OPTIMIZATING AND DIFFUSING A HANDOVER BEHAVIORAL ASSESSMENT TOOL FOR SIMULATION

by

RODNEY CHEN

DISSERTATION

Presented to the Faculty of the Medical School
The University of Texas Southwestern Medical Center
In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF MEDICINE WITH DISTINCTION IN
QUALITY IMPROVEMENT AND PATIENT SAFETY

The University of Texas Southwestern Medical Center
Dallas, TX

ACKNOWLEDGMENTS

Thank you to the Office of Quality, Safety and Outcomes Education for their continued support!

~Eleanor Phelps, Dr. Gary Reed, Veronica Garcia, Virginia Pineda-Dow, Annette Irving, Mary Baldwin, Mozhdeh Sadighi, and Patrice Griffith~

Thank you to the Department of Anesthesiology for your invaluable mentorship and support.

~Dr. Philip Greilich, Dr. Meghan Michael, Dr. Isaac Lynch~

Thank you to the QEP team for the many opportunities to explore new and exciting aspects of QI, patient safety, and course implementation. I look forward to our continued work in handover education and beyond.

~Dr. Philip Greilich, Eleanor Phelps, Richard Preble, Mozhdeh Sadighi~

ABSTRACT

DIFFUSING AND OPTIMIZING A HANDOVER BEHAVIORAL ASSESSMENT TOOL FOR SIMULATION

RODNEY CHEN

The University of Texas Southwestern Medical Center, 2021
Supervising Professor: Dr. Philip Greilich, MD, MSc, FASE

Introduction: With multiple simulated and clinical scenarios included in the ongoing Quality Enhancement Plan (QEP), a standardized approach to assessing and trending handover quality across class years could quantify the improvements established through the QEP. This study assesses the utility of the Liang Handover Assessment Tool for Simulation (L-HATS), a valid and reliable behavioral assessment tool tested during the transition to clerkship (T2C) handover module. Here, we use the L-HATS to assess handovers delivered during residency essentials (RE) and COVID-19 telehealth courses, checking for tool reliability in settings other than T2C. In cases where we find the tool to be less reliable, we optimize the L-HATS by improving the observer training course. The study aim is to confirm tool reliability of ICC>0.75, consistent with levels of reliability found during testing in the T2C module.

Methods: We select volunteer observers from a group of medical students who had completed the T2C course, with each observer assigned a set of videos to score for each activity. The primary outcome measure for this study is the two-way random effects ICC, which represents tool inter-rater reliability in each novel activity. An ICC>0.75 is considered good reliability, an ICC 0.5-0.75 is considered moderate reliability, and an ICC<0.5 is considered poor reliability. As the volunteer observer training improves across activities, we assess for observers' intra-rater reliability. Intra-rater reliability is assessed along the same scale used for inter-rater reliability.

Results: RE inter-rater reliability was 0.561 [0.167, 0.953], with each of six observers scoring four videos. COVID-19 telehealth inter-rater reliability was 0.644 [0.244, 0.964], with five observers each scoring four videos. The intra-rater reliability calculated for the telehealth course ranged from 0.105 [-0.361, 0.863] to 0.667 [0.020, 0.971].

Conclusion: This study demonstrates moderate levels of reliability in both the RE and telehealth courses. However, neither novel activity could match the reliability scores calculated during original L-HATS testing, suggesting that the tool is less reliable in settings outside of the T2C course. Future studies might increase the number of graded videos per handover activity, to narrow the confidence intervals found in the present study. Moreover, we find that a universally flexible assessment tool is difficult to design, suggesting that each new learning activity may require a uniquely tailored behavioral assessment tool.

TABLE OF CONTENTS

INTRODUCTION

Problem

In 1999, the Institute of Medicine identified medical error as a leading cause of patient mortality, alerting the medical community to an emerging healthcare crisis.[1] Despite this, the number of reported iatrogenic mortalities has recently risen, escalating medical error from the sixth leading cause of death in the country to the third.[2,3] The prevalence of medically preventable mortality has since prompted many healthcare providers to investigate and identify the most common sources for medical errors. Among these, communication errors are known to be a leading and preventable cause of medical errors.[4,5]

Recently published literature has shown handovers to contribute to up to 80% of communication-related medical errors, while established handover protocols generally improve patient outcomes.[6,7] Efforts to standardize handover protocols have resulted in several popular approaches.[8,9] However, these approaches to handover protocols largely remain institution and department-specific, with many methods remaining unvalidated and unreliable.[10] Clinical research into optimal handover protocol is ongoing with many national organizations, such as the multicentered handoff collaborative (MHC), leading efforts to standardize handovers across the country.[11]

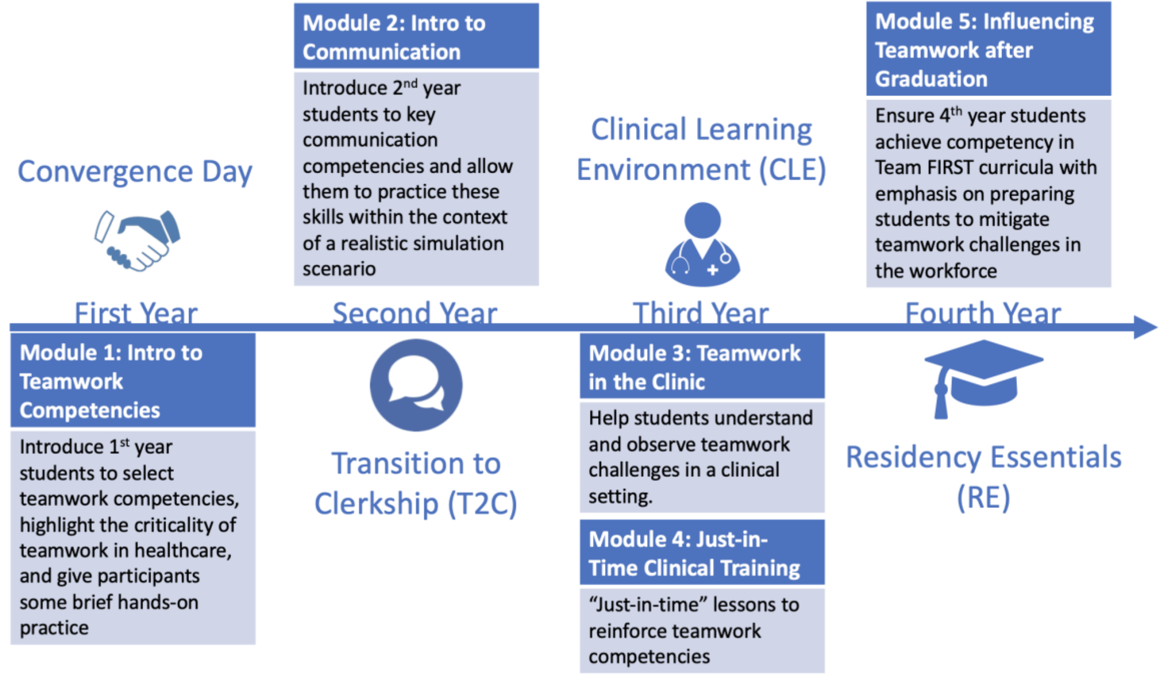
Meanwhile, accrediting bodies, such as the American College of Graduate Medical Education (ACGME), identified medical education as a point for intervention, establishing quality improvement and patient safety education as an essential aspect of graduate medical education (GME).[12,13] Efforts to educate both older medical students and resident level trainees soon followed, with mixed success.[14] With handovers believed to contribute to medical error heavily, healthcare programs have increasingly focused on assessments for handover quality within their institutions. The Handoff CEX, for example, utilizes a behavioral assessment to evaluate simulated and real handoff efficacy.[15,16] However, the literature on the Handoff CEX, like the majority of handover education, focuses heavily on residents. Undergraduate medical education (UME) remains largely untrained in handover protocols, with limited literature reports on fourth-year medical students. First- and second-year health professional education remains an unexplored but potentially important population to train in effective handover protocol.

Available Knowledge

This thesis presents work conducted at the University of Texas Southwestern Medical Center at Dallas (UTSW) as part of the ongoing Quality Enhancement Plan (QEP) project.[17] In 2019, UTSW identified handover education as a critical healthcare skill not routinely taught in its medical school curriculum.[17] Therefore, UTSW leadership tasked the QEP team to design, introduce, and optimize a novel handover curriculum capable of ensuring that healthcare graduates from UTSW are comfortable and capable of sending and receiving quality handovers.

This novel handover curriculum is designed to span the entire length of each student's education, with activities specifically tailored to the medical school's schedule. As such, students will participate in four unique handover modules designed to correspond to each of the four years of medical school, maximizing trainee retention through spaced repetition (Figure 1).[18] Nevertheless, an interprofessional approach to handover education is essential for the QEP project. Therefore, we will adjust other health profession curricula to best accommodate the four learning modules into participating health profession programs based on each program's unique schedule. In the earlier stages of the QEP project, UTSW leadership has asked that interprofessional activities be limited primarily to medical students and physician assistant (PA) students. Therefore, we restrict the population for the current study to medical and physician assistant students.

Figure 1: Timeline for planned handover education activities over four years of medical education. Module 1 occurs during Convergence Day. Module 2 occurs during Transition to Clerkship. Modules 3 and 4 occur in the clinical learning environment. Module 5 occurs during residency essentials.


Convergence Day
First Year
Second Year
Transition to Clerkship (T2C)
Clinical Learning Environment (CLE)
Third Year
Fourth Year
Residency Essentials (RE)

Module 1: Intro to Teamwork Competencies

Introduce 1st year students to select teamwork competencies, highlight the criticality of teamwork in healthcare, and give participants some brief hands-on practice

Module 2: Intro to Communication

Introduce 2nd year students to key communication competencies and allow them to practice these skills within the context of a realistic simulation scenario

Module 3: Teamwork in the Clinic

Help students understand and observe teamwork challenges in a clinical setting.

Module 4: Just-in-Time Clinical Training

"Just-in-time" lessons to reinforce teamwork competencies

Module 5: Influencing Teamwork after Graduation

Ensure 4th year students achieve competency in Team FIRST curricula with emphasis on preparing students to mitigate teamwork challenges in the workforce

In the context of the limited literature describing behavioral tools to assess undergraduate health professionals, the QEP team decided to utilize existing handover assessment tools to design and optimize a novel tool targeted to assess participating students at UTSW.[16,19] We plan to use this new behavioral assessment tool to measure handover quality in all activities, trending student ability to handover patients during each of the four modules.[20] Module 2 of the Transition to Clerkship (T2C) course was selected as an early test for assessment tool usability and reliability.

T2C is an ongoing component of the UTSW medical school curriculum targeted at second-year medical students (MS2s). The course is strategically positioned at the end of the pre-clerkship block of medical student training, focusing on teaching practical aspects of patient care, such as access to patient electronic medical records (EMR), patient triage, and handover simulation. The aim is to equip novice learners with the requisite skills necessary to contribute to clinical teams positively. Module 2 introduces students to a structured approach to handovers through SBAR (Situation, Background, Assessment, and Recommendation).[9] As it's the first opportunity for medical students to practice structured handovers, module 2 was an ideal testing environment for our new assessment tool, termed the Liang Handover Assessment Tool for Simulation (L-HATS). At the end of the handover module, the QEP team calculated inter-rater reliability for the L-HATS, finding acceptable levels of reliability in the setting of module 2.

Rationale/Aims

One year into the QEP project, program leaders have identified a growing need for a valid and reliable assessment tool to trend student participant improvement across the four handover activities (Figure 1). Prior research has already shown that the L-HATS is a valid and reliable tool for behavioral assessment in the context of the T2C course. However, we have yet to assess the utility of the L-HATS tool in contexts outside of module 2. We believe that the tool is sufficiently robust and agnostic to assess handovers in contexts and structures outside those provided during the module 2 simulations.

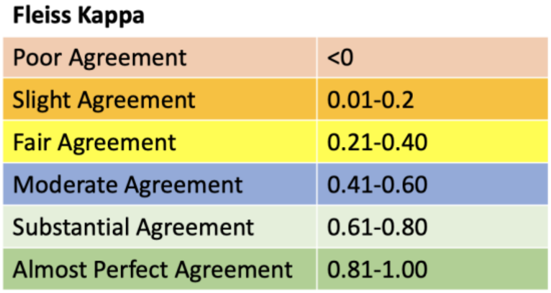
In the present study, we aim to diffuse the L-HATS to other learning environments, assessing the tool's reliability in two other handover activities: module 4, incorporated into the COVID-19 telehealth elective (COVID telehealth), and module 5 of Residency Essentials (RE). We predict that the L-HATS will demonstrate as good reliability in these new activities as that in module 2. This study will allow the researcher team to assess the tool's capacity for use in unique contexts and identify any limitations preventing the tool from being applied to unique scenarios.

Our team determined that the behavioral assessment had good reliability (two-way random effects intra-class correlation coefficient 0.866, 95% CI [0.765, 0.930]). However, a subsequent principal component analysis evaluating each component of the L-HATS demonstrated significant variability in agreement (Figure 2). The research team agreed that the surprising degree of variability is likely due to inadequate observer training during L-HATS testing, which entailed a one-hour Q/A session with the tool creators. We determined that variability in L-HATS components could worsen reliability when used during RE and COVID telehealth. Therefore, if the L-HATS is less reliable when tested in these new contexts, we predict that the drop is due directly to inadequate observer training. We will subsequently improve observer training to improve L-HATS reliability in the context of modules 4 and 5.

Figure 2: Component analysis for the L-HATS conducted using data collected during assessment tool testing. The dataset used to complete the component analysis concluded that the L-HATS had good reliability to intraclass correlation coefficient (ICC) of 0.866.

L-HATS element	Fleiss Kappa	ICC single agreement
Name	0.853	0.858
Age	0.771	0.777
Gender	-0.49	0.116
Problem	0.14	0.241
Course of Events	0.099	0.203
Past Medical/Surgical History	0.814	0.819
Allergies	0.839	0.843
Vitals	0.64	0.648
Access	0.691	0.702
Medications	0.521	0.535
Labs/Imaging	0.483	0.746
Case Specific Items	0.629	0.639
Diagnosis	0.111	0.208
Illness Severity	0.471	0.489
Action Plan	0.438	0.453
Anticipatory Guidance	0.263	0.285
Sender Introduces Self	0.64	0.65
Receiver Introduces Self	0.848	0.853
Sender Introduces Role	0.538	0.556
Sender Introduces Role	0.466	0.486
Handover Readback	0.166	0.176
Questions	0.791	0.797

Primary Concern Identified	-1	0.028
Concludes Handover	0.124	0.97
Organization	0.061	0.173
Specificity	-0.02	0.11



Fleiss Kappa

Poor Agreement	<0
Slight Agreement	0.01-0.2
Fair Agreement	0.21-0.40
Moderate Agreement	0.41-0.60
Substantial Agreement	0.61-0.80
Almost Perfect Agreement	0.81-1.00

METHODS

Context

Here, we assess the L-HATS reliability in two new contexts. The first context, RE, prepares graduating medical students for their future careers as medical residents by reviewing and remediating critical topics covered in medical school (Figure 1). Module 5 will review handover delivery and reception, with emphasis on handover delivery. The handover module spans a period of four days on the final week of February, with each of 224 fourth-year medical students and ten physician assistant students assigned one timeslot out of the four days. During the module, student participants first manage one of two simulated emergent cases with a pre-assigned student partner. Each student then hands their assigned simulated patient over to a faculty evaluator, who assesses handover quality based on a non-L-HATS rubric. Students who do not adequately hand the patient over to their faculty evaluator remediate the handover module at a later date. Participating students include all fourth-year medical students and ten volunteer physician assistant (PA) students. Faculty evaluators include volunteer faculty physicians affiliated with UTSW medical school and were assigned times based on availability (Table 1).

Table 1: Schedule for faculty assignments during RE 2020. Faculty were assigned according to individual availability.

2/24/20 7:45 am – 12:45	2/25/20 7:45 am – 12:45	2/27/20 7:45 am – 12:45	2/28/20 7:45 am – 12:45
D. Sendelbach	D. Sendelbach	D. Sendelbach	D. Sendelbach
B. Rege	M. Sulistio	J. Hernandez	J-A. Nesiama
J. Walsh	H. Katragadda	C. Washington	J. McConnell
A. Mihalic	J. Wagner	L. Agharokh	N. Oakman

The UTSW simulation center hosts the RE activity. Resources provided include the debrief rooms, waiting rooms, and simulation rooms with annexed observation rooms, allowing observers to assess student participants via a two-way mirror. The simulation center faculty and staff record all simulated activities, including both parts of the handover module, using two or more cameras installed in each simulation room. These recordings are automatically uploaded to a cloud-based server, and off-site simulation center staff control access to all recorded handover sessions.

While modules 2 and 5 are both hosted and recorded through the simulation center, these activities are unique. During module 5, students must work individually, and handovers must be performed according to the I-PASS structure (appendix a.1).[8] Unlike in module 2, student participants in module 5 must handover the simulated patient on their own for a grade on a pass/fail basis. Meanwhile, the L-HATS was designed to assess SBAR-based handovers. Testing L-HATS in context of module 5 will test tool viability in assessing non-SBAR-based handovers. Additionally, the course directors designed RE for fourth year medical students, who are more experienced relative to the second-year medical students participating in T2C.

Finally, the RE course occurs once annually, limiting any quality improvement for RE 2020. Instead, any improvements to either L-HATS use or observer training were collected from volunteer observers and applied to the second context.

In the second context, COVID-19 telehealth elective, second- and thirst-year medical students interview post-discharge COVID-positive patients, assessing early signs and symptoms of complications secondary to COVID-19 infection. Medical students present each patient to the attending physician in a SBAR structured handover format.[21] To assist students, the research team hosts weekly SBAR review sessions across the four-week elective, allowing for opportunities to practice sending and receiving handovers of mock patients. Faculty attending physicians from the UTSW internal medicine department volunteered to supervise students during the elective. As schedules could rapidly change, one volunteers are double-booked for each day to ensure that a physician would be able to receive student handovers. We present a detailed process map outlining the day-to-day approach to handing over patients during the COVID-19 elective in Appendix b.1. The QEP team believe that the COVID telehealth environment is an ideal test opportunity to assess for L-HATS reliability in a clinical setting. Therefore, we record all handovers performed during the telehealth elective, planning to assign these videos to trained volunteer observers for scoring using the L-HATS.

Students participating in the telehealth elective meet attendings every business day afternoon via Microsoft Teams, handing patients over purely in a virtual setting. These participating students cannot volunteer as observers for the duration of the elective. However, they can complete the volunteer observer training during the elective and participate as observers for subsequent telehealth elective rotations.

We require L-HATS evaluators for each of the above activities. Early in L-HATS testing, we found that student volunteers who had taken the T2C course would be best suited for evaluator training. The original L-HATS team selected observers based on this criterion, concluding good reliability among these observers. Therefore, we selected evaluators for the present study from the same population of student volunteers as those used in the original L-HATS tests. To improve consistency, we required evaluators volunteering to observe handovers during the COVID-19 telehealth elective to complete a 2-week-long orientation and pre-screen, during which we train each evaluator on the L-HATS in context of each activity, while also assessing for intra- and inter-rater reliability (Appendix a.2).[22,23] Members of the original L-HATS testing team teach each observer training course. Finally, we use an iterative PDSA approach to continuously improve observer training course quality. Assuming that the L-HATS is a robust assessment tool, we anticipate that improvements to the observer training will result in improved calculated reliability, matching those found during assessment tool testing.

Baseline Observer Training

The observer training used during testing for the L-HATS included a simple question and answer session prior to assigning handovers to the volunteer observers. Early elements such as the hour-long question-and-answer session as well as providing the L-HATS and instruction manual for individual review were kept in subsequent versions of the observer training course.

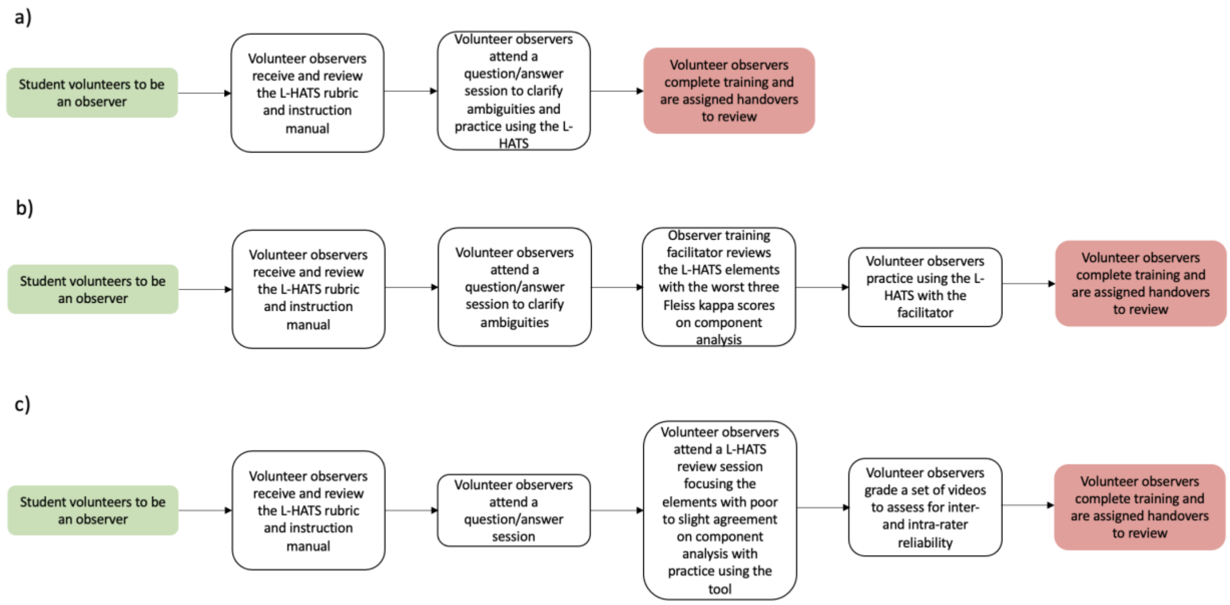
Module 5 Observer Training

Following initial testing, component analysis of the L-HATS demonstrated increased need to focus on select elements in the tool, deemed to have poor agreement (Figure 2). To improve observer reliability, we added an additional component to observer training, reviewing L-HATS items found to have the lowest agreement. As with prior observer training, module 5 training involves an hour-long question and answer session addended with the item-specific review. A total of six volunteers underwent training.

Module 4 COVID-19 Telehealth Observer Training

Lower-than-expected intraclass correlation coefficient (ICC) values from module 5 prompted additional improvement to the L-HATS training module. We decided to design and implement an optional observer training course for the telehealth elective. All volunteer observers complete elements included in the module 5 observer training course. However, we include additional opportunities to practice using the tool as well as a 1.5 week-long reliability assessment, to test individual observer intra-rater reliability. We calculate a two-way random effects ICC to determine the quality of intra-rater reliability.

Figure 3: Iterative improvement of the observer training module. (a) High level process map for the observer training used during L-HATS testing. (b) High level process map for observer training during RE 2020. (c) High level process map for the COVID telehealth course.


a)
Student volunteers to be an observer
Volunteer observers receive and review the L-HATS rubric and instruction manual
Volunteer observers attend a question/answer session to clarify ambiguities and practice using the L-HATS
Volunteer observers complete training and are assigned handovers to review
b)
Student volunteers to be an observer
Volunteer observers receive and review the L-HATS rubric and instruction manual
Volunteer observers attend a question/answer session to clarify ambiguities
Observer training facilitator reviews the L-HATS elements with the worst three Fleiss kappa scores on component analysis
Volunteer observers practice using the L-HATS with the facilitator
Volunteer observers complete training and are assigned handovers to review
c)
Student volunteers to be an observer
Volunteer observers receive and review the L-HATS rubric and instruction manual
Volunteer observers attend a question/answer session
Volunteer observers attend a L-HATS review session focusing the elements with poor to slight agreement on component analysis with practice using the tool
Volunteer observers grade a set of videos to assess for inter- and intra-rater reliability
Volunteer observers complete training and are assigned handovers to review

Intervention(s)

We study the use of the L-HATS in novel simulated and clinical environments. Because the original context for the L-HATS involved assessment of recorded simulation sessions, we record all handovers assessed in the current study for retrospective review by designated observers. During the RE course, the simulation center staff log and store videos through the SIMULATIONiQ© software managed by Education Management Solutions (EMS). In contrast, we host the COVID telehealth elective via Microsoft Teams, recording all sessions through the built-in recording software. A pre-designated research team member then stores video recordings on a shared Microsoft OneDrive. Volunteer observers completing training and signing the corresponding release forms gain access to handover videos either through the simulation center staff or through the designated team member. Once the observer finishes scoring videos, they send a digital copy of their scores to the research team, who perform statistical analysis to test tool reliability.

We continue to improve our observer training module throughout this study, requiring that each volunteer adequately complete the most up-to-date observer training course prior to scoring videos for the corresponding activity. Earlier iterations for L-HATS observer training involved a simple question and answer session to clarify elements of the grading rubric. However, the component analysis demonstrated high degree of inconsistency among L-HATS components, prompting a more standard approach to observer training. Contrasting the initial design, the two week-long training course includes the original question-and-answer session with additional training and review of L-HATS elements in the first three days. We dedicate a week and a half of the observer training to iterative observer testing to confirm tool comprehension and to assess for intra-rater reliability.

Study of the Intervention(s)

To assess for intervention success, we use a combination of inter- and intra-rater reliability values. Additionally, we collect comments from both student participants and volunteer observers through focus groups, ethnographic reports, and one-on-one interviews. However, interventions related to modules 4 and 5 do not directly impact student participant activities. Therefore, feedback collected from student participants in the two study contexts are out of scope of the present study and not taken into consideration.

To ensure adequate power for the present study, we consult third-party statisticians. Per their recommendations, we select four unique patient handovers from a pool of forty-six handovers delivered on the final day of the module 5 activity. We use these four handovers to test intra-rater reliability during the volunteer observer training for the telehealth elective. The observer training requires quick turnover, as extended training time is not viable given volunteer observer schedules. As such, observers score the same four videos three separate times, with a washout period of two days between scores and re-scores.

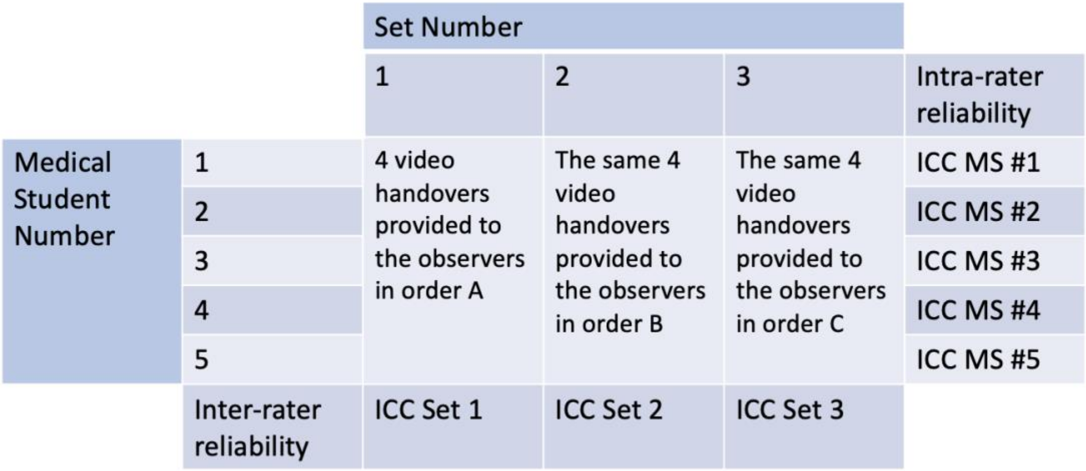
Measures and Analysis

The primary outcome measure is the two-way random effects ICC for each context, to determine tool inter-rater reliability with each novel application. We establish level of reliability according to Portney and Watkins' reliability scale.[24] Good reliability is an ICC > 0.75, moderate reliability is an ICC 0.5-0.75, and poor reliability is an ICC < 0.5. An additional outcome measure includes level of observer satisfaction. We collect verbal volunteer observer feedback at the end of each context in a focus group approach and use these comments to improve subsequent observer training modules.

The primary process measure was the intra-rater reliability calculated during the observer training session. Similar to the calculation for inter-rater reliability, we use a two-way random effects ICC to determine intra-rater reliability, with levels of reliability following guidelines from Portney and Watkins. The research team completes all calculations using the SPSS statistical package for Windows 26 (SPSS Inc, Chicago, IL, USA). Additionally, we collect and review observer feedback on completion of the observer training course.

Five medical student volunteers were recruited to observe the COVID telehealth handovers. During observer training, each volunteer returned rubrics for four unique handovers, completing the scoring activity three times (three total sets). For each set, the four unique handovers were differently ordered, and a wash out period of 2 days was used between each grading activity (appendix a.2). Intra- and inter-rater reliability were subsequently calculated for each volunteer and across each set (Figure 5).

Figure 4: Intra-rater reliability study setup. Five volunteer observers score the same four videos three separate times. The research team allow for a 2-day washout between viewings. Finally, inter- and intra-rater reliability was assessed through calculating for the ICC.



		Set Number			
		1	2	3	Intra-rater reliability
Medical Student Number	1	4 video handovers provided to the observers in order A	The same 4 video handovers provided to the observers in order B	The same 4 video handovers provided to the observers in order C	ICC MS #1
	2				ICC MS #2
	3				ICC MS #3
	4				ICC MS #4
	5				ICC MS #5
	Inter-rater reliability	ICC Set 1	ICC Set 2	ICC Set 3	

Ethical Considerations

The target population for this study includes health profession students attending either the UTSW school of health professions or the medical school. As such, all study participants are protected under the Family Educational Rights and Privacy Act (FERPA).[25,26] All research participants requiring access to either in-person or virtual handovers delivered by students in either the RE or the COVID-19 telehealth setting signed a confidentiality agreement acknowledging that the researcher had read and understood the rules and regulations set forth by FERPA. Similarly, all student attendees for both study settings were required to sign the same form, acknowledging and allowing for the researchers to view and assess student handover performance. All signed forms were verified and stored by a designated UTSW employee.

All COVID-19 positive patients interviewed during the telehealth course are protected under the Health Insurance Portability and Accountability Act (HIPAA).[27] Confidential patient information was discussed exclusively between students and attendings with access to view relevant video clips granted solely to the research team and evaluators. All study participants have previously undergone HIPAA training and are eligible to view protected health records.

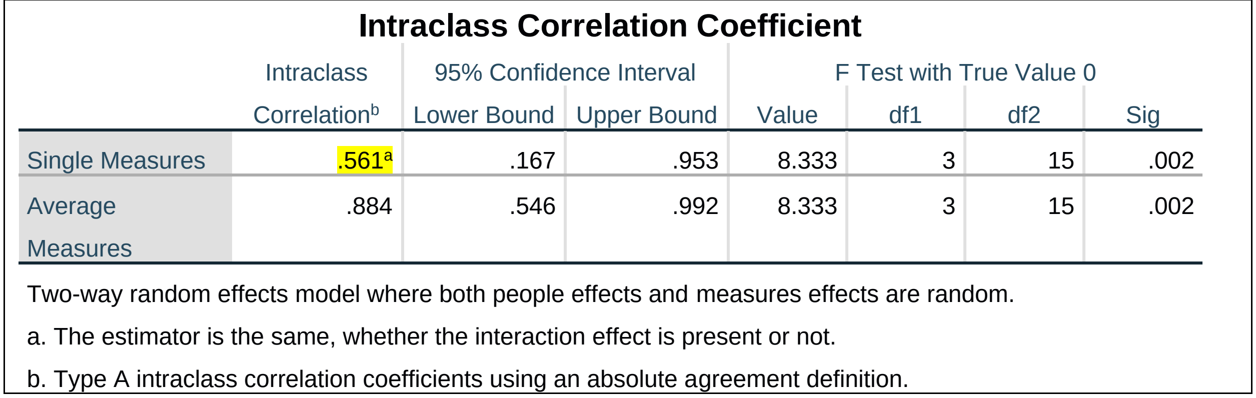
Finally, as this study is under the QEP, it qualifies as a quality improvement project. It does not introduce novel clinical research practices or methodologies and, therefore, is exempt from institutional review (IRB exempt).[28]

RESULTS

Module 5 (Residency Essentials)

A total of six medical student volunteer observers graded one set of four video recorded simulations selected from the final day of RE. We found the two-way random effects ICC to be 0.561 for inter-rater reliability with a 95% CI of 0.167 to 0.953 (Table 2). When asked for feedback, volunteer observers mentioned reasonable ease of use when reviewing videos.

Table 2: RE Inter-rater reliability. The ICC is 0.561 [0.167,0.953].



Intraclass Correlation Coefficient

	Intraclass	95% Confidence Interval		F Test with True Value 0			
	Correlation[b]	Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.561[a]	.167	.953	8.333	3	15	.002
Average Measures	.884	.546	.992	8.333	3	15	.002

Two-way random effects model where both people effects and measures effects are random.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

Module 4 (COVID-19 Telehealth)

A total of five student volunteers graded four pre-selected simulation videos from the telehealth elective. Due to the relatively results calculated during the observer training module, the research team decided to conclude the study prior to assigning video handovers to volunteer observers. Instead, we decided to calculate inter-rater reliability for each set of four videos during the observer training module (Table 3). Using this approach, the averaged two-way ICC was 0.644 with an averaged 95% CI of 0.244 to 0.964.

Table 3: COVID telehealth inter-rater reliabilities calculated for each of the sets scored during the observer training module.

Set #	Two-way random effects ICC
1	0.505 [0.068, 0.945]
2	0.854 [0.535, 0.989]
3	0.574 [0.128, 0.957]

Module 4 (COVID-19 Telehealth) Observer Training

The same five volunteer observers completed all three sets of videos during the observer training, allowing for the research team to calculate intra-rater reliability for each. Volunteer observer intra-rater reliability ranged from 0.105 to 0.667 (Table 4). When asked to review the observer training, volunteers commented that the training was reasonable and comprehensive.

Table 4: COVID telehealth Intra-rater reliability.

Rater #	Two-way random effects ICC
1	0.649 [-0.002, 0.970]
2	0.105 [-0.361, 0.863]
3	0.667 [0.020, 0.971]
4	0.542 [-0.115, 0.957]
5	0.667 [0.020, 0.971]

DISCUSSION

Summary

This study tested and optimized the L-HATS in two unique environments to determine the tool's reliability in settings outside of T2C. We gauged study success through a two-way random effects ICC assessing for inter-rater reliability for each handover activity. An additional process measure studied intra-rater reliability, which we tested exclusively in context of the observer training module for the COVID-19 telehealth elective. The RE reliability calculated for the six participants was 0.561 [0.167, 0.953], which is consistent with moderate reliability, according to the Portney and Watkins scale.[24] However, the confidence interval spans all three levels of the proposed scale, which undermines the actual reliability of the L-HATS in context of RE.

The research team concluded that the ambiguous results found when applying the L-HATS to handovers delivered during RE was a direct result of inadequate observer training and screening ahead of the actual event. To improve tool reliability during the telehealth course, we decided to extend observer training by 1.5 weeks, testing the volunteer observers by having each observer score the same set of four handover videos a total of three times. A washout period of two days was used to limit recall between sets. Intra-rater reliability calculated for the five volunteer observers ranged from 0.105 to 0.667. Notably, none of the observers were able to demonstrate a good level of intra-rater reliability, suggesting that all volunteers had at least a moderate level of variability on grading the same set of videos. Unfortunately, no single observer could match the level of reliability found during L-HATS testing.

Because of the unexpected results from the intra-rater reliability study during observer training, we decided to conclude the study without assigning any further videos to volunteer observers. However, each set of videos scored by the observers was assessed for inter-rater reliability, using the same approach as that for RE. Calculating inter-rater reliability yielded ICC values ranging from 0.505 to 0.854. The confidence interval for each of these calculations remained relatively large, with only the confidence interval for the highest ICC calculated limited to moderate to good reliability. The inter-rater reliability for the other two sets mirrored the results found during RE. A calculated reliability of 0.854 does correlate with good reliability in the setting of the COVID telehealth course, suggesting that further improvement in the volunteer observer training module may result in higher results for reliability.

Finally, we asked volunteer observers for verbal feedback regarding tool ease of use and satisfaction with observer training. We received the feedback in a focus group setting with all participating observers able to verbally voice comments and concerns related to their experience. Feedback received during these sessions were generally positive. Therefore, improvement in observer training largely focused on assessment tool reliability rather than volunteer feedback.

Interpretation

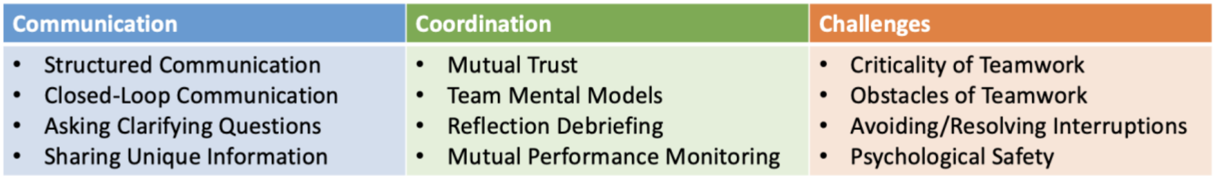
In context of the broader literature, a novel handover assessment tool tailored for undergraduate medical education would supplement the published rubrics.[15,29] Though originally designed to satisfy this niche, the L-HATS demonstrates increasing worsening in reliability as it is introduced to novel handover activities. While it maintains moderate levels of inter-rater reliability when used by the original research team, the L-HATS may demonstrate worse reliability at another program. In its current state, the L-HATS does not adequately diffuse to novel handover activities; though, optimizing observer training did seem to improve reliability.

Limitations

Study limitations include the relatively small number of handovers scored by each observer and the relatively high variability between handover activities. Each volunteer observer scored four to six pre-selected handover videos, which is a relatively low number of videos as compared to the original L-HATS study, where each observer scored thirty videos. The relatively low number of videos for the present study were partially due to logistical challenges in context of the ongoing COVID-19 pandemic and, in the case of the telehealth course, lower-than-expected intra-rater reliability testing, which prompted the researchers to end the study early. Future tests may repeat the study with larger numbers of videos selected for scoring.

One additional limitation included the high variability in handover environments. When comparing the T2C course to RE and COVID telehealth, each activity is drastically different. For example, RE focuses on teaching the I-PASS handover pneumonic versus T2C, which concentrates on SBAR. The telehealth course is a clinical environment and, therefore, not a simulation. Ideally, the current study would be repeated each year to further validate the results reported here. However, because RE only occurs once a year, repeating the present study is challenging.

Figure 5: QEP 12 core competencies. Each of the five modules selectively targets select competencies to cover all 12 competencies by medical school graduation.



Communication	Coordination	Challenges
• Structured Communication	• Mutual Trust	• Criticality of Teamwork
• Closed-Loop Communication	• Team Mental Models	• Obstacles of Teamwork
• Asking Clarifying Questions	• Reflection Debriefing	• Avoiding/Resolving Interruptions
• Sharing Unique Information	• Mutual Performance Monitoring	• Psychological Safety

Conclusions

This study demonstrates that diffusion of the L-HATS to handoff activities outside of T2C has moderate levels of reliability. Further testing during the COVID telehealth course suggests that improved observer training may positively correlate with inter-rater reliability. However, the large confidence intervals calculated throughout this study suggest that the tool's reliability in novel contexts remains ambiguous. The small numbers of video handoffs scored along with fewer-than-desired volunteers contributed to the study's ambiguous results. When compared against the inter-rater reliability calculated during T2C, the reliability for the L-HATS in other contexts is worse. We, therefore, conclude that the L-HATS does not easily diffuse into other handover environments; though, it may retain a moderate degree of inter- and intra-rater reliability in modules 4 and 5.

After several iterations of improving observer training, we were still unsuccessful in improving L-HATS scores, suggesting that there may be flaws to the assessment tool, itself. Applying the L-HATS to contexts outside of module 2 may have highlighted these flaws, resulting in the apparent reduction in tool reliability.

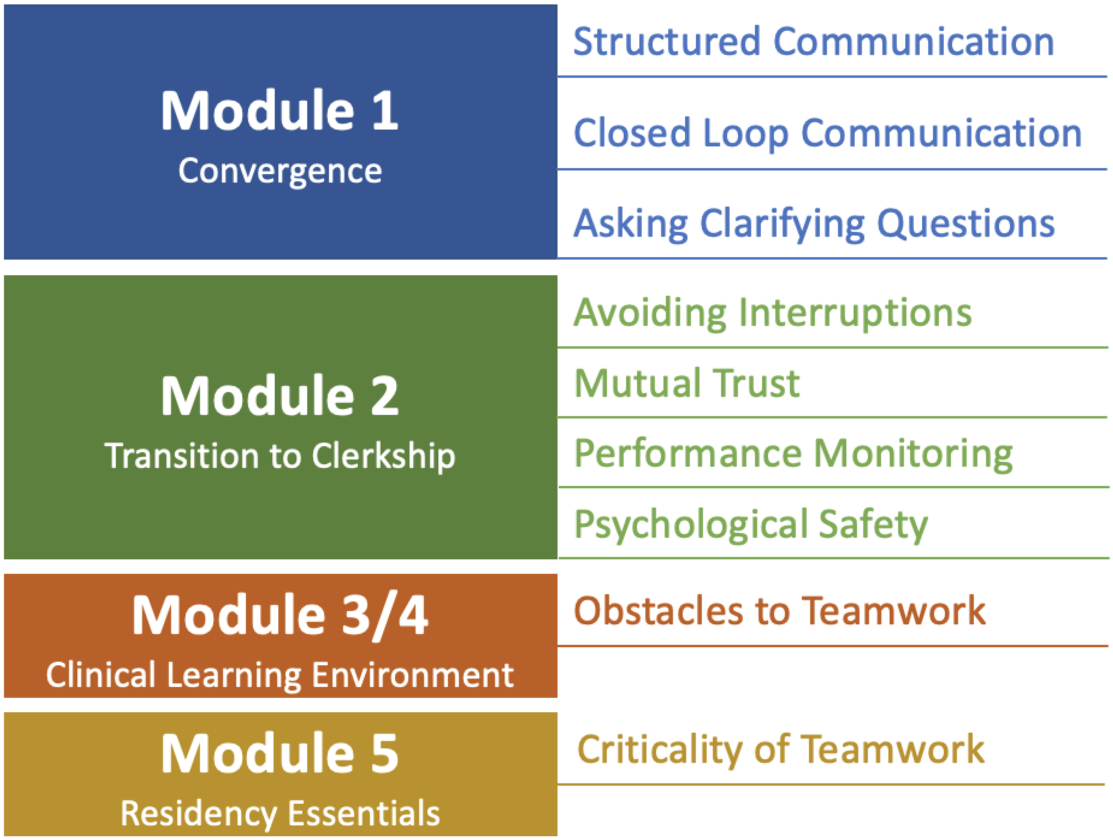
Next Steps

It is currently uncertain whether the L-HATS can readily be diffused to similar handover activities, given the lower levels of calculated reliability when tested in modules 4 and 5. However, handover activities such as RE and the telehealth elective require assessment tools to trend handover quality over time. Without a reliable rubric, each novel activity will require a unique assessment tool, tailored to each setting.

As we'd mentioned, the L-HATS demonstrated high reliability during module 2 of 2019 but likely is limited to that specific activity. One hypothesis for this limitation is that the contexts for each module are simply too different to allow for a universal assessment tool. When reviewing the core competencies that each module targets, we can appreciate the different objectives that each module targets (Figures 5 and 6). However, the module design is a cumulative process, such that module 2 includes those competencies covered in module 1. It may be possible to cumulatively assess classes by appending a basic assessment tool developed specifically for module 1 handover activities.

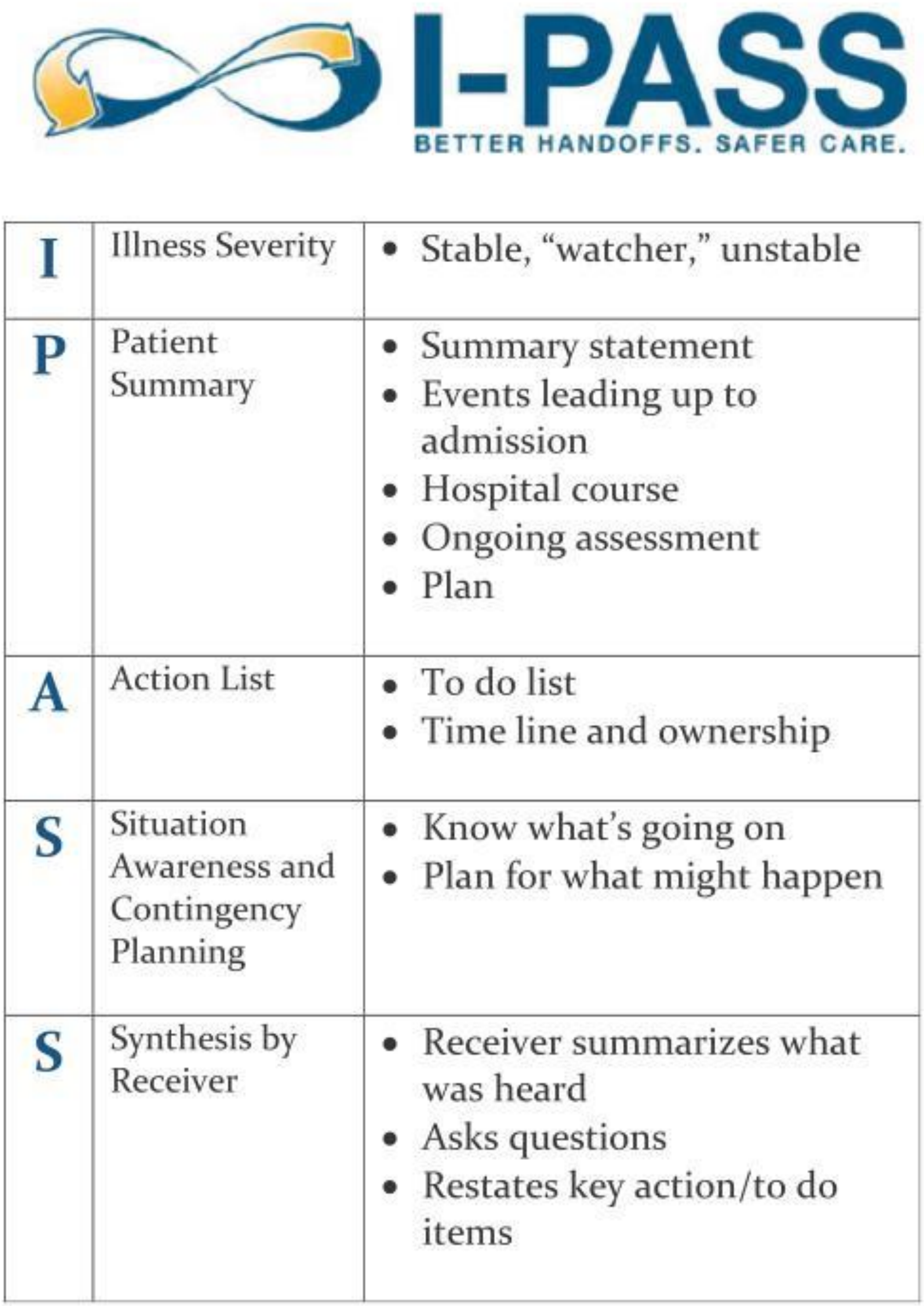
An alternative approach would focus on creating a basic assessment tool centered around elements most commonly found in proposed handover protocols. A recent paper describes the most common features included in published handover pneumonics.[30] Focusing only on the most universally recognized handover features, we could re-design an assessment tool covering these common elements.

Figure 6: Core competencies assigned to each of the five modules.



APPENDIX

Appendix a.1 The I-PASS handover structure taught during RE.



I-PASS

BETTER HANDOFFS. SAFER CARE.

I	Illness Severity	• Stable, “watcher,” unstable
P	Patient Summary	• Summary statement • Events leading up to admission • Hospital course • Ongoing assessment • Plan
A	Action List	• To do list • Time line and ownership
S	Situation Awareness and Contingency Planning	• Know what’s going on • Plan for what might happen
S	Synthesis by Receiver	• Receiver summarizes what was heard • Asks questions • Restates key action/to do items

Appendix a.2 Curriculum for the volunteer observer training module.

Curriculum for student experts training in reliability studies

Learning Objectives

This program will provide medical and health professional students with the ability to reliably review both clinically performed and simulated handovers using the Liang Handover Assessment Tool for Simulation (L-HATS). At the end of the L-HATS training course, the student will be able to:

1. Identify the key elements of an SBAR-based handover
2. Use L-HATS to reliably evaluate handovers in a pre-designated learning activity

Course Requirements

This 3-week training course will be designated as an extracurricular activity. Volunteers will contribute to the Quality Enhancement Plan (QEP) 2019 project, while gaining an in-depth knowledge for elements of an SBAR-based handover. All students are responsible to fulfill the following requirements for the course:

1. Attend and participate in the Q&A training session (online via Zoom)
2. Review and evaluate all handover videos according to the instructions provided
3. Complete the Transition to Clerkship (T2C) simulated handover activity OR talk with the Course Director
4. Sign and Submit the Student Participation Agreement form

Evaluation

Students will be evaluated according to their submitted grades from each of the assigned handover videos. Both inter- and intra-rater reliability analyses will be performed with passing individuals scoring $r > 0.7$ for each analysis performed. Students will have the opportunity to ask questions throughout the course, which will be shared with other trainees. However, L-HATS rubrics must be completed without outside collaboration.

The grade for the L-HATS training course will be pass/fail based on each student’s r scores. Passing students will receive the opportunity for continued participation in L-HATS tool use and refinement. Meanwhile, failing students will receive an opportunity to repeat the training course for a different learning activity.

Week 1

Activities	Materials	Week 1 Deliverables
Activities will focus on introducing L-HATS to students and allowing for opportunities to ask questions about L-HATS: 1. Review the preparatory email to be sent at the start of the week 2. Attend the online Q&A session to review L-HATS with an experienced trainer 3. Practice grading two sample handover videos at (using a tool for just-in-time grading)	1. A digital copy of the Student Participation Agreement form 2. A digital copy of L-HATS 3. A digital copy of the Grader Instruction Manual 4. A digital copy of the learning activity-specific grader guide	1. Signed Student Participation Agreement 2. Graded example videos 1 and 2

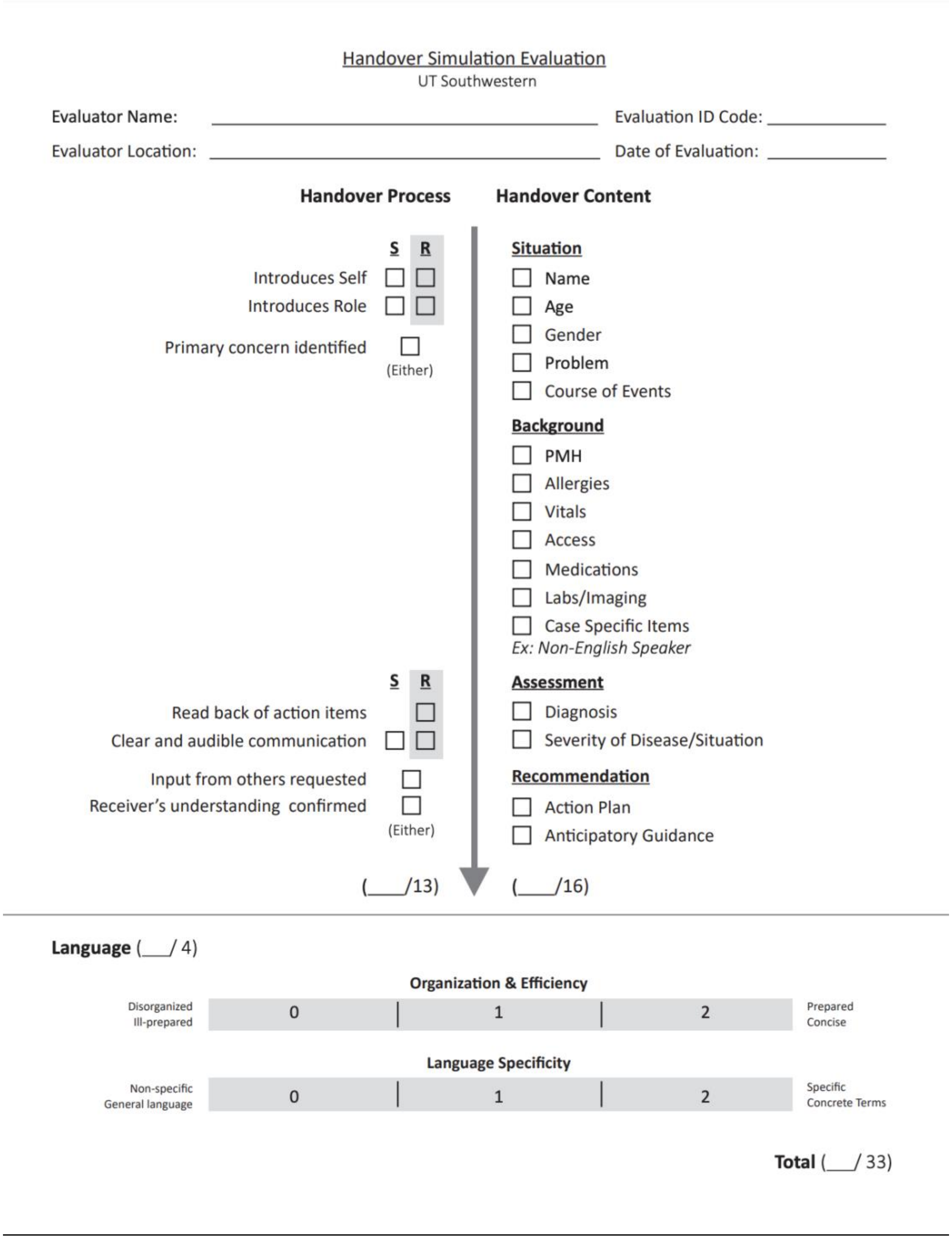
Week 2

Activities	Materials	Week 2 Deliverables
Student grader inter- and intra-rater reliability will be assessed: 1. Review the materials from week 1 2. Watch and grade first set of trial videos (not the example videos)	1. Set 1 of the trial videos 2. Week 1 materials	1. Grades for first set of trial videos from start of week 2

Week 3

Activities	Materials	Week 1 Deliverables
Student grader inter- and intra-rater reliability will be completed: 1. Watch and grade second set of trial videos	1. Set 2 of the trial videos 2. Week 1 materials	3. Grades for second set of trial videos from start of week 3

Appendix a.4 L-HATS used throughout the study.



Handover Simulation Evaluation

UT Southwestern

Evaluator Name: ______________________ Evaluation ID Code: __________

Evaluator Location: ______________________ Date of Evaluation: __________

Handover Process

	S	R
Introduces Self	□	□
Introduces Role	□	□
Primary concern identified	□ (Either)	

	S	R
Read back of action items		□
Clear and audible communication	□	□
Input from others requested	□	
Receiver's understanding confirmed	□ (Either)	

(___/13)

Handover Content

Situation

- □ Name
- □ Age
- □ Gender
- □ Problem
- □ Course of Events

Background

- □ PMH
- □ Allergies
- □ Vitals
- □ Access
- □ Medications
- □ Labs/Imaging
- □ Case Specific Items

Ex: Non-English Speaker

Assessment

- □ Diagnosis
- □ Severity of Disease/Situation

Recommendation

- □ Action Plan
- □ Anticipatory Guidance

(___/16)

Language (___/ 4)

Organization & Efficiency

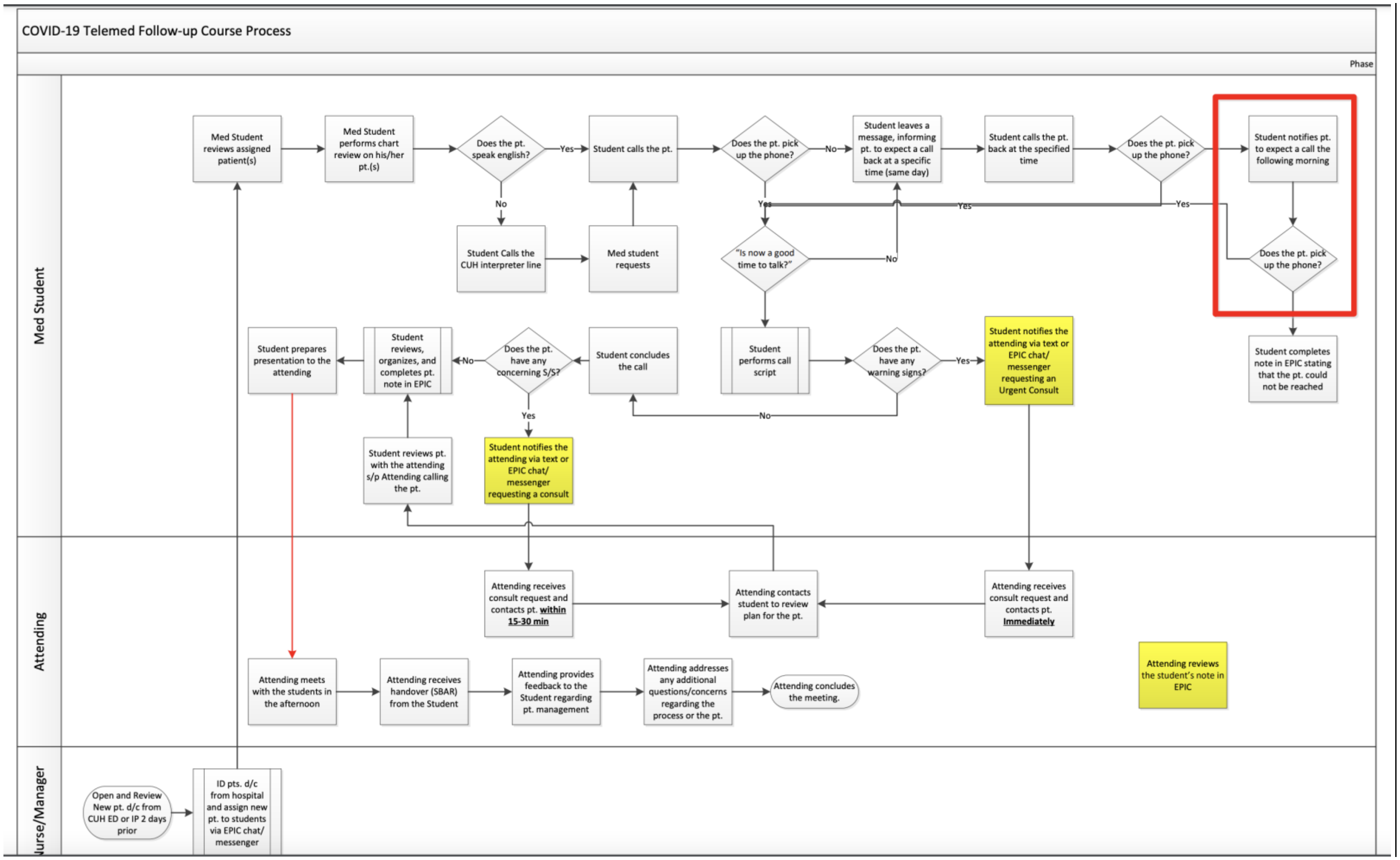
Disorganized Ill-prepared	0	1	2	Prepared Concise

Language Specificity

Non-specific General language	0	1	2	Specific Concrete Terms

Total (___/ 33)

Appendix b.1 Process map for the COVID-19 telehealth elective.


COVID-19 Telemed Follow-up Course Process
Phase
Med Student
Attending
urse/Manager
Open and Review New pt. d/c from CUH ED or IP 2 days prior
ID pts. d/c from hospital and assign new pt. to students via EPIC chat/ messenger
Med Student reviews assigned patient(s)
Med Student performs chart review on his/her pt.(s)
Does the pt. speak english?
Yes
No
Student Calls the CUH interpreter line
Med student requests
Student calls the pt.
Does the pt. pick up the phone?
No
Yes
"Is now a good time to talk?"
No
Student leaves a message, informing pt. to expect a call back at a specific time (same day)
Student calls the pt. back at the specified time
Does the pt. pick up the phone?
Yes
Yes
Student notifies pt. to expect a call the following morning
Does the pt. pick up the phone?
Student completes note in EPIC stating that the pt. could not be reached
Student performs call script
Does the pt. have any warning signs?
Yes
No
Student notifies the attending via text or EPIC chat/ messenger requesting an Urgent Consult
Student concludes the call
Does the pt. have any concerning S/S?
No
Yes
Student notifies the attending via text or EPIC chat/ messenger requesting a consult
Student reviews, organizes, and completes pt. note in EPIC
Student reviews pt. with the attending s/p Attending calling the pt.
Student prepares presentation to the attending
Attending receives consult request and contacts pt. within 15-30 min
Attending contacts student to review plan for the pt.
Attending receives consult request and contacts pt. Immediately
Attending meets with the students in the afternoon
Attending receives handover (SBAR) from the Student
Attending provides feedback to the Student regarding pt. management
Attending addresses any additional questions/concerns regarding the process or the pt.
Attending concludes the meeting.
Attending reviews the student's note in EPIC

REFERENCES

1. In: Kohn LT, Corrigan JM, Donaldson MS, eds. *To Err is Human: Building a Safer Health System.* Washington (DC)2000.
2. Carver N, Gupta V, Hipskind JE. Medical Error. In: *StatPearls.* Treasure Island (FL)2020.
3. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ.* 2016;353:i2139.
4. Bari A, Khan RA, Rathore AW. Medical errors; causes, consequences, emotional response and resulting behavioral change. *Pak J Med Sci.* 2016;32(3):523-528.
5. Murphy JG, Dunn WF. Medical errors and poor communication. *Chest.* 2010;138(6):1292-1293.
6. Joint Commission Center for Transforming Healthcare releases targeted solutions tool for hand-off communications. *Jt Comm Perspect.* 2012;32(8):1, 3.
7. Keebler JR, Lazzara EH, Patzer BS, et al. Meta-Analyses of the Effects of Standardized Handoff Protocols on Patient, Provider, and Organizational Outcomes. *Hum Factors.* 2016;58(8):1187-1205.
8. Starmer AJ, Spector ND, Srivastava R, et al. I-pass, a mnemonic to standardize verbal handoffs. *Pediatrics.* 2012;129(2):201-204.
9. Shahid ST, S. Situation, Background, Assessment, Recommendation (SBAR) Communication Tool for Handoff in Health Care - A Narrative Review. *Safety in Health.* 2018;4(7).
10. Riesenberg LA, Leitzsch J, Little BW. Systematic review of handoff mnemonics literature. *Am J Med Qual.* 2009;24(3):196-204.
11. Greilich PK, J. Multicenter Handoff Collaborative. *APSF Newsletter.* 2017;32(2).
12. Weiss KB, Bagian JP, Wagner R. CLER Pathways to Excellence: Expectations for an Optimal Clinical Learning Environment (Executive Summary). *J Grad Med Educ.* 2014;6(3):610-611.
13. Weiss KB, Wagner R, Bagian JP, Newton RC, Patow CA, Nasca TJ. Advances in the ACGME Clinical Learning Environment Review (CLER) Program. *J Grad Med Educ.* 2013;5(4):718-721.
14. Gordon M, Findley R. Educational interventions to improve handover in health care: a systematic review. *Med Educ.* 2011;45(11):1081-1089.
15. Horwitz LI, Dombroski J, Murphy TE, Farnan JM, Johnson JK, Arora VM. Validation of a handoff assessment tool: the Handoff CEX. *J Clin Nurs.* 2013;22(9-10):1477-1486.
16. Horwitz LI, Rand D, Staisiunas P, et al. Development of a handoff evaluation tool for shift-to-shift physician handoffs: the Handoff CEX. *J Hosp Med.* 2013;8(4):191-200.
17. *The University of Texas Southwestern Medical Center: Quality Enhancement Plan for the Southern Association of Colleges and Schools Commission on Colleges.* The University of Texas Southwestern Medical Center;2019.
18. Kang SHK. Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction. *Behavioral and Brain Sciences.* 2016;3(1):12-19.

19. Frankel A, Gardner R, Maynard L, Kelly A. Using the Communication and Teamwork Skills (CATS) Assessment to measure health care team performance. *Jt Comm J Qual Patient Saf.* 2007;33(9):549-558.
20. Liang T. *Development of the Liang Handover Assessment Tool for Simulation (L-HATS)*: Quality Safety and Outcomes Education, The University of Texas Southwestern Medical Center; 2020.
21. Podder V, Lew V, Ghassemzadeh S. SOAP Notes. In: *StatPearls.* Treasure Island (FL)2020.
22. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64(1):96-106.
23. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol.* 2012;8(1):23-34.
24. Portney LGW, M.P. *Foundations of Clinical Research; Applications of Practice.* 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall; 2009.
25. Family Educational Rights and Privacy Act (FERPA). *J Empir Res Hum Res Ethics.* 2007;2(1):101.
26. Rinehart-Thompson LA. Amendments to FERPA regulations. *J AHIMA.* 2009;80(7):56-57.
27. United S. Health Insurance Portability and Accountability Act of 1996. Public Law 104-191. *US Statut Large.* 1996;110:1936-2103.
28. Harrington L. Quality improvement, research, and the institutional review board. *J Healthc Qual.* 2007;29(3):4-9.
29. O'Toole JK, Stevenson AT, Good BP, et al. Closing the gap: a needs assessment of medical students and handoff training. *J Pediatr.* 2013;162(5):887-888 e881.
30. Nasarwanji MF, Badir A, Gurses AP. Standardizing Handoff Communication: Content Analysis of 27 Handoff Mnemonics. *J Nurs Care Qual.* 2016;31(3):238-244.

VITAE

Rodney Chen (July 22, 1993 – Present) is a fourth-year medical student at UT Southwestern Medical School in Dallas, TX. He graduated with a bachelor’s degree in Chemistry from Princeton University, and his current research interests focus on quality improvement and implementation sciences, specifically longitudinal medical student handover training. In his free time, he enjoys spending time with friends and family traveling the country, hiking, and running.

Permanent Address: 4548 S Lindhurst Ave
Dallas, TX 75229