**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:  Satisfaction  Interest | Median 9 IQR 8-9  Median 9  IQR 8-10 | Median 8 IQR 8-9  Median 9 IQR 8-10 | 0.019  0.515 | Medical knowledge using a 16-item multiple choice test  Self-reported non-technical skills 17-item questionnaire with five scales:  Task management  Teamwork  Situational awareness  Decision making  Coping with stress | Median 10 IQR 8-11  Median 8  IQR 7.3-8.8  Median 7.7  IQR 7.3-8.6  Median 7.7  IQR 7-8.7  Median 7.7  IQR 6.8-8.7  Median 6.5  IQR 5-8 | Median 9 IQR 7-10  Median 7.8  IQR 7-8.3  Median 7.7  IQR 6.8-8.3  Median 7.3  IQR 7-8  Median 7.7  IQR 6.7-8  Median 6.5  IQR 5-7.8 | <0.01  0.098  0.258  0.059  0.48  0.497 | 10-point Likert scale measuring perceived learning transfer | Median 8  IQR 7-10 | Median 8  IQR 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 6  IQR 4-8  Mean 66%  95% CI 57-76 | Median 6.5  IQR 5-8  Mean 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 17.6  95% CI 12.4- 22.7 | Median 9  95% CI 4.9-13.1 | <0.05 | Anesthetists’ Nontechnical Skills (ANTS) score by external rater in simulation (min 15 – max 60) | Median 40  95% CI 37.9 – 42.1 | Median 39.4  95% CI 37.4 – 41.5 | 0.036 |  |  |  |  |
| Brydges 2012 | Self-reported confidence on a 11-point Likert scale (min 0 – max 10) | Mean 7.07  SE 0.34 | Mean 6.33  SE 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 3.7  SE 0.21  Mean 28.59  SE 0.9 | Mean 3.25  SE 0.21  Mean 25.75  SE 1.33 | NR  NR |  |  |  |  |
| Hobgood 2010A |  |  |  |  | 36-item self-reported teamwork attitudes instrument (min 36 – max 180)  12-item teamwork knowledge test  10-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 145.7  SE 1.614  Mean 9.81  SE 0.259  Mean 8  SD 1.683  Mean 19.92  SD 0.144 | Mean 142.9  SE 1.137  Mean 10.3  SE 0.171  Mean 8.053  SD 1.545  Mean 19.5  SD 0.524 | NR  NR  0.238  0.134 |  |  |  |  |
| Hobgood 2010B |  |  |  |  | 36-item self-reported teamwork attitudes instrument (min 36 – max 180)  12-item teamwork knowledge test  10-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 143.2  SE 1.439  Mean 10.2  SE 0.205  Mean 8.667  SD 1.047  Mean 19.88  SD 0.189 | Mean 142.9  SE 1.137  Mean 10.3  SE 0.171  Mean 8.053  SD 1.545  Mean 19.5  SD 0.524 | NR  NR  0.238  0.134 |  |  |  |  |
| Lai 2016 |  |  |  |  | Ottawa Crisis Resource Management Global Rating Scale by external observers in a simulated environment (min 5 – max 35) | Mean 24.3  SD 5.4 | Mean 21.8  SD 4.5 | 0.22 |  |  |  |  |
| McCoy 2017 | % Self-reported favorability of training modality:  Enhances care  Would like more  Deepened insights  Added value  Effective  Beneficial  Would participate | 97%  97%  97%  100%  95%  95%  100% | 97%  94%  100%  100%  95%  100%  95% | NR  NR  NR  NR  NR  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:  Relevance  Usefulness | Mean 6.4  SD 0.9  Mean 6.9  SD 0.3 | Mean 6.6  SD 0.6  Mean 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 4.1  SD 0.95  Mean 19.4  SD 1.9  Mean 23.3  SD 4.9 | Mean 4.4  SD 1.15  Mean 20.5  SD 1.17  Mean 22.7  SD 2.7 | 0.074  0.692  0.883 |  |  |  |  |
| VanderWielen 2014 | Self-reported perceived knowledge on a 10-point Likert scale:  Basic anatomy  Sonoanatomy  Procedure risk | Mean 8  SD 1.3  Mean 7.7  SD 1.4  Mean 7.8  SD 1.7 | Mean 7.8  SD 1.4  Mean 7.4  SD 1.3  Mean 8.3  SD 1.1 | NR  NR  NR | % correctly identifying structures using ultrasound in simulation rated by external observer:  Spinous process  Lamina  Transverse process | Mean 100%  Mean 88.2%  Mean 76.5% | Mean 100%  Mean 72.2%  Mean 77.8% | NR  NR  NR |  |  |  |  |
| 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 management  Teamwork  Situational awareness  Decision making  Coping with stress | Median 8  IQR 7-9  Median 6.8  IQR 6-7.3  Median 7  IQR 6-8  Median 7.3  IQR 6-7.7  Median 6.3  IQR 5.4-7.7  Median 6  IQR 5-8 | Median 8  IQR 7-9  Median 6.5  IQR 6-7  Median 7  IQR 6-8  Median 7.3  IQR 5-7.3  Median 6.7  IQR 5-7.7  Median 6  IQR 4-7 | 0.277  0.32  0.562  0.88  0.891  0.206 | 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 0.28  Mean 25.5  SE 2.13 | Mean 3.63  SE 0.23  Mean 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 7.9  Mean 0.79  SD 0.15 | Mean 114  SD 14.1  Mean 0.6  SD 0.1 | NR  NR |  |  |  |  |
| 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  NR  NR |  |  |  |  |

NR: Not reported

SD: Standard deviation

IQR: Interquartile range