Assessment instrument validation for critical clinical competencies: pediatric-neonatal intubation and cholinergic crisis management

Andreatta P, Klotz J
University of Minnesota Medical School
Minneapolis, MN
pandreat@umn.edu, jjklotz@umn.edu

Madsen JM, Hurst CG
U.S. Army Medical Research Institute for Chemical Defense (USAMRICD)
Aberdeen, MD
james.m.madsen2.mil@mail.mil, charles.g.hurst.civ@mail.mil

Talbot TB
Telemedicine & Advanced Technology Research Center (TATRC)
Fort Detrick, MD
thomas.talbot@tatrc.org

ABSTRACT

Background: Military and civilian first-responders must be able to recognize and effectively manage casualties that necessitate immediate application of critical clinical competencies. Two examples of these critical competencies are the clinical management of injuries resulting from nerve agents and difficult intubation, especially for pediatric or neonatal patients. The opportunity to learn and practice the necessary skills for these rare, but urgent, situations is complicated by the limited ability to replicate essential situational factors that influence performance in the applied clinical environment. Simulation-based training may resolve some of these challenges, however it is imperative that evidence be captured to document the achievement of performance competencies in the training environment that transfer to applied clinical care. The purpose of this study was to establish psychometric characteristics for competency assessment instruments associated with two such critical competencies: management of cholinergic crisis and pediatric-neonatal intubation. **Methods:** To inform the development of assessment instruments, we conducted comprehensive task analyses across each performance domain (knowledge, performance). Expert review confirmed content validity. Construct validity was established using the instruments to differentiate between the performance abilities of practitioners with variable experience (novice through expert). Purposively selected first-responder subjects for pediatric-neonatal intubation (N=214) and cholinergic crisis management (N=123) were stratified by level of experience performing the requisite clinical competencies. All subjects completed knowledge and performance assessments. Reliability was established using test-retest (Pearson correlation) and internal consistency (Cronbach's alpha) for knowledge and performance assessments. **Results:** Significantly higher scores for subjects with greater levels of experience, compared to those with less experience established construct validity for each assessment instrument ($p < .01$). Significant correlations between test-retest outcomes indicated measurement reliability $p < .01$. Cronbach's alpha for knowledge and performance scores demonstrated excellent internal consistency. **Conclusions:** Psychometric evidence establishes the value of assessment for identifying and remedying critical competency performance gaps.

ABOUT THE AUTHORS

Pamela Andreatta, EdD, PhD is an Associate Professor and SimPORTAL/CREST Director of Human Performance Research in Medical Education at the University of Minnesota Medical School, and Adjunct Associate Professor of Obstetrics & Gynecology at the University of Michigan Medical School.

Jessica Klotz, BSc is a research associate with SimPORTAL/CREST at the University of Minnesota Medical School.

James M. Madsen, MD, MPH, FCAP, FACOEM, COL, MC-FS, USA is Deputy Chief of the Chemical Casualty Care Division (CCCD) U.S. Army Medical Research Institute for Chemical Defense (USAMRICD); Acting Consultant in Chemical Casualty Care to the Surgeon General of the Army; Adjunct Associate Professor of Preventive Medicine and Biometrics, Assistant Professor of Military and Emergency Medicine, and Assistant Professor of Pathology at the Uniformed Services University of the Health Sciences (USUHS).

Charles G. Hurst, MD,COL (ret.), MC, USA is Director of the Chemical Casualty Care Division (CCCD) U.S. Army Medical Research Institute for Chemical Defense (USAMRICD).

Thomas B. Talbot, MD, MS, FAAP is the Principal Medical Expert at the University of Southern California Institute for Creative Technologies. A pediatrician, he is an Associate Research Professor in Medical Education at USC Keck School of Medicine, and Chief Scientist of the Armed Forces Simulation Institute for Medicine at the US Army Telemedicine and Advanced Technology Research Center.

Assessment instrument validation for critical clinical competencies: pediatric-neonatal intubation and cholinergic crisis management

Andreatta P, Klotz J
University of Minnesota Medical School
Minneapolis, MN
pandreat@umn.edu, jjklotz@umn.edu

Madsen JM, Hurst CG
U.S. Army Medical Research Institute for Chemical Defense (USAMRICD)
Aberdeen, MD
james.m.madsen2.mil@mail.mil, charles.g.hurst.civ@mail.mil

Talbot TB
Telemedicine & Advanced Technology Research Center (TATRC)
Fort Detrick, MD
thomas.talbot@tatrc.org

INTRODUCTION

The increased burden of global conflict and the war on terrorism has correspondingly increased the necessity for both military and civilian medical personnel to rapidly and effectively respond to mass disasters and crises-related injuries (Brennan, Waeckerle, Sharp, & Lillibridge, 1999; Cannard, 2006; Kadivar & Adams, 1991; Khan, Morse, & Lillibridge, 2000; Macilwain, 1993; Morita et al., 1995; Murray, 2006; Ohbu et al., 1997; Okumura et al., 1996; Rodgers, 1998; Sidell, 1997; Waeckerle, 2000). Clinicians who respond to crises must be able to provide the highest level of care to their patients, however the opportunities for contextually relevant training are limited. The relatively rare occurrence of these situations coupled with the ethical responsibility to provide timely and accurate care to patients in crisis makes it difficult to use the applied clinical environment for training. The result is that it is extremely difficult for medical personnel to acquire and maintain the skills they need to competently perform in these situations.

To address this challenge, simulation-based training methods are sometimes used as a proxy for clinical contexts; the assumption being that knowledge and skills gained using high fidelity simulation will transfer to applied performance on humans. There is both empirical and theoretical support for this reasoning. Experiential learning theory supports the premise that it is an individual's processing of stimuli from the environment and the resulting cognitive structures that produce adaptive behavior, rather than the stimuli themselves (Bruner, 1990; Dewey, 1933/1998; Harnad, 1982; Vygotsky, 1978). Other factors, such as situational variables, emotions, and consequences, all play an important role in the production of overt, adaptive behavior and these factors are typically absent from traditional learning environments that rely purely on intellectual discourse (Doyle, 1997). The theory is that through direct engagement in a contextually relevant environment, learners will create mental representations such as attitudes, mental models, scripts, and schemas that correspondingly transform related behaviors.

This theoretical framework serves all clinical training programs well, but particularly crisis and emergency medical training. Training that focuses solely on knowledge components misses the vital mastery of associated skills and affective elements that are required of trainees in these contexts. Especially in a mass casualty environment, cognitive dissonance resulting from affective overload can interfere in the application of knowledge and skills (Arad et al., 1993; Berkenstadt, Arad, Nahtomi, & Atsmon, 1999; Fine & Kobrick, 1987; Hendler et al., 2000; Humara, 1999; King & Frelin, 1984; Mandler, 1979; Smith & Nye, 1989; Smolander, Louhevaara, & Korhonen, 1985; VanGemmert & VanGalen, 1997). Cognitive dissonance refers to the psychological effects that occur when an individual perceives a logical inconsistency among his/her cognitions and needs to accommodate new information in order to reconcile current beliefs and reality. In a disaster situation, the environment and scale of the crisis situation may be unlike anything the clinician has experienced or trained for, and that contradiction could lead to dissonance that manifests as anxiety, anger, stress and other negative emotions that adversely affect performance. Therefore, training associated with the development of crises and mass disaster abilities is frequently conducted using contextually relevant factors as part of the training environment (e.g. mock disaster drills), to achieve performance mastery that is less susceptible to stressor-related dissonance (Andreatta et al., 2010; Dandoy & Goldstein, 1990; Filner, 2009; Inzana, Driskell, Salas, & Johnston, 1996).

Knowledge-based evidence and self-reported affirmations from both trainees and trainers suggest that although these instructional methods are perceived to be effective for learning how to manage clinical crises, this evidence omits contextually-based performance assessment (Murray, 2006). Simulation-based training may resolve some of these

challenges, however it is imperative that evidence be captured to document achievement of performance competencies in the training environment that transfer to applied clinical care. The use of simulated patient mannequins for cardiopulmonary resuscitation training is likely the most documented application of simulation-based methods in clinical education, and there is substantial evidence to support their training benefits in those contexts (Andreatta, Saxton, Thompson, & Annich, 2010; Campbell, Barozzino, Farrugia, & Sgro, 2009; Eppich, Adler, & McGaghie, 2006; Falck, Escobedo, Baillargeon, Villard, & Hall, 2003; Halamek et al., 2000; Hall et al, 2005; Mayo, Hackney, Mueck, Ribaudo, & Schneider, 2004; Overly, Sudikoff, & Shapiro, 2007; Vadodaria, Gandhi, & McIndoe, 2004). Although simulation technology has impressive capabilities and great potential for clinical training, at present it largely remains unverified for high-stakes clinical applications such as responding to a mass casualty or other crisis situation.

Medical education has historically relied on intangible measures of clinical performance and at present, there is little scientific evidence to unequivocally assert effectiveness of clinical training. The lack of accepted standards of performance and measurement for most clinical processes and procedures is a substantial obstacle to evaluating training outcomes. A benefit of simulation is that it provides a consistent platform to support the development of quantitative metrics that will facilitate the establishment of performance standards (Andreatta & Gruppen, 2009). Valid data are critical for any substantive scientific inquiry and there are well-established methods for developing the metrics to capture relevant data (American Psychological Association, 1999). Although the validation process for confirming metrics associated with clinical performance of a targeted task or procedure is time consuming and resource intensive, it is nonetheless mandatory for the accurate assessment and evaluation of any construct. Without valid metrics, any examination of training effectiveness remains qualitative conjecture.

This presents a challenge for training military and civilian medical personnel to be able to recognize and effectively manage casualties during crisis situations, where timely and competent clinical performance is essential, but evidence of competency is undefined and unmeasured. Two examples of these critical competencies are the clinical management of injuries resulting from nerve agents (cholinergic crisis) and difficult intubation, especially for pediatric or neonatal patients. Nerve agents are typically associated with mass casualties and require medical personnel to enter an unstable environment while wearing protective gear that simultaneously overtaxes and limits their perceptual capacities (Arad et al., 1993; Berkenstadt et al., 1999; Fine & Kobrick, 1987; Hendler et al., 2000; King & Frelin, 1984; Smolander et al., 1985). Intubation of adult patients is a complicated and stressful procedure for clinicians, however intubation of pediatric and neonatal patients is especially challenging. In addition to anatomical and physiological differences between adults, children and neonates, affective elements associated with the need to intubate a child or infant during a crisis situation are substantial, and success rates for pediatric and neonatal intubation remain sub-optimal (Bismilla et al., 2010; Falck et al., 2003; Leone, Rich, & Finer, 2005; O'Donnell, Kamlin, Davis, & Morley, 2006; Sanders et al., 2013).

The purpose of this study was to establish psychometric characteristics for competency assessment instruments associated with the management of cholinergic crisis and ability to perform pediatric and neonatal intubation. The study objectives were to determine the psychometric characteristics that describe the measurement accuracy and constraints of assessment instruments for 1) clinical management of cholinergic crisis and 2) pediatric and neonatal intubation.

METHODS

Institutional review and approvals were secured from the University of Michigan, University of Minnesota, and U.S. Army Medical Research and Materiel Command (HRPO). All study related activities took place at the University of Michigan and the University of Minnesota. The study involved clinical activities and therefore a purposive sample of clinicians with varying levels of expertise was recruited. All subjects completed written informed consent prior to participating in study related activities.

Assessment Metrics: Clinical Management of Cholinergic Crisis

We performed a comprehensive literature review to examine assessment mechanisms and performance standards for the clinical management of cholinergic crisis. This literature informed the development of a comprehensive task analysis that included essential knowledge and performance skills required to accurately manage a nerve agent

casualty. We used this comprehensive task analysis to create a knowledge assessment instrument designed to test knowledge associated with managing a nerve agent casualty. The knowledge test is comprised of 23 multiple choice and short-answer questions, with a maximum score of 39 total points. We also referenced the task analysis to create a semi-objective performance assessment instrument with behavioral markers for assessing the clinical management of nerve agent casualties at multiple points, and in aggregate with a maximum total score of 45 points.

We collected evidence of measurement accuracy and consistency for each of the assessment instruments. Five experts in the clinical management of cholinergic crisis mass casualty events reviewed the assessment instruments for content accuracy, phrasing, sequencing, and scoring. To establish construct validity, we recruited adult, English speaking first responder medical personnel from a 100-mile radius around Minneapolis, Minnesota to participate in a simulated cholinergic crisis event (N=123). We categorized subjects by whether or not they had prior training in the management of chemical casualties: *No Experience* and *Experience.* All subjects completed the knowledge test independently, however two pairs of trained raters scored performance assessments while the subjects managed simulated casualties (actors). Raters were trained until they achieved and maintained inter-rater agreement above .90. We used analysis of variance statistics to compare the scores for knowledge and performance assessments between experience levels to determine if there were statistically significant differences such that those with more experience scored better.

After the first assessment cycle, all subjects were asked to complete a second assessment cycle to collect test-retest reliability evidence (Pearson correlation) for knowledge and performance assessments; order effects were controlled for through randomization within (item order) and between (instrument order) assessment administrations. Internal consistency was also calculated for knowledge and performance assessments (Cronbach's alpha).

All statistical analyses were conducted using SPSS-Statistics v21 (IBM, Armonk, New York, USA). Statistical significance was set at $p < .05$.

Assessment Metrics: Pediatric and Neonatal Intubation

We consulted the published literature to identify assessment tools, methods, and performance standards for pediatric or neonatal intubation (Klotz, Dooley-Hash, House, & Andreatta, in press). We identified several procedural checklists, however they did not include psychometric properties to convey information about measurement accuracy or reproducibility constraints. These checklists informed our process of instrument development, but we relied heavily on the American Heart Association Pediatric Advanced Life Support and Neonatal Resuscitation Program clinical guidelines and protocols to develop comprehensive task analyses of the knowledge and skills required to perform pediatric and neonatal intubation (American Heart Association, 2006; American Heart Association & American Academy of Pediatrics, 2006). We used these task analyses data to create a knowledge test comprised of 10 short-answer questions about pediatric and neonatal intubation, with a maximum score of 34 total points. We then created two semi-objective performance assessment instruments with behavioral markers to assess the ability to intubate pediatric and neonatal patients at multiple procedural points, and in aggregate. Each performance assessment has a maximum total score of 46 points.

We collected evidence of measurement accuracy and consistency for each of the assessment instruments. Five experts in pediatric and neonatal intubation reviewed all assessment instruments for the accuracy of content, phrasing, sequencing, and scoring.

We recruited adult, English speaking medical personnel from a 100-mile radius around Ann Arbor, Michigan to participate in pediatric and neonatal intubation activities designed to establish construct validity of measurement (N=214). We categorized subjects according to prior pediatric or neonatal intubation experience: *Low* (no independent experience intubating a pediatric or neonatal patient), *Moderate* (independent experience intubating a pediatric or neonatal patient), and *High* (routinely and independently intubates pediatric or neonatal patients).

All subjects completed the knowledge test and were scored by two pairs of trained raters while performing intubation on a pediatric mannequin simulator (SimBaby™, Laerdal Medical Corporation, Stavanger, Norway), and neonatal mannequin simulators (SimNewB™, Laerdal Medical Corporation, Stavanger, Norway). Inter-rater consistency between the four physician raters was .90. We used analysis of variance statistics to compare the scores for the knowledge test and both performance assessments between experience levels to determine if greater experience corresponded with more favorable scores.

After the first assessment cycle, all subjects completed a second assessment cycle to collect test-retest reliability evidence (Pearson correlation) for the knowledge and two performance assessments; order effects were controlled through randomization within (item order) and between (instrument order) assessment administrations. Internal consistency was also calculated for all assessment instruments (Cronbach's alpha).

Statistical analyses were conducted using SPSS-Statistics v21 (IBM, Armonk, New York, USA). Statistical significance was set at $p < .05$.

RESULTS

The five experts from each competency area (cholinergic crisis management and pediatric-neonatal intubation) reviewed and, after minor modifications, agreed to the accuracy of content, phrasing, sequencing, and scoring of all instruments.

Cholinergic Crisis Management

Construct validity was supported for assessment of cholinergic crisis management by statistically significant differences between subjects with and without prior training in the clinical management of a cholinergic event (Figure 1). Scores for greater levels of experience exceeded those of lower experience levels for knowledge assessment ($F(1, 122) = 8.89$, $p = .003$), and performance assessment ($F(1, 122) = 170.34$, $p = .000$). Significant correlations between test-retest outcomes for both assessment instruments indicated measurement reliability ($p < .001$). Test-retest correlations for knowledge and performance were .97 and .98, respectively. The assessment instruments demonstrated excellent internal consistency (Cronbach's alpha) for knowledge (.69) and performance (.90).

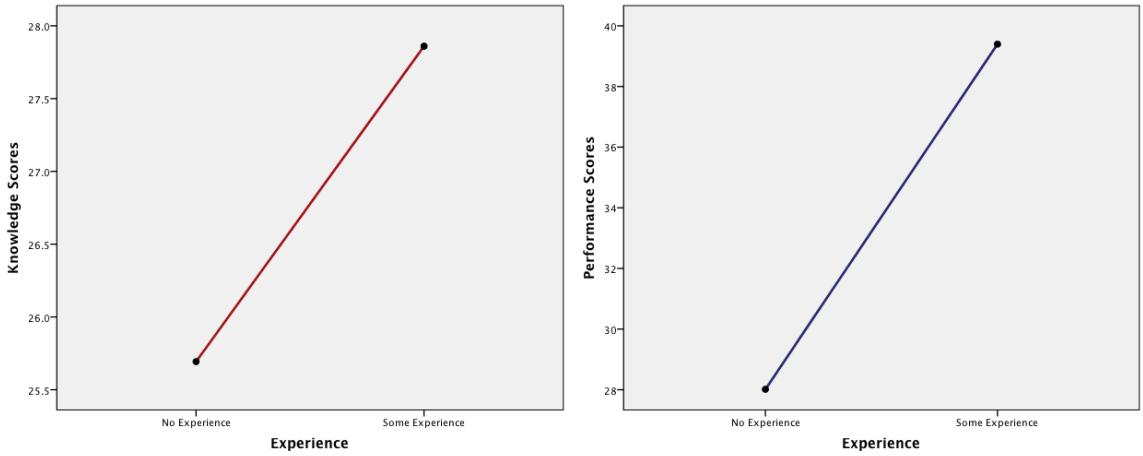

Figure 1: Construct validity for knowledge and performance assessments.

Pediatric-Neonatal Instruments

For pediatric and neonatal intubation, construct validity was supported by statistically significant differences between subjects with varying levels of experience, such that the scores for greater levels of experience exceeded those of lower experience levels (Figure 2). Between group ANOVA outcomes for experience levels were as follows: knowledge assessment ($F(2,212) = 107.45$, $p = .000$), pediatric performance assessment)$F(2,212) = 115.18$, $p = .000$), and neonatal performance assessment ($F(2,212) = 111.89$, $p = .000$). Bonferroni post-hoc analyses confirmed statistical significance between all experience levels for each assessment instrument ($p = .000$). Significant correlations between test-retest outcomes for all assessment instruments indicated measurement reliability ($p < .01$). Test-retest correlations were knowledge (.99), pediatric performance (.96), and neonatal Performance (.96). The assessment instruments demonstrated excellent internal consistency (Cronbach's alpha) for knowledge (.92), pediatric performance (.84) and neonatal performance (.85).

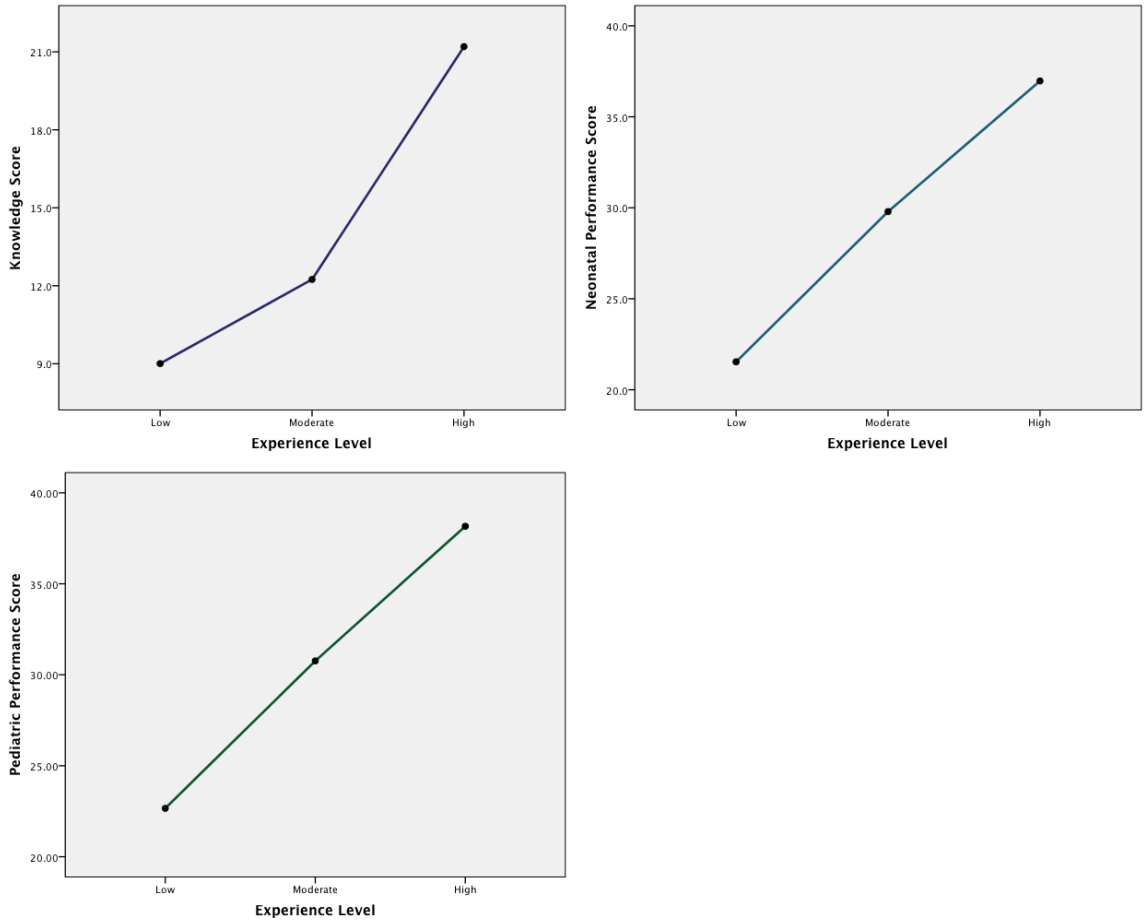

Figure 2: Construct validity for pediatric-neonatal knowledge and performance assessments.

DISCUSSION

Constrained opportunities for medical personnel to practice skills associated with the management of mass disaster and rare event casualties make it difficult to acquire and maintain these critical competencies. Although training and certification programs may help providers learn knowledge and skills, without evidence of performance abilities it is difficult to determine the extent to which that knowledge transfers into applied practice. Gaps between training outcomes and applied performance data suggest that there is room for improvement in how we facilitate training to assure clinical competency (Irvine & Martin, 2014; Nestel, Groom, Eikeland-Husebø, & O'Donnell, 2011). Our results demonstrate the value of using assessment instruments with psychometric validity and reliability evidence for providing feedback about performance competency. The data derived from our assessment instruments provided sufficient detail to differentiate between the experience levels of clinicians and demonstrated similar findings to previous work describing performance gaps between interns, senior residents, and attending physicians (Falck et al., 2003; Sanders et al., 2013).

Implications

The importance of being able to measure performance to the extent that it provides feedback to clinicians about their abilities at any point in time, where improvement is necessary, and when refreshment training or practice is recommended, provides valuable information to clinicians and training administrators. Training programs could then benefit from tailoring requirements or content to a clinician's competence in order to make best use of resources and training time. Performance data also facilitates standard setting, the metrics of which are useful for developing instructional opportunities that support deliberate practice with performance feedback. With performance standards, individuals can hone their performance until they achieve a predetermined level of mastery, or competency threshold. Simulation-based training methods provide on-demand options for prescribed and standardized instruction that facilitates the deliberate practice that allows clinicians to develop and improve their performance abilities. Improvements in the technological characteristics of mannequin simulators; such as anatomical, physiological, and tissue fidelity; would substantially add to their prospective benefits. (Klotz et al., in press).

Limitations and Next Steps

A limitation of this study is that we did not attempt to evaluate performance during applied clinical care, so it remains to be determined whether competent performance in the simulated context translates to competent performance managing actual patients in crisis. This is an important next step to establish the predictive validity of these assessment instruments, but it was beyond the scope of this study. Theoretically the mental models created from a relevant training environment will transfer to a clinical care setting, and future work will characterize the extent of transfer in different clinical contexts.

CONCLUSIONS

The results of this study confirm the developed assessment instruments yield sufficient evidence of psychometric character to support the measurement of competency in each of the designated performance domains: 1) the clinical management of a mass casualty cholinergic crisis event and 2) pediatric and neonatal intubation. Assessment instruments that are supported by strong psychometric evidence, such as the validity and reliability of measurement, facilitate standard setting for critical competencies. Precision assessment instruments and performance standards help identify individual performance gaps for remediation, as well as inform the development of on-demand instructional programs to effectively train medical personnel.

ACKNOWLEDGEMENTS

The authors acknowledge the support of: Cameron Johnson and Laszlo Alberti at the University of Minnesota Medical School; The City of Plymouth Fire Department, Plymouth, MN; Stephen Barnes, MD, and Alex Bukoski, DVM, PhD, ACVAA at The University of Missouri School of Medicine and College of Veterinary Medicine, Columbia, MO; Natalie Dowd at Synaptic Design, Minneapolis, MN; Joe Hauptman, DVM, and Bea Biddinger, LVT, VTS(ECC) at Michigan State University College of Veterinary Medicine, East Lansing, MI.. This work was funded by the United States Army Medical Research and Materiel Command (USAMRMC) and supported by Telemedicine & Advanced Technology Research Center (TATRC) as part of the Joint Program Committee 1 Medical Training and Health Information Sciences (JPC-1).

REFERENCES

American Heart Association. (2006). 2005 American Heart Association (AHA) guidelines for cardiopulmonary resuscitation (CPR) and emergency cardiovascular care (ECC) of pediatric and neonatal patients: Pediatric advanced life support. *Pediatrics, 117*(5), e1005-e1028. http://dx.doi.org/10.1542/peds.2006-0346

American Heart Association & American Academy of Pediatrics. (2006). 2005 American Heart Association (AHA) guidelines for cardiopulmonary resuscitation (CPR) and emergency cardiovascular care (ECC) of pediatric and neonatal patients: Neonatal resuscitation guidelines. *Pediatrics*, *117*(5), e1029-e1038. http://dx.doi.org/10.1542/peds.2006-0349

American Psychological Association (1999). *The standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Andreatta, P. B., & Gruppen, L. D. (2009). Conceptualizing and classifying validity evidence for simulation. *Medical Education*, *43*(11), 1028-1035. http://dx.doi.org/10.1111/j.1365-2923.2009.03454.x

Andreatta, P., Saxton, E., Thompson, M., & Annich, G. (2010). Simulation-based mock codes significantly correlate with improved pediatric patient cardiopulmonary arrest survival rates. *Pediatric Critical Care Medicine*, *12*(1), 33-38. http://dx.doi.org/10.1097/PCC.0b013e3181e89270

Andreatta, P. B., Maslowski, E., Petty, S., Shim, W., Marsh, M., Hall, T., … Frankel, J. (2010). Virtual reality triage training provides a viable solution for disaster-preparedness. *Academic Emergency Medicine*, *17*(8), 870-876. http://dx.doi.org/10.1111/j.1553-2712.2010.00728.x

Arad, M., Berkenstadt, H., Zehlinger, J., Laor, A., Shemer, J., & Atsmon, J. (1993). The effects of continuous operation in a chemical protective ensemble on the performance of medical tasks in trauma management. *Journal of Trauma and Acute Care Surgery, 35*(5), 800-804. http://dx.doi.org/10.1097/00005373-199311000-00025

Berkenstadt, H., Arad, M., Nahtomi, O., & Atsmon, J. (1999). The effect of a chemical protective ensemble on intravenous line insertion by emergency medical technicians. *Military Medicine, 164*(10), 737-739.

Bismilla, Z., Finan, E., McNamara, P. J., LeBlanc, V., Jefferies, A., & Whyte, H. (2010). Failure of pediatric and neonatal trainees to meet Canadian Neonatal Resuscitation Program standards for neonatal intubation. *Journal of Perinatology*, *30*(3), 182-187. http://dx.doi.org/10.1038/jp.2009.152

Brennan, R. J., Waeckerle, J. F., Sharp, T. W., & Lillibridge, S. R. (1999). Chemical warfare agents: Emergency medical and emergency public health issues. *Annals of Emergency Medicine*, *34*(2), 191-204. http://dx.doi.org/10.1016/S0196-0644(99)70229-5

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Campbell, D. M., Barozzino, T., Farrugia, M., & Sgro, M. (2009). High-fidelity simulation in neonatal resuscitation. *Paediatrics & Child Health*, *14*(1), 19-23.

Cannard, K. (2006). The acute treatment of nerve agent exposure. *Journal of the Neurological Sciences*, *249*(1), 86-94. http://dx.doi.org/10.1016/j.jns.2006.06.008

Dandoy, A. C., & Goldstein, A. G. (1990). The use of cognitive appraisal to reduce stress reactions: A replication. *Journal of Social Behavior and Personality*, *5*, 275-285.

Dewey, J. (1933/1998). *How we think* (rev. ed.). Boston, MA: Houghton Mifflin Company.

Doyle, J. K. (1997). The cognitive psychology of systems thinking. *System Dynamics Review*, *13*(3), 253-265.

Eppich, W. J., Adler, M. D., & McGaghie, W. C. (2006). Emergency and critical care pediatrics: Use of medical simulation for training in acute pediatric emergencies. *Current Opinion in Pediatrics*, *18*(2), 266-271. http://dx.doi.org/10.1097/01.mop.0000193309.22462.c9

Falck, A. J., Escobedo, A. J., Baillargeon, J. G., Villard, L. G., & Hall, C. A. (2003). Proficiency of pediatric residents in performing neonatal endotracheal intubation. *Pediatrics*, *112*(6), 1242-1247. http://dx.doi.org/10.1542/peds.112.6.1242

Filner, B. H.R. 4269: Battlefield Excellence through Superior Training Practices Act. (introduced 2009, December 10) Retrieved from http://www.thomas.gov/cgi-bin/bdquery/z?d111:h.r.04269

Fine, B. J., & Kobrick, J. L. (1987). Effect of heat and chemical protective clothing on cognitive performance. *Aviation, Space, and Environmental Medicine*, *58*(2), 149-54.

Halamek, L. P., Kaegi, D. M., Gaba, D. M., Sowb, Y. A., Smith, B. C., Smith, B. E., & Howard, S. K. (2000). Time for a new paradigm in pediatric medical education: Teaching neonatal resuscitation in a simulated delivery room environment. *Pediatrics*, *106*(4), e45. http://dx.doi.org/10.1542/peds.106.4.e45

Hall, R. E., Plant, J. R., Bands, C. J., Wall, A. R., Kang, J., & Hall, C. A. (2005). Human patient simulation is effective for teaching paramedic students endotracheal intubation. *Academic Emergency Medicine*, *12*(9), 850-855. http://dx.doi.org/10.1111/j.1553-2712.2005.tb00961.x

Harnad, S. (1982). Neoconstructivism: A unifying theme for the cognitive sciences. In T. Simon & R. Scholes (Eds.), *Language, mind and brain* (pp. 1-11). Hillsdale, NJ: Erlbaum.

Hendler, I., Nahtomi, O., Segal, E., Perel, A., Wiener, M., & Meyerovitch, J. (2000). The effect of full protective gear on intubation performance by hospital medical personnel. *Military Medicine*, *165*(4), 272-274.

Humara, M. (1999). The relationship between anxiety and performance: A cognitive-behavioral perspective. *Athletic Insight*, *11*(2), 1-14.

Inzana, C. M., Driskell, J. E., Salas, E., & Johnston, J. H. (1996). Effects of preparatory information on enhancing performance under stress. *Journal of Applied Psychology*, *81*(4), 429-435. doi: 10.1037/0021-9010.81.4.429

Irvine, S., & Martin, J. (2014). Bridging the gap: From simulation to clinical practice. *The Clinical Teacher*, *11*(2), 94-98. http://dx.doi.org/10.1037/0021-9010.81.4.429

Kadivar, H., & Adams, S. C. (1991). Treatment of chemical and biological warfare injuries: Insights derived from the 1984 Iraqi attack on Majnoon Island. *Military Medicine*, *156*(4), 171-177.

Khan, A. S., Morse, S., & Lillibridge, S. (2000). Public-health preparedness for biological terrorism in the USA. *Lancet*, *356*(9236), 1179-1182. http://dx.doi.org/10.1016/S0140-6736(00)02769-0

King, J. M., & Frelin, A. J. (1984). Impact of the chemical protective ensemble on the performance of basic medical tasks. *Military Medicine*, *149*(9), 496-501.

Klotz, J. J., Dooley-Hash, S. L., House, J. B., & Andreatta, P. B. Pediatric and neonatal intubation training gap analysis: instruction, assessment, and technology. *Simulation in Healthcare* (in press).

Leone, T. A., Rich, W., & Finer, N. N. (2005). Neonatal intubation: Success of pediatric trainees. *The Journal of Pediatrics*, *146*(5), 638-641. http://dx.doi.org/10.1016/j.jpeds.2005.01.029

Macilwain, C. (1993). Study proves Iraq used nerve gas. *Nature*, *363*(6424), 3. http://dx.doi.org/10.1038/363003b0

Mandler, G. (1979). Thought processes, consciousness, and stress In V. Hamilton & D. M. Warburton (Eds.), *Human stress and cognition: An information processing approach* (pp. 179-201). New York, NY: John Wiley & Sons.

Mayo, P. H., Hackney, J. E., Mueck, T., Ribaudo, V., & Schneider, R. F. (2004). Achieving house staff competence in emergency airway management: Results of a teaching program using a computerized patient simulator. *Critical Care Medicine*, *32*(12), 2422-2427. http://dx.doi.org/10.1097/01.CCM.0000147768.42813.A2

Morita, H., Yanagisawa, N., Nakajima, T., Shimizu, M., Hirabayashi, H., Okudera, H., … Mimura, S. (1995). Sarin poisoning in Matsumoto, Japan. *Lancet*, *346*(8970), 290-293. http://dx.doi.org/10.1016/S0140-6736(95)92170-2

Murray, C. K. (2006). Development of a deployment course for graduating military internal medicine residents. *Military Medicine*, *171*(10), 933-936.

Nestel, D., Groom, J., Eikeland-Husebø, S., & O'Donnell, J. M. (2011). Simulation for learning and teaching procedural skills: The state of the science. *Simulation in Healthcare*, *6 Suppl*, S10-S13. http://dx.doi.org/10.1097/SIH.0b013e318227ce96

O'Donnell, C. P., Kamlin, C. O., Davis, P. G., & Morley, C. J. (2006). Endotracheal intubation attempts during neonatal resuscitation: Success rates, duration, and adverse effects. *Pediatrics*, *117*(1), e16-e21. http://dx.doi.org/10.1542/peds.2005-0901

Ohbu, S., Yamashina, A., Takasu, N., Yamaguchi, T., Murai, T., Nakano, K., … Hinohara, S. (1997). Sarin poisoning on Tokyo subway. *Southern Medical Journal*, *90*(6), 587-593. http://dx.doi.org/10.1097/00007611-199706000-00002

Okumura, T., Takasu, N., Ishimatsu, S., Miyanoki, S., Mitsuhashi, A., Kumada, K., … Hinohara, S. (1996). Report on 640 victims of the Tokyo subway sarin attack. *Annals of Emergency Medicine*, *28*(2), 129-135. http://dx.doi.org/10.1016/S0196-0644(96)70052-5

Overly, F. L., Sudikoff, S. N., & Shapiro, M. J. (2007). High-fidelity medical simulation as an assessment tool for pediatric residents' airway management skills. *Pediatric Emergency Care*, *23*(1), 11-15. http://dx.doi.org/10.1097/PEC.0b013e31802c61d5

Rodgers, J. C. (1998). Chemical incident planning: A review of the literature. *Accident and Emergency Nursing*, *6*(3), 155-159. http://dx.doi.org/10.1016/S0965-2302(98)90031-5

Sanders, R. C., Jr., Giuliano, J. S., Jr., Sullivan, J. E., Brown, C. A., 3rd, Walls, R. M., Nadkarni, V., … the National Emergency Airway Registry for Children Investigators and Pediatric Acute Lung Injury and Sepsis Investigators Network. (2013). Level of trainee and tracheal intubation outcomes. *Pediatrics*, *131*(3), e821-e828. http://dx.doi.org/10.1542/peds.2012-2127

Sidell, F. R. (1997). Nerve agents. In F.R. Sidell, E. T. Takafuji, & D. R. Franz (Eds.), Medical Aspects of Chemical and Biological Warfare, (pp. 129-180) *Textbook of Military Medicine, Part 1. Warfare, Weaponry, and the Casualty, Vol. 3.* Washington, DC: Walter Reed Army Medical Center Borden Institute; Falls Church, VA: U.S. Army Office of The Surgeon General; Fort Sam Houston, TX: U.S. Army Medical Department Center and School; Fort Detrick Frederick, MD: U.S. Army Medical Research and Materiel Command; & Bethesda, MD: Uniformed Services University of the Health Sciences.

Smith, R., & Nye, S. (1989). Comparison of induced affect and covert rehearsal in the acquisition of stress management coping skills. *Journal of Counseling Psychology*, *36*(1), 17-23. http://dx.doi.org/10.1037/0022-0167.36.1.17

Smolander, J., Louhevaara, V., & Korhonen, O. (1985). Physiological strain in work with gas protective clothing at low ambient temperatures. *American Industrial Hygiene Association Journal*, *46*(12), 720-723. http://dx.doi.org/10.1080/15298668591395625

Vadodaria, B. S., Gandhi, S. D., & McIndoe, A. K. (2004). Comparison of four different emergency airway access equipment sets on a human patient simulator. *Anaesthesia*, *59*(1), 73-79. http://dx.doi.org/10.1111/j.1365-2044.2004.03456.x

VanGemmert, A., & VanGalen, G. (1997). Stress, neuromotor noise, and human performance: A theoretical perspective. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(5), 1299-1313. http://dx.doi.org/10.1037/0096-1523.23.5.1299

Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Waeckerle, J. F. (2000). Domestic preparedness for events involving weapons of mass destruction. *Journal of the American Medical Association*, *283*(2), 252-254. http://dx.doi.org/10.1001/jama.283.2.252