

# Cleft Lip Repair Competence Can Be Evaluated with High-fidelity Simulation

Carolyn R. Rogers-Vizena, MD\*†‡  
 Caroline A. Yao, MD§¶  
 Georgios D. Sideridis, PhD‡||  
 Lindsey Minahan, BS†  
 Francesca Y. L. Saldanha,  
 MBBChir\*  
 Katie A. Livingston, ME†  
 Peter H. Weinstock, MD,  
 PhD†‡\*\*

**Background:** Demonstrating competency before independent practice is increasingly important in surgery. This study tests the hypothesis that a high-fidelity cleft lip simulator can be used to discriminate performance between training levels, demonstrating its utility for assessing procedural competence.

**Methods:** During this prospective cohort study, participants performed a unilateral cleft lip repair on a high-fidelity simulator. Videos were blindly rated using the Objective Structured Assessment of Technical Skills (OSATS) and the Unilateral Cleft Lip Repair Competency Assessment Tool (UCLR). Digital measurement of symmetry was estimated. Influence of training level and cumulative prior experience on each score was estimated using Pearson r.

**Results:** Participants (n = 26) ranged from postgraduate year 3 to craniofacial fellow. Training level correlated best with UCLR (R = 0.4842, P = 0.0122\*) and more weakly with OSATS (R = 0.3645, P = 0.0671), whereas cumulative prior experience only weakly correlated with UCLR (R = 0.3450, P = 0.0843) and not with OSATS (R = 0.1609, P = 0.4323). UCLR subscores indicated marking the repair had little correlation with training level (R = 0.2802, P = 0.1656), whereas performance and result did (R = 0.5152, P = 0.0071\*, R = 0.4226, P = 0.0315\*, respectively). Correlation between symmetry measures and training level was weak.

**Conclusions:** High-fidelity simulation paired with an appropriate procedure-specific assessment tool has the construct validity to evaluate performance for cleft lip repair. Simply being able to mark a cleft lip repair is not an accurate independent assessment method nor is symmetry of the final result. (*Plast Reconstr Surg Glob Open* 2022;10:e4435; doi: 10.1097/GOX.0000000000004435; Published online 22 July 2022.)

## INTRODUCTION

Simulation of varying fidelity has long been a mainstay of training and assessing performance for health care teams in situations such as codes. As simulation is increasingly integrated into teaching both technical and

nontechnical surgical skills, interest in using simulation for the purpose of individual evaluation has grown. The Fundamentals of Laparoscopic Surgery (FLS) (Society of American Gastrointestinal and Endoscopic Surgeons, Los Angeles, CA) program has one of the most well-known examples in surgical training of simulation applied to both teaching and structured assessment. A unique component of the FLS platform is the deliberate practice of technical skills on a laparoscopic box trainer in addition to the cognitive curriculum. The subsequent FLS assessment includes both evaluation of knowledge and manual skills to measure technical competence with core laparoscopic

From the \*Department of Plastic and Oral Surgery, Boston Children's Hospital, Boston, Mass.; †Boston Children's Hospital Simulator Program (SIMPeds), Boston, Mass.; ‡Harvard Medical School, Boston, Mass.; §Division of Plastic Surgery, Shriners' Hospital for Children, Los Angeles, Calif.; ¶Operation Smile, Virginia Beach, Va.; ||Institutional Centers for Clinical & Translational Research, Boston Children's Hospital, Boston, Mass.; and \*\*Department of Anesthesia and Critical Care, Boston Children's Hospital, Boston, Mass.

Received for publication May 13, 2022; accepted May 24, 2022.

Presented at the American Cleft Palate-Craniofacial Association Virtual Meeting, April 2021.

Copyright © 2022 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 \(CCBY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/GOX.0000000000004435

**Disclosure:** The simulator used in this paper is proprietary technology developed by the Boston Children's Hospital Simulator Program. Carolyn R. Rogers-Vizena, Lindsey Minahan, Francesca Y. L. Saldanha, and Peter H. Weinstock each were involved in some element of the development process. The other authors have no current financial interest to declare. This work was supported by a National Endowment for Plastic Surgery grant from the Plastic Surgery Foundation.

Related Digital Media are available in the full-text version of the article on [www.PRSGlobalOpen.com](http://www.PRSGlobalOpen.com).

surgical maneuvers.<sup>1,2</sup> This program has demonstrated such great importance that successful completion of FLS and its partner program, Fundamentals of Endoscopic Surgery (Society of American Gastrointestinal and Endoscopic Surgeons, Los Angeles, CA), is an American Board of Surgery requirement for graduating general surgery residents. This is just one example of the growing sentiment that demonstrating competence before independent practice is vital for graduating residents and credentialing surgeons.

The abstract nature of many plastic surgical procedures makes developing and validating simulators more challenging than for other specialties. High-fidelity simulation is unique in that it allows the operator to use real instruments to perform varied maneuvers requiring both judgment and skill. Thus, it may be uniquely advantageous for evaluating abstract procedures in plastic surgery such as cleft lip repair. To address this, through collaboration with engineers and special effects experts within our hospital-based simulator program, we developed a high-fidelity unilateral cleft lip surgical simulator. A pilot study involving practicing cleft surgeons demonstrated preliminary face and content validity of the simulator by assessing user experience and comparing simulator surface change with real patients.<sup>3</sup> A subanalysis in a more recent educational study suggested that the simulator may have the construct validity to discriminate between learners of different levels.<sup>4</sup> However, before one can propose that this or any other simulator is an accurate means to evaluate someone's ability, its construct validity must be more rigorously assessed, something that has been generally lacking in simulation of plastic surgical procedures.

This study fills a void in rigorously validated simulators in plastic surgery by deliberately evaluating the construct validity of a high-fidelity unilateral cleft lip surgical simulator. The central hypothesis is that the cleft lip simulator can discriminate surgical performance between training levels, demonstrating its utility for assessing surgeon competence. If correct, it would warrant the simulator's broader incorporation into plastic surgery evaluation and credentialing. This has value not just for surgical training. In the world of cleft lip and palate care, documenting ability before participation would also be useful for surgeons involved in cleft-related humanitarian missions,<sup>5</sup> where participation by surgeons of variable skill has been a criticism.

## METHODS

This prospective educational study was conducted under IRB-exemption status (IRB-P00035608) from April 2019 to January 2021. Residents and fellows were recruited from three residency programs rotating at Boston Children's Hospital and from the Boston Children's Hospital craniofacial/pediatric plastic surgery fellowship program. Participants had access to the same instructional videos created by the first author to prepare for the simulated procedure as they would a normal case. Participation was voluntary.

Before beginning the simulation, each participant read the same introductory statement orienting him or

## Takeaways

**Question:** Can high-fidelity simulation be effectively used to evaluate cleft lip repair surgical ability?

**Findings:** In a group of trainees of varying levels operating on a high-fidelity cleft lip simulator, there was significant moderate correlation between cleft lip-specific skill and training level. In the same group, correlation between training level and both global technical skill and surface symmetry was weak.

**Meaning:** High-fidelity cleft lip simulation paired with an appropriate procedure-specific skills checklist can be effectively used to evaluate surgical performance. However, global evaluations of technical skill and surface measures of symmetry alone are insufficient measures to discriminate ability.

her to the session to facilitate consistent understanding regardless of prior simulation experience. Participants then performed a cleft lip repair from start to finish on the high-fidelity simulator. Procedures were assisted one-on-one by the same surgeon; no guidance or coaching was offered. Simulated operations were videotaped from a single overhead view (Fig. 1) without audio to allow for anonymous rating at the end of the study. Repaired simulators were 3D imaged using a TRIOS 3 intraoral scanner (3Shape Global) to create high-resolution 3D digital surface scans of the lip and nose for objective measurement.

## Video Evaluation

Procedural videos were blindly scored by two independent raters, both of whom are practicing cleft surgeons. The mean score for each independent item was used in the analyses. Two rating scales were used: the modified Objective Structured Assessment of Technical Skills (OSATS) and the Unilateral Cleft Lip Repair Competency Assessment Tool (UCLR). The modified OSATS is a four-point global assessment of surgical skill specifically abbreviated for use in reviewing intraoperative video.<sup>6</sup> (See appendix, Supplemental Digital Content 1, which displays Modified OSATS, <http://links.lww.com/PRSGO/C101>.) The UCLR is an 18-item cleft lip repair skills checklist that is agnostic to eponymous repair technique with subscores



Fig. 1. Frontal view used for rating simulated cleft lip repair videos.

for marking, performance, and result.<sup>7</sup> (See appendix, Supplemental Digital Content 2, which displays the UCLR, <http://links.lww.com/PRSGO/C102>.)

### Measurement of Lip/Nasal Symmetry

As a purely objective assessment of result, digital measurements of median distance and root-mean-square deviation (RMSD) between 3D surfaces were used as indicators of symmetry, with greater symmetry assumed to represent an objectively better outcome. These measurements were calculated from the measured distances between a 3D digital surface scan of each repaired simulator and its mirror image. Digital 3D scans of each repaired simulator were exported as .stl files and imported into 3-matic software (Materialise, Belgium). To consistently set the plane of symmetry, a contour surface of the outer edge of the replaceable simulator was used for alignment, and superfluous data were eliminated. Each scan was then mirrored about the YZ plane, and the scan and its mirror image were analyzed for symmetry against one another by utilizing the part comparison tool in 3-matic software (Materialise, Belgium), which measures the absolute value of the closest distance between all triangle nodes of the two surfaces. Median distance measured between the two surfaces was used as one numerical indicator of asymmetry because of the nonnormal distribution of data generated. RMSD was selected as another measure given its increasing acceptance as a tool for evaluating facial symmetry (Fig. 2).<sup>8,9</sup>

### Statistical Analysis

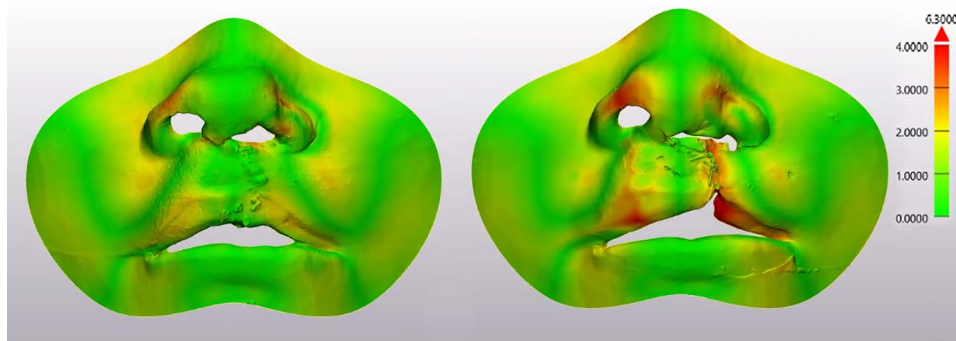
Descriptive statistics were tabulated. Influence of training level and cumulative prior experience on each score was estimated using Pearson  $r$  with the expectation that despite individual variation within a category, in a larger cohort there would be moderate correlation with score if simulation had appropriate construct validity. Several analytical means were involved to estimate interrater reliability. First, intraclass correlation coefficient (ICC) was estimated, which reflects amount of variance explained

by one rater in relation to another.<sup>10</sup> Two indices of ICC were utilized, namely consistency and agreement, using a single-measurement and a two-way mixed-effects model. Conventions for ICCs are as follows: 0–0.50, poor; 0.50–0.75, moderate; 0.75–0.90, good; and greater than 0.90, excellent. Additionally, the concordance correlation coefficient<sup>11,12</sup> was estimated, which is the product of Pearson's  $r$  with a measure of accuracy (deviance of least squares line from 45°) with values below 0.90 reflecting poor agreement. Finally, Cohen's<sup>13</sup> weighted kappa was used for which estimates between 0.40 and 0.59 show weak agreement, between 0.60 and 0.79 show moderate agreement, between 0.80 and 0.90 show strong agreement, and above 0.90 show almost perfect agreement.

## RESULTS

Participants ranged from postgraduate year (PGY) 3 integrated and PGY6 independent residents through craniofacial or pediatric plastic surgery fellow. For comparative analysis, integrated and independent residents are grouped together based on their analogous position in the plastic surgery rotation schedule, and hereafter, independent residents will be referred to by their integrated-equivalent PGY. Fellows were assigned the most senior position of PGY7 based on the integrated residency structure, regardless of actual PGY. Prior cleft lip repair experience ranged from 0 to 45 repairs (median = 5.5) with the majority having participated in eight or fewer procedures (Table 1). Of note, one participant was unable to fully complete repair.

Our primary interest was whether surgical performance correlated with training level (Fig. 3) because positive correlation would suggest that simulation is a useful tool for discriminating ability. We found that total UCLR score had significant positive correlation with training level ( $R = 0.4842$ ,  $P = 0.0122^*$ ), while OSATS had nonsignificant weaker correlation ( $R = 0.3645$ ,  $P = 0.0671$ ), although the latter is in the range of a medium effect size based on the work of Cohen<sup>14</sup> and can be highlighted for that purpose as the  $P$  value was



**Fig. 2.** Example symmetry measurements. These color maps are examples of the output from 3-matic after overlaying a 3D scan of the repaired simulator with its mirror image. Areas of overlap between the two images are represented in green, while increasing distance between the two images caused by asymmetry shifts the color from green to yellow to red (scale in mm at far right). A more symmetrical or better cleft lip repair is shown on the left with a smaller median distance (0.8674 mm) and a smaller RMS value (1.1242 mm). A more asymmetrical or worse cleft lip repair is shown on the right with a larger median distance (1.1458 mm) and larger RMSD (1.4742 mm).

**Table 1. Demographics**

Participants	Number (%)
Training level	
Integrated 3	4 (15.4)
Integrated 4/independent 6	4 (15.4)
Integrated 5/independent 7	8 (30.8)
Integrated 6/independent 8	5 (19.2)
Craniofacial or pediatric plastic surgery fellow	5 (19.2)
Gender	
Male	20 (76.9)
Female	6 (23.1)
Prior experience with cleft simulation	
Yes	5 (19.2)
No	21 (80.8)
Total prior experience*	
0–2	6 (23.1)
3–5	7 (26.9)
6–8	6 (23.1)
9+	7 (26.9)
	Number
Type of cleft lip repair performed	
Mohler	21
Millard	4
Mulliken	1
	Mean ± SD
Duration of simulated procedure (min)	81 ± 36
Total participants	26

\*Prior experience is defined as the sum of all cleft lip repairs that an individual observed, assisted with, and/or performed independently.

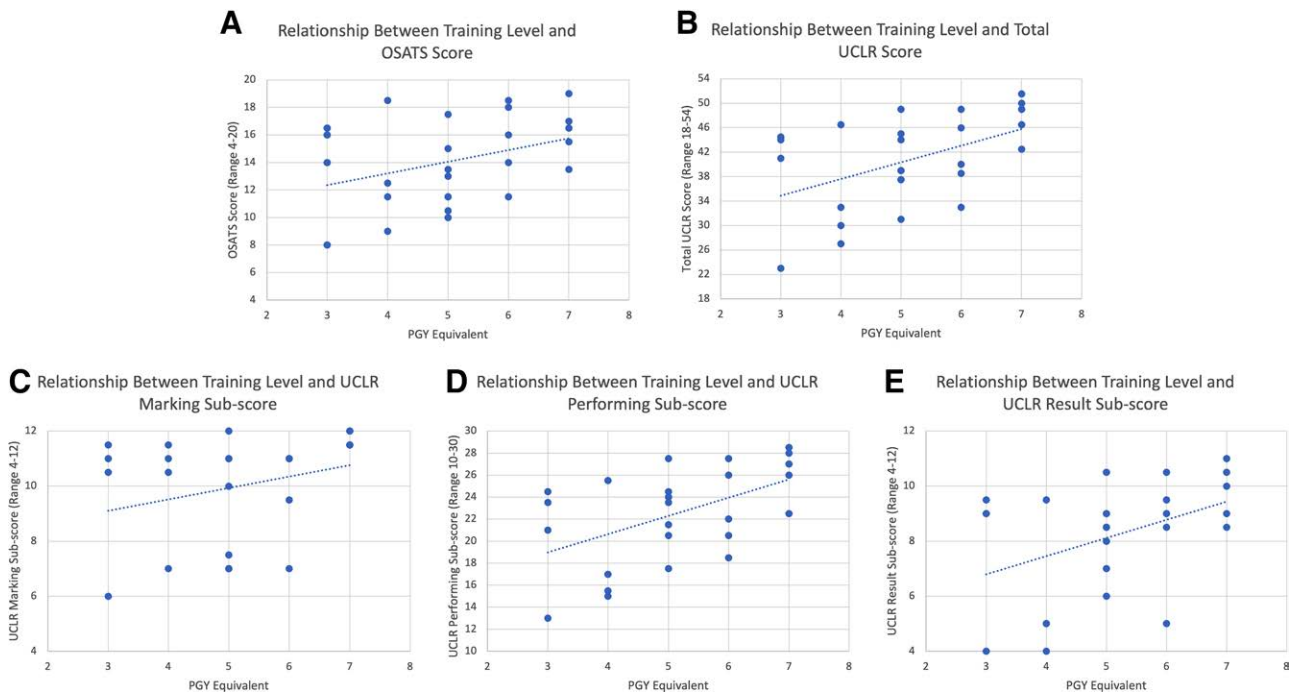
heavily influenced by low statistical power. Diving into the cleft lip-specific score a little further, training level had little correlation with accurately marking the repair ( $R = 0.2802$ ,  $P = 0.1656$ ), but both performing the repair ( $R = 0.5152$ ,  $P = 0.0071^*$ ) and result ( $R = 0.4226$ ,  $P = 0.0315^*$ ) correlated significantly with training level (Table 2).

We were also interested in whether prior experience with cleft lip repair correlated with surgical performance

(Fig. 4) and found minimal correlation for OSATS ( $R = 0.1609$ ,  $P = 0.4323$ ), the global assessment, and weak non-significant correlation for the UCLR score ( $R = 0.3450$ ,  $P = 0.0843$ ), although the latter is again in the range of a medium effect size. Interestingly, although nonsignificant, marking the repair correlated better with prior experience ( $R = 0.3392$ ,  $P = 0.0900$ ) than it did with training level.

Finally, we looked at how surface symmetry correlated with training level and prior experience. For median difference, there was a nonsignificant negative correlation but of a medium effect size such that more senior trainees tended toward more symmetric results than junior trainees ( $R = -0.3659$ ,  $P = 0.0667$ ), but no correlation between symmetry and prior experience ( $R = 0.0162$ ,  $P = 0.9355$ ). For RMSD, there was little correlation with training level ( $R = -0.2821$ ,  $P = 0.1626$ ) and no correlation with prior experience ( $R = -0.0361$ ,  $P = 0.8610$ ).

Table 3 shows the results of interrater reliability for the OSATS and UCLR scales using total scores and UCLR subscales. Results indicate moderate levels of interrater reliability for total UCLR score using ICC and kappa as well as for the UCLR's marking subscore. For UCLR performing subscore, there were moderate levels using ICC consistency and agreement. For UCLR result subscore, moderate levels of interrater reliability were evidence using ICC consistency only. However, to further evaluate whether UCLR and the objective symmetry index were measuring similar constructs, we looked for correlation between the UCLR results score and the symmetry measurements and found significant correlation between the two for both median difference ( $R = -0.6416$ ,  $P = 0.0004^*$ ) and RMSD ( $R = -0.6292$ ,  $P = 0.0006$ ). Finally, estimates of interrater reliability for

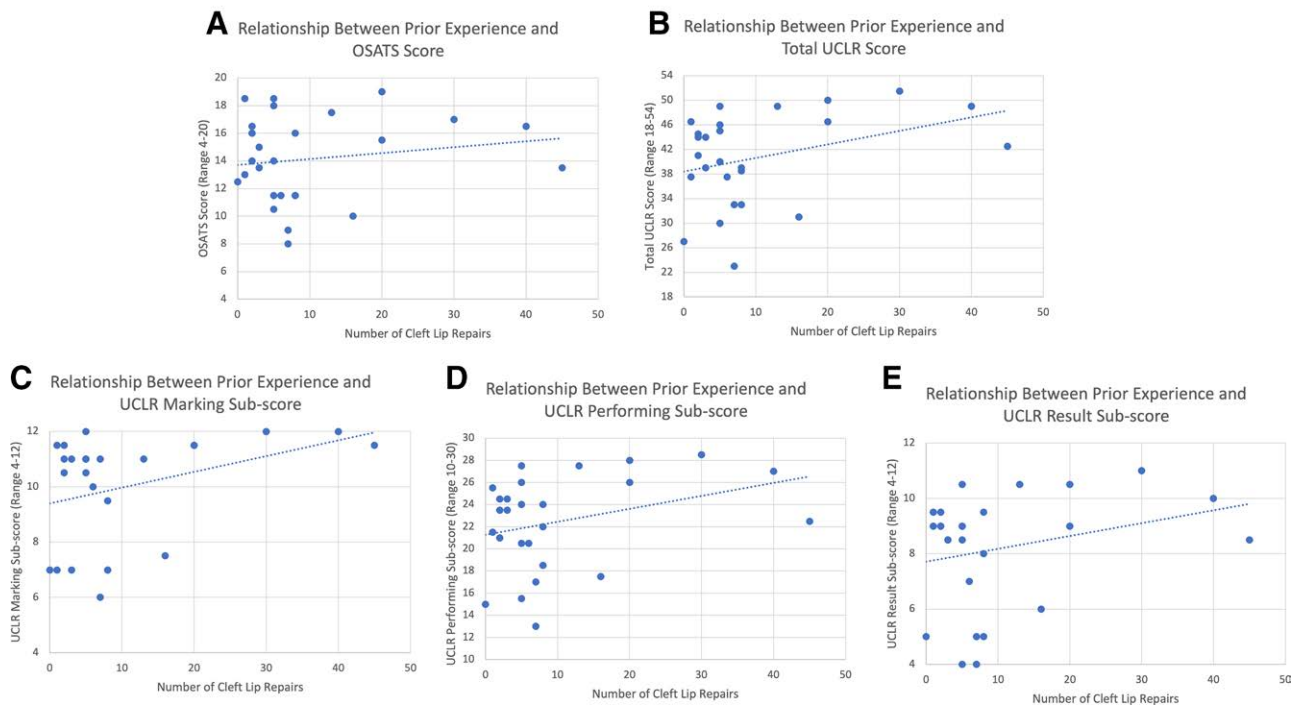


**Fig. 3.** Correlation between structured rating scales and participant training level. A, Relationship between training level and OSATS. B, Relationship between training level and total UCLR score. C, Relationship between training level and UCLR Marking subscore. D, Relationship between training level and UCLR Performing subscore. E, Relationship between training level and UCLR result subscore.

**Table 2. Correlation between Assessment Measures (Pearson r)**

Variable	Training Level	Prior Experience	Median Difference	RMSD
RMSD	R = -0.2821; P = 0.1626	R = -0.0361; P = 0.8610	<b>R = 0.8315; P &lt; 0.0001*</b>	—
Median difference	R = -0.3659; P = 0.0667	R = 0.0162; P = 0.9355	—	—
OSATS score	R = 0.3645; P = 0.0671	R = 0.1609; P = 0.4323	—	—
Total UCLR score	<b>R = 0.4842; P = 0.0122*</b>	R = 0.3450; P = 0.0843	—	—
UCLR marking subscore	R = 0.2802; P = 0.1656	R = 0.3392; P = 0.0900	—	—
UCLR performance subscore	<b>R = 0.5152; P = 0.0071*</b>	R = 0.3211; P = 0.1097	—	—
UCLR result subscore	<b>R = 0.4226; P = 0.0315*</b>	R = 0.2634; P = 0.1943	<b>R = -0.6416; P = 0.0004*</b>	<b>R = -0.6292; P = 0.0006*</b>

\*P-value < 0.05 was used as a threshold for significance.



**Fig. 4.** Correlation between structured rating scales and participant total prior experience with cleft lip repair. A, Relationship between prior experience and OSATS. B, Relationship between prior experience and total UCLR score. C, Relationship between prior experience and UCLR marking subscore. D, Relationship between prior experience and UCLR performing subscore. E, Relationship between prior experience and UCLR result subscore.

OSATS were below acceptable levels across all reliability indicators. The amounts of variance explained by the agreements between the two raters ranged between 10.2% for OSATS and 57.8% for UCLR’s marking subscore.

### DISCUSSION

With the growing emphasis on safety, simulation is increasingly used for training and evaluating modern surgeons in a risk-free environment. In general surgery, two examples are the Advanced Trauma Life Support classes (American College of Surgeons, Chicago, IL) used to teach

procedural skills and verify competence of trauma teams and the previously referenced FLS program that teaches and tests minimally invasive surgical techniques. In orthopedic surgery, arthroscopy simulation is used for resident teaching and evaluation<sup>15,16</sup> and is now a mandatory part of certification by the Swiss Orthopaedics Board.<sup>17</sup> Unlike those specialties, simulation has played a little formal role in plastic surgery training and no role in the process of evaluating residents and credentialing attending surgeons. We have largely relied on proxy measures of competence, such as meeting minimum case numbers during

**Table 3. Interrater Reliability Estimates of UCLR Total and Subscores and OSATS Total Score**

Rating Scale	ICC Consistency/95% CI	ICC Agreement/95% CI	Concordance Coefficient/95% CI	Weighted Kappa/95% CI
OSATS score	0.290/(-0.103 to 0.604)	0.217/(-0.0956 to 0.522)	0.210/(-0.0552 to 0.448)	0.233/(-0.012 to 0.478)
Total UCLR score	<b>0.686/(0.413–0.846)</b>	<b>0.618/(0.237–0.820)</b>	0.608/(0.342–0.784)	<b>0.600/(0.3498–0.8507)</b>
Marking subscore	<b>0.728/(0.480–0.868)</b>	<b>0.726/(0.482–0.866)</b>	0.194/(-0.009 to 0.382)	<b>0.718/(0.541–0.895)</b>
Performance subscore	<b>0.682/(0.407–0.843)</b>	<b>0.584/(0.137–0.811)</b>	0.575/(0.309–0.758)	0.548/(0.334–0.762)
Result subscore	<b>0.507/(0.156–0.744)</b>	0.420/(0.033–0.695)	0.410/(0.107–0.644)	0.410/(0.077–0.743)

Estimates in bold are at the moderate level of reliability. P-value < 0.05 was used as a threshold for significance.

residency and passing oral examinations where cases are discussed rather than performed. Although there is value in minimum experience benchmarks and evaluating cognitive elements of plastic surgery, neither technical ability nor surgical judgment is fully elucidated with these methods. There is a need for comprehensive, multifaceted assessment in plastic surgery.<sup>18</sup>

The arrival of high-fidelity simulation allows synchronous evaluation of cognitive ability and both technical and non-technical skills in a way that is uniquely suited to plastic surgery. High-fidelity simulation allows the operator to use real surgical equipment to physically perform abstract maneuvers that require a level of judgment and skill not feasible on digital or low-fidelity simulators. This provides hands-on experience that can be translated directly to patient care and creates situational stress that better prepares the learner for the real world. High-fidelity simulation has been introduced as an educational tool for multiple plastic surgery procedures, including cleft palate repair,<sup>19</sup> cleft lip repair,<sup>3,20</sup> rhinoplasty,<sup>21</sup> carpal tunnel release,<sup>22</sup> migraine injections,<sup>23</sup> and more. However, before broadly incorporating any simulator into educational curricula or using it to evaluate competency, one must ensure that it possesses the capability to discriminate between operators of varying ability, and thus, is a valid test instrument. This study fills a void in validated plastic surgery simulators by demonstrating the construct validity of a unilateral cleft lip simulator.

In this study, we evaluated cleft lip repair performance and end result for a cohort of residents of varying training levels. This was done under the assumption that although individual variation in knowledge and technical ability can be expected within individuals at any given training level, when looking at a larger cohort, progressive improvement would be anticipated. Taken as a whole, there was an expected moderate level of correlation between cleft lip-specific performance and training level. This suggests that the high-fidelity cleft lip simulator does indeed have proper levels of convergent validity as it correlated well with relevant constructs. Thus, our hypothesis was supported that high-fidelity simulation has the construct validity to be used to evaluate surgeon competence with cleft lip repair, as evidenced by correlation between training level and a skills checklist designed specifically for the procedure being assessed.

There were some surprising findings. For one thing, marking the repair correlated less with training level than performing the repair and final result. This may be because even relatively junior participants could mark a unilateral cleft lip repair reasonably well but often omitted or incorrectly performed more advanced, yet still important, maneuvers. The lack of impact of marking may also result from the fact that the study was limited to residents rotating at a major children's hospital who were rarely complete novices. Interns and PGY 2 residents completely lacking exposure to cleft lip surgery were not included. This unexpected study finding was particularly interesting because demonstrating cleft lip repair marking has historically been asked in the context of board examinations. Our results indicate that simply being able to mark a cleft lip repair is not the best way to assess competence, which gives support to the role of high-fidelity simulation in plastic surgery assessment.

Similarly, we were disappointed that the objective measures of surface symmetry correlated poorly with training level. A challenge to digitally measuring result is the fact that an "ideal" immediate cleft lip repair shape is not universally agreed upon, and while purely objective, symmetry may not be optimal, especially for the nose. Outputting a single, completely objective measure of result would have been a less onerous evaluation tool to implement on a large scale than observing the procedure from start to finish. However, our results indicate that to effectively use high-fidelity simulation for competence assessment, it must be paired with a procedure-specific rating tool that takes into consideration all elements of the procedure including expert rating of the end result, not just final lip/nasal symmetry.

With the growing excitement around competency-based graduation,<sup>24,25</sup> we were very interested in how experience correlated with performance and outcome. We were surprised to find that experience correlated considerably less with performance and outcome than did training level. This may be related to the greatly skewed experience of our cohort, but the results do call into question the value of arbitrary "case numbers" as an indicator of readiness for graduation. Additionally, the expected within-PGY score variation seen here provides modest credence for the concept of competency-based graduation since some junior residents were able to operate on par with craniofacial fellows, whereas some senior participants could have used a bit more training.

Taken as a whole, our results support the use of high-fidelity cleft lip simulation for objective evaluation of cleft lip repair surgical competency. Although we do not yet know conclusively how performance on a simulator correlates with operating on a real patient, our results suggest potential applications for evaluating readiness for graduation, credentialing surgeons, and verifying ability of surgeons volunteering for cleft-related humanitarian missions. While it is important to note that our results are only directly applicable to this specific simulator, they give credibility to a broader role for high-fidelity simulation in plastic surgery evaluation. In a growing simulation landscape containing many high-fidelity simulators relevant to plastic surgery, with thoughtful validation, one could foresee a future in which plastic surgeons undergo similar structured evaluation to their counterparts in other surgical specialties with potential to enhance patient safety.

## LIMITATIONS

Despite significant results found, this is a relatively small sample size, and thus, some results are limited by statistical power. Recruitment was limited by an in-person-research hiatus mid-study due to COVID-19. Additionally, we recognize that our objective measure of result, asymmetry, is imperfect in the sense that absolute symmetry would not be considered ideal by all surgeons. For example, some either intentionally overcorrect position of the lower lateral cartilage or simply avoid operating on the nose entirely, leaving a purposefully asymmetric result. We explored the alternative of comparing simulators to a "golden standard" repaired by experts but did not pursue this because again, an "ideal" result remains subject

to personal opinion. A third limitation was the poor interrater reliability for the OSATS scale. This may be related to the fact that OSATS was originally designed for use in person where the operator could be observed and heard. Thus, it may not be an optimal assessment in this anonymized environment. Another limitation is that the participants in this group were relatively inexperienced, with a few having substantial prior cleft lip case volume and the rest having very limited experience. This uneven distribution may have limited our ability to truly measure how performance correlates with experience. Another limitation is the simulator itself. Although it provides myriad possible movements much like an actual cleft lip/nasal deformity, silicone by nature retains some shape, and thus, a truly optimal cleft lip repair is likely impossible. Finally, the video used in the study included a narrow field of view without audio. This was a necessary part of the blinding process to avoid operators being recognized, but precluded the evaluation of nontechnical skills, such as communication and use of assistants, which are clearly an important part of surgical competency.

### CONCLUSIONS

High-fidelity cleft lip simulation has value for evaluating surgeon competence. This has implications for screening skill in the context of graduating residents, credentialing surgeons, and verifying ability of surgeons participating in cleft missions. With this in mind, one could imagine a future where rather than talking about operating during examinations, one actually performs procedures on a variety of well-validated simulators.

Carolyn R. Rogers-Vizena, MD

Department of Plastic and Oral Surgery  
Boston Children's Hospital  
300 Longwood Ave  
Boston, MA 02115

E-mail: carolyn.rogers@childrens.harvard.edu

### ACKNOWLEDGMENTS

The authors greatly appreciate the assistance of Ms. Tamia Hargrove in digitizing repaired simulators and Mr. Gregory Loan and Mr. Stephen Wilson for refinements in design and fabrication of the simulator that made this work possible. We also thank the residents and fellows who participated in this work and continue to be a joy to both teach and learn from every day.

### REFERENCES

- Peters JH, Fried GM, Swanstrom LL, et al; SAGES FLS Committee. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*. 2004;135:21–27.
- Bilgic E, Kaneva P, Okrainec A, et al. Trends in the Fundamentals of Laparoscopic Surgery (FLS) certification exam over the past 9 years. *Surg Endosc*. 2018;32:2101–2105.
- Rogers-Vizena CR, Saldanha FYL, Hosmer AL, et al. A new paradigm in cleft lip procedural excellence: creation and preliminary digital validation of a lifelike simulator. *Plast Reconstr Surg*. 2018;142:1300–1304.
- Saldanha FYL, Loan GJ, Calabrese CE, et al. Incorporating cleft lip simulation into a “bootcamp-style” curriculum. *Ann Plast Surg*. 2021;86:210–216.
- Schneider WJ, Politis GD, Gosain AK, et al. Volunteers in plastic surgery guidelines for providing surgical care for children in the less developed world. *Plast Reconstr Surg*. 2011;127:2477–2486.
- C-SATS. Objective Structured Assessment of Technical Skills (OSATS). Johnson & Johnson Co. Available at <https://www.csats.com/osats/>. Accessed September 26, 2019.
- Rogers-Vizena CR, Sideridis GD, Patel KG, et al. A competency assessment tool for unilateral cleft lip repair. *Plast Reconstr Surg Glob Open*. 2020;8:e2954.
- Taylor HO, Morrison CS, Linden O, et al. Quantitative facial asymmetry: using three-dimensional photogrammetry to measure baseline facial surface symmetry. *J Craniofac Surg*. 2014;25:124–128.
- Linden OE, Taylor HO, Vasudavan S, et al. Three-dimensional analysis of nasal symmetry following primary correction of unilateral cleft lip nasal deformity. *Cleft Palate Craniofac J*. 2017;54:715–719.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
- Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45:255–268.
- Lin LI-K. A note on the concordance correlation coefficient. *Biometrics*. 2000;56:324–325.
- Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70:213–220.
- Cohen J. A power primer. *Psychol Bull*. 1992;112:155–159.
- Bishop ME, Ode GE, Hurwit DJ, et al. The arthroscopic surgery skill evaluation tool global rating scale is a valid and reliable adjunct measure of performance on a virtual reality simulator for hip arthroscopy. *Arthroscopy*. 2021;37:1856–1866.
- Beaudoin A, Larrivée S, McRae S, et al. Module-based arthroscopic knee simulator training improves technical skills in naive learners: a randomized trial. *Arthrosc Sports Med Rehabil*. 2021;3:e757–e764.
- VirtaMed. Better surgeons through mandatory simulator exam: “we will not take a step backwards.” VirtaMed, Zurich, Switzerland. Available at <https://www.virtamed.com/en/news/better-surgeons-through-mandatory-simulator-exam-we-will-not-take-step-backwards/>. Published March 30, 2017. Accessed October 8, 2021.
- Gosman A, Mann K, Reid CM, et al. Implementing assessment methods in plastic surgery. *Plast Reconstr Surg*. 2016;137:617e–623e.
- Podolsky DJ, Fisher DM, Wong KW, et al. Evaluation and implementation of a high-fidelity cleft palate simulator. *Plast Reconstr Surg*. 2017;139:85e–96e.
- Podolsky DJ, Wong Riff KW, Drake JM, et al. A high fidelity cleft lip simulator. *Plast Reconstr Surg Glob Open*. 2018;6:e1871.
- Global Technologies. Rhinoplasty surgical training simulator. Global Technologies, Davie, Fla. Available at <https://www.gtsimulators.com/products/rhinoplasty-surgical-training-simulator-ar361>. Accessed on October 8, 2021.
- Mentone Educational Centre. CANCEP carpal tunnel surgery simulator. Mentone Educational Centre, Victoria, Australia. Available at <https://www.mentone-educational.com.au/simulation/surgical-simulation/orthopaedic-surgery/upper-limb/procedural-simulators/cancep-carpal-tunnel-surgery-simulator>. Accessed on: October 8, 2021.
- Laufer S, Kempton SJ, Maciolek K, et al. A multi-layered needle injection simulator. *Stud Health Technol Inform*. 2016;220:205–208.
- Knox ADC, Gilardino MS, Kasten SJ, et al. Competency-based medical education for plastic surgery: where do we begin? *Plast Reconstr Surg*. 2014;133:702e–710e.
- Nguyen VT, Losee JE. Time- versus competency-based residency training. *Plast Reconstr Surg*. 2016;138:527–531.