**Supplemental Digital Content 2:**

**Table 1: Post-Intervention Outcomes**

**Table 2: Longest Follow-up Outcomes**

**Table 1: Post-Intervention Outcomes**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study ID**  | **Kirkpatrick Level 1 Measures** | **Post-Intervention Results**  | ***P*** | **Kirkpatrick Level 2 Measures**  | **Post-Intervention Results**  | ***P*** | **Kirkpatrick Level 3 Measures** | **Post-Intervention Results**  | ***P*** |
| **Active** | **Observer** | **Active** | **Observer** | **Active** | **Observer** |
| Baxter 2012 |  |  |  |  | Seven-point Likert scale assessing decision-making, collaboration, communication, crisis management and global impression of performance by external observer in a simulated environment | Mean 5.04 SD 0.48 | Mean 4.74 SD 0.88 | NS |  |  |  |  |
| Blanie 2018 | 10-point Likert scale measuring:SatisfactionInterest | Median 9 IQR 8-9Median 9IQR 8-10 | Median 8 IQR 8-9Median 9 IQR 8-10 | 0.0190.515 | Medical knowledge using a 16-item multiple choice testSelf-reported non-technical skills 17-item questionnaire with five scales:Task managementTeamworkSituational awarenessDecision makingCoping with stress | Median 10 IQR 8-11Median 8IQR 7.3-8.8Median 7.7IQR 7.3-8.6Median 7.7IQR 7-8.7Median 7.7IQR 6.8-8.7Median 6.5IQR 5-8 | Median 9 IQR 7-10Median 7.8IQR 7-8.3Median 7.7IQR 6.8-8.3Median 7.3IQR 7-8Median 7.7IQR 6.7-8Median 6.5IQR 5-7.8 | <0.010.0980.2580.0590.480.497 | 10-point Likert scale measuring perceived learning transfer | Median 8IQR 7-10 | Median 8IQR 7-10 | 0.701 |
| Boncyk 2016 |  |  |  |  | 10-point Likert scale of self-reported perception of knowledge% accuracy of identifying structures using ultrasound in simulation rated by external observer | Median 6IQR 4-8Mean 66%95% CI 57-76 | Median 6.5IQR 5-8Mean 73%95% CI 65-80 | 0.42≥0.14 |  |  |  |  |
| Bong 2017 | Self-reported modified Depression and Anxiety Stress Scale to include only questions relating to anxiety of stress (min 0 – max 72) | Median 17.695% CI 12.4- 22.7 | Median 995% CI 4.9-13.1 | <0.05 | Anesthetists’ Nontechnical Skills (ANTS) score by external rater in simulation (min 15 – max 60) | Median 4095% CI 37.9 – 42.1 | Median 39.495% CI 37.4 – 41.5 | 0.036 |  |  |  |  |
| Brydges 2012 | Self-reported confidence on a 11-point Likert scale (min 0 – max 10) | Mean 7.07SE 0.34 | Mean 6.33SE 0.36 | NR | Global Rating Score of technical skills by external rater in simulation (min 0 – max 5)Procedural checklist completed by external rater in simulation (min 0 – max 35) | Mean 3.7SE 0.21Mean 28.59SE 0.9 | Mean 3.25SE 0.21Mean 25.75SE 1.33 | NRNR |  |  |  |  |
| Hobgood 2010A |  |  |  |  | 36-item self-reported teamwork attitudes instrument (min 36 – max 180)12-item teamwork knowledge test10-item Standardized patient evaluation of teamwork skills in simulated environment (min 0 – max 50)20-item May High Performance Teamwork Scale by an external observer in a simulated environment (min 0 – max 100) | Mean 145.7SE 1.614Mean 9.81SE 0.259Mean 8SD 1.683Mean 19.92SD 0.144 | Mean 142.9SE 1.137Mean 10.3SE 0.171Mean 8.053SD 1.545Mean 19.5SD 0.524 | NRNR0.2380.134 |  |  |  |  |
| Hobgood 2010B |  |  |  |  | 36-item self-reported teamwork attitudes instrument (min 36 – max 180)12-item teamwork knowledge test10-item Standardized patient evaluation of teamwork skills in simulated environment (min 0 – max 50)20-item May High Performance Teamwork Scale by an external observer in a simulated environment (min 0 – max 100) | Mean 143.2SE 1.439Mean 10.2SE 0.205Mean 8.667SD 1.047Mean 19.88SD 0.189 | Mean 142.9SE 1.137Mean 10.3SE 0.171Mean 8.053SD 1.545Mean 19.5SD 0.524 | NRNR0.2380.134 |  |  |  |  |
| Lai 2016 |  |  |  |  | Ottawa Crisis Resource Management Global Rating Scale by external observers in a simulated environment (min 5 – max 35) | Mean 24.3SD 5.4 | Mean 21.8SD 4.5 | 0.22 |  |  |  |  |
| McCoy 2017 | % Self-reported favorability of training modality: Enhances careWould like moreDeepened insightsAdded value Effective Beneficial Would participate  | 97%97%97%100%95%95%100% | 97%94%100%100%95%100%95% | NRNRNRNRNR NR NR | Medical knowledge  | Mean 96.8 95% CI 94.5-98.6 | Mean 96.8 95% CI 94.8-98.9 |  |  |  |  |  |
| Nilsson 2014 |  |  |  |  | Standardized evaluation of management of post-partum hemorrhage by external observer in simulated environment | Median 40%95% CI 29.5-47 | Median 35%95% CI 25-42 | 0.852 |  |  |  |  |
| Semler 2015 | 7-point Likert scale self-reported learner satisfaction:RelevanceUsefulness | Mean 6.4 SD 0.9Mean 6.9 SD 0.3  | Mean 6.6 SD 0.6Mean 6.8 SD 0.6  |  | 7-point overall teamwork behavior score by external observer in a simulated environment (min 1 – max 7)Adherence to ACLS algorithm scored by external observer in a simulated environment (min 0 – max 30)Recognize, Respond, Reassess score by external observer in a simulated environment (min 0 – max 39 points) | Mean 4.1SD 0.95Mean 19.4SD 1.9Mean 23.3SD 4.9 | Mean 4.4SD 1.15Mean 20.5SD 1.17Mean 22.7SD 2.7 | 0.0740.6920.883 |  |  |  |  |
| VanderWielen 2014 | Self-reported perceived knowledge on a 10-point Likert scale:Basic anatomySonoanatomyProcedure risk | Mean 8SD 1.3Mean 7.7SD 1.4Mean 7.8SD 1.7 | Mean 7.8SD 1.4Mean 7.4SD 1.3Mean 8.3SD 1.1 | NRNRNR | % correctly identifying structures using ultrasound in simulation rated by external observer:Spinous processLaminaTransverse process | Mean 100%Mean 88.2%Mean 76.5% | Mean 100%Mean 72.2%Mean 77.8% | NRNRNR |  |  |  |  |
| Weiler 2018 |  |  |  |  | Critical thinking rated by external observer in a simulated scenario (min 0 – max 5) | Median 3 | Median 3 | 0.541 |  |  |  |  |

NS: Not significant. The authors did not report an exact value.

IQR: Interquartile range

NR: Not reported

CI: Confidence interval

SE: Standard error

SD: Standard deviation

ACLS: Advanced cardiac life support

A Low-fidelity simulation vs. observed simulation

B High-fidelity simulation vs. observed simulation

**Table 2: Longest Follow-up Outcomes**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study ID**  | **Kirkpatrick Level 1 Measures** | **Post-Intervention Results**  | ***P*** | **Kirkpatrick Level 2 Measures**  | **Post-Intervention Results**  | ***P*** | **Kirkpatrick Level 3 Measures** | **Post-Intervention Results**  | ***P*** |
| **Active** | **Observer** | **Active** | **Observer** | **Active** | **Observer** |
| Blanie 2018 | 10-point Likert scale measuring the degree of interest  | Median 9 IQR 7-10 | Median 8 IQR 7-9 | 0.211 | Medical knowledge using a 16-item multiple choice test Self-reported non-technical skills 17-item questionnaire with five scales: Task managementTeamworkSituational awarenessDecision making Coping with stress | Median 8 IQR 7-9 Median 6.8 IQR 6-7.3Median 7 IQR 6-8Median 7.3 IQR 6-7.7 Median 6.3IQR 5.4-7.7 Median 6 IQR 5-8 | Median 8 IQR 7-9Median 6.5 IQR 6-7 Median 7 IQR 6-8Median 7.3 IQR 5-7.3 Median 6.7 IQR 5-7.7 Median 6IQR 4-7  | 0.2770.320.5620.880.8910.206 | 10-point Likert scale measuring perceived learning transfer | Median 8 IQR 7-9  | Median 7 IQR 5-8  | 0.036 |
| Boncyk 2016 |  |  |  |  | % correctly identifying structures using ultrasound in simulation rated by external observer | Mean 49%95% CI 40-58 | Mean 55% 95% CI 47-64 | NR  |  |  |  |  |
| Brydges 2012 |  |  |  |  | Global Rating Score by external rater in simulation (min 0 – max 5)Procedural checklist completed by external rater in simulation (min 0 – max 35) | Mean 3.05 SE 0.28Mean 25.5 SE 2.13 | Mean 3.63 SE 0.23Mean 27.41 SE 1.57 | NR NR  |  |  |  |  |
| Nasr Esfahani 2014 |  |  |  |  | Self-reported empathy on the Jefferson Scale of Empathy (min 0 – max 140)External observer rated empathy in a simulated environment (min 0 – max 1)  | Mean 125 SD 7.9Mean 0.79 SD 0.15 | Mean 114SD 14.1 Mean 0.6 SD 0.1  | NRNR  |  |  |  |  |
| VanderWielen 2014 |  |  |  |  | % correctly identifying structures using ultrasound in a simulated environment rated by external observer:Spinous process Lamina Transverse process  | Mean 93.3% Mean 66.6%Mean 13.3%  | Mean 100%Mean 47.5% Mean 17.6%  | NR NRNR |  |  |  |  |

NR: Not reported

SD: Standard deviation

IQR: Interquartile range