handoffs; workplace-based assessment; pediatric residents; handoff communication

Introduction

In the wake of the 2011 Accreditation Council for Graduate Medical Education (ACGME) Duty Hour Standards, patient handoffs are frequent,¹ yet handoff skills have not been routinely taught.² The need for safer patient handoffs has been well established.^{3–6} From 2010–2012, the Joint Commission identified communication failures as root causes of 59–82% of “sentinel events” – serious preventable adverse events.⁷ These findings prompted the Agency for Healthcare Research and Quality^{8,9} and the Joint Commission³ to designate the patient handoff as a priority target in nationwide efforts to improve patient safety. Despite national emphasis on handoffs, one study of internal medicine trainees revealed that only a quarter of verbal patient handoffs included key handoff elements: the patient’s clinical condition, hospital course, and tasks to complete.¹⁰ Limited familiarity with patients may increase the risk that residents omit or inaccurately convey important handoff data, circumstances that may be exacerbated by duty hour changes.^{10–12} Perhaps in recognition of current underperformance, patient handoffs were recently included in the Association of American Medical Colleges’ recommendations for Entrustable Professional Activities (EPAs) for entering residency.¹³

Best practices for handoffs have been established, based on demonstrated reductions in medical errors following implementation of a handoff program.¹⁴ Previous studies have found that team training reduces medical errors^{15–17} and that structured communication between clinicians^{18,19} and computerized handoff tools^{20–22} show promise in improving handoffs. Previously, implementation of a resident handoff bundle (RHB) was associated

with reductions in medical errors,²³ improvements in the verbal (oral) handoff environment (e.g., handoffs more often occurred in a quiet, private location),²³ and improvements in the completeness of written handoffs.²³ This single institution study informed the development of the subsequent multicenter I-PASS study, which also found that implementation of a structured handoff program was associated with reduced medical errors.¹⁴ However, the effect of the RHB on the content and process of verbal handoffs was not formally evaluated. We hypothesized that RHB implementation would be associated with higher quality verbal patient handoffs among resident physicians. Such an association would provide evidence for a causal relationship between RHB implementation and the observed reduction of medical errors by providing a mechanistic explanation.

To determine whether RHB implementation was associated with improved verbal communication during patient handoffs, we set out to: (1) develop a Verbal Handoff Assessment Tool (VHAT) to measure the quality of verbal handoffs; (2) develop evidence that would support the validity of using scores generated from the VHAT to draw conclusions about resident handoff skills; and (3) compare the quality of verbal handoffs performed before and after implementation of the RHB.

Methods

Following approval by the Boston Children's Hospital institutional review board, we conducted a prospective pre-post study of verbal patient handoffs on two general pediatric units, embedded within a study designed to evaluate the association between the RHB and medical errors.²³ All resident physicians in the Boston Combined Residency Program received training in patient handoffs and were asked to follow new verbal and written handoff processes.

Participants

Pediatric residents rotating on the two study units who provided voluntary written informed consent were eligible to participate. Post-graduate year 1 (PGY-1) physicians provided direct patient care and were supervised by PGY-3 physicians. We provided small incentives to participating residents (e.g. cookies and gift cards).²³

Intervention and Setting

Details of RHB development and of the two study units have been published previously.²³ The two study units, referred to as "Unit 1" and "Unit 2," were geographically distinct inpatient pediatric hospital wards that admitted general pediatric and specialty pediatric patients. Briefly, the RHB intervention consisted of: (1) team training based in part on the TeamSTEPPS™ method⁸; (2) implementation of verbal handoff process changes; (3) training in best practices for verbal and written handoffs, including use of a standardized mnemonic, "SIGNOUT?" for verbal handoffs¹⁸; and (4) introduction of a computerized handoff tool integrated into the electronic medical record (Cerner PowerChart) on Unit 1. Unit 2 continued to use a free-text printed handoff document throughout the study period.

The RHB was introduced at a Boston Combined Residency Program retreat on October 1, 2009. Major verbal handoff process changes implemented were (1) team handoffs, (2) use of

a private location, and (3) use of a quiet location. In a team handoff, the departing senior resident conducts the verbal handoff in the presence of the departing intern and the receiving intern and senior. This replaced the previous practice of separate intern and senior resident handoffs.

Audio recordings

To capture verbal handoff content and process, trained research assistants audio recorded a convenience sample of handoff sessions conducted before and after RHB implementation. Each audio recording captured one “handoff session,” in which one resident was the primary speaker, or “giver.” Pre-intervention sessions were recorded July–September 2009. RHB implementation on October 1, 2009 was followed by a 4-week “wash in” period, during which no sessions were recorded. Post-intervention handoff sessions were recorded October 28, 2009–January 2010. Ten patient handoffs given by a single resident were scored for each handoff session (see Decision study results, Figure 2). Raters assigned scores by listening to audio recordings of handoffs, rather than observing live handoff sessions. In the post-intervention period, speakers were PGY-3 trainees for all team handoffs, as specified by the RHB.²³ Because we hypothesized that PGY-3 trainees may have superior handoff skills due to greater experience, the change to a team handoff system could result in better handoff scores in the post-intervention period merely due to the elimination of PGY-1 givers. As such, for this analysis, audio recorded handoffs given by PGY-1 trainees in the pre-intervention period were purposefully under-sampled to reduce possible confounding by training year and to improve comparability between the pre- and post-intervention periods. However, to allow for comparisons between baseline PGY-1 and PGY-3 handoffs, PGY-1 giver handoffs were not entirely excluded from the analysis.

Assessment Measure: Verbal Handoff Assessment Tool

We evaluated audio recordings using a Verbal Handoff Assessment Tool (VHAT; Appendix A) developed for this study. To develop the VHAT, we conducted a literature review through which we identified a core set of observable “best practices” in handoff behaviors, including assignment of illness severity, use of a descriptive patient summary, assignment of tasks or “action items,” guidance for contingencies, and open communication that encourages and addresses questions about handoff content. We iteratively refined the VHAT based on expert consensus (D. West, N. Spector, A. Starmer, C. Landrigan, T. Sectish and the I-PASS Scientific Oversight Committee) and pilot tested it, making subsequent refinements to improve the tool’s usability and performance.

Outcome Measures: VHAT Scores

We designed the VHAT to be used with each verbal patient handoff, with a minimum score of 0 and a maximum scaled score of 500. The VHAT score assigned to each patient handoff was the sum of subscores from five key domains reflecting core handoff behavior best practices (see Supplementary Digital Appendix).^{5,24–28} Subscores were transformed to equally weighted 100 point scaled subscores to eliminate fractional scoring (see Appendix). To determine the optimal number of patient handoffs to score for each handoff session, we performed Decision studies (see Analysis section and Figure 2).

To compare overall handoff performance before and after RHB implementation, mean VHAT scores for each time period were calculated by summing individual patient handoff scores and dividing them by the total number of patient handoffs scored during the relevant time period. To better understand which core handoff behaviors, if any, changed after implementation of the RHB, we compared resident cohort performance of specific handoff behaviors: illness severity assessment (VHAT component 1, see Appendix), use of a complete patient summary statement (component 2a), inclusion of action items (component 3a), contingency planning (component 4a) and open communication (component 5). We defined the dichotomous variable “good contingency planning” as a component 4a score of 67 or 100 (versus 0 or 33). We defined the dichotomous variable “open communication” as a component 5 score of 67 or 100 (versus 0 or 33).

Analysis

We summarized relevant demographic characteristics, including patient census, patient complexity, resident gender, and training level of residents giving handoff during the pre- and post-intervention periods using proportions for discrete variables and means with standard deviations for continuous variables. We used All Patient Refined Diagnosis Related Groups (APR-DRG) to measure patient complexity. For continuous variables that demonstrated skewed distributions, we calculated medians and ranges.

We used Generalizability Theory to establish the reliability of scores derived from the VHAT.^{29,30} We hypothesized that the primary sources of score variance would be variation between raters and variation in resident performance. Because many clinical skills are context dependent, we hypothesized that variations in resident performance may be due in part to patient factors (captured as a resident x patient interaction term). To test this hypothesis, we performed a nested Generalizability study (G-study) with rater and patient (nested within giver) as facets (Figure 1), which allowed us to determine the contributions of rater and patient (and interactions) to residents’ mean VHAT score variance. Two of the authors were trained as raters (AMF and AJS) and independently rated 50 patient handoffs conducted by 10 residents. We then used the variance data from the G-study to perform a Decision study (D-study) in which we varied the number of raters and individual patient handoff VHAT assessments. The goal was to determine the conditions required to generate a sufficiently reliable sample of patient handoffs from an individual handoff session (Generalizability coefficient [G-coefficient] 0.7) and for both relative (G-coefficient 0.8) and criterion-based (Phi 0.8) decisions about resident handoff skills.³⁰

We used *t*-tests to compare means of continuous variables. For continuous variables that demonstrated skewed distributions, we used non-parametric statistical tests (Wilcoxon rank sum). We used Fisher’s exact test to compare discrete variables, such as the component handoff behaviors. To assess the association between implementation of the RHB and VHAT scores, we used a generalized estimating equations model that accounted for clustering of verbal handoffs by resident physician, assuming an exchangeable correlation structure with robust standard errors. Inpatient unit, unit census, patient complexity, and resident giver gender and training level were considered plausible confounding factors in the relationship between the RHB and VHAT scores, and were included in a multivariate model of VHAT

scores. Analyses were performed using Stata, version 10.0 (StataCorp, College Station, TX) and GENOVA.³¹

Results

From the initial G-study (Figure 1) and subsequent D-studies, we determined how to minimize VHAT score variance (Figure 2). A single rater scoring at least 9 individual patient handoffs per handoff session yielded a G-coefficient above our target of 0.7 (Figure 2). Based on these results, we employed a single rater to score the first 10 audible patient handoffs during the resident physician's handoff session (G-coefficient=0.74). For the purposes of skill assessment, the average score of 15 patient handoffs per resident by a single rater resulted in a G-coefficient >0.8 (Figure 1). A score based on the average of 21 patient handoffs per resident achieved a Phi >0.8 with a single rater, or 18 patient handoffs per resident if 2 raters score each handoff.

Overall, 240 patient handoffs conducted by 24 residents during 24 handoff sessions were scored (Tables 1 & 2). We analyzed 24 of 33 (73%) eligible audio recordings featuring unique givers. Twelve of 21 pre-intervention recordings featured pediatric interns; we scored 3 of 12 selected at random. All other recordings featured upper year trainees. APR-DRG weight was available for 224/240 patient handoffs (93%) and varied little, though on Unit 1 patient complexity increased in the post-intervention period (Table 1). Female resident physicians were givers of the majority of verbal patient handoffs, which reflects the overall gender composition of the pediatric residency program.

VHAT scores increased significantly in the post-intervention period (mean 142 vs. 191, possible score 0–500, $p<0.0001$; Table 2), though analysis by inpatient unit demonstrated significant gains only on Unit 2 (mean 151 vs. 247, $p<0.0001$; Unit 1: mean 129 vs. 151, $p=0.17$; Table 2). Contrary to our hypothesis that PGY-3 givers would have superior skills as compared to PGY-1 givers prior to RHB implementation, pre-intervention handoff scores did not differ between PGY-1 and PGY-3 givers (Units 1&2: PGY-1 mean 149 vs. PGY-3 mean 140, $p=0.52$; Unit 2: PGY-1 mean 149 vs. PGY-3 mean 153, $p=0.80$). Comparing VHAT scores among only PGY-3 givers pre- and post-intervention likewise did not alter results (Units 1& 2: mean 140 vs. 191, $p=0.0001$; Unit 1: see above (no PGY-1 givers); Unit 2: mean 153 vs. 247, $p<0.0001$). Open communication, defined as a question and answer during the handoff or a specific prompt for “Any questions?” (VHAT component 5, Supplemental Digital Appendix) increased significantly following RHB implementation (Table 2). Adoption of best handoff practices was inconsistent: post-intervention, 42% of residents incorporated illness severity into verbal patient handoffs (0/12 pre-RHB vs. 5/12 post-RHB $p=0.06$), but only 19% of post-RHB verbal patient handoffs included illness severity (0/120 pre-RHB vs. 23/120 post-RHB, $p<0.001$).

Implementation of the RHB was associated with a 63-point increase in VHAT score (possible score range 0–500, $p=0.005$, 95% CI (18–108); Table 3) when accounting for inpatient unit, patient complexity, and clustering by resident physician. Resident physician training year, resident gender, and unit census were not associated with VHAT scores on univariate analysis even when using a permissive significance level ($p=0.1$), but were

included in the multivariate model because they represented hypothesized confounders. Neither resident gender nor resident physician training level nor unit census was independently predictive of VHAT score in this multivariate model (Table 3). Inpatient unit was significantly associated with VHAT scores, even when accounting for patient complexity (β 64, $p=0.009$, CI (16, 112); Table 3). APR-DRG weight was significantly associated with VHAT scores, but the effect size was small (β 4, $p=0.001$, CI (2–7)).

Discussion

We developed a Verbal Handoff Assessment Tool (VHAT) to measure the quality of verbal patient handoffs and demonstrated that verbal handoffs improved significantly following implementation of a resident handoff bundle (RHB). Implementation of the RHB was associated with a 63-point increase in VHAT score when adjusting for clustering by resident, inpatient unit, and patient complexity as measured by APR-DRG weight. This reflects substantive improvements in verbal handoff quality.

A prior report found a clinically meaningful reduction in medical errors following RHB implementation,²³ a finding replicated in a multisite handoff study.¹⁴ We hypothesized that improvements in verbal patient handoff quality were an important part of this observed reduction in medical errors. Interestingly, although total errors decreased on both units, Starmer et al. found that rates of preventable adverse events declined more on Unit 2 (4.3 versus 0.8 per 100 admissions, $p < 0.008$) than on Unit 1 (2.5 versus 2.0 per 100 admissions, $p=0.64$), which is directionally consistent with our finding that verbal handoff quality improved on Unit 2, but not Unit 1.²³ Our study findings lend weight to the hypothesis that medical error reduction following implementation of the RHB was mediated, at least in part, by improved verbal handoffs, and that verbal handoff quality is a patient safety matter.

Inpatient unit assignment was strongly associated with VHAT score, even when accounting for patients' illness complexity and clustering of patient handoffs by resident physician. Following implementation of the RHB, higher absolute scores and greater score change was observed on Unit 2, which cares for "complex care" children as well as general pediatric patients. Although APR-DRG weights were comparable on the two units, residents on unit 2 may have perceived their patients as sicker, leading to better adoption of new verbal handoff behaviors on this unit post-RHB. Verbal contingency planning increased on both units after implementation of the RHB. Relative gains were more marked on Unit 1, which experienced more than a doubling of verbal contingency planning relative to pre-intervention levels. Unit 1's computerized handoff document contained a prompt and dialog box for contingency planning, whereas the handoff document utilized on Unit 2 remained in a free-text format, with no visual prompt. The robust relative gains in contingency planning on Unit 1 suggest that written document structure affects verbal handoff content, a finding consistent with another published report.¹⁰ Otherwise, there were no known differences in resident team structure or unit leadership culture that readily explain the unit-specific differences in VHAT scores.

While we observed significant improvements following our intervention, handoff scores remained fairly low even in the post-intervention period, with mean scores of <250 out of

500 possible points. Modest reinforcement of the handoff mnemonic and verbal handoff processes during the post-intervention period may have contributed to suboptimal adoption of new verbal handoff practices. Based on this experience, we enhanced our training in the subsequent multisite study, and integrated strategies to train and encourage faculty supervisors to reinforce handoff elements and processes.^{32–34} Subsequently, we observed much better adoption of verbal handoff practices.¹⁴

The move to competency-based medical education using the framework of milestones³⁵ and EPAs^{36,37} requires tools to assess trainees' patient handoff skills. Considered in the context of the most current validity framework,³⁸ our findings provide important initial evidence to support the validity of using VHAT-generated scores to help fill this assessment gap. The rigorous process that we used to create the VHAT provides strong evidence of content validity. The observation that VHAT scores increased, as hypothesized, after residents were trained in handoff skills provides additional important validity evidence (i.e. relationship to other variables). The G- and D-studies provide strong evidence of internal structure validity and demonstrate how to administer and interpret VHAT scores for high-stakes decision-making. Due to the large resident-patient interaction effect (Figure 1), observing more verbal patient handoffs improves VHAT score reliability. Because raters contributed little to score variance in this study, the burden of utilizing additional raters was not outweighed by the small efficiency gains in terms of number of handoffs needed to produce reliable scores (Figure 2). It should be noted that our raters (AMF and AJS) participated in tool development, so rater factors may contribute more substantially to score variance if a more diverse group of clinical faculty serve as handoff raters. Our findings indicate that the average score generated by one rater observing a resident handing off at least 15 patients is sufficiently reliable for high-stakes relative ranking of handoff skills (G-coefficient = 0.8). For absolute or criterion-based decisions, such as competency or entrustment decisions, a minimum of 18–21 patients per resident should be assessed, depending on the number of raters (Phi = 0.8). Practically, this translates into observing or recording each resident handing off patients on two or three occasions for either formative or summative assessment. This degree of observation is not unduly burdensome for faculty assessors or residents.

When considering whether sufficient validity evidence has been gathered, we must consider the uses and consequences of VHAT scores. The observation that group handoff scores increased after training and that this change was associated with a significant reduction in medical errors provides important but preliminary validity evidence. To provide stronger validity evidence, future studies should focus on prospective appraisal of VHAT scores for making high-stakes assessments of individual residents' verbal patient handoff skills and determining whether higher VHAT scores are associated with reduced rates of preventable medical errors. One important step that remains before the VHAT can be used to guide entrustment decisions is to determine what minimum VHAT score indicates acceptable handoff quality.

Our study has several limitations. Resident rotation schedules rendered it impossible to obtain pre- and post-intervention measurements for the same residents; we addressed this by conducting between groups rather than within subjects comparisons. Similarly, our study design cannot account for resident learning over time—handoff skills may have improved

simply due to increasing experience. However, the majority of residents assessed were PGY-3 level, whose verbal handoff habits were likely established prior to the study period. In addition, clinical details of the handoff (such as the nature of winter versus summer ailments) made rater blinding impossible. However, results of our generalizability study indicate rater factors had relatively little impact on score variance in this sample, and resident assessment for entrustment decisions will likewise be unblinded. Another potential limitation is the generalizability of pre-intervention scores from this resident cohort. Data presented here reflect resident verbal handoff practices before recent attention to handoff skills, which are now acknowledged as important skills for entering residency.¹³ We hope that handoff training that begins in medical school will lead to superior handoff skills among residents, such that the pre-intervention scores documented here will be lower than those seen in subsequent cohorts.

In conclusion, RHB implementation was associated with improvements in verbal patient handoffs as determined by VHAT scores, and this association was robust when adjusted for resident, inpatient unit, and patient factors. This suggests that a resident handoff bundle contributes to reductions in medical error in part by improving verbal patient handoff quality. The VHAT shows promise as a tool that could be used to help make competency and entrustment decisions about resident handoff skills.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the I-PASS Scientific Oversight Committee for its contributions to the development of the VHAT, Andrew Ho, PhD, of the Harvard Graduate School of Education for helpful feedback regarding G-study design, and the resident physicians of the Boston Combined Residency Program in Pediatrics for their participation.

Funding/Support: This study was supported by the Controlled Risk Insurance Company Risk Management Foundation Grant Program as well as a grant from the Boston Children's Hospital Program for Patient Safety and Quality Research Grant Program. Dr. Starmer was supported by grants T32 HP10018, a National Research Service Award in Pediatrics and 1K12HS019456-01 from the Oregon Comparative Effectiveness Research K12 Program through the Agency for Healthcare Research and Quality. Dr. Landrigan is partially supported by the Child Health Corporation of America for his work as a member of the PRIS Network Executive Council.

Abbreviations

ACGME	Accreditation Council for Graduate Medical Education
APR-DRG	All Patient Refined Diagnosis Related Groups
EPAs	Entrustable Professional Activities
PGY	Post-Graduate Year
RHB	Resident Handoff Bundle
VHAT	Verbal Handoff Assessment Tool

References

1. Desai SV, Feldman L, Brown L, et al. Effect of the 2011 vs 2003 duty hour regulation-compliant models on sleep duration, trainee education, and continuity of patient care among internal medicine house staff: A randomized trial. *JAMA Intern Med.* 2013; 173(8):649–655. [PubMed: 23529771]
2. 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.* 2008; 17(1):6–10. [PubMed: 18245212]
3. Joint Commission Center for Transforming Healthcare. Hand-off communications. <http://www.centerfortransforminghealthcare.org/projects/detail.aspx?Project=1>
4. Singh H, Thomas EJ, Petersen LA, Studdert DM. Medical errors involving trainees: A study of closed malpractice claims from 5 insurers. *Arch Intern Med.* 2007; 167(19):2030–2036. [PubMed: 17954795]
5. Arora V, Johnson J, Lovinger D, et al. Communication failures in patient sign-out and suggestions for improvement: A critical incident analysis. *Qual Saf Health Care.* 2005; 14(6):401–407. [PubMed: 16326783]
6. Petersen LA, Brennan TA, O'Neil AC, et al. Does housestaff discontinuity of care increase the risk for preventable adverse events? *Ann Intern Med.* 1994; 121(11):866–872. [PubMed: 7978700]
7. Joint Commission T. Sentinel event data: Root causes by event type, 2004–2012. 2013. http://www.jointcommission.org/assets/1/18/Root_Causes_Event_Type_04_4Q2012.pdf
8. Clancy CM, Tornberg DN. TeamSTEPS: Assuring optimal teamwork in clinical settings. *Am J Med Qual.* 2007; 22(3):214–217. [PubMed: 17485563]
9. McDonald KM, Sundaram V, Bravata DM, et al. Closing the quality gap: A critical analysis of quality improvement strategies. Agency for Healthcare Research and Quality Technical Reviews. 2007:7.
10. Horwitz LI, Moin T, Krumholz HM, et al. What are covering doctors told about their patients? analysis of sign-out among internal medicine house staff. *Qual Saf Health Care.* 2009; 18(4):248–255. [PubMed: 19651926]
11. Block L, Habicht R, Wu AW, et al. In the wake of the 2003 and 2011 duty hours regulations, how do internal medicine interns spend their time? *J Gen Intern Med.* 2013; 28(8):1042–1047. [PubMed: 23595927]
12. Sen S, Kranzler HR, Didwania AK, et al. Effects of the 2011 duty hour reforms on interns and their patients: A prospective longitudinal cohort study. *JAMA Intern Med.* 2013; 173(8):657–62. discussion 663. [PubMed: 23529201]
13. Englander, R, Aschenbrener, CA, Call, SA., et al., editors. MedEdPortal. Association of American Medical Colleges; 2014. Core entrustable professional activities for entering residency. Resource ID 887 ed<https://www.mededportal.org/icollaborative/resource/887> [accessed 15.08.27]
14. Starmer AJ, Spector ND, Srivastava R, et al. Changes in medical errors after implementation of a handoff program. *N Engl J Med.* 2014; 371(19):1803–1812. [PubMed: 25372088]
15. Morey JC, Simon R, Jay GD, et al. Error reduction and performance improvement in the emergency department through formal teamwork training: Evaluation results of the MedTeams project. *Health Serv Res.* 2002; 37(6):1553–1581. [PubMed: 12546286]
16. Wallin CJ, Meurling L, Hedman L, et al. Target-focused medical emergency team training using a human patient simulator: Effects on behaviour and attitude. *Med Educ.* 2007; 41(2):173–180. [PubMed: 17269951]
17. Pronovost P, Berenholtz S, Dorman T, et al. Improving communication in the ICU using daily goals. *J Crit Care.* 2003; 18(2):71–75. [PubMed: 12800116]
18. Horwitz LI, Moin T, Green ML. Development and implementation of an oral sign-out skills curriculum. *J Gen Intern Med.* 2007; 22(10):1470–1474. [PubMed: 17674110]
19. Haig KM, Sutton S, Whittington J. SBAR: A shared mental model for improving communication between clinicians. *Jt Comm J Qual Patient Saf.* 2006; 32(3):167–175. [PubMed: 16617948]
20. Van Eaton EG, Horvath KD, Lober WB, et al. A randomized, controlled trial evaluating the impact of a computerized rounding and sign-out system on continuity of care and resident work hours. *J Am Coll Surg.* 2005; 200(4):538–545. [PubMed: 15804467]

21. Petersen LA, Orav EJ, Teich JM, et al. Using a computerized sign-out program to improve continuity of inpatient care and prevent adverse events. *Jt Comm J Qual Improv.* 1998; 24(2):77–87. [PubMed: 9547682]
22. Volpp KG, Grande D. Residents' suggestions for reducing errors in teaching hospitals. *N Engl J Med.* 2003; 348(9):851–855. [PubMed: 12606742]
23. Starmer AJ, Sectish TC, Simon DW, et al. Rates of medical errors and preventable adverse events among hospitalized children following implementation of a resident handoff bundle. *JAMA.* 2013; 310(21):2262–2270. [PubMed: 24302089]
24. Wohlauer MV, Arora VM, Horwitz LI, et al. The patient handoff: A comprehensive curricular blueprint for resident education to improve continuity of care. *Acad Med.* 2012; 87(4):411–418. [PubMed: 22361791]
25. Cohen MD, Hilligoss PB. The published literature on handoffs in hospitals: Deficiencies identified in an extensive review. *Qual Saf Health Care.* 2010; 19(6):493–497. [PubMed: 20378628]
26. Riesenber LA, Leitzsch J, Little BW. Systematic review of handoff mnemonics literature. *Am J Med Qual.* 2009; 24(3):196–204. [PubMed: 19269930]
27. Riesenber LA, Leitzsch J, Massucci JL, et al. Residents' and attending physicians' handoffs: A systematic review of the literature. *Acad Med.* 2009; 84(12):1775–1787. [PubMed: 19940588]
28. Patterson ES, Roth EM, Woods DD, et al. Handoff strategies in settings with high consequences for failure: Lessons for health care operations. *Int J Qual Health Care.* 2004; 16(2):125–132. [PubMed: 15051706]
29. Kreiter CD, Yin P, Solow C, Brennan RL. Investigating the reliability of the medical school admissions interview. *Adv Health Sci Educ Theory Pract.* 2004; 9(2):147–159. [PubMed: 15141132]
30. Shavelson, RJ., Webb, NM. Generalizability theory. In: Kemp-Leonard, K., editor. *Encyclopedia of social measurement.* Oxford, UK: Elsevier; 2004. p. 99-105.
31. Brennan, RL. Generalizability theory. New York, NY: Springer; New York: Springer; 2001. p. 538
32. Starmer AJ, O'Toole JK, Rosenbluth G, et al. Development, implementation, and dissemination of the I-PASS handoff curriculum: A multisite educational intervention to improve patient handoffs. *Acad Med.* 2014; 89(6):876–884. [PubMed: 24871238]
33. Sectish TC, Starmer AJ, Landrigan CP, Spector ND. I-PASS Study Group. Establishing a multisite education and research project requires leadership, expertise, collaboration, and an important aim. *Pediatrics.* 2010; 126(4):619–622. [PubMed: 20876168]
34. O'Toole JK, West DC, Starmer AJ, et al. Placing faculty development front and center in a multisite educational initiative: Lessons from the I-PASS handoff study. *Acad Pediatr.* 2014; 14(3): 221–224. [PubMed: 24767774]
35. Schumacher DJ, Spector ND, Calaman S, et al. Putting the pediatrics milestones into practice: A consensus roadmap and resource analysis. *Pediatrics.* 2014
36. Aylward M, Nixon J, Gladding S. An entrustable professional activity (EPA) for handoffs as a model for EPA assessment development. *Acad Med.* 2014
37. Englander R, Carraccio C. From theory to practice: Making entrustable professional activities come to life in the context of milestones. *Acad Med.* 2014
38. Downing SM. Validity: On meaningful interpretation of assessment data. *Med Educ.* 2003; 37(9): 830–837. [PubMed: 14506816]

What's New

This study presents evidence of improved verbal handoff quality following implementation of a handoff training program and provides important validity evidence to support using scores from the Verbal Handoff Assessment Tool to make competency and entrustment decisions.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.**Schema of Nested Generalizability Study**

The subsample of 50 handoffs utilized for the G-study (5 unique patients handed off by 10 resident handoff givers) were each rated by 2 raters, yielding 5 VHAT scores for each giver from each rater (one per patient), for a total of 100 handoff ratings. Each raw handoff score, $X_{(p|g)r}$ is composed of the sum of the population mean score (μ) and variance components (v) due to each of the measured facets, as well as residual measurement error.

Figure 2.**Sources of VHAT Score Variance and Decision Study Results**

Increasing the number of patient handoffs scored increases the reliability of mean VHAT scores. The graph shows the reliability coefficients for relative error variance (G-coefficient) and absolute error variance (Phi). If a threshold reliability of 0.8 is used for both relative ranking (G-coefficient) and absolute (criterion-referenced) decisions (Phi), under conditions of a single rater, then at least 15 patient handoffs should be scored for each resident to establish relative ranking, and at least 21 patients should be scored for high-stakes criterion-based competency decisions.

Table 1
Patient, Handoff and Unit Characteristics – pre- and post-Resident Handoff Bundle

	Unit 1 & 2 combined			Unit 1: RHB + computerized tool			Unit 2: RHB		
	Pre-RHB (s=12; n=120)	Post-RHB (s=12; n=120)	p-value	Pre-RHB (s=5; n=50)	Post-RHB (s=7; n=70)	p-value	Pre-RHB (s=7; n=70)	Post-RHB (s=5; n=50)	p-value
Average Daily Unit Census <i>mean±SD</i>	14.3±2.8	16.8±2.8	<0.0001	14.2±2.3	17.3±2.6	<0.0001	14.4±3.1	16.2±3.0	0.002
APR-DRG Weight for Handoff Patients <i>median (range)</i>	1.0 (0.3–19.0)	1.2 (0.4–13.4)	0.05	0.8 (0.3–13.4)	1.3 (0.4–13.4)	0.03	1.1 (0.3–19.0)	1.2 (0.5–11.0)	0.51
Patient Handoffs conducted by PGY-1	30/120 (25%)	0/120 (0%)	<0.0001	0/50 (0%)	0/70 (0%)	N/A	30/70 (43%)	0/50 (0%)	<0.001
Patient Handoffs conducted by female resident physician	90/120 (75%)	80/120 (67%)	0.20	40/50 (80%)	40/70 (57%)	0.01	50/70 (71%)	40/50 (80%)	0.39

Abbreviations: RHB=Resident Handoff Bundle; s=session; n=patient

Table 2

Verbal Patient Handoffs Pre- and Post-Resident Handoff Bundle

	Units 1 & 2			Unit 1: RHB + computerized tool			Unit 2: RHB		
	Pre-RHB (n=120)	Post-RHB (n=120)	p-value	Pre-RHB (n=50)	Post-RHB (n=70)	p-value	Pre-RHB (n=70)	Post-RHB (n=50)	p-value
Mean VHAT Score ^a ±SD	142±67	191±104	<0.0001	129±71	151±94	0.17	151±64	247±93	<0.0001
Illness Severity Assessed	0% (0/120)	19% (23/120)	<0.001	0% (0/50)	0.04% (3/70)	0.27	0% (0/70)	40% (20/50)	<0.001
Complete Patient Summary Statement	34% (41/120)	47% (56/120)	0.07	40% (20/50)	41% (29/70)	>0.99	30% (21/70)	54% (27/50)	0.01
Inclusion of Action Items	38% (46/120)	42% (50/120)	0.69	38% (19/50)	41% (29/70)	0.85	39% (27/70)	42% (21/50)	0.71
Good Contingency Planning	24% (29/120)	41% (49/120)	0.009	14% (7/50)	33% (23/70)	0.02	31% (22/70)	52% (26/50)	0.04
Open Communication	48% (57/120)	65% (78/120)	0.009	38% (19/50)	51% (36/70)	0.19	54% (38/70)	84% (42/50)	0.001

Abbreviations: RHB=Resident Handoff Bundle;

^aMean VHAT Score refers to the mean of *n* patient handoff scores

Table 3

Multivariate Model of Change in Handoff Score

Variable	Change in VHAT Score	<i>p</i> -value	95% CI for Change in VHAT Score
RHB	63	0.006	(18, 108)
Inpatient Team	64	0.009	(16, 111)
APR-DRG Weight	4	0.001	(2, 7)
PGY-3 Training Level ^a	29	0.11	(−6, 64)
Male Gender ^b	8	0.78	(−48, 64)
Unit Census	−2	0.55	(−10, 5)

Abbreviations: VHAT=Verbal Handoff Assessment Tool

^aPGY-1 is reference group^bFemale gender is reference group