

PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

A Simulator-Based Tool That Assesses Pediatric Resident Resuscitation Competency

Marisa B. Brett-Fleegler, Robert J. Vinci, Debra L. Weiner, Sion Kim Harris,
Mei-Chiung Shih and Monica E. Kleinman

Pediatrics 2008;121:e597-e603; originally published online Feb 18, 2008;
DOI: 10.1542/peds.2005-1259

The online version of this article, along with updated information and services, is
located on the World Wide Web at:

<http://www.pediatrics.org/cgi/content/full/121/3/e597>

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2008 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™



A Simulator-Based Tool That Assesses Pediatric Resident Resuscitation Competency

Marisa B. Brett-Fleegler, MD^a, Robert J. Vinci, MD^b, Debra L. Weiner, MD, PhD^a, Sion Kim Harris, PhD^c, Mei-Chiung Shih, PhD^d,
Monica E. Kleinman, MD^e

Divisions of ^aEmergency Medicine and ^cCritical Care Medicine, ^cClinical Research Program, Children's Hospital Boston, Boston, Massachusetts; ^bDivision of Pediatric Emergency Medicine, Boston Medical Center, Boston, Massachusetts; ^dDepartment of Health Research and Policy, Stanford University, and VA Palo Alto Cooperative Studies Programs Coordinating Center, Palo Alto, California

The authors have indicated they have no financial relationships relevant to this article to disclose.

ABSTRACT

BACKGROUND. Competency in pediatric resuscitation is an essential goal of pediatric residency training. Both the exigencies of patient care and the Accreditation Council for Graduate Medical Education require assessment of this competency. Although there are standard courses in pediatric resuscitation, no published, validated assessment tool exists for pediatric resuscitation competency.

OBJECTIVE. The purpose of this work was to develop a simulation-based tool for the assessment of pediatric residents' resuscitation competency and to evaluate the tool's reliability and preliminarily its validity in a pilot study.

METHODS. We developed a 72-question yes-or-no questionnaire, the Tool for Resuscitation Assessment Using Computerized Simulation, representing 4 domains of resuscitation competency: basic resuscitation, airway support, circulation and arrhythmia management, and leadership behavior. We enrolled 25 subjects at each of 5 different training levels who all participated in 3 standardized code scenarios using the Laerdal SimMan universal patient simulator. Performances were videotaped and then reviewed by 2 independent expert raters.

RESULTS. The final version of the tool is presented. The intraclass correlation coefficient between the 2 raters ranged from 0.70 to 0.76 for the 4 domain scores and was 0.80 for the overall summary score. Between the 2 raters, the mean percent exact agreement across items in each domain ranged from 81.0% to 85.1% and averaged 82.1% across all of the items in the tool. Across subject groups, there was a trend toward increasing scores with increased training, which was statistically significant for the airway and summary scores.

CONCLUSIONS. In this pilot study, the Tool for Resuscitation Assessment Using Computerized Simulation demonstrated good interrater reliability within each domain and for summary scores. Performance analysis shows trends toward improvement with increasing years of training, providing preliminary construct validity.

www.pediatrics.org/cgi/doi/10.1542/peds.2005-1259

doi:10.1542/peds.2005-1259

Results from this study were presented at the annual meeting of the Pediatric Academic Societies; May 3, 2004; San Francisco, CA.

Key Words

assessment, competency, medical education, pediatric resuscitation, simulation

Abbreviations

ACGME—Accreditation Council For Graduate Medical Education
TRACS—Tool for Resuscitation Assessment Using Computerized Simulation
ICC—intraclass correlation coefficient

Accepted for publication Jul 30, 2007

Address correspondence to Marisa B. Brett-Fleegler, MD, Division of Emergency Medicine, Main South Basement, Room CB0120, Children's Hospital Boston, 300 Longwood Ave, Boston, MA 02115. E-mail: marisa.brett@childrens.harvard.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275). Copyright © 2008 by the American Academy of Pediatrics

COMPETENCY IN PEDIATRIC resuscitation is an essential objective of pediatric residency training.¹ It is mandated not only by residency review committee regulations² but by the exigencies of patient care, given low survival rates from pediatric cardiopulmonary arrest.^{1,3,4} In September 1999, the Accreditation Council for Graduate Medical Education (ACGME) formally identified and endorsed 6 general competencies for medical education, among them competencies in direct patient care, medical knowledge, and professionalism; residency review committees mandated the implementation and assessment of these competencies as training program requirements effective July 2002. The current evolution of physician assessment is toward performance evaluation, symbolized by the pyramid described by Miller,⁵ emphasizing a progression from "knows" to "knows how," "shows how," and "does." This highest level corresponds with competency in the performance of a skill, beyond and distinct from the knowledge of how to do so.

Because resuscitation is a rare event in pediatrics, residency curricula must rely on simulated experiences rather

than actual patient care to teach this topic and determine competency. High-fidelity medical simulation offers a realistic training environment that poses no threat to patient safety. Long recognized and used in aviation training, simulation was initially adopted for medical training and crisis management in the field of anesthesia^{6,7} but has since been used in an increasingly diverse group of medical specialties, including emergency medicine, surgery, and obstetrics-gynecology.⁸⁻¹² The use of simulation has been extended beyond training to assessment,¹³ and the ACGME assessment toolbox recognizes simulations and models as the most desirable assessment methods for medical procedures in patient care.¹⁴

Objective assessment requires reliable and valid tools. Reliability refers to the reproducible assessment of the same performance under different circumstances. Interrater reliability in particular refers to consistency in ratings between different examiners. Interrater reliability is an indication of the generalizability of the tool, that is, whether the tool would have the same properties when applied by multiple users. Elements of validity include the selection of appropriate subject matter (content validity), appropriate variation in concert with expectations (construct validity), and accurate comparison with the established gold standard for the behavior in question (criterion validity). Simulator-based rating systems have been used in anesthesia with demonstration of good reliability and strong construct validity.¹⁵⁻¹⁸

Although there are standard courses in pediatric resuscitation, no published, validated assessment tool exists for the performance of pediatric resuscitation; a neonatal resuscitation specific-tool has recently been developed.¹⁹ The need for such a tool is illustrated by deficiencies in the resuscitation knowledge and skills of pediatric house officers.^{13,20-24} We sought to develop a reliable and valid tool for the assessment of pediatric residents' resuscitation competency using simulation. We present our novel tool, the Tool for Resuscitation Assessment Using Computerized Simulation (TRACS), that we have developed and report the results of a pilot study to assess its reliability and validity. Our study's primary goal was to conduct an initial assessment of the TRACS' interrater reliability. Our secondary objective was to preliminarily evaluate the TRACS' construct validity by comparing performance at different training levels.

METHODS

This study was approved by the Children's Hospital Boston Committee on Clinical Investigation.

Tool Development

Using the American Heart Association Pediatric Advanced Life Support curriculum,²⁵ combined with standard resuscitation management, we delineated 4 domains of resuscitation performance (basic resuscitation, airway support, circulation and arrhythmia management, and leadership behavior) and identified specific objective elements for each domain. We elected to use a checklist rather than global rating style for this tool in an

effort to formally evaluate specific item-level competency and avoid any masking of deficits that could potentially occur with a global rating score. To provide content validity, these behaviors were reviewed by a panel of 13 experts (including Drs Vinci and Weiner) in the fields of pediatric emergency medicine, pediatric critical care, and medical education using a modified Delphi process.²⁶ The panel was asked to confirm that the selected items should be expected of pediatric residents, to eliminate inappropriate items, and to add any missing items. Modifications were made to these selected items after trial runs in the simulator suite to these selected items to clarify them and facilitate scoring. The final version of the TRACS used 72 yes-or-no (performed or not performed) items: basics (6 items), airway (32 items), circulation (25 items), and leadership (9 items). The number of items in each domain is not meant to precisely reflect its relative importance but rather to gauge the complexity of its mastery, hence the increased number of items in the more procedural airway and circulation domains. Although leadership is perhaps an infinitely complex task, only a basic competency in leadership is being assessed here. The TRACS is shown in Fig 1.

The tool developers (Drs Brett-Fleegler and Kleinman) and 2 physician expert raters (Drs Vinci and Weiner) met in a faculty development session to standardize scoring. A practice-session videotape was reviewed and scored. Each item was discussed to elaborate the criteria required for a pass ("yes") for that item. These comments were codified in a guide for use in reviewing the remainder of the videotapes. Raters were instructed not to score items if performance could not be assessed from the videotape.

Subject Enrollment

We recruited a convenience sample from a pool of all of the available members of 5 training groups: intern, junior resident, senior resident, recent graduate (pediatric residency graduates in their first year postresidency), and senior fellow (second- and third-year fellows in pediatric critical care or emergency medicine). We enrolled 25 subjects, 5 from each of the 5 training levels. All of the subjects in the first 4 groups were current residents or graduates of the Boston Combined Residency in Pediatrics based at Children's Hospital Boston and Boston Medical Center. The recent graduate pool included chief residents, first-year fellows, and other pediatricians still working clinically at these hospitals.

Sessions

We conducted sessions in the Children's Hospital Boston simulator suite between July and October 2003. In each session, subjects provided written informed consent and completed a training information questionnaire. After simulator suite orientation, each subject participated in 3 standardized code scenarios written by the authors (Drs Brett-Fleegler and Kleinman) using the Laerdal SimMan universal patient simulator (Laerdal Corporation, Stockholm, Sweden). At the beginning of each scenario, a

TRACS: Tool for Resuscitation Assessment Using Computerized Simulation			
BASICS			
TASK GROUP	TASK		
H&P	Elicits essential information about patient and situation (attends to age, inciting event, signs/symptoms provided); solicits additional basic information such as PMH/medications	Y	N
	Performs/directs pertinent PE	Y	N
Monitors	Ensures cardiorespiratory and O ₂ saturation monitors placed within one minute	Y	N
Access	Attempts IV access/directs team member to attempt IV access	Y	N
Labs	Orders appropriate laboratory testing	Y	N
	Asks for lab results	Y	N
AIRWAY			
TASK GROUP	TASK		
Assessment	Assesses airway	Y	N
	Assesses breathing	Y	N
Basic interventions	Performs general airway maneuvers - positions patient	Y	N
	Provides any oxygen support	Y	N
	Provides optimal level of oxygen	Y	N
Bag-valve mask ventilation	Initiates positive-pressure ventilation in timely manner	Y	N
	Selects correct mask and bag size ^a	Y	N
	Correctly connects to oxygen source	Y	N
	Bags at appropriate rate ^a	Y	N
	Adequate chest rise (seal and bagging technique)	Y	N
	Confirms efficacy by auscultation or looking at chest wall (self or directs team member)	Y	N
Endotracheal intubation	Initiates team efforts for endotracheal intubation in timely manner	Y	N
	Preoxygenates patient	Y	N
	Selects appropriate size endotracheal tube	Y	N
	Selects appropriate size laryngoscope blade	Y	N
	Places stylet correctly (if stylet is used; use not required) ^a	Y	N
	Ensures suction available ^a		
	Asks for cricoid pressure	Y	N
	Appropriate endotracheal tube insertion technique (patient position, avoids teeth)	Y	N
	Places endotracheal tube in trachea	Y	N
	Secures tube/directs team member to secure tube	Y	N
	Time to intubation less than 8 minutes from start (see log) ^b	Y	N
Intubation assessment	Auscultates patient/ asks team member to auscultate patient	Y	N
	Assesses oxygenation	Y	N
	Checks end-tidal CO ₂	Y	N
	Asks for portable chest x-ray to confirm tube placement	Y	N
Airway RSI	Selects premedication(s) appropriately ^a	Y	N
	Doses premedication(s) appropriately ^a	Y	N
	Selects sedative/hypnotic agent(s) appropriately	Y	N
	Doses sedative/hypnotic agent(s) appropriately	Y	N
	Selects paralytic appropriately	Y	N
	Doses paralytic appropriately	Y	N
CIRCULATION AND ARRHYTHMIAS			
TASK GROUP	TASK		
Basics	Assesses blood pressure (asks/looks for reading)	Y	N
	Palpates central pulses	Y	N
CPR	Initiates appropriately	Y	N
	Correct hand placement ^a	Y	N
	Correct compression to ventilation rate ^a	Y	N
	Time to initiation of CPR less than 1 minute (see log) ^b	Y	N
Management	Initiates IV fluids appropriately	Y	N
	Selects isotonic fluid	Y	N
Arrhythmias	Recognizes presence of abnormal (non-sinus) rhythm on initial presentation	Y	N
	Identifies first abnormal rhythm correctly (wide-complex SVT)	Y	N
	Identifies second abnormal rhythm correctly (ventricular fibrillation)	Y	N
	Recognizes indication to defibrillate	Y	N
	Doses energy correctly (may be PALS or ACLS based dose)	Y	N
	Charges defibrillator	Y	N
	Places paddles correctly	Y	N
	Confirms personnel clear before charging	Y	N
	Releases charge	Y	N
	Administers 3 shocks before proceeding	Y	N
	Time to deliver defibrillation (x3) less than 4 minutes from onset of ventricular fibrillation (see log) ^b	Y	N
	Administers shock before medications	Y	N
	Recognizes indication for first line medication (epinephrine)	Y	N
	Uses correct concentration of epinephrine (1:10 000)	Y	N
	Uses correct dose of epinephrine (0.1 cc/kg or 1mg)	Y	N
	Repeats shock after medication	Y	N
	Repeats pulse check after CPR/resuscitation medications	Y	N
BEHAVIOR			
Professionalism	Has professional attitude toward patient	Y	N
	Has professional attitude towards team members	Y	N
Leadership	Assumes leadership of code	Y	N
	Assigns roles	Y	N
	Utilizes personnel effectively	Y	N
	Communicates effectively with team	Y	N
	Assumes adequate responsibility when in non-leader roles (airway, circulation)	Y	N
Management	Performs tasks in appropriate sequence/prioritizes well	Y	N
	Intermittently summarizes/ maintains global view	Y	N

FIGURE 1

The TRACS. H&P indicates History and physical; PMH, Past medical history; PE, Physical exam; IV, intravenous; RSI, Rapid sequence intubation; CPR, cardiopulmonary resuscitation; SVT, Supraventricular tachycardia; PALS, Pediatric advanced life support; ACLS, Advanced cardiac life support; ^a Items eliminated from further analysis (see text); ^b measurements taken directly from the computer log recorded by SimMan.

brief scripted vignette with pertinent clinical information was provided (see Fig 2). At the time of this study, only an adult mannequin was available for use, so all of the

scenarios involved patients age 12 years and older. Subjects functioned in 3 different code team roles, 1 for each scenario. In the first scenario, subjects served as the code team leader for a 14-year-old male near-drowning victim. Next, they managed the airway of a 12-year-old girl in status asthmaticus. Finally, they provided circulation and arrhythmia management for a 15-year-old girl with a tricyclic antidepressant overdose. Subjects performed the resuscitation along with a code team consisting of 3 to 5 additional health care providers (physicians, nurses, respiratory therapists, and/or pharmacists). These additional personnel were familiar with the clinical scenario and its expected management; all were trained through practice sessions to provide assistance to the subject when requested to do so but were explicitly instructed not to direct care in order to isolate the performance of the subject. All of the sessions were videotaped for evaluation. No code cards were allowed, because our intent was to assess competency independent of such aids. Afterward, subjects received feedback from faculty and answered a second questionnaire soliciting their feedback.

Scoring

The 2 raters independently rated the videotaped performance of all 25 of the subjects using the TRACS. Each subject's 30-minute videotape was viewed once, to approximate real-time use. Because both raters are faculty members of the Boston Combined Residency in Pediatrics, they were not blind to the subjects' training levels. Rater scores were reviewed for each individual item; an a priori decision mandated the elimination of any item not scored in $\geq 50\%$ of subjects by both raters. The score in each domain was determined by the percentage of answered items scored as "yes" if $\geq 75\%$ of items were answered in that domain. If $< 75\%$ of items were answered, the domain score was set to missing. Because the number of items in each domain is meant to be reflective of the complexity of tasks involved, we chose to calculate an overall summary score that was a weighted average of the 4 domain scores, in which each domain was given a weight proportional to the number of items scored in that domain. Final subject scores were the average of the 2 raters' scores.

Statistical Analyses

To evaluate interrater reliability, intraclass correlation coefficients (ICCs) were calculated for domain and summary scores. These have been used frequently to assess interrater reliability when the variables being compared have ordinal or continuous score distributions.²⁷ For each individual, dichotomous item within the domains, we calculated the most commonly used measure of agreement for nominal or categorical variables, the Cohen's κ .²⁸ However, Cohen's κ value has a number of weaknesses that limit its usefulness in certain circumstances. Specifically, the κ value is highly affected by skewed distributions, which cause it to have extremely low values, and it is unable to be calculated at all when there are rows or columns missing from a table.²⁹⁻³¹

FIGURE 2

Clinical vignette for scenario 1: subject leads resuscitation. T indicates temperature (C); HR, heart rate; RR, respiratory rate (breaths per minute); BP, blood pressure; Sao₂, arterial oxygen saturation; wt, weight; BLS, basic life support; IV, intravenous.

A 14-year-old boy presents to the emergency department after a lake accident; witnessed to flounder and go under. There was no diving involved.

- Vital signs on arrival: T 36; HR 140; RR 6; BP 60/40; SaO₂ 85%; estimated wt 50 kg (if asked).
- Pale and unresponsive. Poor respiratory effort and aeration, weak pulse, extremities cool.
- Prehospital care: BLS only. Received bag-valve-mask ventilation. No IV access.

Several of the TRACS items had extremely skewed distributions (mostly positive ratings) or had 0 rows or columns because of 1 rater answering all “yes” or all “no”; in addition, some items had a 0 on the diagonal resulting in a negative κ value. Therefore, we also chose to examine some raw agreement indices, including the overall percentage of agreement (proportion of all of the cases that received the same rating by both raters), the percentage of agreement for positive ratings, and the percentage of agreement on negative ratings. These latter 2 indices are analogous to sensitivity and specificity and are useful because they provide more specific information about the direction and type of rating inconsistencies. In addition, high agreement proportions for both positive and negative ratings would indicate that the observed level of agreement is higher than would occur by chance.^{29,30} To summarize item-level reliability across domains and for the overall tool, we then calculated the mean percentage of the exact agreement across items within each of the 4 domains and across all of the items in the tool.

To preliminarily evaluate construct validity, we assessed whether the TRACS could differentiate between trainees at different levels of training, although the study was not powered for this comparison. The domain and overall summary scores were compared among the 5 training levels using the nonparametric Kruskal-Wallis test. Because the scores are expected to increase as the level of training increases, a nonparametric trend test, the Jonckheere-Terpstra test, which tests for monotone increasing trends, was also conducted.³² All of the tests were 2 sided.

RESULTS

Training Data

Subjects’ intended career paths and mock resuscitation experience are shown in Table 1. Of note is the signifi-

cant variation in exposure to mock codes and relatively limited code team leadership experience.

Tool Performance

Item Elimination

Seven of the original 72 items were not scored for 50% of subjects by ≥ 1 rater and were, therefore, eliminated from current analyses. Four items, 2 each from the airway and circulation domains, were infrequently scored, because the objective was performed by a team member instead of the subject. Two additional items in the airway domain referred to the administration of premedications for endotracheal intubation, which were infrequently given. Because this was deemed clinically acceptable by the raters, they independently elected not to score those items rather than rating them as incorrect. Another airway domain item (ensures suction available) was infrequently scored because of a typographical error on the scoring sheet. These 7 items were eliminated from the TRACS, leaving a total of 65 items for analysis: 6 in basic resuscitation, 27 in airway, 23 in circulation, and 9 in leadership behavior.

Scoring

The response rate for a given item is the percentage of subjects scored. After elimination of items as above, response rates for the 65 items by rater 1 ranged from 76% to 100%, with 55 of 65 items scored for $\geq 95\%$ of subjects; for rater 2, response rates ranged from 68% to 100%, with 55 of 65 items scored for $\geq 95\%$ of subjects. Reasons for inability to score a subject included difficulty visualizing relevant activity on the videotape and performance of the item by another team member. All of the domains were scored for 24 of the 25 subjects by both raters, given that $\geq 75\%$ of items were answered in each domain; basic resuscitation was not scored for 1 subject by either rater because of a videotaping error.

TABLE 1 Training Characteristics (N = 25)

Group (n = 5 for All Groups)	Career Path (n)	No. of Previous Mock Codes		Previous Simulator Experience, n/N
		Participation Median (Range)	Leadership Median (Range)	
Intern	Fellowship (3)	1.5 (0.0 to ≥ 11.0)	0.0 (0.0 to 3.5)	2/5
	Undecided (2)			
Junior resident	Fellowship (4)	5.5 (3.5 to 5.5)	1.5 (0.0 to 1.5)	2/5
	Undecided (1)			
Senior resident	Fellowship (3)	5.5 (3.5 to 7.5)	1.5 (1.5 to 3.5)	4/5
	Primary care (2)			
Recent graduate	Fellowship (4)	5.5 (3.5 to 5.5)	1.5 (0.0 to 1.5)	3/5
	Hospitalist (1)			
Senior fellow	Fellowship (5)	≥ 11 (5.5 to ≥ 11.0)	9.5 (1.5 to ≥ 11.0)	5/5

In evaluating domain and summary scores, the ICC between the 2 raters ranged from 0.70 to 0.76 for the 4 domain scores and was 0.80 for the overall summary score (see Table 2). In comparing individual item scoring between raters, κ values were 0 or negative for 19 of the 65 items. The mean of the remaining κ was 0.51. In evaluating all of the items (including those with incalculable κ values), the mean percentage of exact agreement across items in each domain was 85.5% (range: 77.3%–95.7%) for basics, 81.2% (range: 52.0%–100.0%) for airway, 81.0% (range: 25.0%–100.0%) for circulation, and 85.1% (range: 52.0%–100.0%) for leadership and averaged 82.1% across all of the items in the tool. The mean percentage of positive agreement overall was 86.5%, and the mean percentage of negative agreement overall was 48.2%.

Group Performance

Table 3 shows the mean domain and summary scores for each of the 5 groups. The Kruskal-Wallis test indicates statistically significant differences between the groups for the airway domain ($P[r] \leq .05$) but not for the other domains or the overall summary score. With regard to trends, increasing scores with increased training were seen in all of the domains except for basics and were statistically significant using the Jonckheere-Terpstra trend test for the airway and summary score trends ($P = .01$).

Session Evaluation

On a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), subjects agreed that the session was realistic (mean score: 4.2) and useful (mean score: 4.9) and that the feedback was useful (mean score: 4.8). All of the subjects agreed or strongly agreed that they would like to participate in similar sessions in the future.

DISCUSSION

The goal of the TRACS is to serve as an objective, reliable, and valid performance measure for pediatric resuscitation. In this initial study of tool performance, the TRACS demonstrated strong interrater reliability, with ICCs ≥ 0.70 for all of the domain and summary scores. An average percentage of exact agreement of $\geq 81.0\%$ across all of the items also supports good interrater reliability. The percentage of positive agreement of 86.5% supports good sensitivity of the tool. In considering those items with very low κ values, the majority have low agreement in negative ratings. These items will be revised in future iterations of the tool, and guidelines for

reviewers will be clarified to improve agreement. Consideration will be given to eliminating the leadership behavior section of the tool, because it is the least objective element of the tool, and although critically important, perhaps it is better suited to another assessment modality. We, thus, anticipate even better reliability in future iterations of this tool.

Reviewing videotaped performances to establish interrater reliability has ample precedent.^{18,33–35} The use of 2 reviewers provides appropriate opportunity to compare interrater reliability across a range of performances and is supported in the literature, as long as rater training is provided and objectives are well defined.^{33,36,37}

The validity of an assessment tool depends on content, construct, and criterion validity. Content validity of the TRACS is based on its roots in the widely accepted Pediatric Advanced Life Support curriculum and is supported by expert panel review. Construct validity for this tool is supported by the performance analysis that shows trends toward improved performance with increasing years of training. More vigorous differences based on training were likely not seen because of the small sample size used in this pilot study and the marked heterogeneity of mock code experience within that sample. Other pertinent variations in training are almost certainly present and likely underlie this performance variation in subjects at the same level of training, thus emphasizing the importance of formal assessment. Among these may be the varied career interests of these subjects; should this be the case, formal assessment and training are all the more important to ensure that even those trainees not intending a career in critical care possess basic competency in resuscitation. No standard exists on which to base criterion validity. Recent research on previously validated oral board and objective structured clinical examination-based assessment tools used in emergency medicine has supported their validity in a simulator setting,^{38,39} but comparable tools do not exist that are specific for pediatric resuscitation.

In considering the limitations of this study, it should be noted that, whereas the use of 2 reviewers provides reasonable opportunity to assess interrater reliability, comparisons of additional raters would strengthen this assessment. The lack of blinding of our raters to the training status of the subjects is another potential limitation to our study but is expected to be the usual situation in routine use. In addition, the heterogeneity of scores at each level suggests that the raters were not unduly influenced by a priori assumptions about skill level. Including additional, blinded raters may be useful in evaluating the next version of this tool. We will also try to improve interrater agreement by targeting individual items with lower interrater agreement. The majority of these items were those involving assessment by the trainee, such as assessment of oxygenation status or blood pressure. These items are less observable and ask the rater to evaluate a thought process that may not be verbalized by the trainee. This highlights the challenges in assessing the performance of higher-order cognitive tasks. In our convenience sample, selection bias may have occurred for volunteers with either strong or weak

TABLE 2 Scoring Comparison Between 2 Raters: ICCs

Domain	No. of Items	Intraclass Correlation	
		ICC	95% Confidence Interval
Basics	6	0.73	0.47–0.87
Airway	27	0.76	0.53–0.89
Circulation	23	0.70	0.42–0.85
Leadership	9	0.74	0.49–0.88
Overall	65	0.80	0.60–0.91

TABLE 3 TRACS Scores: Mean Percentage (SD)

Domain	Intern (n = 5), Mean (SD)	Junior (n = 5), Mean (SD)	Senior (n = 5), Mean (SD)	Recent Graduate (n = 5), Mean (SD)	Senior Fellow (n = 5), Mean (SD)	<i>p</i> ^a	<i>p</i> ^b
Basics	58.7 (26.0)	81.7 (10.9)	73.3 (18.1)	56.3 (27.1)	71.7 (20.9)	.35	.92
Airway	66.3 (12.9)	75.7 (9.8)	76.0 (12.2)	72.9 (13.5)	89.7 (6.0)	.05	.01
Circulation	79.6 (6.4)	78.7 (17.8)	80.0 (11.0)	85.1 (8.1)	84.2 (9.5)	.87	.38
Leadership	76.5 (28.6)	93.3 (7.2)	91.1 (10.8)	88.6 (14.0)	95.6 (7.2)	.63	.24
Summary	71.7 (8.9)	79.7 (11.0)	79.3 (10.8)	79.0 (11.9)	86.9 (7.1)	.17	.01

^a *P* value is from the Kruskal-Wallis test.

^b *P* value is from the Jonckheere-Terpstra test.

resuscitation skills. In reviewing the intended careers of participants, using this as a proxy for resuscitation skills, our population seems diverse, because it includes both future critical care and noncritical care providers and, most importantly, reflects the target trainee population. With regard to our secondary objective of assessment of construct validity, the primary limitation of this study is its small sample size. The study was not powered to detect significant intergroup differences.

A reliable and valid assessment tool for pediatric resuscitation competency is important in addressing the mandate of the ACGME patient care competency. In graduate medical education, learning may be assessment driven,⁴⁰ and poor clinical performers may overestimate their capabilities.⁴¹ Thus, assessment of pediatric resuscitation competency is essential to both facilitate learning and to safeguard patient care. This type of assessment tool will also be useful for both formative and summative purposes in physician training. Specific, formative feedback is highly valued by trainees; individual items from the TRACS can be used to direct constructive feedback regarding areas needing improvement. As a summative tool, the TRACS will enable the evaluation of residents over time and provide residency programs with objective measurements of trainee competency as they advance to positions of greater responsibility for patient care. Ultimately, trends seen in groups of residents may provide feedback at a programmatic level and guide curriculum modification.

Simulators offer a unique and valuable training and assessment environment. The subjects in our study found the simulator experience and feedback useful and rated these sessions as more helpful than mock resuscitations in a nonsimulator setting. This is consistent with previous experience in which simulators have been well received by trainees at many levels and in diverse settings.^{42–47} In many ways, simulation mirrors the original bedside teaching model while also incorporating the principles of problem-based learning.^{8,43} Simulators combine an interactive learning opportunity with a realistic setting and the chance to manage rare events; we feel this justifies the logistic and economic challenges of developing simulator capabilities, as is de facto supported by the rapid expansion of simulation centers⁴⁸ and the explosive development of simulation as an educational environment.^{49,50}

We intend to revise the TRACS on the basis of the strengths and limitations discovered in this pilot study. The next evaluation phase for the TRACS will involve

further assessment of its interrater reliability and its construct validity via a study powered to evaluate intergroup differences. A reliable and valid assessment tool is crucial to the training and evaluation of pediatric residents in resuscitation.

ACKNOWLEDGMENTS

We thank Liana Kappus and Sarah Aldridge for assistance with the simulator suite. We gratefully acknowledge Kate Anastasio, RT, Alana Arnold, PharmD, Allison Beitel, MD, James Bennett, RPh, Rachel Bent, RN, Susan Hamilton-Bruno, RN, MS, CCRN, Steven Bin, MD, Ari Cohen, MD, Brenda Dodson, PharmD, Andrew Fine, MD, MPH, Eric Fleegler, MD, MPH, Kathleen Gura, PharmD, BCNSP, Joanne Kalogerakos, PharmD, Anupam Kharbanda, MD, Grace Lee, PharmD, Jason Levy, MD, Peter Lutz, RPh, Andrew Mandic, RN, Shannon Manzi, PharmD, Stephanie Masick, RT, Neeta Mehta, RN, Kshitij Mistry, MD, Sean O'Neill, PharmD, Richard Simon, RPh, Rosamond Veator, RN, and Kristopher Young, PharmD, for participation as members of the code team. We also acknowledge the members of our expert panel: Eric Fleegler, MD, MPH, for editorial and statistical input; Henry Bernstein, DO, for educational expertise; and Jennifer Doyle, MA, for guidance regarding assessment methodology.

REFERENCES

- Schoenfeld PS, Baker MD. Management of cardiopulmonary and trauma resuscitation in the pediatric emergency department. *Pediatrics*. 1993;91(4):726–729
- Accreditation Council for Graduate Medical Education. Outcome project: general competencies. Available at: www.acgme.org/outcome/comp/compCPRL.asp. Accessed May 29, 2007
- Lewis, JK, Minter MG, Eshelman SJ, Witte MK. Outcome of pediatric resuscitation. *Ann Emerg Med*. 1983;12(5):297–299
- Schindler MB, Bohn D, Cox PN, et al. Outcome of out-of-hospital cardiac or respiratory arrest in children. *N Engl J Med*. 1996;335(20):1473–1479
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9 suppl):S63–S67
- Denson JS, Abrahamson S. A computer-controlled patient simulator. *JAMA*. 1969;208(3):504–508
- Abrahamson S, Denson JS, Wolf RM. Effectiveness of a simulator in training anesthesiology residents. *J Med Educ*. 1969;44(6):515–519
- Forrest F, Taylor M. High level simulation in medical education. *Hosp Med*. 1998;59(8):653–655
- Ellis C, Hughes G. Use of human patient simulation to teach

- emergency medicine trainees advanced airway skills. *J Accid Emerg Med.* 1999;16(6):395–399
10. Patel RM, Crombleholme WR. Using simulation to train residents in managing critical events. *Acad Med.* 1998;73(5):593
 11. Sica GT, Barron DM, Blum R, Frenna TH, Raemer DB. Computerized realistic simulation: a teaching module for crisis management in radiology. *AJR Am J Roentgenol.* 1999;172(2):301–304
 12. Chapman DM, Rhee KJ, Marx JA, et al. Open thoracotomy procedural competency: validity study of teaching and assessment modalities. *Ann Emerg Med.* 1996;28(6):641–647
 13. Overly FL, Sudikoff SN, Shapiro MJ. High-fidelity medical simulation as an assessment tool for pediatric residents' airway management skills. *Pediatr Emerg Care.* 2007;23(1):11–15
 14. Accreditation Council for Graduate Medical Education. Competencies: suggested best methods for evaluation. Available at: www.acgme.org/Outcome/assess/ToolTable.pdf. Accessed April 13, 2004
 15. Byrne AJ, Jones JG. Responses to simulated anaesthetic emergencies by anaesthetists with different durations of clinical experience. *Br J Anaesth.* 1997;78(5):553–556
 16. Devitt JH, Kurrek MM, Cohen MM. Testing internal consistency and construct validity during evaluation of performance in a patient simulator. *Anesth Analg.* 1998;86(6):1160–1164
 17. Devitt JH, Kurrek MM, Cohen MM, Cleave-Hogg D. The validity of performance assessments using simulation. *Anesthesiology.* 2001;95(1):36–42
 18. Morgan PJ, Cleave-Hogg DM, Guest CB, Herold J. Validity and reliability of undergraduate performance assessments in an anesthesia simulator. *Can J Anaesth.* 2001;48(3):225–233
 19. Lockyer J, Singhal N, Fidler H, Weiner G, Aziz K, Curran V. The development and testing of a performance checklist to assess neonatal resuscitation megacode skill. *Pediatrics.* 2006;118(6):e1739–1744
 20. Brady R, Raftos J. Emergency management skills of South Australian paediatric trainees. *J Paediatr Child Health.* 1997;33(2):113–116
 21. Durojaiye L, O'Meara M. Improvement in resuscitation knowledge after a one-day paediatric life-support course. *J Paediatr Child Health.* 2002;38(3):241–245
 22. Jewkes F, Phillips B. Resuscitation training of paediatricians. *Arch Dis Child.* 2003;88:118–121
 23. Nadel FM, Lavelle JM, Fein JA, Giardino AP, Decker JM, Durbin DR. Assessing pediatric senior residents' training in resuscitation: fund of knowledge, technical skills, and perception of confidence. *Pediatr Emerg Care.* 2000;16(2):73–76
 24. White JR, Shugerman R, Brownlee C, Quan L. Performance of advanced resuscitation skills by pediatric housestaff. *Arch Pediatr Adolesc Med.* 1998;152(12):1232–1235
 25. American Heart Association. *PALS Provider Manual.* Hazinski MF, ed.: American Heart Association; 2002
 26. Delbecq AL, Van de Ven AH, Gustafson DH. Group techniques for program planning. In: Filley AC, ed. *Management Applications Series.* Glenview, AZ: Scott, Foresman and Company; 1975
 27. Shrout PE, Fleiss JL. Intra-class correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86(2):420–428
 28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174
 29. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol.* 1990;43(6):551–558
 30. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol.* 1990;43(6):543–549
 31. Byrt T, Bishop J, Carlin JB. Bias, prevalence, and kappa. *J Clin Epidemiol.* 1993;46(5):423–429
 32. Hollander M, Wolfe D. *Nonparametric Statistical Methods.* 2nd ed. New York, NY: John Wiley and Sons; 1998
 33. Devitt JH, Kurrek MM, Cohen MM. Testing the raters: inter-rater reliability of standardized anaesthesia simulator performance. *Can J Anaesth.* 1997;44(9):924–928
 34. Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology.* 1998;89(1):8–18
 35. Liu P, Miller E, Herr G, Hardy C, Sivarajan M, Willenkin R. Videotape reliability: a method of evaluation of a clinical performance examination. *J Med Educ.* 1980;55(8):713–715
 36. Morgan PJ, Cleave-Hogg D, Guest CB. A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. *Acad Med.* 2001;76(10):1053–1055
 37. Winckel CP, Reznick RK, Cohen R, Taylor B. Reliability and construct validity of a structured technical skills assessment form. *Am J Surg.* 1994;167(4):423–427
 38. Gordon JA, Tancredi DN, Binder WD, Wilkerson WM, Shaffer DW. Assessment of a clinical performance evaluation tool for use in a simulator-based testing environment: a pilot study. *Acad Med.* 2003;78(10 suppl):S45–S47
 39. Gordon JA, Tancredi DN, Binder WD, et al. Assessing global performance in emergency medicine using a high-fidelity patient simulator: a pilot study. *Acad Emerg Med.* 2003;10(5):472.
 40. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet.* 2001;357(9260):945–949
 41. Hodges B, Regehr G, Martin D. Difficulties in recognizing one's own incompetence: novice physicians who are unskilled and unaware of it. *Acad Med.* 2001;76(10 suppl):S87–S89
 42. Block EF, Lottenberg L, Flint L, Jakobsen J, Liebnitzky D. Use of a human patient simulator for the advanced trauma life support course. *Am Surg.* 2002;68(7):648–651
 43. Euliano TY. Small group teaching: clinical correlation with a human patient simulator. *Adv Physiol Educ.* 2001;25(1–4):36–43
 44. Gordon JA. The human patient simulator: acceptance and efficacy as a teaching tool for students. The Medical Readiness Trainer Team. *Acad Med.* 2000;75(5):522
 45. Gordon JA, Wilkerson WM, Shaffer DW, Armstrong EG. "Practicing" medicine without risk: students' and educators' responses to high-fidelity patient simulation. *Acad Med.* 2001;76(5):469–472
 46. Gordon JA, Pawlowski J. Education "on demand": the development of a simulator-based medical education service. *Acad Emerg Med.* 2002;77(7):751–752
 47. Reznick M, Smith-Coggins R, Howard S, et al. Emergency medicine crisis resource management (EMCRM): pilot study of a simulation-based crisis management course for emergency medicine. *Acad Emerg Med.* 2003;10(4):386–389
 48. Morgan PJ, Cleave-Hogg D. A worldwide survey of the use of simulation in anesthesia. *Can J Anaesth.* 2002;49(7):659–662
 49. Eppich WJ, Adler MD, McGaghie WC. Emergency and critical care pediatrics: use of medical simulation for training in acute pediatric emergencies. *Curr Opin Pediatr.* 2006;18(3):266–271
 50. Weinstock PH, Kappus LJ, Kleinman ME, Grenier B, Hickey P, Burns JP. Toward a new paradigm in hospital-based pediatric education: the development of an onsite simulator program. *Pediatr Crit Care Med.* 2005;6(6):635–641

A Simulator-Based Tool That Assesses Pediatric Resident Resuscitation Competency

Marisa B. Brett-Fleegler, Robert J. Vinci, Debra L. Weiner, Sion Kim Harris,
Mei-Chiung Shih and Monica E. Kleinman

Pediatrics 2008;121:e597-e603; originally published online Feb 18, 2008;
DOI: 10.1542/peds.2005-1259

Updated Information & Services	including high-resolution figures, can be found at: http://www.pediatrics.org/cgi/content/full/121/3/e597
References	This article cites 45 articles, 12 of which you can access for free at: http://www.pediatrics.org/cgi/content/full/121/3/e597#BIBL
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): Office Practice http://www.pediatrics.org/cgi/collection/office_practice
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://www.pediatrics.org/misc/Permissions.shtml
Reprints	Information about ordering reprints can be found online: http://www.pediatrics.org/misc/reprints.shtml

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

