Evaluation of Three Sources of Validity Evidence for a Synthetic Thoracoscopic Esophageal Atresia/Tracheoesophageal Fistula Repair Simulator

Katherine A. Barsness, MD,[1,2] Deborah M. Rooney, PhD,[3] Lauren M. Davis, BA,[4] and Ellie O'Brien, BS[4]

Abstract

Purpose: Thoracoscopic esophageal atresia (EA)/tracheoesophageal fistula (TEF) repair is technically challenging. We have previously reported our experiences with a high-fidelity hybrid model for simulation-based educational instruction in thoracoscopic EA/TEF, including the high cost of the tissue for these models. The purposes of this study were (1) to create a low-cost synthetic tissue EA/TEF repair simulation model and (2) to evaluate the content validity of the synthetic tissue simulator.

Materials and Methods: Review of the literature and computed tomography images were used to create computer-aided drawings (CAD) for a synthetic, size-appropriate EA/TEF tissue insert. The inverse of the CAD image was then printed in six different sections to create a mold that could be filled with platinum-cured silicone. The silicone EA/TEF insert was then placed in a previously described neonatal thorax and covered with synthetic skin. Following institutional review board–exempt determination, 47 participants performed some or all of a simulated thoracoscopic EA/TEF during two separate international meetings (International Pediatric Endosurgery Group [IPEG] and World Federation of Associations of Pediatric Surgeons [WOFAPS]). Participants were identified as ''experts,'' having 6–50 self-reported thoracoscopic EA/TEF repairs, and ''novice,'' having 0–5 self-reported thoracoscopic EA/TEF repairs. Participants completed a self-report, six-domain, 24-item instrument consisting of 23 5-point rating scales and one 4-point Global Rating Scale. Validity evidence relevant to test content and response processes was evaluated using the many-facet Rasch model, and evidence of internal structure (interitem consistency) was estimated using Cronbach's alpha.

Results: A review of the participants' ratings indicates there were no overall differences across sites (IPEG versus WOFAPS, $P = .84$) or experience (expert versus novice, $P = .17$). The highest observed averages were 4.4 (Value of Simulator as a Training Tool), 4.3 (Physical Attributes—chest circumference, chest depth, and intercostal space), and 4.3 (Realism of Experience—fistula location). The lowest observed averages were 3.5 (Ability to Perform—closure of fistula), 3.7 (Ability to Perform—acquisition target trocar sites), 3.8 (Physical Attributes—landmark visualization), 3.8 (Ability to Perform—anastomosis and dissection of upper pouch), and 3.9 (Realism of Materials—skin). The Global Rating Scale was 2.9, coinciding with a response of ''this simulator can be considered for use in neonatal TEF repair training, but could be improved slightly.'' Material costs for the synthetic EA/TEF inserts were less than $2 U.S. per insert.

Conclusions: We have successfully created a low-cost synthetic EA/TEF tissue insert for use in a neonatal thoracoscopic EA/TEF repair simulator. Analysis of the participants' ratings of the synthetic EA/TEF simulation model indicates that it has value and can be used to train pediatric surgeons, especially those early in their learning curve, to begin to perform a thoracoscopic EA/TEF repair. Areas for model improvement were identified, and these areas will be the focus for future modifications to the synthetic EA/TEF repair simulator.

[1]Division of Pediatric Surgery, Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois.
[2]Departments of Surgery and Medical Education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.
[3]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, Michigan.
[4]Innovations Laboratory, Northwestern Simulation, Center for Education in Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

Background

W<small>E HAVE PREVIOUSLY PRESENTED</small> validity evidence in support of using a novel thoracoscopic esophageal atresia (EA)/tracheoesophageal fistula (TEF) repair simulator as an educational tool for pediatric surgical education.[1,2] The most significant limitation to widespread adoption of the simulator is the cost and need for preparation of the fetal bovine tissue. Additionally, animal tissue is not allowed in the majority of hospital-based simulation laboratories because of the risk of cross-contamination of human surgical instruments and potential infectious diseases. Therefore, we sought (1) to create a low-cost synthetic tissue insert for thoracoscopic EA/TEF repair and (2) to examine validity evidence to support or refute its use in pediatric surgical education.

Recently, Cook et al.[3] promoted the use of the Standards framework as best practice during evaluation of validity evidence of simulation-based assessment. In the same work, they also encouraged medical education researchers to broaden the sources of validity evidence examined to include evidence relevant to response processes. Using methods used in previous work,[1,2,4] we applied the Standards framework to evaluate validity evidence relevant to three sources—test content, internal structure, and response processes.

Materials and Methods

Study setting

After review and exempt determination by the Ann and Robert H. Lurie Children's Hospital of Chicago Institutional Review Board, data were collected during advanced minimally invasive skills courses offered at the 23rd Annual International Pediatric Endosurgery Group (IPEG) and the Fourth World Federation of Associations of Pediatric Surgeons (WOFAPS) meetings. In total, 47 pediatric surgeons contributed to this study: 29 participants from IPEG and 18 participants from WOFAPS. Participants were categorized as "experienced" or "novice," based on self-reported experience with thoracoscopic EA/TEF repair. Fourteen surgeons were identified as "experienced," having a mean of 27 (range, 6–50) self-reported thoracoscopic EA/TEF repairs. Thirty surgeons were identified as "novice," having a mean of 0.5 (range, 0–5) self-reported thoracoscopic EA/TEF repairs. Three participants did not report prior experience with thoracoscopic EA/TEF repair.

Simulator

As previously described,[1,2,5] the external surround of the EA/TEF repair simulator was assembled using a term neonatal rib cage (right side only), a stabilizing base, and synthetic skin overlay. The simulator was completed with a platinum-cured silicone insert modeled into the appearance of a classic C-type EA/TEF (proximal EA with distal TEF). In order to create the silicone insert, computer-aided design (CAD) models of the EA and TEF were first designed in SolidWorks. Dimensions for the models were consistent with that of a 50th percent term neonate. The inverse of these models was used to create molds for both parts. The CAD images of the molds were exported and three-dimensionally printed. The external molds were printed on a powder printer. Both molds contained an inner portion designed to make the
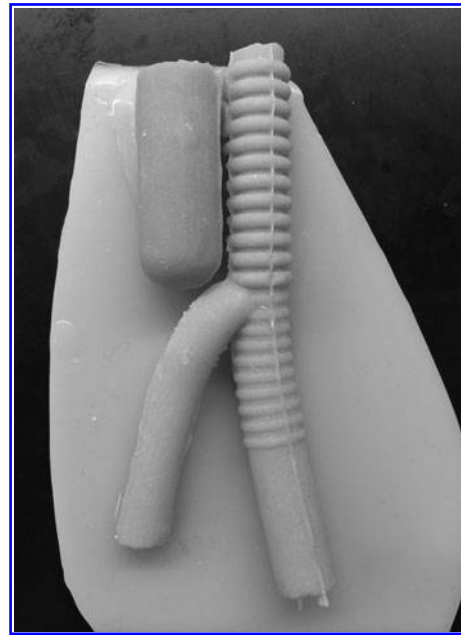
FIG. 1. Silicone insert for a synthetic thoracoscopic esophageal atresia/tracheoesophageal fistula repair simulator.

structures luminal. These parts were printed in a digitally blended material. One ounce of platinum-cured silicone was tinted a red color, poured into the molds, and cured for 2 hours. One ounce of platinum-cured silicone was tinted a flesh color and poured into a flat sheet to create the base of the cartridge. Upon curing, the EA and TEF were removed from the molds and secured to the cartridge base using platinum-cured silicone gel (Fig. 1). Modifications of the originally described simulator included a replaceable rib insert for ribs 3–7, made out of two digitally blended materials allowing for flexibility in the ribcage and palpation (Fig. 2). Participants
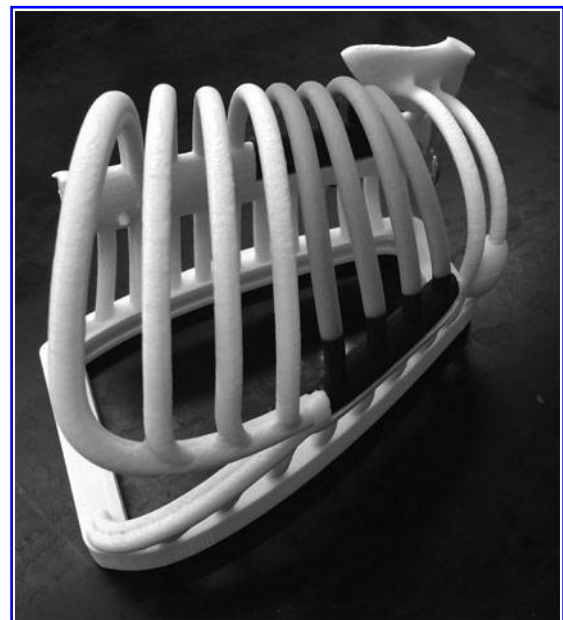
FIG. 2. Three-dimensionally printed, composite construction, right-sided ribcage for a full-term neonate.

were provided with 3-mm instruments and a 4-mm telescope (Karl Storz Endoscopy-America, Segundo, CA).

Measures and rating procedures

All participants completed a self-report survey following their experience with the simulator. The 24-item survey consisted of 23 five-point rating scales measuring the simulator's quality across six domains (Physical Attributes, Realism of Materials, Realism of Experience, Ability to Perform Task, Value, and Relevance) and 1 4-point Global Rating Scale to measure participants' overall impression of the simulator.

Analyses

In order to evaluate validity evidence, we used the Standards for Educational and Psychological Testing (the Standards), the guide developed jointly by the American Education Research Association, American Psychological Association, and the National Council on Measurement in Education.[6] The current Standards framework identified five different sources of validity evidence: (a) test content, (b) internal structure, (c) response processes, (d) relationships to other variables, and (e) consequences of testing. We used this work to evaluate three sources of validity evidence: test content, response processes, and internal structure.

To analyze validity evidence from these sources, we used methods from both modern measurement and classical test theories. Similar to methods used in previous work to evaluate evidence of test content, we used a many-facet Rasch model to analyze three Rasch indices: observed averages, point-measure correlation, and Rasch item-fit indices.[7] To evaluate validity evidence relevant to response processes, we examined Rasch fit statistics to identify rating differences across participant sites and level of experience. Analyses of the self-report survey measures were performed using Facets software version 3.68 (2011; Linacre software, Beaverton, OR). To evaluate evidence relevant to internal structure, we estimated interitem consistency using Cronbach's alpha. Statistical analysis was performed using IBM SPSS statistical software (version 22.0; IBM Corp., Armonk, NY).

Results

Evidence relevant to test content (Table 1)

Observed averages. For the survey items, the overall observed average for all participants was 4.0. In descending order, the combined observed averages of the six domains

TABLE 1. OBSERVED AVERAGES, ITEM OUTFIT STATISTICS, AND POINT-MEASURE CORRELATIONS ACROSS 7 DOMAINS AND 24 ITEMS

Domain	Item	Observed average	Outfit MS	Point-measure correlation
Physical attributes		4.1[a]		
1	Chest circumference	4.3	1.03	.45
2	Chest depth	4.3	1.05	.50
3	Intercostal space	4.3	1.01	.44
4	Landmark—tactile	3.8	1.27	.44
5	Landmark—visual	4.0	.92	.54
6	Scale of tissue	4.2	1.13	.44
Realism of materials		4.1[a]		
7	Overall impression	4.0	1.02	.47
8	Skin	3.9	1.18	.48
9	Ribs	4.1	1.19	.43
10	Tissue	4.1	1.02	.50
Realism of experience		4.2[a]		
11	Chest wall resistance	4.1	.59	.60
12	Anatomy	4.1	1.56	.39
13	Location of fistula	4.3	1.40	.42
14	Upper pouch anatomy	4.0	1.74	.41
15	Expected experience	3.8	1.39	.43
Ability to perform task		3.6[a]		
16	Trocar locations	3.7	.98	.41
17	Place trocars	3.9	1.06	.36
18	Closure of fistula	3.5	1.35	.45
19	Dissection upper pouch	3.8	2.14	.25
20	Repair atresia	3.2	1.80	.41
Value		4.1[a]		
21	As a training tool	4.4	1.19	.36
22	As a testing tool	4.1	1.02	.49
Relevance				
23	To practice	4.1	1.42	.29
Global				
24	Overall	2.9	1.01	.50

[a]Observed average of each domain.
MS, mean square.

were 4.2 (Realism of Experience), 4.1 (Relevance to Practice), 4.1 (Value), 4.1 (Physical Attributes), 4.1 (Realism of Materials), and 3.6 (Ability to Perform Task). Closer examination indicated the highest-rated items from the survey were Value of the Simulator as a Training Tool (4.4), Physical Attributes—chest circumference, chest depth, and intercostal space (4.3), and Realism of Experience—realism of location of ''fistula'' (4.3), whereas the lowest ratings were associated with Ability to Perform—closure of fistula (3.2) and Ability to Perform—anastomosis (3.5). The observed average of the opinion Global Rating Scale was 2.9 (out of 4.0), indicating that, on average, participants believed the synthetic thoracoscopic EA/TEF repair simulator could be considered for training, but could be improved slightly.

Point-measure correlations. For the survey, all of the 24 items had positive point-measure correlations (range, 0.29–0.62). For the purpose of this work, positive point-measure correlations offer evidence the raters' scores align with their observations, so that we can make reasonable inferences about the quality of the simulator.

Rasch item-fit indices. For all but three items, the item Outfit mean square (MS) values fell between 0.5 and 1.5, suggesting a reasonable amount of variability in responses. For items 5 (Physical Attributes—scale of tissue), 14 (Realism of Experience—upper pouch), and 19 (Ability to Perform Task—dissection of upper pouch), item Outfit MS values were 1.80, 1.74, and 2.14, respectively.

Evidence relevant to response processes

There were no overall differences in observed averages between sites (IPEG versus WOFAPS, $P = .84$) or self-reported experience with thoracoscopic EA/TEF repair (experienced versus novice, $P = .17$). These findings indicate no overall rating differences across these participant groups and support evidence relevant to response processes. A closer examination of site fit statistics indicated high agreement in the IPEG participants (Outfit MS, 0.94), with decreased agreement across the WOFAPS participants (Outfit MS, 1.71). Rasch person-fit statistic identifies responses of raters who may be inconsistent or unexpected, and inconsistencies may highlight problematic response patterns such as carelessness or item bias that can interfere with the measurement of the construct that is intended. Closer examination of participants with an Outfit MS over 1.5 indicated that six of the seven are associated with the WOFAPS site.

Evidence relevant to internal structure

Interitem consistency of the 19 items relevant to simulator quality (items 1–15 and 21–24) was estimated to be high ($\alpha = 0.89$). Interitem consistency of the five items relevant to participants' ability to perform the critical tasks using the simulator (items 16–20) was also estimated to be high ($\alpha = 0.88$). This index offers a measure of control and, when adequately high, indicates these assessment items are grouped appropriately and measure the same general construct. This allows us to make inferences from our findings with a high degree of confidence and offers evidence of internal structure.

Simulator costs

The base rib cage and rib 3–7 insert materials costs were approximately $200 U.S. The cost to three-dimensionally print the EA and TEF molds was approximately $100 U.S., with the final silicone cartridge costing $1.44 U.S.

Discussion

As the number of simulation-based educational tools continues to grow, so too does the cost of simulation. A recent cost analysis after implementation of a tissue-based American College of Surgeons/Association of Program Directors in Surgery surgical skills curriculum for general surgery residents determined that the annual operation cost for the 35-module curriculum was more than $110,000 U.S., or $3150 U.S. per training resident.[8] Although real tissues may be critical for advanced skills, surgical learners early in the learning curve may only need to focus on specific tasks that are amenable to low-cost synthetic materials. For example, a learner with poor intracorporeal suturing skills hardly needs real tissue to refine the motions of his or her hands and the instruments. Pediatric surgical training in the United States is notable for occurring after the completion of general surgery residency. Although the majority of the trainees have the necessary skills for laparoscopy in adults, it is often the limitations of the constrained spaces that are the initial challenges in converting from adult, to pediatric, to infant laparoscopy. It is in this setting that we sought to create a low-cost synthetic EA/TEF simulation model for use early in the learning curve of thoracoscopic EA/TEF repair and to evaluate three sources of validity evidence to support or refute its use in pediatric surgical education.

The highest rated items from the survey were Value of the Simulator as a Training Tool (4.4), Physical Attributes—chest circumference, chest depth, and intercostal space (4.3), and Realism of Experience—realism of location of ''fistula'' (4.3). The same items were the highest rated items for our previous work. Yet, the ratings are lower than the ratings of the real tissue EA/TEF repair simulator (Value, 4.8; Physical Attributes, 4.7; Realism of fistula location, 4.7). It is interesting that the lower scores were on a newly refined rib cage that incorporated better flexibility of the ribs (including the option for an open procedure), improved scapular placement, and other structural changes. However, we did not change the dimensions of the chest; therefore the circumference, depth, and intercostal spacing were identical. Given that these items did not change structurally, but the ratings decreased, we attribute the lower ratings to different participant pools. In fact, the participants in this work come from two different international pediatric meetings, compared with a single U.S. pediatric surgery meeting.

The lowest ratings were associated with Ability to Perform—closure of fistula (3.2) and Ability to Perform—completion of the anastomosis (3.5). In the setting of courses that are geared specifically toward pediatric surgeons with little or no minimally invasive experience, these results are expected. With a mean number of thoracoscopic EA/TEF repairs of less than one (median, zero), the simulated operation was an entirely new experience for the majority of the course participants. Additionally, these specific tasks are some of the most challenging to perform in a minimally invasive setting. Yet, despite the low ratings, the overall global opinion rating for the synthetic EA/TEF repair simulator was

consistent with ''could be considered for training, but could be improved slightly.'' Of the pooled data, three of the Outfit MS results were high, indicating that there was more than expected variation in the ratings. These areas, including the anatomy/dissection of the upper pouch and performance of the anastomosis, will be the focus of subsequent structural modifications to the model. It is interesting that the highest individual Outfit MS results were identified for the WOFAPS participants. These results likely relate to the marked variability of baseline minimally invasive surgery skills noted in that particular participant group. Because WOFAPS is not as focused as heavily on minimally invasive surgery as is IPEG, the participants had different expectations of an advanced course than did the IPEG participants.

For the purpose of this study, high observed averages from the survey suggest high perceived value for the simulator, whereas positive point-measure correlations and reasonable item fit indices attest to the ''psychometric soundness'' of the survey's items. These findings support the assumption that participants' ratings reflect the intended concepts—perceived value of the simulator and quality of performance during an EA/TEF repair—and higher perceived value aligns with higher scores. Additional indices, the Rasch mean-squared item fit statistics, are used to identify problematic measurement conditions, such as multidimensionality and poorly written items, by indicating discrepancies between observed scores from expected values. A negative unweighted mean-squared fit index indicates less variability than the Rasch model predicted, whereas a positive value suggests more variability than expected,[9] and values over 1.5 indicate concerning amount of variability that may degrade the quality of measures.[10] Although findings do support evidence of content validity, the higher Outfit MS indices for the three items do indicated decreased agreement for those particular items and should be used as a guide for simulator refinement.

The single most expensive item of the simulator is the three-dimensionally printed rib cage, at approximately $200 U.S. each. However, the same rib cage can be used multiple times (at least 20–30) without structural damage. The cost of the synthetic tissue is far lower ($<$2 U.S. per insert) than the cost of fetal bovine tissue ($90 U.S. per tissue block). In addition to the cost advantage, the synthetic tissue is easily transportable throughout the world, can be used within a hospital-based simulation laboratory, and does not require any cold storage or tissue prep time to use. In a comprehensive, proficiency-based curriculum for thoracoscopic EA/TEF repair, the synthetic tissue would be ideal for novice learners working to master the critical technical skills. Once the learner has demonstrated competence with the skills, then he or she would advance to comprehensive cognitive, technical, and nontechnical skills training using the real tissue simulator in an immersive operating room environment.

There are several limitations related to the interpretation and applications of the findings presented in this study. First, the novice study participants were a heterogeneous group of pediatric surgeons with variable training and clinical practice experiences, representing numerous different countries around the world. Additionally, the two study recruitment locations are expected to increase the heterogeneity given that one meeting (IPEG) was heavily weighted toward minimally invasive surgical techniques, whereas the other meeting (WOFAPS) would appeal to pediatric surgeons with

all levels of interest, or lack therein, for minimally invasive surgical techniques. As such, the item results with higher or lower variability than expected likely are a direct reflection of highly diverse interests, experiences, and skills among the study participants. Next, because of course time constraints, many of the novice participants were unable to complete all tasks on the simulator, limiting the novice participants' ability to fully evaluate the model within the Ability to Perform domain. Finally, this article only begins to address the full content validity of measures of performance for a simulated thoracoscopic EA/TEF repair. Additional measures of validity evidence will need to be collected.

In conclusion, we have successfully created a low-cost, realistic, and valuable synthetic thoracoscopic EA/TEF repair simulation model. Initial validity evidence relevant to test content, response processes, and internal structure supports further structural refinement, but the current version could be used in pediatric surgical education. Additional validity evidence will be collected from a refined model.

Acknowledgments

Disclosure Statement

No competing financial interests exist.

References

1. Barsness KA, Rooney DM, Davis LM. Collaboration in simulation: The development and initial validation of a novel thoracoscopic neonatal simulator. J Pediatr Surg 2013;48:1232–1238.
2. Barsness KA, et al. Validation of measures from a thoracoscopic esophageal atresia/tracheoesophageal fistula repair simulator. J Pediatr Surg 2014;49:29–32; discussion 32–33.
3. Cook DA, et al. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. Adv Health Sci Educ Theory Pract 2014;19:233–250.
4. Barsness KA, Rooney DM, Davis LM. The development and evaluation of a novel thoracoscopic diaphragmatic hernia repair simulator. J Laparoendosc Adv Surg Tech A 2013;23:714–718.
5. Davis LM, Barsness KA, Rooney DM. Design and development of a novel thoracoscopic tracheoesophageal fistula repair simulator. Stud Health Technol Inform 2013;184: 114–116.
6. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association, 1999.
7. Wolfe ESJ, Smith EV Jr. Instrument development tools and activities for measure validation using Rasch models: Part II—Validation activities. J Appl Measure 2007;8:243–290.
8. Henry B, Clark P, Sudan R. Cost and logistics of implementing a tissue-based American College of Surgeons/

Association of Program Directors in Surgery surgical skills curriculum for general surgery residents of all clinical years. Am J Surg 2014;207:201–1208.

9. Smith AB, et al. Rasch fit statistics and sample size considerations for polytomous data. BMC Med Res Methodol 2008;8:33.
10. Linacre J. A User's Guide to Facets. Chicago: Winsteps, 2010.

Address correspondence to:
Katherine A. Barsness, MD
Division of Pediatric Surgery
Ann and Robert H. Lurie Children's Hospital of Chicago
225 East Chicago Avenue
Box 63
Chicago, IL 60611

E-mail: kbarsness@luriechildrens.org

This article has been cited by:

1. Julien Rod, Jean-Baptiste Marret, Jules Kohaut, Yves Aigrain, Jean Philippe Jais, Philine de Vries, Stephen Lortat-Jacob, Jean Breaud, Thomas Blanc. 2017. Low-Cost Training Simulator for Open Dismembered Pyeloplasty: Development and Face Validation. *Journal of Surgical Education* . [Crossref]

2. Benjamin Langridge, Sheikh Momin, Ben Coumbe, Evelina Woin, Michelle Griffin, Peter Butler. 2017. Systematic Review of the Use of 3-Dimensional Printing in Surgical Teaching and Assessment. *Journal of Surgical Education* . [Crossref]

3. David C. van der Zee. 2017. Endoscopic surgery in children – the challenge goes on. *Journal of Pediatric Surgery* **52**:2, 207-210. [Crossref]

4. Deie Kyoichi, Ishimaru Tetsuya, Takazawa Shinya, Harada Kanako, Sugita Naohiko, Mitsuishi Mamoru, Fujishiro Jun, Iwanaka Tadashi. 2017. Preliminary Study of Video-Based Pediatric Endoscopic Surgical Skill Assessment Using a Neonatal Esophageal Atresia/Tracheoesophageal Fistula Model. *Journal of Laparoendoscopic & Advanced Surgical Techniques* **27**:1, 76-81. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]

5. Schwab Ben, Rooney Deborah M., Hungness Eric S., Barsness Katherine A.. 2016. Preliminary Evaluation of a Laparoscopic Common Bile Duct Simulator for Pediatric Surgical Education. *Journal of Laparoendoscopic & Advanced Surgical Techniques* **26**:10, 831-835. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]

6. Satoshi Ieiri, Takahiro Jimbo, Yuta Koreeda, Satoshi Obata, Munenori Uemura, Ryota Souzaki, Yo Kobayashi, Masakatsu G. Fujie, Makoto Hashizume, Tomoaki Taguchi. 2015. The effect of forceps manipulation for expert pediatric surgeons using an endoscopic pseudo-viewpoint alternating system: the phenomenon of economical slow and fast performance in endoscopic surgery. *Pediatric Surgery International* **31**:10, 971-976. [Crossref]

7. Katherine A. Barsness. 2015. Trends in technical and team simulations: Challenging the status Quo of surgical training. *Seminars in Pediatric Surgery* **24**:3, 130-133. [Crossref]