|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2. Validation results** | | | | | | | |
|
| **Assessment tools** | **Paper(s) included in systematic review** | **Paper(s) validation evidence** | **Proposed use** | **Scoring** | **Generalization** | **Extrapolation** | **Implications** |
| CTS (Clinical Teamwork Scale) | Fransen et al 2012 | Guise et al 2008 | Assess teamwork performance in interprofessional teams; consisting of gynaecologists, midwives, residents, and nurses; in high-fidelity simulated obstetric emergency scenarios. | -Details on item development provided. The tool was originally validated using scripted performances by actors in simulated standardized scenarios. -Details on raters, and their training, were provided. The raters were intentionally not extensively trained in the use of the tool, as the intention was that the tool would be easy to use without formal training. -No content validity evidence provided. | -Overall agreement among raters, measured by kappa statistics, was substantial, and agreement among teamwork levels was high. Concordance (Kendall coefficient) and correlation (Pearson coefficient) was high. -Internal consistency not reported. -Reliability of the tool's ratings was also examined by estimating the variance of each component based on generalizability theory. | -No expert-novice analysis, or factor analysis, was reported. -The tool was originally validated in a simulated scenario, performed by scripted actors, and therefore the authenticity of the context could be challenged. Observed performance might not reflect desired real-life clinical performance. | -Implications of using the assessment tool was not reported |
| SNAPPI pre-defined scoring rubric | Weller et al 2014 | Weller et al 2014 | Assess call-outs by teams; consisting of nurses, anesthesiologists, and anesthetic technicians; in simulated post-anesthesia care unit crisis scenarios. | -Details on item development provided. Items were created by consensus among five specialist anaesthesiologists with expertise in teamwork and communication training. -Details on rater training provided. The piloting of the tool was also detailed. -No content validity evidence provided. | -Pre/Post testing in 'test setting' -Details for sample size calculations provided. -Inter-rater agreement, using intra-class correlation was good, in both the pilot, and in the study proper. -Internal consistency not reported. | -No factor analysis, or expert-novice analysis, were reported. -Simulations were held in simulated PACU environments, to resemble real-life PACU environments, and to provide authenticity. Insufficient details about the simulations were reported to verify this assumption. | -Implications of using the assessment tool was not reported |
| Teamwork and Patient Care Measure | Fernandez et al 2013 | Fernandez et al 2013 | Assess teamwork behaviours and patient care performance in code teams, consisting of medical students, emergency medicine residents, and confederate nurse actors. | -Details on item development were provided. -Raters were trained, but no details were provided. -Items content validated by external teamwork and clinical subject matter experts. | -Inter-rater reliability using Cohen's kappa showed substantial agreement for teamwork behaviour, and excellent agreement for patient behaviour. -Internal consistency not reported. | -Higher experience correlated with higher teamwork process measures and higher patient care measures. -Analysis of covariance showed that the particular scenario used for assessment did not influence training outcomes. Analysis of covariance was also used to assess the effect of the training intervention on teamwork behaviors and patient care performance. -No factor analysis was reported. | -Implications of using the assessment tool was not reported |
| ANTS (Anesthetist's Non-Technical Skills) | Jankouskas et al 2011, Lee et al 2012 | Fletcher et al 2003, Jankouskas et al 2007 & 2011, Lee et al 2012 | Assess teamwork parameters in interprofessional teams; consisting of paediatric nurses, paediatric residents and anaesthesiology residents; in simulated pediatric crisis scenarios. | -Details on item development were provided. The tool was originally created to assess NTS and teamwork behaviours of anesthesiologists. The literature review focused on relevant, general teamwork behavior; the tool was created by anesthesiologists; and rated by anesthesiologists. -Raters received minimal training in rating NTS, and in the use of the tool, to show the tool's ease-of-use. -A panel of expert anesthesiologists provided item content validity, when using the tool to assess anesthesiologists. | -Pre/Post testing in 'test setting' -Internal consistency was high, and interrater agreement was reasonable.  -The authors argue that interrater agreement likely would be higher with more rater training. | -No factor analysis, or expert-novice analysis, were reported. -High-fidelity simulations were used, with the assumption that it correlates to real-life clinical practice. Insufficient details were provided to verify this assumption. | -Implications of using the assessment tool was not reported |
| ORTAS (Operating Room Teamwork Assessment Scale) | Paige et al 2014 | Paige et al 2009, Paige et al 2014 | Assess team-based performance of interprofessional operative teams, consisting of medical and nursing students, in operating room scenarios | -Details on item development provided. -Items were content validated by a panel of relevant content experts. -Details about rater training provided. No details on the selection of raters, nor any other descriptive details about the raters were reported. | -Pre/Post testing in 'test setting' -Internal consistency, using Cronbach’s alpha, of the scales and subscales, was found to be good and very good respectively. -Interrater reliability was not reported. | -Scales and subscales were validated by using factor analysis. -No expert-novice analysis was reported. -Increase in teamwork performance scores post intervention was found both in simulated scenarios and in real-life elective surgical cases, supporting authenticity of the scoring. | -Implications of using the assessment tool was not reported |
| TEAM (Team Emergency Assessment Measure) | Rubio-Gurung et al 2014, Couto et al 2015 | Cooper et al 2010, Cooper et al 2014 | Developed to measure teamwork performance in interprofessional teams, for both simulated and actual emergencies. | -Details on item development provided. -Details on raters were not provided. -Face and content validity assessed by an international panel of resuscitation experts. Level of agreement was determined by content validity index. | -Internal consistency was high, and inter-rater reliability was fair. | -Comparison between simulation and actual emergencies -Construct validity was assessed by using component factor analysis, showing one construct, "teamwork", to explain the majority of the variance. -No expert-novice analysis was reported. -Item development, and rating, was done by experienced clinicians and academics who were resuscitation officers, from relevant fields, supporting authenticity of the scoring. -Since TEAM was developed to be used in actual emergencies, it might come at the expense of the generalization (to maximize extrapolation). | -Implications of using the assessment tool was not reported |
| KidSIM Team Performance Scale | Sigalet et al 2013 | Sigalet et al 2013 | To evaluate team performance of undergraduate, interprofessional health professionals in a simulation-based interprofessional curriculum. | -Details on item development were provided. -Details on rater background and training were provided. -No evidence for content validity was provided. | -Interrater agreement and internal consistency, using Cronbach's alpha, were excellent. | -Construct validity was assessed by factor analysis with varimax rotation, resulting in a three-factor solution accounting for the majority of the variance. -No expert-novice analysis was reported. | -Implications of using the assessment tool was not reported |
| NOTECHS (Nontechnical skills evaluation rating scale) | Powers et al 2008 | Moorthy et al 2006, Powers et al 2008 | Originally developed to assess non-technical skills in aviation, Moorthy et al used the tool to assess NTS in a crisis scenario during a simulated saphenofemoral junction high-tie procedure. Powers et al used the tool to assess NTS in surgeons during a novel, interprofessional, simulated team training for minimally invasive surgery. | -Details on item development was not provided. -Details on raters were not provided. -Expert surgeons participating in the minimal invasive study assessed the content validity of the laparoscopic crisis endosuite scenario. -Content validity of the assessment tool was not reported. | -Interrater reliability and internal consistency was determined by Cronbach's alpha, and was found to be high, in both studies, indicating good reliability. | -No factor analysis was reported. -Expert-novice analysis using Tukey–Kramer Studentized range test was done, favoring higher scores for experts in all of the components of the NTS score, with greater variability in the novice group. -Details on the authenticity of the simulated environment, and the simulated abdominal model, were provided. -Face validity assessed by participant questionnaire. | -Implications of using the assessment tool was not reported |
| TLIS (Teamwork Leadership Interpersonal Skills) & ECCS (Emergency Clinical Care Skills) | Pascual et al 2011 | Pascual et al 2011, https://www.facs.org/education/program/resident-skills | Assess pre- and posttraining non-technical skills performance of surgical critical care fellows and in advanced practioners, in simulated ICU emergency scenarios. | -Details on item development was not provided. -Details on raters were not provided. -The authors reported that the tools were developed from an established curriculum. The curriculum was said to be validated, but no further details were provided. -No content validation evidence was reported. | -Pre/Post testing in 'test setting' -Interrater reliability for all scores was excellent. -Internal consistency not reported. | -No expert-novice analysis, or factor analysis, was reported. -The simulated scenarios were descibed in detail, and were created by relevant field experts to provide authenticity. -Face validity of the curriculum was evaluated by participant questionnaire. | -The authors discuss the possible effects assessment by TLIS might have on future ICU staffing -Implications of using the assessment tool was not reported |
| NOTSS (Non-technical Skills for Surgeons) | Abdelshehid et al 2013, Lee et al 2012 | Yule et al 2006 & 2008, Abdelshehid et al 2013, Lee et al 2012 | To be used by post-graduate surgeons in theatre to assess non-technical skills of interprofessional surgical teams. | -Details on item development provided. -Details on rater training provided. -Content validity was derived from the systematic development process with subject matter experts. | -System sensitivity, internal structure, and inter-rater reliability was reported, and found to be high. -When analysing, a generally accepted criterion for an acceptable level of agreement was used. -Intraclass correlation coefficient was also calculated, following the convetion for acceptable reliability. | -No expert-novice analysis, or factor analysis, was reported. -Differences in agreement based on the surgical field of the raters was observed. Also There were substantial differences in rater agreement between different scenarios. | -Arguments for good investment value in using NOTSS were presented, as short rater training still reached a high level of agreement and sensitivity -Implications of using the assessment tool was not reported |
| TBR (Team Behavioral Rater) | Frengley et al 2011 | Weller et al 2011 | Measure teamwork behaviours in interprofessional teams in the field of critical care medicine. | -Details on item development and developers provided. -Details about pilot testing, and rater training provided. -No content validity evidence provided. | -Internal consistency using Cronbach's alpha was "good" and "excellent" for the factors identified in the factor analysis. -The analysis of Variance Components found that more subjective items seemed to be more affected by the rater. -Interrater reliability was not reported. | -Construct validity was supported by improved performance of the teams over time, and superior performance by specialists versus trainees. -Exploratory Factor Analysis showed that items clustered around three factors: “Leadership and Team Coordination”, “Mutual Performance Monitoring”, and “Verbalising Situational Information”. | -Implications of using the assessment tool was not reported |
| TPOT (Trauma Team Performance Observation Tool) | Capella et al 2010 | Capella et al 2009 & 2010 | To assess the team performance of interprofessional teams, during trauma resuscitations. | -No validation evidence provided. | -No further validation evidence provided. | -Pre/Post testing in 'nontest setting' -No further validation evidence provided. | -No validation evidence provided. |
| SAGAT (Situation Awareness Global Assessment Technique) | Morgan et al 2015 | Hogan et al 2006, Morgan et al 2015 | Assessment of interprofessional, medical teams' situation awareness. | -Details about initial item development and content validation provided. Details for tool modification was also provided. -Details on raters were not provided. -Content validity evidence provided for trauma emergency setting, but not for obstetric emergencies. -Scoring process and scenario development was explained. | -Interrater reliability using Cronbach’s alpha was acceptable. -Internal consistency not reported. | -Experts scored higher than novices, supporting construct validity. -No factor analysis was reported. -SAGAT was compared to a traditional checklist, and Pearson's correlation supported that both measurements measured the same construct. | -Implications of using the assessment tool was not reported |
| CEPTE (Clinical Emergency Predaredness Team Evaluation) | Kennedy et al 2013 | Kennedy et al 2013 | Assessment of team performance and skill in pediatric emergency situations. | -Details on item development was not provided. -The contents of CEPTE was provided. -Details on the raters, and their training, was provided. -No content validity evidence provided. | -Pre/Post testing in 'test setting' -Interrater agreement was reported to be substantial, but the method of calculation was unclearly reported. -Internal consistency not reported. | -No expert-novice analysis, or factor analysis, was reported. -Technical details about the simulations were provided, arguing for the authenticity of the simulations. -Team performance scores improved from baseline. In situ simulations also showed improvement in team performance. This supports a measurement of actual team performance. | -Implications of using the assessment tool was not reported |
| "Checklist of Expected Actions", and "Health Failure Modes Effects Analysis" (HFMEA) | Daniels et al 2008 | Daniels et al 2008, DeRosier et al 2002 | Assess the clinical and behavioral performance of the OB residents during interprofessional, simulated, obstetric emergencies | -Details on item development was not provided. -Details about raters and their training provided. -Details on how the tool was modified for the study was provided. -No evidence for content validity was provided. | -Interrater variability (using Savr) showed substantial agreement across all the checklist items. -Internal consistency not reported. | -No expert-novice analysis, or factor analysis, was reported. -Technical details about the simulations and simulated setting was provided. -No discussion about the authenticity of the simulated scenarios. -Transferability of the perceived learning into real-life situations was not assessed. | -Implications of using the assessment tool was not reported |
| Videorecordings analysed qualitatively for language patterns | Minehart et al 2012 | Minehart et al 2012 | -Analyze the verbal communication strategies used in a simulated maternal-fetal crisis. -Based on the assumption that language patterns in the OR describe or explain clinical decision making or teamwork | -Details for the method of coding was provided and explained.  -No details about the coders were provided. -Demographic data of the participants were not reported. | -No arguments for sampling sufficiency. -No analysis of the nurses communication strategies was done, which might have yielded more wholesome insight into interprofessional communication. -The authors only reported quantitative results; no argument for how language patterns relates to collaborative performance could be made. | -No expert-novice analysis, or factor analysis, was reported. -Since the studied situations are not usually handled by novices in real-life clinical practise, the choice of excluding pre-graduates provides authenticity. | -Implications of using the assessment tool was not reported |
| "Template Analysis Approach", Qualitative analysis with a priori coding and template analysis | Muller-Juge et al 2014 | Muller-Juge et al 2014 | Describe physicians' and nurses' characteristics and behaviours that contribute to teamwork quality in the setting of a simulated hospital internal medicine ward | -Development of a priori codes detailed. -Details about the coders and their training were provided, and they were of relevant professions. | -No argument for sampling sufficiency was provided. -Data triangulation was done through audio- and video-recordings; coding differences were solved by consensus to provide validity to the analysis. -The authors did not provide any arguments for the presence or absence of observer effect, or how this was dealt with. | -No expert-novice analysis, or factor analysis, was reported. -“High-fidelity” simulations were used to reproduce several clinical situations, and to provide authenticity. The simulated scenarios were clinically relevant and varied. -The authors argue that their choice of sampling would best represent real-life clinical practice of most internal medicine wards, where the situation is assessed by junior staff-members before a senior staff-member is called.  -No evidence to support that the observed performance correlates to relevant real-life practise was provided. | -Implications of using the assessment tool was not reported |
| TPDSCI (Team Performance During Simulated Crisis Instrument) & CRM checklist (Crisis Resource Management), validation not reported | Calhoun et al 2014 | Calhoun et al 2011 & 2014 | Assess team performance in interprofessional teams in paediatric emergency simulations. | -Details on item development were provided. -Details on rater training were provided. Details about rater selection were not provided. -No content validity evidence provided. | -Internal consistency was calculated using Cronbach’s alpha and a composite Cronbach alpha to account for intrasubject correlation. Overall cronbach’s alpha was 0.72 (acceptable), and the composite cronbach’s alpha was 0.69 (questionable). -Interrater reliability (calculated by ICC) for most items was good or excellent, except for the professionalism domain, which had poor interrater reliability. | -No expert-novice analysis, or factor analysis, was reported. -No evidence for construct validity, and the authors explain the reasons behind this. -The tools ease-of-use was supported by high interrater reliability, despite no prior training. | -Implications of using the assessment tool was not reported |
| KSNPS (Kramer and Schmalenberg Nurse-Physician Scale), Collaboration & Satisfaction with Patient Care Decisions (CSPCD) and Clinical Practice Group Cohesion (GC). | Messmer et al 2008 | Kramer & Schmalenberg 2002 & 2003, Messmer et al 2008 | Determine the level of nurse-physician collaboration during simulation training. | -Items were developed by qualitatively identifying themes of nurse-physician collaboration in interviews conducted with nurses. -No argument for sampling sufficiency. -No evidence for content validity was provided. -Details on item creation and validation for the NACHRI tools were not provided. | -Alpha reliability for the NACHRI tools was reported to be good (>0.80). However, the origin of this statement could not be verified, as we could not gain access to the referenced study. -Internal consistency, or interrater reliability was not reported for KSNPS. | -No expert-novice analysis, or factor analysis, was reported. -No evidence for construct validity was provided. | -Implications of using the assessment tool was not reported |
| Kalamazoo Essential Elements Communication Skills | Calhoun et al 2010 | Calhoun et al 2009 & 2010 | Assess communication of interdisciplinary teams in simulated pediatric scenarios | - Details on item development were provided. - Details about rater selection or training were not provided. -No content validity evidence provided. | -Internal consistency was calculated as Cronbach's alpha, and was found to be "good" for all items. -Interrater reliability was not reported. | -Factorial analysis provided construct validity evidence by showing that all the dimensions of the tool contributed to a single measured construct. -No expert-novice analysis was reported. | -Implications of using the assessment tool was not reported |