



Featured Article

# A Review of Currently Published Evaluation Instruments for Human Patient Simulation

Suzie Kardong-Edgren, PhD, RN, Katie Anne Adamson, MSN, RN,  
Cynthia Fitzgerald, PhD, FNP

Washington State University, Spokane, WA 99210

## KEYWORDS

evaluation;  
simulation;  
review;  
human patient  
simulation;  
evaluation instruments

**Abstract:** The lack of reliable and valid instruments to evaluate simulation learning outcomes is inhibiting the adoption and progress of simulation in nursing education. This article (a) discusses the importance of learning domains in evaluation, (b) reviews current challenges in designing simulation evaluation instruments, and (c) provides a review of currently published instruments for simulation evaluation.

## Cite this article:

Kardong-Edgren, S., Adamson, K. A., & Fitzgerald, C. (2010, January). A review of currently published evaluation instruments for human patient simulation. *Clinical Simulation in Nursing*, VOL(6), e25-e35. doi:10.1016/j.ecns.2009.08.004.

© 2010 International Nursing Association for Clinical Simulation and Learning. Published by Elsevier Inc. All rights reserved.

Recent reviews of new nursing graduates suggest that traditional methods of clinical instruction are not particularly successful (Benner, Sheets, Uris, Malloch, Schwed, & Jamison, 2002; Del Bueno, 2005; Gaba, 2004). Clinical education has traditionally consisted of a “relatively unsystematic apprenticeship process” (Gaba, 2004, p. i2). The expectation was that if a student spent enough time in a clinical situation, the student would eventually “get it” (Dunn, 2004). The lack of alternatives to traditional clinical instruction delayed rigorous evaluation of that model although its effectiveness was increasingly in question (Dunn, 2004; Gaba, 2004).

Innovations in technology have expanded the options for teaching and learning in clinical and nursing education. Human patient simulation (HPS) provides the ability to produce clinical experiences encompassing the affective, cognitive, and psychomotor domains, all essential to nursing practice (Jeffries & Norton, 2005). HPS also provides the opportunity to evaluate all students using the

same scenario under the same controlled conditions. Thus simulation has generated a need for reliable and valid clinical evaluation tools to measure student learning in the simulation setting. Increasingly, stakeholders such as risk managers and grant funding sources are interested in patient outcomes. A novel simulation evaluation tool might reflect not only student performance but also patient outcomes. This “gold standard” has not been seen in the literature to date.

Current National Council of State Boards of Nursing research demonstrates higher levels of support for the use simulation than for traditional clinical experiences (Spector, 2006). However, the relative lack of reliable and valid evaluation instruments measuring learning outcomes and/or the effectiveness of HPS as a teaching strategy may be inhibiting its adoption and progress in nursing education. Researchers and institutions around the world are engaged in developing tools for evaluating simulations and measuring learning outcomes from HPS. Purchasing,

maintaining, training for HPS, and adoption of a new teaching paradigm are time and resource intensive. The evaluation of learning outcomes from HPS is perhaps held to a higher standard than traditional clinical teaching is, despite research indicating that traditional clinical education methods are not working particularly well.

### Key Points

- We suggest a moratorium on the indiscriminate development of new evaluation tools.
- Few nursing faculty who use simulation are also skilled in instrument development.
- We do not want to build a mile-wide and inch-deep evaluation landscape.

One way to enhance current research related to the effectiveness of HPS in nursing education is to develop reliable and valid instruments to measure performance in simulation settings. The knowledge, values, and abilities that are essential to nursing practice encompass the affective, cognitive, and psychomotor learning domains (Jeffries & Norton, 2005;

Oermann & Gaberson, 2006). Therefore, research aimed at assessing the student in simulation should address how well he or she demonstrates learning in these domains. It is important to evaluate currently available instruments that measure these global aspects of learning. Ideally, evaluation tools should provide educators with information about cognitive, psychomotor, and affective learning outcomes for both individuals and groups.

The purposes of this article are to (a) discuss the importance of learning domains in evaluation, (b) review current challenges in designing simulation evaluation tools, and (c) provide a sampling of clinical simulation tools currently found in the literature that show promise for further refinement and development. We suggest a moratorium on the indiscriminate development of new evaluation tools and propose instead the use of existing tools and the reuse of currently existing tools to build reliability and validity of available tools. This will allow the community of scholars and educators to discern what is not being captured and what is still needed. For this article, tools are organized into five categories: tools to measure (a) cognitive (Table 1), (b) psychomotor (Table 2), and (c) affective learning outcomes (Table 3); (d) interdisciplinary evaluation tools (Table 4); and (e) tools currently under development (Table 5). For this article, it is assumed that there is general agreement that evaluation, taken in the context of HPS, refers to the process of appraising student performance (cognitive, psychomotor, and affective) in a simulation setting.

## Method

Recently published articles in nurse and medical education journals and the two major simulation journals (*Simulation in Healthcare* and *Clinical Simulation in Nursing*) were

used for this article. In addition, instrument developers met at conferences or on the Society for Simulation in Health and International Nursing Association for Clinical Simulation and Learning Listservs were contacted for information about their tools and permission to include them in the article. We do not claim to have produced a comprehensive list of evaluation instruments. These instruments are but current representative samples of many different ways of evaluating clinical abilities. A comprehensive literature review of current simulation evaluation instruments is planned for a future article.

## Learning Domains

Recognition of the challenge of measuring complex learning outcomes in the cognitive, affective, and psychomotor domains is not new. Kolb and Shugart (1984) asserted that evaluation in nursing education is “complicated by the problem of trying to evaluate each domain separately when, in most instances, several behaviors occur simultaneously” (p. 84) but that “simulation tools can be developed which will measure skills in all three behavioral domains” (p. 86). Our research indicates a broad range of sophistication in instrument development at this time. A brief review of each of the domains underscores their collective importance for evaluation in nursing education.

Learning in the cognitive domain may include the acquisition and recall of facts and figures, concepts, and principles (Anderson et al., 2001). Lower level cognitive learning outcomes have been identified as a simple and “quick” target for student learning assessment. The classic lecture followed by a multiple-choice or short-answer test serves as an example of cognitive evaluation. In contrast, HPS offers the opportunity for teaching and evaluation of higher level cognitive functions such as application, synthesis, and evaluation of nursing knowledge.

Psychomotor learning includes the acquisition of technical skills, and it may also incorporate cognitive and affective learning (Jeffries & Norton, 2005). HPS allows educators to teach and evaluate psychomotor skills in a setting that is more realistic than a traditional skills station (such as an IV arm) yet safer than a genuine patient care setting (at the bedside of a critically ill patient).

Learning in the affective domain includes what Oermann and Gaberson (2006, p. 16) described as the development of “values, attitudes, and beliefs [that are] consistent with the standards of professional nursing practice.” Evaluation of learning outcomes in this domain involves identifying whether students have knowledge of these values, attitudes, and beliefs and whether students have internalized these values, attitudes, and beliefs to influence their professional nursing behaviors. HPS offers students the opportunity to demonstrate how their affective learning translates into practice in a simulated patient care scenario. Simulation holds great promise for teaching and

**Table 1** Cognitive Evaluation Tools for Simulation

| Article                                                   | Tool                                                                                                                                                                                                                                             | Reliability & Validity                                                                                                                                                                                                                                                                                                                            | Special Notes                                                                                                                                                                                                                                                                                                                                                                                                                |
|-----------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Clark (2006)                                              | Author designed and copyrighted: <i>Clinical Simulation Grading Rubric</i> for obstetrics. Tool can be modified to fit any scenario.                                                                                                             |                                                                                                                                                                                                                                                                                                                                                   | Originally written for trauma obstetric scenario. Five of Bloom's six cognitive domain categories: knowledge, comprehension, application, analysis, synthesis. Also incorporates Benner's five levels of nursing experience: novice, advanced beginner, competent, proficient, expert. Interrater reliability easy to establish with this tool.<br>Contact: <a href="mailto:mariko.clark@mwsu.edu">mariko.clark@mwsu.edu</a> |
| Gore, Hunt, & Raines (2008)                               | Faculty-devised weighted scoring tool for simulation objectives for beginning clinical student nurses                                                                                                                                            | Not evaluated for reliability or validity                                                                                                                                                                                                                                                                                                         | Good example of how to devise a form specific to objectives of a simulation                                                                                                                                                                                                                                                                                                                                                  |
| Herm, Scott, & Copley (2007)                              | Check-off tool used with nursing students for critical elements: head-to-toe assessment, pain assessment, documentation, patient safety, therapeutic communication, planning prioritizing, implementing interventions, medication administration | Not evaluated for reliability or validity                                                                                                                                                                                                                                                                                                         | Shows side-by-side evaluations of actual clinical evaluation tool and simulation evaluation tool. Students able to pass clinical tool could not perform in simulation on evaluation tool.                                                                                                                                                                                                                                    |
| Hoffman, O'Donnell, & Kim (2007)                          | Basic Knowledge Assessment Tool 6 (BKAT-6): 100-item paper-and-pencil test measures basic recall and application of information in critical care for new graduates or nurses new to critical care                                                | Validated by expert panel. Cronbach's alpha = .80                                                                                                                                                                                                                                                                                                 | Instrument citation: Toth, J.C., & Ritchey, K. (2001). <i>The Basic Knowledge Assessment Tool, version six (BKAT-6) for adult critical care nursing</i> . Washington, DC: Catholic University of America.                                                                                                                                                                                                                    |
| Kuiper, Heinrich, Matthias, Graham, & Bell-Kotwell (2008) | Outcome Present State Test [OPT] Model Debriefing tool designed based on the OPT model and clinical reasoning Web worksheets. Used with 44 undergraduate senior BSN students in a medical-surgical course.                                       | Used since 2003, with refinement continuing. Interrater reliability with this version (Kendall's coefficient: $W = .703$ , $X^2(24) = .573$ , $p = .000$ ) from Kautz, Kuiper, Bartlett, Buck, & Williams, (2007) interrater reliability on this version: .87. Validation of subsection scores shows differences between students ( $p = .001$ ). | Instrument included in article                                                                                                                                                                                                                                                                                                                                                                                               |
| Lasater (2007)                                            | Developed based on Tanner's (2006) clinical judgment model                                                                                                                                                                                       | In development; 53 observations to date with tool. No reliability or validity reported yet.                                                                                                                                                                                                                                                       | Copy of instrument included in article. Developed for evaluation of a single clinical simulation experience with a student nurse directing a team in patient care.                                                                                                                                                                                                                                                           |

(continued on next page)

**Table 1** (continued)

| Article                                          | Tool                                                                                                                                                                                                                                                                                                          | Reliability & Validity                                                                                    | Special Notes                                                                                                                                                   |
|--------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Radhakrishnan, Roche, & Cunningham (2007)        | Clinical Simulation Evaluation Tool designed to measure “safety, basic assessment skills, prioritization, problem-focused assessment, ensuing interventions, delegation and communication in a complex-two patient scenario [p.2]” for second-degree senior nursing students.                                 | Instructor designed. No reliability or validity reported.                                                 | Used with senior nursing students in scenarios involving two patients. Copy of instrument included in article.                                                  |
| Todd, Manz, Hawkins, Parsons, & Hercinger (2008) | Simulation Evaluation Instrument designed using American Academy of Colleges of Nursing core competencies, including 22 behaviors in categories of critical thinking, communication, assessment, and technical skills. Tool tested with senior-level students ( $N = 72$ ) in groups of 4 to 5 in a scenario. | Content validity determined by expert panel and review of literature. Interrater reliability: .85 to .89. | Tool provides a group grade focused on collaboration skills. Extensive explanation of tool development in article. Good for those considering designing a tool. |

evaluating learning in each of these domains. However, challenges still exist.

### Challenges in Instrument Development

Measuring student performance in HPS is similar in complexity to evaluating student performance in any clinical setting. Initial learning outcomes from HPS have focused on the “low-hanging fruit” of self-reported satisfaction and confidence. While easily obtained, these measurements will “not result in a comprehensive and valid assessment of the overall impact of simulation experiences” (Prion, 2008, p. 71). A systematic review by Davis et al. (2006) explored the limitations of self-assessment data and cautioned that such data should be used with great care. Their review of 725 articles comparing self-assessment to external assessment of competency identified 17 studies for inclusion in the final review. Thirteen of these studies demonstrated little, no, or an inverse relationship between self-perception and an external reviewer’s perception. Their findings were true across a large number of professions, in that the least skilled practitioners in the fields assessed were the most self-confident.

Instrument development is a time-consuming and difficult process that requires multiple skill sets and conditions (Stewart & Archibold, 1997). The development of effective evaluation tools follows an organized, rigorous process culminating in a clearly defined end point. Few nursing faculty members who use simulation are also skilled in instrument development. Attempting to train faculty in both simulation

and tool development may contribute to an excessive cognitive load for faculty members. On a positive note, there are emerging doctoral programs in nursing education that are providing the next round of simulation evaluation developers.

HPS evaluation instruments ideally should include some measures for each of the three domains of student performance: cognitive, psychomotor, and affective. An instrument measuring any one of these alone will not be accurate in evaluating overall performance.

Thus, instruments that each measure one or a part of the outcome or outcomes of interest might be employed for a comprehensive evaluation. Further, if outcome expectations for each of these domains have not been clearly defined, any instrument developed to measure student performance will fall short.

### A Word about Reliability and Validity Reporting

All authors should report information about how validity and reliability of an instrument were established, when they are reporting on an instrument in a publication. This will provide future users of the instrument at least minimal information on which to move its development forward. In the articles selected for this review, some data on instrument reliability and validity are reported, although most are not. This article does not attempt to evaluate the psychometrics of any of these instruments. Where psychometric data are not available, an opportunity exists for collaboration among researchers to establish the validity and reliability of existing instruments.

**Table 2** Psychomotor Evaluation Tools for Simulation

| Citation                                                  | Tool                                                                                                                                                                                                         | Reliability & Validity                                                                                                                                                                                                      | Special Notes                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Gibbons et al. (2002)                                     | Twenty-two-item checklist for critical assessment skills with <i>yes, no, not applicable</i> check marks. Used with 27 advanced-practice nursing student participants.                                       | No reliability or validity data reported.                                                                                                                                                                                   | Copy of instrument is included in the article.                                                                                                                                                                                                                                                                                                                                                                                   |
| Murray, Boulet, Ziv, Woodhouse, Kras, & McAllister (2002) | Item analysis of essential behaviors for a specific scenario. Scoring weight from 1 to 4 for each item, based on importance in patient care. Designed for medical simulation. Scoring explanation extensive. | Constructed by expert panel of 3 physicians. Raters consistent in score assignment.                                                                                                                                         | Sample of items for analysis included in article. Extensive explanation of item analysis and computing statistics for a given scenario. May be of interest to those interested in this methodology.                                                                                                                                                                                                                              |
| Rosen, Salas, Silvestri, Wu, & Lazzara (2008)             | Simulation module for assessment of residents targeted event responses (SMARTER). How-to article with sample forms and evaluations.                                                                          | Twenty-nine medical faculty and residents participated in reviewing scenario performance with use of tool. Beginning reliability and validity statistics reported as percentages. Rigorous evaluation statistics to follow. | Uses Accreditation Council for Graduate Medical Education (ACGME) core outcome competencies to derive learning objectives for simulations. Provides a planning model to develop scenarios to test competencies. Knowledge, skills, and attitudes are evaluated, with sample critical events and explicit targeted expected responses. Sample check list and application of the idea to ACGME core competency modeled in article. |

## Validity

Several types of internal instrument validity can be measured by the developers of instruments as they evaluate, revise, and publish new instruments. These include, at a minimum, *content validity* (the appropriateness of sample items and comprehensiveness of the measurement) and *construct validity* (the process of establishing that a particular action adequately represents the concept being evaluated). Criterion validity (a measure of how well any one item or cluster of items in an instrument predicts success on all other measures) may be more difficult for beginning tool developers to address.

## Reliability

Reliability refers to the consistency of results and helps to ensure that the instrument is assessing what the developers intended it to measure. Reliability establishes that the instrument will yield similar results in different circumstances. Reliability provides the consistency that makes validity possible. Tool developers often describe internal

consistency using Cronbach's alpha. It is more difficult to describe interrater or interobserver reliability, that is, to establish that two observers are consistent in their observations, something that is significantly important in evaluating often rapidly evolving simulation scenarios. Several instruments are being developed for use in multiple clinical scenario settings, among them the [Lasater \(2007\)](#) and [Clark \(2006\)](#) instruments. It is best to choose and use an evaluation instrument with known reliability and validity, if one is available, or to work with an instrument developer to devise one.

## Instruments Evaluating the Cognitive Domain

After a careful review of each instrument and article, we grouped all instruments by domains that appeared to fit most clearly. Some of the tools described, especially those in the cognitive group ([Table 1](#)), could fit into multiple domains. The cognitive instruments appeared to be the most comprehensive, with the greatest potential for reaching across learning domains. Several instruments are reported in the literature for evaluating the cognitive domain ([Table 1](#)). [Herm, Scott, and Copley \(2007\)](#) presented a comprehensive

**Table 3** Affective Evaluation Tools for Simulation

| Citation                           | Tool                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Reliability & Validity                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Special Notes                                                                                                              |
|------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| Abdo & Ravert (2006)               | Nineteen-item student satisfaction survey; 4-point Likert-type scale (1 = <i>Strongly disagree</i> , 2 = <i>Disagree</i> , 3 = <i>Agree</i> , and 4 = <i>Strongly agree</i> ). Instrument has three subscales: realism ( $n = 3$ ), transferability ( $n = 3$ ), and value ( $n = 6$ ) for nursing students.                                                                                                                                                                                                                                                                                                                                                                                                                                             | Reliability for overall student satisfaction instrument for sample was .86. Reliability of the subscales for sample: coefficient alphas of .41 for realism, .78 for transferability, and .69 for value. Initial factor analysis demonstrated between 3 and 6 subscales per scree plot and eigenvalues > 1. Overall alpha was .86, but underlying subscales had low alphas, which could indicate too few items or that the subscales were not purposed correctly, given the initial factor analysis. | Instrument originally developed by Feingold, Calaluze, & Kallen (2004)                                                     |
| Arnold et al. (2009)               | Emergency response performance tool (ERPT) measures performance on ventricular tachycardia scenario. Eleven-item confidence tool developed for use with emergency response scenarios for multidisciplinary teams.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 41 RNs divided into 3 groups with varying levels of critical care experience. Interrater reliability high for majority of ERPT items. Measures time to perform required task also. Cronbach's alpha reported at .92 for confidence tool.                                                                                                                                                                                                                                                            | Describes instrument development and testing. Samples of two instruments included in article.                              |
| Dobbs, Sweitzer, & Jeffries (2006) | Simulation Design Scale 20-item tool developed by the NLN/Laerdal multisite project group to measure how well certain constructs from the Jeffries (2005) simulation model were embedded in the simulation. Satisfaction With the Teaching Methodology five-item subscale with a 5-point Likert-type response scale ranging from <i>Strongly agree</i> to <i>Strongly disagree</i> . Self-Confidence in Learning assessed student confidence in caring for an insulin-managed diabetic client. Subscale contains eight items on the same response scales. Self-Perceived Judgment Performance 17-item, Likert-type scale based on items asking students to rate their simulation performance on assessment, decision-making, and problem-solving skills. | Simulation Design Scale Cronbach's alpha for this study was .92. Satisfaction With the Teaching Methodology Cronbach's alpha was .94. Self-Confidence in Learning Cronbach's alpha was .85. Self-Perceived Judgment Performance Cronbach's alpha was .92.                                                                                                                                                                                                                                           | Instruments available from <a href="http://www.nln.org">www.nln.org</a> . Further reliability and validity data available. |

(continued on next page)

**Table 3** (continued)

| Citation                                | Tool                                                                                                                                                                                                                                                                              | Reliability & Validity                                         | Special Notes                                      |
|-----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|----------------------------------------------------|
| Lambton, Pauly O'Neill, & Dudham (2008) | Ten-question Likert-type survey. Four self-report questions. Tool focus is student confidence, communication and collaboration, learning opportunities, error recognition, and transfer from simulation to clinical. Used with second-semester junior pediatric nursing students. | Questions developed by three pediatric nursing content experts | Tool included in article                           |
| McCausland, Curran, & Cataldi (2004)    | Ten-item student nurse evaluation of simulation experience. Scale from 1 to 5.                                                                                                                                                                                                    | Not reported                                                   | Instrument included in article                     |
| Mole & McLafferty (2004)                | Fourteen-item questionnaire with a 4-response Likert-type scale questionnaire written to match scenario aims; 10 to 12 senior nursing students take care of five simulated surgical ward patients for 90 minutes.                                                                 | No reliability or validity reported                            | Self-report questionnaire included in article      |
| Schoening, Sittner, & Todd (2006)       | Faculty designed 10-item evaluation of scenario; 4-point Likert-type scale. Qualitative questions and narrative data collected also. Used with 60 BSN second-semester junior nursing students in a high-risk obstetrics scenario.                                                 | Peer reviewed by two doctorally prepared nurse educators       | Student self-report instrument included in article |

instrument for evaluating critical scenario elements, including head to toe assessment, pain assessment, documentation, patient safety, therapeutic communication, planning and prioritizing, intervention implementation, and medication administration. A unique feature of Herm et al. is a side-by-side comparison of the simulation evaluation tool with the tool used in the clinical environment. Herm et al. noted that students able to pass the clinical evaluation tool could not successfully pass the simulation event.

Clark (2006) used Bloom's taxonomy and Benner's novice-to-expert levels of experience (as cited in Clark, 2006) to develop a cognitive evaluation tool for an obstetrical trauma scenario. Its strong framework has allowed modification to fit multiple scenarios for both undergraduate and graduate nursing students (Lu Sweeney, 2009, personal communication). Hoffman, O'Donnell, and Kim (2007) provided an example of a cognitive tool, the Basic Knowledge Assessment Tool, based on recall and application of information in the adult critical care areas, for use in evaluating nursing knowledge after participation in critical care scenarios. Gore, Hunt, and Raines (2008) developed a scoring tool based on simulation objectives and provided an example of writing a tool based on specific scenario

objectives. Kuiper, Heinrich, Matthias, Graham, and Bell-Kotwell (2008) developed the Outcome Present State Model debriefing tool, based on the model of the same name. Lasater's (2007) Clinical Judgment Evaluation tool provides a methodical framework to evaluate concepts in Tanner's (2006) work. This tool has been critiqued as cumbersome, spawning revisions that attempt to address this issue (Quint & Kardong-Edgren, personal communication, June, 2009; Walker, personal communication, September, 2008). Radhakrishnan, Roche, and Cunningham (2007) designed the Clinical Simulation Evaluation Tool to evaluate student competencies in "safety, basic assessment, prioritization, problem-focused assessment, ensuing interventions, delegations and communication" (p. 2). Todd, Manz, Hawkins, Parsons, and Hercinger (2008) used the American Association of Colleges of Nursing's core new graduate competencies to design a tool that provides a group grade in simulation.

### Instruments Evaluating the Psychomotor Domain

Psychomotor tools are not as commonly discussed in the literature (Table 2). This may be because the vast majority

**Table 4** Group Evaluation Tools for Simulation

| Citation                                                     | Tool                                                                                                                                                             | Reliability & Validity                                                                                                                                           | Special Notes                                                                                                                      |
|--------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
| Kim, Neilipovitz, Cardinal, Chiu, & Clinch (2006)            | Ottawa Crisis Resource Management Global Rating Scale: 7-point Likert-type scale tool for assessment of performance in 5 categories of crew resource management. | Acceptable interrater reliability and construct validity. Initial reporting of data suggests that tool distinguishes varying levels of crew resource management. | Tool included in article                                                                                                           |
| Malec et al. (2007)                                          | Sixteen-item paper-and-pencil tool designed to be used by participants to measure crew resource management skills and their team performance                     | Cronbach's alpha: .85. Construct validity and many other metrics reported.                                                                                       | Tool included in article                                                                                                           |
| Millward & Jeffries (2001)                                   | Team SurveyObjective independent index of team effectiveness. Team is rated by team and independently.                                                           | Detailed construct and concurrent validity addressed in detail. Split half coefficient .93. Subscales analyzed; reliability coefficients between .70 and .93     | Team Survey was developed by Millward & Ramsay in 1998. Focuses on short-term performance. Can be used with a wide range of teams. |
| University of Aberdeen Industrial Psychology Research Center | Anesthesia Non-Technical Skills observation rating tool. Used for team evaluation                                                                                | Full database and Web site devoted to this tool. Downloadable, user friendly. Article about tool available in PDF.                                               | <a href="http://www.abdn.ac.uk/iprc/ants.shtml">http://www.abdn.ac.uk/iprc/ants.shtml</a>                                          |

of psychomotor skills are learned on task trainers in a skills lab. Learning evaluated by simulation focuses on more comprehensive skill sets in a broad range of domains that may incorporate basic psychomotor skills but do not focus on them. Herm et al. (2007) and Murray et al. (2002) provided examples of how this might be done. Gibbons et al. (2002) provided a model of a tool for evaluating critical assessment skills. Rosen, Salas, Silvestri, Wu, & Lazzara (2008) used the Accreditation Council for Graduate Medical Education core competency outcomes to develop a measurement tool for evaluating the responses of medical residents in emergency scenarios. This model might provide a starting point for more advanced nursing care scenarios in emergency medicine.

### Instrument Evaluating the Affective Domain

Table 3 lists tools grouped in the affective domain. The National League for Nursing (NLN) simulation group devised three tools for evaluating self-confidence and satisfaction with learning in a simulation scenario and a simulation design evaluation tool ([www.nln.org/research/toolsandinstruments.htm](http://www.nln.org/research/toolsandinstruments.htm)). Reliability and validity of these tools are widely reported and consistent in the literature. The tools are available free on request from the NLN if the requesting agency is a member of the NLN and at a nominal fee for nonmembers. Dobbs, Sweitzer, and Jeffries (2006) and Kardong-Edgren, Starkweather, and Ward (2008) reported research using these tools. Abdo and Ravert

(2006) devised a 19-item modification of the student satisfaction survey used originally by Feingold, Callaluce, and Kallen (2004). Other researchers (Arnold et al., 2009; Lambton, Pauly O'Neill, & Dudum, 2008; McCausland, Curran, & Cataldi, 2004; Mole & McLafferty, 2004; Schoening, Sittner, & Todd, 2006) have also provided samples of tools in the affective domain. Further affective tools are being developed (see Table 5) by Ackerman (personal communication, June 2009), Reese (personal communication, June 2009), and Reed (personal communication, June 2009).

### Group Evaluation Instruments

Large class sizes and clinical groups necessitate most simulations' being written and designed for use by more than one student at a time (Table 4). Thus, educators need tools to measure learning outcome for both individuals and groups of students. Much like Rosen et al. (2008), Todd, Manz, Hawkins, Parsons, and Hercinger (2008) used their disciplines, in this case nursing's core competencies, to devise a tool designed to provide a group simulation grade.

Interdisciplinary scenarios are becoming more common, both in the hospital and in the educational setting. Several new tools are emerging for evaluating such scenarios. The University of Aberdeen Industrial Psychology Research Center has reported findings of organizations using its Anesthesia Non-Technical Skills (ANTS; n.d.) tool. It provides a full database and Web site devoted to the tool. The tool is downloadable in a PDF format (from [www.abdn.ac](http://www.abdn.ac).



**Table 5** Tools in Development

| Citation                        | Tool                                                                                                                                                          | Reliability & Validity                                                                                       | Special Notes                                                                                                                                                                                                                                                                                                                                                                                                           |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ackerman, A.                    | Simulation Evaluation Tool for Nursing                                                                                                                        | In development. Contact author for details.                                                                  | E-mail: <a href="mailto:Ackerman@Msmc.Edu">Ackerman@Msmc.Edu</a>                                                                                                                                                                                                                                                                                                                                                        |
| Cicero, T., & Mikasa, A.        | Simulation evaluation tool designed for objective, efficient use by faculty and nursing students. Tool can be used for all simulations across the curriculum. | Reliability and validity not reported. Contact authors.                                                      | Copyrighted simulation evaluation tool based on fusion of clinical course objectives and American Association of Colleges of Nursing BSN competencies. Student performance is leveled under each objective using Bloom's Taxonomy. Professional behaviors are also evaluated.<br>E-mail: <a href="mailto:cicero@seattleu.edu">cicero@seattleu.edu</a> or <a href="mailto:mikasaa@seattleu.edu">mikasaa@seattleu.edu</a> |
| Quint, S., & Kardong-Edgren, S. | Modified Lasater tool for simulation or clinical evaluation for nursing students. Modifications to Lasater clinical judgment model tool.                      | Tested with 81 students over 2.5 years. Ongoing data analysis. Reliability coefficient .83. Contact authors. | Tested with medical—surgical students in simulations and as clinical evaluation tool.<br>E-mail: <a href="mailto:sedgren@wsu.edu">sedgren@wsu.edu</a>                                                                                                                                                                                                                                                                   |
| Reese, C.                       | Student's Perception of Effective Teaching in Clinical Simulations Scale for nursing                                                                          | Reliability and validity pending; contact author.                                                            | E-mail: <a href="mailto:Cynthia.Reese@llcc.edu">Cynthia.Reese@llcc.edu</a>                                                                                                                                                                                                                                                                                                                                              |
| Reed, S.                        | Student perception of debriefing tool for nursing. In development.                                                                                            | Reliability and validity not reported; contact author.                                                       | E-mail: <a href="mailto:Shelly-Reed@byu.edu">Shelly-Reed@byu.edu</a>                                                                                                                                                                                                                                                                                                                                                    |
| Sweeney, L.                     | Modification of Clark (2006) tool, Sweeney-Clark's Simulation Performance Evaluation Rubric; can be used with any scenario                                    | Spring 2009 data on 66 nursing students; Cronbach's alpha on all items > .86                                 | E-mail: <a href="mailto:lu.sweeney@dominican.edu">lu.sweeney@dominican.edu</a>                                                                                                                                                                                                                                                                                                                                          |
| Walker, M. B.                   | Two-page thorough model for adaptation of simulation evaluation tool. Significant modification of Lasater clinical judgment model tool.                       | No validity reported; reliability coefficient = .83.                                                         | Tool included in accompanying folder;<br>e-mail: <a href="mailto:m.walker@tcu.edu">m.walker@tcu.edu</a>                                                                                                                                                                                                                                                                                                                 |

[uk/iprc/ants.shtml](#)) and concentrates on nontechnical skills in crew resource management. Malec et al. (2007) developed a tool for a group evaluation of team management and performance. Millward and Jeffries (2001) developed a survey that is scored by an independent reviewer evaluating team performance. Kim, Neilipovitz, Cardinal, Chiu, and Clinch (2006) reported on the Ottawa Crisis Resource Management Global Rating Scale, a 7-point Likert-type scale tool for performance assessment in five categories of crew resource management.

### Instruments in Development

Seven instruments, either new or modifications of existing instruments, are listed in Table 5. Some are being developed as part of student doctoral dissertations. Other

developers are pursuing their own research interests. All will need participants willing to help with evaluation and refinement of these instruments in multiple settings.

### Recommendations

This article provides information on 22 simulation evaluation tools, available within published articles, for evaluating clinical simulations. Most of them do not report reliability and validity figures. Further use and development of these published simulation evaluation instruments are of the highest importance. Multiple uses and reuses of these instruments in multiple regions by various nursing programs will help determine reliability and validity in multiple settings.

Tools coming closest to addressing the three learning domains simultaneously appear to be Herm et al. (2007), Lasater (2007), Radhakrishnan et al. (2007), and Todd et al. (2008). Reuse and enhanced validation of reliability and validity figures in multiple venues with multiple users would make significant contributions to the science of nursing simulation.

Nursing faculty who use simulation would be wise to actively seek others skilled in tool evaluation to use and refine currently published tools in order to build reliability and validity data and then publish and present this information. Perhaps the International Nursing Association for Clinical Simulation and Learning Web site might be a central and easily accessible location for a database, readily available to members, of current tools, reliability and validity data, and tools in production. A moratorium on self-report and satisfaction instrumentation development is suggested as the literature does not suggest that these data provide particularly useful information (Davis et al., 2006). Experimentation with the use of multiple instruments is required to evaluate a simulation to capture all the learning domains at once.

## Conclusions

It is the hope of the authors that users of simulation evaluation instruments will use and reuse currently available instruments and participate in multisite studies using these instruments. In this way, large sample sizes in more than one geographic location will provide valid data for reliability and validity statistics for tools, moving simulation science forward. We do not want to build a mile-wide and inch-deep evaluation landscape, a phenomenon that arguably has occurred in the clinical evaluation literature. Further refinement of both simulation and clinical evaluation instruments is a high priority, made easier by the advent of sophisticated simulation techniques. Questions remain about actions in simulation carrying over into the clinical arena. The ability to see actions and reactions to clinical events and the ability to evaluate them without risking patient safety are major steps forward in evaluating future clinical actions.

## References

- Abdo, A., & Ravert, P. (2006). Student satisfaction with simulation experiences. *Clinical Simulation in Nursing Education*, 2(1). Retrieved on 29 September, 2009 from <http://download.journals.elsevierhealth.com/pdfs/journals/1876-1399/PIIS1876139909004393.pdf>.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., & Pintrich, P. R. et al. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Anesthesia Non-Technical Skills (ANTS) Observation Rating Tool. (n.d.). University of Aberdeen Industrial Psychology Research Center. Retrieved May 25, 2009, from [www.abdn.ac.uk/iprc/ants.shtml](http://www.abdn.ac.uk/iprc/ants.shtml)
- Arnold, J. J., Johnson, L. M., Tucker, S. J., Malec, J. F., Henrickson, S. E., & Dunn, W. F. (2009). Evaluation tools in simulation learning: Performance and self-efficacy in emergency response. *Clinical Simulation in Nursing*, 5(1). Doi:10.1016/j.ecns.2008.10.003.
- Benner, P., Sheets, V., Uris, P., Malloch, K., Schwed, K., & Jamison, D. (2002). Individual, practice, and systems causes of errors in nursing. *Journal of Nursing Administration*, 32(10), 509-523.
- Clark, M. (2006). Evaluating an obstetric trauma scenario. *Clinical Simulation in Nursing Education*, 2, e13-e16. Retrieved on 29 September 2009 from <http://download.journals.elsevierhealth.com/pdfs/journals/1876-1399/PIIS18761399090044587.pdf>.
- Davis, D., Mazmanian, P., Fordis, M., Harrison, R. V., Thorpe, K., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence. *Journal of the American Medical Association*, 296(9), 1094-1102.
- Del Bueno, D. (2005). A crisis in critical thinking. *Nursing Education Perspectives*, 26(5), 278-282.
- Dobbs, C., Sweitzer, V., & Jeffries, P. (2006). Testing simulation design features using an insulin management simulation in nursing education. *Clinical Simulation in Nursing Education*, 2(1), e17-e22. Retrieved on 29 September 2009 from <http://download.journals.elsevierhealth.com/pdfs/journals/1876-1399/PIIS1876139909004423.pdf>.
- Dunn, W. (2004). Educational theory: does simulation really fit? In W. Dunn (Ed.), *Simulators in critical care and beyond* (pp. 15-19). Des Plaines, IL: Society of Critical Care Medicine.
- Feingold, C. E., Callaluce, M., & Kallen, M. A. (2004). Computerized patient model and simulated clinical experiences: Evaluation with baccalaureate nursing students. *Journal of Nursing Education*, 43(4), 156-163.
- Gaba, D. (2004). The future vision of simulation in health care. *Quality and Safety in Health Care*, 13(2), 2-10.
- Gibbons, S., Adamo, G., Padden, D., Ricciardi, R., Graziano, M., Levine, E., et al. (2002). Clinical evaluations in advanced nursing education: Using standardized patients in health assessment. *Journal of Nursing Education*, 41(5), 215-221.
- Gore, T., Hunt, C., & Raines, K. (2008). Mock hospital unit simulation: A teaching strategy to promote safe patient care. *Clinical Simulation in Nursing*, 4(5). doi:10.1016/j.ecns.2008.08.006.
- Herm, S., Scott, K., & Copley. (2007). Sim'ational revelations. *Clinical Simulation in Nursing Education*, 3(1), e25-e30. Retrieved on 29 September 2009 from <http://download.journals.elsevierhealth.com/pdfs/journals/1876-1399/PIIS1876139909004678.pdf>.
- Hoffman, R., O'Donnell, J., & Kim, Y. (2007). The effects of human patient simulators on basic knowledge in critical care nursing with undergraduate senior nursing students. *Simulation in Healthcare*, 2(2), 110-114.
- Jeffries, P. R., (2005). A framework for designing, implementing, and evaluating simulations used as teaching strategies in nursing. *Nursing Education Perspectives*, 26(2), 96-103.
- Jeffries, P. R., & Norton, B. (2005). Selecting learning experiences to achieve curriculum outcomes. In D. M. Billings & J. A. Halstead (Eds.), *Teaching in nursing: A guide for faculty* (2nd ed., pp. 187-212). St. Louis, MO: Elsevier.
- Kardong-Edgren, S., Starkweather, A., & Ward, L. (2008). The integration of simulation into a clinical foundations of nursing course: Student and faculty perspectives. *International Journal of Nursing Education Scholarship*, 5(1), Article 26. Retrieved September 21, 2009, from <http://www.bepress.com/ijnes/vol5/iss1/art26>
- Kim, J., Neillipovitz, D., Cardinal, P., Chiu, M., & Clinch, J. (2006). A pilot study using high-fidelity simulation to formally evaluate performance in the resuscitation of critically ill patients: The University of Ottawa Critical Care Medicine, High-Fidelity Simulation, and Crisis Resource Management I Study. *Critical Care Medicine*, 34(8), 2167-2174.
- Kolb, S. E., & Shugart, E. B. (1984). Evaluation: Is simulation the answer. *Journal of Nursing Education*, 23(2), 84-86.
- Kuiper, R. A., Heinrich, C., Matthias, A., Graham, M., & Bell-Kotwell, L. (2008). Debriefing with the OPT model of clinical reasoning during high fidelity patient simulation. *International Journal of Nursing Education*

- Scholarship*, 5(1), Article 17. Retrieved on 29 September 2009 from <http://www.bepress.com/ijnes/vol5/iss1/art17>.
- Lambton, J., Pauly O'Neill, S., & Dudum, T. (2008). Simulation as a strategy to teach clinical pediatrics within a nursing curriculum. *Clinical Simulation in Nursing*, 4(3), Retrieved on 29 September from doi:10.1016/j.ecns.2008.08.001.
- Lasater, K. (2007). Clinical judgment using simulation to create an assessment rubric. *Journal of Nursing Education*, 46(11), 496-503.
- Malec, J., Torsher, L., Dunn, W., Wiegman, D., Arnold, J., Brown, D., et al. (2007). The Mayo high performance teamwork scale: Reliability and validity for evaluating key crew resource management skills. *Simulation in Healthcare*, 2(1), 4-10.
- McCausland, L., Curran, C., & Cataldi, P. (2004). Use of a human simulator for undergraduate nurse education. *International Journal of Nursing Education Scholarship*, 1(1), Article 23. Retrieved 29 September 2009 from <http://www.bepress.com/ijnes/vol1/iss1/art23>.
- Millward, L., & Jeffries, N. (2001). The team survey: A tool for health care team development. *Journal of Advanced Nursing*, 35(2), 276-287.
- Mole, L., & McLafferty, I. (2004). Evaluating a simulated ward exercise for third year student nurses. *Nurse Education in Practice*, 4, 91-99.
- Murray, D., Boulet, J., Ziv, A., Woodhouse, J., Kras, J., & McAllister, J. (2002). An acute care skills evaluation for graduating medical students: A pilot study using clinical simulation. *Medical Education*, 36(9), 833-841.
- Oermann, M. H., & Gaberson, K. B. (2006). *Evaluation and testing in nursing education* (2nd ed., Table 1.1 on pages 12-13). New York: Springer.
- Prion, S. (2008). A practical framework for evaluating the impact of clinical simulation experiences in prelicensure nursing education. *Clinical Simulation in Nursing*, 4(5), 65-78.
- Radhakrishnan, K., Roche, J., & Cunningham, H. (2007). Measuring clinical practice parameters with human patient simulation: A pilot study. *International Journal of Nursing Education Scholarship*, 4(1), Article 8. Retrieved 29 September, 2009 from <http://www.bepress.com/ijnes/vol4/iss1/art8>.
- Rosen, M., Salas, E., Silvestri, S., Wu, T., & Lazzara, E. (2008). A measurement tool for simulation-based training in emergency medicine: The simulation module for assessment of resident targeted event responses (SMARTER) approach. *Simulation in Healthcare*, 3(3), 170-179.
- Schoening, A., Sittner, B., & Todd, M. (2006). Simulated clinical experience: Nursing student's perceptions and the educator's role. *Nurse Educator* 31(6), 253-258.
- Spector, N. (2006). Systematic review of nursing education outcomes: An evolving review. NCSBN, April, 2006; Chicago, IL. Retrieved on 29 September, 2009 from [https://www.ncsbn.org/Final\\_Sys\\_Review\\_04\\_06.pdf](https://www.ncsbn.org/Final_Sys_Review_04_06.pdf).
- Stewart, B. J., & Archbold, P. G. (1997). A new look for measurement validity. *Journal of Nursing Education*, 36(3), 99-101.
- Tanner, C. (2006). Thinking like a nurse: A research-based model of clinical judgment in nursing. *Journal of Nursing Education*, 45(6), 204-211.
- Todd, M., Manz, J., Hawkins, K., Parsons, M., & Hercinger, M. (2008). The development of a quantitative evaluation tool for simulation in nursing education. *International Journal of Nursing Education Scholarship*, 5(1), Article 41. Retrieved 29 September 2009 from <http://www.bepress.com/ijnes/vol5/iss1/art41>.