Published in final edited form as:

IJSE Trans Healthc Syst Eng. 2018 ; 88(2): 110–116. doi:10.1080/24725579.2017.1418767.

Developing an objective assessment of surgical performance from operating room video and surgical imagery

Leah K. Taylor¹, Geb W. Thomas², Matthew D. Karam³, Clarence D. Kreiter⁴, and Donald D. Anderson⁵

¹Department of Orthopedics and Rehabilitation, Department of Biomedical Engineering, University of Iowa, Iowa City, IA, USA

²Department of Mechanical and Industrial Engineering, Department of Orthopedics and Rehabilitation, University of Iowa, Iowa City, IA, USA

³Department of Orthopedics and Rehabilitation, University of Iowa Hospitals and Clinics, Iowa City, IA, USA

⁴Department of Family Medicine, University of Iowa, Iowa City, IA, USA

⁵Department of Orthopedics and Rehabilitation, Department of Biomedical Engineering, Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, IA, USA

Abstract

An unbiased, repeatable process for assessing operating room performance is an important step toward quantifying the relationship between surgical training and performance. Hip fracture surgeries offer a promising first target in orthopedic trauma because they are common and they offer quantitative performance metrics that can be assessed from video recordings and intraoperative fluoroscopic images.

Hip fracture repair surgeries were recorded using a head-mounted point-of-view camera. Intraoperative fluoroscopic images were also saved. The following performance metrics were analyzed: duration of wire navigation, number of fluoroscopic images collected, degree of intervention by the surgeon's supervisor, and the tip-apex distance (TAD). Two orthopedic traumatologists graded surgical performance in each video independently using an Objective Structured Assessment of Technical Skill (OSATS).

Wire navigation duration correlated with weeks into residency and prior cases logged. TAD correlated with cases logged. There was no significant correlation between the OSATS total score and experience metrics. Total OSATS score correlated with duration and number of fluoroscopic images.

Our results indicate that two metrics of hip fracture wire navigation performance, duration and TAD, significantly differentiate surgical experience. The methods presented have the potential to provide truly objective assessment of resident technical performance in the OR.

Keywords

Objective assessment; surgical skills; performance metrics; wire navigation

Background

Medical training is in flux. The number of hours residents are allowed to work has decreased while concerns about medical errors have increased (Durkin, *et al*, 2008; Ahmed, *et al*, 2014). A lack of training can lead to errors in the operating room. The medical education community measures the competence of new surgeons by subjectively analyzing their educational and practical experience and, more recently, with simulation. The American Board of Orthopedic Surgery requires applicants for certification to have successfully completed an accredited residency in which their surgical skills have been appropriately certified by a Program Director. The American Board of Surgery additionally requires applicants for certification in general surgery to have satisfactorily completed a Fundamentals of Laparoscopic Surgery program, which relies upon a validated laparoscopic trainer to simulate and assess performance.

Evidence for the benefits of simulation in training technical skills is accumulating, and lab-based training is quickly finding its way into orthopedic residency programs (Karam, *et al*, 2012; Karam, *et al*, 2013). However, surgery performed in the operating room (OR) on a live patient is a unique experience, and there is no substitute for assessing performance in that environment. Furthermore, establishing skill transfer from a simulated setting to the OR is paramount in the development of simulator-based competency assessments (Mayne *et al*, 2016). Determining an unbiased, repeatable process for assessing surgical performance in the OR is a critical next step toward that goal.

Video recordings of surgery provide a platform for assessing surgical performance in the OR (Beard *et al*, 2005; Birkmeyer *et al*, 2013; Guerlain *et al*, 2004), but the community needs to decide how to assess OR performance from those recordings and what surgeries are best to record. The Objective Structured Assessment of Technical Skills (OSATS) is one widely used method for assessing surgical skill. A global rating scale is used to structure expert evaluation of task performance working from a list of operative competencies that are each rated on a 5-point Likert scale, anchored by behavioral descriptors. Recent studies have highlighted shortcomings in the objectivity of OSATS evaluations and in its ability to evaluate surgical outcome (Mayne *et al*, 2016; Anderson *et al*, 2016; Hopmans *et al*, 2014; Van Hove *et al*, 2010). Improved methods are needed to assess skills competency in the setting of direct patient care, especially considering that 63.5% of surgical errors are technical in nature (Fabri *et al*, 2008).

The surgical treatment of hip fractures offers a promising target for OR performance assessment because it is common (Mayne *et al*, 2016) and involves unbiased performance metrics that may be reliably quantified from video recordings and intraoperative fluoroscopic images (Thomas *et al*, 2015). The ability to navigate a wire within bone using fluoroscopy (wire navigation) is a challenging, generalizable skill, and a critical step in fixation of these fractures with devices such as the Dynamic Hip Screw (DHS) or Telescoping Screw Plate (TSP). The success of a DHS surgery depends upon a mechanically beneficial placement of the implant, reflected by the tip-apex distance (TAD) that estimates the distance from the tip of the implant to the apex of the femoral head (Baumgaertner *et al*, 1995). A TAD ≥ 25 mm significantly decreases the risk of implant cut-out from the bone

(Kumar *et al*, 2007). Adept OR performance of wire navigation involves balancing precision of wire placement with procedure duration and the amount of fluoroscopy used. The duration correlates with both cost and patient outcomes (Edwards *et al*, 2008), and adverse effects of radiation exposure are evident (Giordano *et al*, 2011; Hynes *et al*, 1992; Maxon *et al*, 1977).

A publication in *Journal of Surgical Education* (Taylor *et al*, 2016) introduced the grading methodology and assessed the reliability of the technique presented. The reliability assessment includes the TAD measure for seven cases and video metrics for two cases made by multiple graders; these cases are included in the fifteen presented here. The aforementioned publication did not include data related to the experience level of the surgeon.

The present study addresses the following question: Do measures of hip fracture wire navigation performance, assessed from intraoperative video and fluoroscopic images, correlate with surgical experience?

Methods

This prospective study included hip fractures (intertrochanteric and femoral neck) scheduled for open reduction and internal fixation using a DHS or TSP over a period of 20 months. The DHS construct uses a single large screw (Figure 1-A), and the TSP uses three smaller parallel screws (Figure 1-B) to stabilize the hip fracture. These devices were chosen because the placement of a guide wire in the proximal femur is a critical early step shared by both procedures, a task completed using the assistance of a drill guide and fluoroscopic visualization (Figure 1-C&D). The local IRB approved the protocol, and informed consent was obtained from patients and surgeons.

Data were collected for 18 cases, but three were subsequently excluded from analysis; one because it involved a very unstable fracture that confounded wire placement, another because the supervising surgeon was emergently called away during the wire navigation, and a third because the supervising surgeon was using the TSP for the first time, which substantially increased the duration of the surgery and added discussion time related to the unfamiliar implant. One staff surgeon and 13 different residents functioned as the operating surgeons (1 resident completed 2 procedures, both as a third-year resident). The number of weeks into residency and the number of DHS and TSP cases previously logged by each surgeon were used as indicators of surgical experience. The number of cases came from the resident's ACGME case log. The staff surgeon's number of logged cases (24) was taken from a billing log felt to most appropriately represent his considerable prior case experience.

Hip fracture cases were recorded in the OR using a GoPro® Hero3+Silver Edition (GoPro; San Mateo, CA) head-mounted point-of-view camera, and all intraoperative fluoroscopic images were automatically saved. A team member viewed and cropped the video to include only the wire navigation portion of the surgery. Data collected from the video were: duration of wire navigation (time from first fluoroscopic image with the guide wire to the final image

of the wire ready for implant), number of fluoroscopic images collected, and the degree of intervention by the surgeon's supervisor.

The degree of supervisory intervention metric was created as an indicator of resident skill and readiness for independent practice, but it is also unavoidably sensitive to the interaction style of each supervisor. Any OR performance assessment involving residents must account for intervention by the supervising surgeon that influences the progression of the procedure. The degree of intervention of the resident's supervisor was measured by tallying the weight of all interventions, where each intervention was given a weight of: 1 for supervisor instructions to someone other than the operating surgeon (e.g. requesting a fluoroscopic image from the radiology technician), 2 for supervisor instructions to the operating surgeon (e.g. instructions to alter the wire trajectory), and 3 for the supervisor handling an instrument in a capacity exceeding that of an assistant (e.g. physically controlling the wire guide: Figure 2). For trend analysis purposes, the supervisors were grouped into two categories; a senior-level resident or a staff surgeon.

The TAD was measured on the final anterior-posterior and lateral fluoroscopic images of the wire placement using OsiriX[®] software (OsiriX Foundation; Geneva, Switzerland) (Johnson *et al.*, 2008). In a prior investigation of the reliability of our measurement methods, we found that the inter-rater reliability of the TAD was 0.97, and that of the video metrics (duration, number of images, and supervisory intervention score) was 0.99 (Taylor *et al.*, 2016). The TAD measure was excluded for three cases where the supervising surgeon took control of the drill to place the guide wire in the final position. A supervisory intervention score was not assigned for the one case in which a staff surgeon acted as the operating surgeon.

To relate the objective metrics of performance assessment to expert perception of skill, two fellowship-trained traumatologists independently graded each video using a modified OSATS scoring system (Karam *et al.*, 2015). Intraoperative fluoroscopic images were inset into the videos at the point in the surgery where obtained, and the videos were viewed on a secure website. The OSATS scores reflected only the wire navigation portion of the surgery. The average of the total OSATS score (maximum 45 points) from both raters was used for each case.

Statistical Analysis

Duration, number of fluoroscopic images obtained, supervision intervention score, and TAD were each correlated with the experience metrics (weeks in residency and the number of previously logged cases). As data were not normally distributed, Spearman rank correlations were used to assess all relationships. For measures found to have significant correlation, the Mann-Whitney U test compared differences in novice vs. expert performance: novice (PGY 2 and 3, n=10) and expert (PGY 5 and staff surgeon, n=5). Cronbach's α was used to assess the inter-rater reliability of OSATS grading. Statistical significance was defined as $p < 0.05$, and p-values are reported to highlight important correlations. It is worth noting that, although corrections for multiple comparisons are important in strict experimental design, they were not conducted in this study. Statistical analysis was completed with SPSS Software (version 23, IBM, Armonk, NY). A composite score was computed by summing the average standardized values of the four performance metrics using SAS (version 9, SAS

Institute, Cary, NC). For the four cases where a TAD or supervision intervention score was not available, the composite score was computed by summing the average of the three remaining standardized performance values. Correlations between the composite score and experience were calculated using the same approach as for the individual performance metrics.

Results

Patient ages ranged from 55 to 97 years with a mean (standard deviation) of 79.3 (11.5) years. Patient BMI was 25.6 (6.9), and 67% of the patients were female. The participants included 13 residents, 10 male and 3 female, and one male staff surgeon. The residents ranged in experience from their second to fifth year in residency, with an average of 158 weeks (s.d. 50 weeks) into their residency at the time of the procedure. For the one participating staff member, 328 weeks had elapsed since the start of residency; that is, the staff surgeon was in his first year of practice as a staff surgeon.

The Appendix presents the experience and performance metrics for the analyzed cases. The number of previous cases logged for residents ranged from 1 to 13 with an average of 5.7 (4.1) cases. Table 1 presents the correlation coefficients relating performance to experience metrics. Wire navigation duration was significantly correlated with both weeks into residency -0.66 ($p < 0.01$) and prior cases logged -0.59 ($p = 0.02$) (Figure 3). The number of fluoroscopic images and the supervision intervention score did not correlate with either experience metric. TAD was significantly correlated with cases logged -0.67 ($p = 0.02$) but not weeks into residency (Figure 4). A Mann-Whitney U analysis indicated that the wire navigation duration for the novice group was significantly higher than for the experts $U=9$ ($p = 0.05$). There was no significant difference in the TAD between the two experience groupings. The composite performance metric significantly correlated to both weeks into residency -0.55 ($p = 0.03$) and cases logged -0.66 ($p = 0.01$) (Figure 5).

The inter-rater reliability of the two traumatologists' OSATS total scores was 0.71. There was no significant correlation between the OSATS total score and experience metrics (correlation coefficients: weeks into residency 0.43, $p = 0.11$ and cases logged 0.43, $p = 0.11$). Total OSATS score significantly correlated with duration -0.52 ($p = 0.05$) and number of fluoroscopic images -0.83 ($p < 0.001$), but correlated neither with supervision intervention score -0.27 ($p = 0.36$) nor TAD -0.30 ($p = 0.34$).

Discussion

The results from the evaluation of video recordings of 15 hip fracture cases performed by staff and residents indicate that two individual metrics of hip fracture wire navigation performance, duration and TAD, significantly differentiate surgical experience. A composite score integrating multiple individual performance metrics also provided strong correlations with surgical experience. Duration of the wire navigation had the strongest correlation to surgical experience, yielding significant negative correlation to both experience metrics. This is consistent with previous studies finding that greater surgical experience was associated with shorter operations (Birkmeyer *et al*, 2013; Leong *et al*, 2008). The TAD also exhibited a

significant negative correlation with the number of previous cases logged. This finding agrees with that of a simulated setting in which experienced surgeons obtained better wire placement than novices (Thomas *et al*, 2015).

The stronger correlation seen between performance and the number of previous cases logged, rather than to the point in residency, is consistent with previous findings that surgical skill is not directly related to duration of surgical practice, but to the number of surgeries performed. For relatively complex procedures such as wire navigation, morbidity and mortality have been shown to correlate with the number of surgeries performed (Birkmeyer *et al*, 2013). The lack of TAD correlation to point in residency could be due to an uneven distribution of experience levels (mostly PGY 3 residents). It should also be noted that all residents performed wire navigation with a TAD ≤ 25 mm, suggesting residents are meeting recommended placement accuracy.

There was a modest trend for more experience to correlate with fewer fluoroscopic images taken during the wire navigation phase of the surgery, consistent with prior work that showed an increase in surgical experience results in a decrease of radiation exposure (Giordano *et al*, 2011; Bar-On *et al*, 2010; Blattert *et al*, 2004; Palácio *et al*, 2014). The lack of statistical significance in our findings could be due to the high variability in the number of images used by the junior level residents (PGY 2 and 3); the number ranged from 19 to 79. We would expect to see a stronger correlation with an increase in sample size, specifically adding to the number of senior-level cases.

The supervision intervention score did not significantly correlate with any of the performance metrics, but negatively correlated with weeks into residency -0.49 ($p = 0.08$). This negative trend supports the hypothesis that as surgeon experience increases, the supervision intervention score decreases, most likely because the resident is more prepared for independent practice. Interestingly, although an increase in supervision intervention score did significantly predict increased duration, there was no corresponding improvement in the TAD.

Other scales used to assess supervisory intervention on surgical performance, such as the O-SCORE, rely upon the supervising surgeon to grade his/her intervention (Gofton *et al*, 2012; Miskovic *et al*, 2011). Such an approach may provide insight regarding the supervisor's role in helping residents gain technical proficiency, but directly involving the supervisor limits objectivity. We adopted a novel method of objectively quantifying involvement of the supervising surgeon from the point of view of an outsider, which shows potential as a discriminating factor of surgical skill.

The composite performance score correlated with weeks into residency more highly than did any individual performance metric except duration. The composite score correlated with previous cases logged with significance greater than any of the individual performance metrics. These findings suggest that the composite metric is the most reliable correlate of experience, and may be the most attractive target for establishing competency standards.

Although there were trends toward correlations between OSATS score and experience metrics, no correlations were significant. The total OSATS score correlated significantly

with duration and number of fluoroscopic images, but not to TAD. As the clinical importance of TAD is well established, the inability of OSATS to distinguish wire placement performance is concerning. This failure of OSATS to correlate with a surgical outcome measure is consistent with work that has shown that OSATS fails to correlate to radiographic assessment of reduction (Mayne *et al*, 2016) and quality of fixation (Anderson *et al*, 2016). The inability of OSATS to accurately measure the key surgical outcome (TAD) suggests that new measures of performance, such as those introduced in this study, are needed to capture skills that affect outcomes.

The main limitation of this study is that a relatively small number of cases were analyzed, all from a single institution. We are currently working towards extending the use of this performance assessment approach to additional institutions. Variation in teaching habits between supervising surgeons (e.g., proclivity toward intervention) was not accounted for in our analysis because of the small sample size. Only one resident acted as the operating surgeon for multiple cases (two), one at 126 weeks into residency and one at 132 weeks. Including cases for individual residents at multiple time points in residency could provide insight into individual learning curves. The wire navigation portion of hip fracture surgery is most often performed by a PGY 3 resident at the University of Iowa; therefore, the range of experience level is limited. The performance of the one staff surgeon had a disproportionate impact on the findings, but we included these data because the benefit of including an additional case outweighed the risk of disrupting outliers, and the scores associated with the staff surgeon demonstrate the effect of surgical experience on performance. Inclusion of more experienced surgeons, specifically practicing orthopedic surgeons, would strengthen the study by more definitively establishing expert metric standards.

During the course of this study, the hospital introduced the TSP implant. This is not expected to affect our results, however, since the study focuses on a procedure common to both operations. As the two implants included were not identical, differences in performance metric scores due to implant type were possible. However, we saw no indication of implant bias impacting performance. The two traumatologists who rated the videos with OSATS were among the attending surgeons on the cases and were familiar with the residents, so a truly blinded grading was not possible. As the purpose of including the OSATS was merely to compare the presented metrics with the gold standard of surgical skill evaluation, we did not feel a completely blinded assessment was essential. The value of the experience metric of previously logged cases is dependent upon diligence of the residents reporting case involvement. No modifications were made for the variance in fracture severity and degree of displacement of the cases. For this reason, performance metrics did not account for case difficulty.

Conclusions

Our results indicate that two individual metrics of hip fracture wire navigation performance, duration and TAD, significantly differentiate surgical experience. A composite score incorporating multiple performance metrics also provided strong correlations with surgical experience. The methods presented have the potential to provide truly objective assessment of technical performance in the OR, a critical step towards ensuring surgeon competency.

Acknowledgments

This work was funding in part with grants from The American Board of Orthopaedic Surgery and the Agency for Healthcare Research and Quality under award number R18HS022077.

We would also like to thank Andres Alfaro for his help in editing and preparing the article.

References

1. Ahmed N, Devitt KS, Keshet I, Spicer J, Imrie K, Feldman L, Cools-Lartigue J, Kayssi A, Lipsman N, Elmi M, Kulkarni AV, Parshuram C, Mainprize T, Warren RJ, Fata P, Gorman S, Feinberg S, Rutka J. A systematic review of the effects of resident duty hour restrictions in surgery: impact on resident wellness, training, and patient outcomes. *Annals of Surgery*. 2014; 259(6):1041–1053. [PubMed: 24662409]
2. Anderson DD, Long S, Thomas GW, Putnam MD, Bechtold JE, Karam MD. Objective structured assessments of technical skills (OSATS) does not assess the quality of the surgical result effectively. *Clinical Orthopaedics and Related Research*. 2016; 474:874–881.
3. Bar-On E, Weigl DM, Becker T, Katz K, Konen O. Intraoperative C-arm radiation affecting factors and reduction by an intervention program. *Journal of Pediatric Orthopaedics*. 2010; 30:320–323. [PubMed: 20502229]
4. Baumgaertner MR, Curtin SL, Lindskog DM, Keggi JM. The value of the tip-apex distance in predicting failure of fixation of peritrochanteric fractures of the hip. *Journal of Bone and Joint Surgery*. 1995; 77:1058–1064. [PubMed: 7608228]
5. Beard JD, Jolly BC, Newble DI, Thomas WE, Donnelly J, Southgate LJ. Assessing the technical skills of surgical trainees. *British Journal of Surgery*. 2005; 92:778–782. [PubMed: 15810048]
6. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, Dimick J, Banerjee M, Birkmeyer NJO. Surgical skill and complication rates after bariatric surgery. *New England Journal of Medicine*. 2013; 369:1434–1442. [PubMed: 24106936]
7. Blattert TR, Fill UA, Kunz E, Panzer W, Weckbach A, Regulla DF. Skill dependence of radiation exposure for the orthopaedic surgeon during interlocking nailing of long-bone shaft fractures: a clinical study. *Archives of Orthopaedic & Trauma Surgery*. 2004; 124:659–664. [PubMed: 15365718]
8. Durkin ET, McDonald R, Munoz A, Mahvi D. The impact of work hour restrictions on surgical resident education. *Journal of Surgical Education*. 2008; 65(1):54–60. [PubMed: 18308282]
9. Edwards C, Counsell A, Boulton C, Moran CG. Early infection after hip fracture surgery: risk factors, costs, and outcome. *Journal of Bone and Joint Surgery*. 2008; 90-B:770–777.
10. Fabri PJ, Zayas-Castro JL. Human error, not communication and systems, underlies surgical complications. *Surgery*. 2008; 144:557–565. [PubMed: 18847639]
11. Giordano BD, Grauer JN, Miller CP, Morgan TL, Rehtine GR. Radiation exposure issues in orthopaedics. *Journal of Bone and Joint Surgery*. 2011; 93:e69. [PubMed: 21498490]
12. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa surgical competency operating room evaluation (O-SCORE): a tool to assess surgical competence. *Academic Medicine*. 2012; 87:1401–1407. [PubMed: 22914526]
13. Guerlain S, Turrentine B, Adams R, Calland JF. Using video data for the analysis and training of medical personnel. *Cognition, Technology & Work*. 2004; 6:131–138.
14. Hopmans CJ, Hoed PTd, Laan Lvd, Harst Ecd, Elst Mvd, Mannaerts GHH, Dawson I, Timman R, Wijnhoven BPL, Ijermans JNM. Assessment of surgery residents' operative skills in the operating theater using a modified objective structured assessment of technical skills (OSATS): a prospective multiocenter study. *Surgery*. 2014; 156:11.
15. Hynes DE, Conere T, Mee MB, Cashman WF. Ionising radiation and the orthopaedic surgeon. *Journal of Bone and Joint Surgery*. 1992; 74:332–334. [PubMed: 1587871]
16. Johnson LJ, Cope MR, Shahrokhi S, Tamblin P. Measuring tip–apex distance using a picture archiving and communication system (PACS). *Injury*. 2008; 39:786–790. [PubMed: 18539280]

17. Karam MD, Kho JY, Yehyawi TM, Ohrt GT, Thomas GW, Jonard B, Anderson DD, Marsh JL. Application of surgical skill simulation training and assessment in orthopaedic trauma. *Iowa Orthopaedic Journal*. 2012; 32:76–82. [PubMed: 23576925]
18. Karam MD, Thomas GW, Koehler DM, Westerlind BO, Lafferty PM, Ohrt GT, Marsh JL, Van Heest AE, Anderson DD. Surgical coaching from head-mounted video in the training of fluoroscopically guided articular fracture surgery. *Journal of Bone and Joint Surgery*. 2015; 97:1031–1039. [PubMed: 26085538]
19. Karam MD, Westerlind B, Anderson DD, Marsh JL. Development of an orthopaedic surgical skills curriculum for post-graduate year one resident learners - the University of Iowa experience. *Iowa Orthopaedic Journal*. 2013; 33:178–184. [PubMed: 24027480]
20. Kumar AJ, Parmar VN, Kolpattil S, Humad S, Williams SC, Harper WM. Significance of hip rotation on measurement of 'tip apex distance' during fixation of extracapsular proximal femoral fractures. *Injury*. 2007; 38:792–796. [PubMed: 17296198]
21. Leong JJH, Leff DR, Das A, Aggarwal R, Reilly P, Atkinson HDE, Emery RJ, Darzi AW. Validation of orthopaedic bench models for trauma surgery. *Journal of Bone and Joint Surgery*. 2008; 90-B:958–965.
22. Maxon HR, Thomas SR, Saenger EL, Buncher CR, Kereiakes JG. Ionizing irradiation and the induction of clinically significant disease in the human thyroid gland. *The American Journal of Medicine*. 1977; 63:967–978. [PubMed: 343589]
23. Mayne IP, Brydges R, Moktar J, Murnaghan ML. Development and assessment of a distal radial fracture model as a clinical teaching tool. *Journal of Bone and Joint Surgery*. 2016; 98:410–416. [PubMed: 26935464]
24. Miskovic D, Wyles SM, Carter F, Coleman MG, Hanna GB. Development, validation and implementation of a monitoring tool for training in laparoscopic colorectal surgery in the English National Training Program. *Surgical Endoscopy*. 2011; 25:1136–1142. [PubMed: 20835723]
25. Palácio EP, Ribeiro AA, Gavassi BM, Di Stasi GG, Galbiatti JA, Junior AD, Mizobuchi RR. Exposure of the surgical team to ionizing radiation during orthopedic surgical procedures. *Revista Brasileira de Ortopedia*. 2014; 49:227–232. [PubMed: 26229805]
26. Taylor L, Thomas GW, Karam MD, Kreiter CD, Anderson DD. Assessing wire navigation performance in the operating room. *Journal of Surgical Education*. 2016; 73(5):780–787. [PubMed: 27184177]
27. Thomas GW, Johns BD, Kho JY, Anderson DD. The validity and reliability of a hybrid reality simulator for wire navigation in orthopedic surgery. *IEEE Transactions on Human-Machine Systems*. 2015; 45:119–125.
28. Van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J. Objective assessment of technical surgical skills. *British Journal of Surgery*. 2010; 97:972–987. [PubMed: 20632260]

Appendix

Table A1

Experience and performance metrics for 15 hip fracture cases listed in chronological order. Case 8 was completed by a staff surgeon (therefore no supervisor or supervision intervention score is shown). For cases 1, 4, and 10 the supervising surgeon placed the final guide pin. Cases 10 and 13 were completed by the same resident.

			<u>Experience Metrics</u>			<u>Performance Metrics</u>			
<u>Case</u>	<u>Implant</u>	<u>Supervisor</u>	<u>Weeks into residency</u>	<u>#DHS Logged</u>	<u>Duration (min)</u>	<u>Fluoro Images</u>	<u>Supervision Intervention</u>	<u>TAD (mm)</u>	<u>Composite Performance</u>
1	DHS	Resident	109	5	14.7	40	107	*	0.311
2	DHS	Resident	114	3	20.3	44	54	17	0.379
3	DHS	Staff	227	10	12.4	58	7	17	−0.123

			<u>Experience Metrics</u>			<u>Performance Metrics</u>			
<u>Case</u>	<u>Implant</u>	<u>Supervisor</u>	<u>Weeks into residency</u>	<u>#DHS Logged</u>	<u>Duration (min)</u>	<u>Fluoro Images</u>	<u>Supervision Intervention</u>	<u>TAD (mm)</u>	<u>Composite Performance</u>
4	DHS	Resident	135	5	9.5	19	27	*	-1.083
5	DHS	Staff	146	9	9.1	25	6	9	-1.194
6	DHS	Staff	150	1	16.8	26	43	23	0.136
7	DHS	Resident	103	1	19.3	68	65	17	0.739
8	TSP	*	328	24	7.1	37	*	10	-0.987
9	TSP	Staff	226	12	11.3	32	45	12	-0.608
10	DHS	Resident	126	2	17.8	55	142	*	1.100
11	TSP	Staff	232	7	9.7	45	43	25	0.104
12	DHS	Resident	132	1	11.5	37	127	19	0.255
13	TSP	Staff	132	3	13.0	79	107	20	0.908
14	TSP	Staff	242	13	11.1	27	65	12	-0.587
15	TSP	Resident	137	8	20.1	41	137	11	0.483

* Indicates field not applicable to case

Figure 1.

Postoperative radiographs of the two implants: (A) DHS and (B) TSP. Intraoperative images showing the guide with guide wire placed on the lateral cortex of the femur for the (C) DHS and (D) TSP. The semi-transparent red arrows in (C) and (D) denote the insertion point of the guide wire.

Figure 2.

Examples of two supervision intervention behaviors. (A) The supervisor (hands indicated with stars) is handling the Bennett retractors in an assisting manner (elevation and retraction), and no supervision impact is counted. (B) The supervisor (on the left, hand indicated with a star) is taking control of the guide, which constitutes handling an instrument and is tallied with a weight of 3.

A **B****Figure 3.**

Plots of *duration* vs. (A) weeks into residency and (B) number of previous cases logged. Results of linear regression are shown as a general indicator of the relationship. The “*” indicates the p-value is significant ($p<0.05$).

Figure 4.

Plots of *TAD* vs. (A) weeks into residency and (B) number of previous cases logged. Results of linear regression are shown as a general indicator of the relationship. The “*” indicates the p-value is significant ($p < 0.05$).

Figure 5. Plots of the *composite performance metric* vs. (A) weeks into residency and (B) number of previous cases logged. Results of linear regression are shown as a general indicator of the relationship. The “*” indicates the p-value is significant ($p < 0.05$).

Table 1

Correlation coefficients relating performance metrics to surgeon experience. Significant correlations ($p < 0.05$) are bolded.

		Experience Metrics	
Performance Metrics		Weeks into residency	Cases Logged
Duration (min)	Correlation Coefficient	-.661	-.587
	Significance, p-value	0.007	0.021
	n	15	15
Fluoroscopic Images	Correlation Coefficient	-0.34	-0.268
	Significance, p-value	0.216	0.335
	n	15	15
Supervision Intervention	Correlation Coefficient	-0.485	-0.346
	Significance, p-value	0.079	0.226
	n	14	14
TAD (mm)	Correlation Coefficient	-0.201	-.669
	Significance, p-value	0.531	0.017
	n	12	12
Composite Performance Metric	Correlation Coefficient	-0.549	-0.656
	Significance, p-value	0.034	0.008
	n	15	15
OSATS	Correlation Coefficient	0.044	0.092
	Significance, p-value	0.886	0.765
	n	13	13