

Validation of a self-efficacy instrument and its relationship to performance of crisis resource management skills

Jennifer L. Plant · Sandrijn M. van Schaik · Diane C. Sliwka ·
Christy K. Boscardin · Patricia S. O’Sullivan

Received: 28 September 2010 / Accepted: 11 January 2011 / Published online: 25 January 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Self-efficacy is thought to be important for resuscitation proficiency in that it influences the development of and access to the associated medical knowledge, procedural skills and crisis resource management (CRM) skills. Since performance assessment of CRM skills is challenging, self-efficacy is often used as a measure of competence in this area. While self-efficacy may influence performance, the true relationship between self-efficacy and performance in this setting has not been delineated. We developed an instrument to measure pediatric residents’ self-efficacy in CRM skills and assessed its content validity, internal structure, and relationship to other variables. After administering the instrument to 125 pediatric residents, critical care fellows and faculty, we performed an exploratory factor analysis within a confirmatory factor analysis as well as a known group comparison. The analyses specified four factors that we defined as: situation awareness, team management, environment management, and decision making. Pediatric residents reported lower self-efficacy than fellows and faculty in each factor. We also examined the correlation between self-efficacy and performance scores for a subset of 30 residents who led video recorded simulated resuscitations and had their performances rated by three observers. We found a significant, positive correlation between residents’ self-efficacy in situation awareness and environment management and their overall performance of CRM skills. Our findings suggest that in a specific context, self-efficacy as a form of self-assessment may be informative with regards to performance.

J. L. Plant (✉)

Department of Pediatrics, University of California Davis, 2516 Stockton Blvd, Sacramento,
CA 95817, USA

e-mail: jennifer.plant@ucdmc.ucdavis.edu

S. M. van Schaik

Department of Pediatrics, University of California San Francisco, San Francisco, CA, USA

D. C. Sliwka · C. K. Boscardin · P. S. O’Sullivan

Department of Medicine, University of California San Francisco, San Francisco, CA, USA

C. K. Boscardin · P. S. O’Sullivan

Office of Medical Education, University of California San Francisco, San Francisco, CA, USA

Keywords Crisis resource management skills · Mock codes · Non-technical skills · Performance assessment · Self-assessment · Self-efficacy · Simulated resuscitations · Simulation

Introduction

Self-efficacy is an individual's confidence in his ability to perform a specific task in a given domain. It is thought to affect the initiation of behavior, the amount of effort expended and the persistence of behavior in spite of challenges and negative experiences (Bandura 1977). It differs from self-confidence in that self-efficacy is context-specific rather than a stable personality trait and it is therefore thought to have a direct effect on performance in specific contexts (Bandura 1997). Self-efficacy is believed to be of particular importance in the context of resuscitation because of its influence on the development of and access to the associated knowledge and skills (Maibach et al. 1996). These skills include both procedural skills and crisis resource management skills (CRM skills), the generic behavioral skills needed to safely and effectively manage medical crises such as resuscitations (Gaba et al. 1994).

Based on this theory, multiple studies have examined the effects of resuscitation training on self-efficacy and used self-efficacy as a surrogate measure for performance (Nadel et al. 2000b; Reznick et al. 2003; van Schaik et al. 2008). Competency assessment of resuscitation skills, especially CRM skills, is challenging. A few validated performance evaluation instruments have been developed for use by trained observers (Fletcher et al. 2003; Kim et al. 2006), however direct observation of actual resuscitations is nearly impossible given the rarity of the events in pediatrics (van Schaik et al. 2008) and simulations can be both time consuming and resource intensive if designed with adequate fidelity and generalizability (Kane 1992). Measuring self-efficacy appears relatively more straightforward and is therefore attractive; however, to date, there is limited evidence for a correlation between self-efficacy and performance. In the context of resuscitation, self-efficacy in medical knowledge and procedural skills has not correlated with performance in those areas (Nadel et al. 2000a; Wayne et al. 2006). These findings parallel a larger body of research highlighting physicians' general inability to self-assess accurately (Colthart et al. 2008; Davis et al. 2006). Little is known about the relationship between self-efficacy and performance of CRM skills.

Many of the studies on self-efficacy and self-assessment have been criticized for their lack of validated instruments (Ward et al. 2002). The development of an instrument is a rigorous process involving conceptual analysis of the domain of functioning, drafting and piloting the instrument, and statistical analysis of results including factor analysis and internal consistency reliability (Shea and Fortna 2002). Throughout this process, the different components of validity such as content, response process, internal structure, relationship to other variables, and consequences can be examined (Downing 2003). Few studies measuring self-efficacy have incorporated such rigorous methods.

Given the proposed importance of self-efficacy for resuscitation proficiency, the practical advantages of using self-efficacy as an adjunct for competency assessment, and the lack of data on the relationship between self-efficacy and performance of CRM skills, we believe that the role of self-efficacy in CRM skills merits further investigation. The first step in advancing our understanding is valid measurement of self-efficacy in this area. Therefore, the aim of this study was to develop and validate an instrument to measure

self-efficacy in CRM skills and to examine the correlation between measured self-efficacy and performance during simulated resuscitations.

Methods

We developed an instrument to measure pediatric residents' self-efficacy in CRM skills and examined evidence of validity of the resulting scores. We focused on three different sources of validity: content validity, internal structure, and relationship to other variables (Downing 2003). Content validity was established and assessed during the instrument design process, internal structure was investigated using factor analysis and internal consistency, and relationship to other variables was examined through a correlational, non-experimental design (Shea and Fortna 2002) (Fig. 1). This approach to instrument development has been established by others (e.g. Fletcher et al. 2004; Holmes and Shea 1998; Wang et al. 2003; Yudkowsky et al. 2006).

Participants and settings

Participants included a convenience sample of pediatric residents, pediatric critical care fellows, and pediatric critical care faculty affiliated with residency and fellowship training programs at two U.S. pediatric teaching hospitals during the 2008 calendar year. The internal review boards at both institutions approved the study.

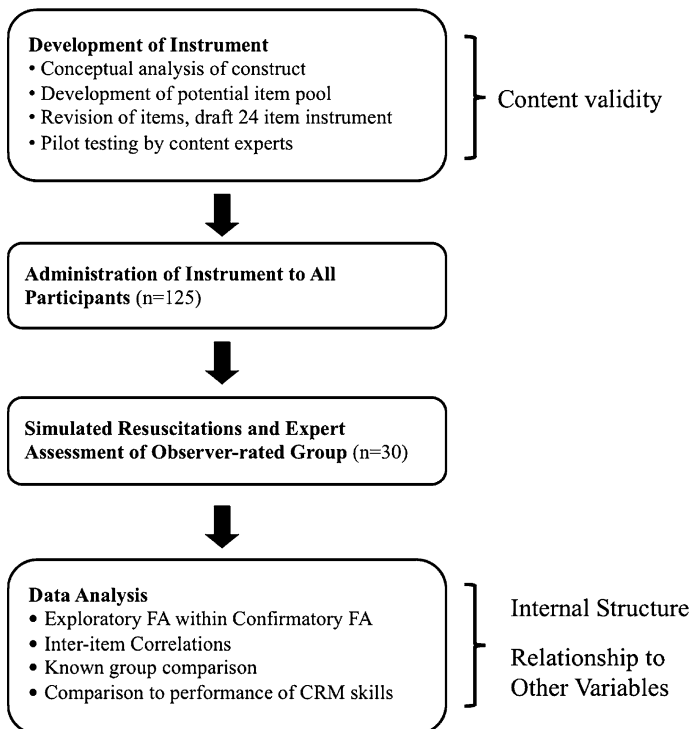


Fig. 1 Study methods

Instruments

Self-efficacy instrument

We designed a paper-and-pencil instrument to measure self-efficacy in CRM skills on a 5-point Likert rating scale. As recommended in work describing the construction of self-efficacy scales with content validity, the design process began with a conceptual analysis of the relevant domain of functioning (Bandura 2006; Shea and Fortna 2002). We defined crisis resource management skills as the generic behavioral skills needed to safely and effectively manage medical crises such as resuscitations (Gaba et al. 1994). This skill set includes cognitive skills such as decision-making and situation awareness as well as interpersonal skills such as team working, communication, and leadership. They are also referred to as non-technical skills to contrast them with technical or procedural skills (Fletcher et al. 2002). We reviewed prior work defining the construct and published instruments measuring skill performance. Specifically, we focused on 3 resources: (1) the seminal work of Gaba et al. (1994) who adapted the aviation industry's principles of *crew* resource management to develop the concept of *crisis* resource management in anesthesia; (2) work to develop the "Anaesthetists' Non-Technical Skills" (ANTS) system, an instrument for evaluating anesthesiologists' performance of non-technical skills (Fletcher et al. 2002, 2003, 2004) that has proven relevance to critical care medicine (Reader et al. 2006) and (3) the application of the construct to critical care medicine in the Ottawa Crisis Resource Management Global Rating Scale (Ottawa GRS) (Kim et al. 2006). Based on this literature review and personal experience facilitating simulated resuscitations, two investigators (JP, SvS) developed a comprehensive list of behaviors associated with CRM skills. We organized our table of specifications using Fletcher et al.'s (2003) categories of task management, team working, situation awareness, and decision making with the intent of including at least 4 items from each of category in the instrument. We wrote a potential item pool of 30 questions. Two other study investigators (CB, POS) reviewed the items and we rephrased or eliminated problematic items. We then constructed a 24-item draft instrument. Three physicians with expertise in pediatric critical care and anesthesia pilot tested the draft and offered comments on content and organization leading to only minor changes in the instrument.

Observer rating instruments

To assess performance of CRM skills, we used both the ANTS system (Fletcher et al. 2003) and the Ottawa GRS (Kim et al. 2006). The ANTS includes four "skills categories" (task management, team working, situation awareness, and decision making) with a total of 15 items. Fletcher et al. established the internal consistency of the items that fall under each of their four scores and demonstrated a best fit for 13 of 15 of these items, but did not report direct evidence of the scores' discreteness. The items are rated on a scale of four, although as others have done recently, we adapted the instrument to a scale of seven to increase the range of possible scores (Yee et al. 2005). Fletcher et al. reported an inter-rater agreement of 0.56–0.65 at the category level. The Ottawa GRS instrument is divided into five specific skills (leadership, problem solving, situation awareness, resource utilization, and communication) and includes an overall performance score. Kim et al. provided a theoretical justification for their choice of five scores in the Ottawa GRS, but no data to support their treatment as distinct constructs. Each item is rated on a 7-point scale. In the pilot study to establish inter-rater reliability, they reported intraclass correlation coefficients for single measures of 0.24–0.63 for the specific skill scores and 0.59–0.61 for the overall score.

Study procedures

The self-efficacy instrument was administered to all participants. After completing the instrument, a subset of residents (the observer-rated group) led simulated resuscitations as part of the residency program curriculum. Only second and third year residents were eligible to lead simulated resuscitations and participate in this part of the study. The simulation sessions followed a structured format. One faculty instructor (SvS) wrote all case scenarios. Scenarios differed per session, but were constructed in a standardized manner with three learning objectives per scenario specific to its medical content. Each session included interprofessional teams, occurred in situ, and demanded a similar level of CRM skills. They utilized medium fidelity mannequins (ALS Baby Trainer with Heartsim 200 and MegaCode Kid VitalSim, Laerdal Medical, Wappingers Falls, NY) and were video recorded. Three independent, trained observers (JP, DS, SvS) viewed the videos and scored the residents' performance of CRM skills on both the ANTS and Ottawa GRS instruments.

Statistical analysis

The statistical analysis included examination of: (1) instrument internal structure and (2) instrument relationship to other variables including known group comparison and comparison to performance.

Instrument internal structure

Since this was an explorative study, we initially performed an exploratory factor analysis (EFA) to determine optimal representation of the data. We included approximately 5 subjects per item in order to ensure stability of the factor analysis (Streiner 1994). Using principal axis factoring, we determined the number of factors observed based on sampling adequacy as assessed with the KMO statistic and on the eigenvalues greater than 1. We eliminated items with loadings <0.4 (Stevens 2002). To determine the goodness of fit and the strength of the parameter estimates and to consider the prior work in crisis resource management that suggested a two factor model (cognitive and interpersonal skills Fletcher et al. 2002), we conducted an EFA within the confirmatory factor analysis (CFA) framework. This method of EFA within CFA has been described elsewhere in the literature (Brown 2006). The subsequent evaluation of the factor analytic models was based on relevant fit indices including: Chi square test/ $df < 2$, Root Mean Square Error of Approximation <0.08 , as well as Tucker Lewis Nonnormed Fit Index and Comparative Fit Index both >0.85 (Hu and Bentler 1999). We examined the average inter-item correlation by calculating Cronbach's alpha for each factor derived from the factor analysis. A value >0.70 was considered adequate internal consistency (Nunnally 1978).

We calculated mean scores for each participant for each factor on the self-efficacy instrument for subsequent validity analyses.

Instrument relationship to other variables: known group comparison

With analysis of variance, we compared the self-efficacy factor scores reported by pediatric residents with those of the pediatric critical care fellows and faculty. This analysis was based on the assumption that pediatric residents have less experience with pediatric emergencies and therefore lower self-efficacy in CRM skills than pediatric critical care

fellows and faculty. We did not evaluate the second and third year residents as separate groups due to the small numbers in each group. In addition, the study took place during the transition between academic years, therefore a separation between second and third year residents would not truly reflect a year difference in experience for all participants.

Instrument relationship to other variables: comparison to performance

We calculated mean observer scores for the four category scores on the ANTS instrument and each score on the Ottawa GRS instrument and found high inter-item correlations for both. For the purposes of our analysis, we therefore calculated composite observer scores for these two instruments by averaging the mean observer scores for the four category scores on the ANTS instrument and averaging the mean observer scores for the five specific scores (excluding the overall score) on the Ottawa GRS instrument. In the process, we eliminated the item “Decision Making: Balancing risks and selecting options” from the ANTS instrument due to a consistent inability to rate this item in our simulations. We assessed inter-rater reliability by calculating type III intraclass correlation coefficients for average measures and considered >0.8 to be good inter-rater reliability (Landis and Kock 1977). In order to evaluate the relationship between self-efficacy and performance of CRM skills, we calculated Pearson’s correlation coefficients for self-efficacy factor scores with composite observer scores for those residents completing the simulations.

We used SPSS Version 16.0 for all statistical calculations other than the factor analysis that we performed with Mplus Version 4.1. A p value < 0.05 was considered significant for all calculations except the Pearson’s correlations. For these correlations between self-efficacy and observer scores, we performed a Bonferroni correction, setting $p = 0.025$ to maintain the familywise error rate at the 0.1 level.

Results

A total of 125 study participants completed the self-efficacy instrument: 31 first year pediatric residents (PGY-1), 34 PGY-2’s, 35 PGY-3’s, 3 pediatric chief residents, as well as 13 fellows and 9 faculty members in pediatric critical care. Thirty pediatric residents (14 PGY-2’s, 16 PGY-3’s) from one institution participated in video recorded simulated resuscitations (observer-rated group).

Instrument internal structure

The EFA specified a four factor model based on sampling adequacy ($KMO = 0.88$) and eigenvalues exceeding 1. In this model, four items were deleted from the final analysis due to low correlations with all other items in the data. The CFA of the remaining 20 items supported the four factor model solution. Each of the model fit indices indicated acceptable fit: χ^2 test = 1.85, RSMEA = 0.08, TLI = 0.89, and CFI = 0.91. The factor loadings were all significant with no negative residual variances. We defined these factors based on past models of CRM skills and our understanding of the construct as: situation awareness, team management, environment management, and decision making. Table 1 contains the items, item loadings, communalities, eigenvalues, and reliability coefficient for each factor.

Table 1 Factor analysis of the self-efficacy instrument

Factor	Defined as	Items	Item loadings	Communalities (h^2)	Eigenvalues (% of variance)	Cronbach's alpha
1	Situation awareness	Recognize clinical deterioration	0.49	0.44	10.06 (45.7)	0.82
		Anticipate events	0.86	0.71		
		Plan	0.80	0.76		
		Gather information	0.47	0.53		
2	Team management	Delegate tasks	0.69	0.66	2.08 (9.4)	0.88
		Coordinate team members	0.82	0.78		
		Identify and utilize skills of team	0.74	0.65		
		Instruct and correct team members	0.56	0.67		
3	Environment management	Communicate plan to team	0.56	0.51	1.16 (5.3)	0.77
		Access additional resources	0.40	0.39		
		Ensure requested interventions taken place	0.68	0.68		
		Provide reassurance and encouragement	0.62	0.68		
4	Decision making	Remain calm	0.54	0.46	1.00 (4.5)	0.91
		Take charge	0.56	0.75		
		Make decisions	0.55	0.70		
		See big picture	0.48	0.58		
		Identify possible interventions	0.57	0.66		
		Decide on most appropriate intervention	0.75	0.68		
		Prioritize	0.62	0.67		
Re-evaluate and change plan as needed	0.58	0.71				

Instrument relationship to other variables: known group comparison

As expected, pediatric residents reported significantly lower mean self-efficacy scores for all four factors measured on our self-efficacy instrument than pediatric critical care fellows and faculty (Table 2).

Instrument relationship to other variables: comparison to performance

The mean self-efficacy scores for the observer-rated group ($n = 30$) and the mean composite observer scores are shown in Table 3. The intraclass correlation for the three observers was 0.86 for the composite ANTS score and 0.85 for the composite Ottawa GRS score.

The correlations between residents' self-efficacy scores and composite observer scores are listed in Table 4. Significant moderate positive correlations were observed between

Table 2 Comparison of mean self-efficacy scores for residents versus fellows/faculty

Factor	Residents (<i>n</i> = 103) Mean ± SD	Fellows/faculty (<i>n</i> = 22) Mean ± SD	<i>p</i> value
Situation awareness	3.75 ± 0.51	4.48 ± 0.38	<0.001
Team management	3.23 ± 0.64	3.99 ± 0.59	<0.001
Environment management	3.54 ± 0.53	4.36 ± 0.49	<0.001
Decision making	3.18 ± 0.57	4.31 ± 0.54	<0.001

Table 3 Resident scores on the self-efficacy, anaesthetists' non-technical skills, and Ottawa global rating scale instruments

Instrument	Score	Scale	Mean ± SD (<i>n</i> = 30)
Self-efficacy instrument	Situation awareness	1–5	3.78 ± 0.51
	Team management	1–5	3.33 ± 0.56
	Environment management	1–5	3.49 ± 0.42
	Decision making	1–5	3.33 ± 0.60
ANTS instrument	Composite score	1–7	5.18 ± 0.99
Ottawa GRS instrument	Composite score	1–7	5.12 ± 1.00

residents' self-efficacy in situation awareness and environment management and the composite ANTS score (0.52 and 0.44 respectively, $p < 0.025$) as well as between these two self-efficacy scores and the composite Ottawa GRS score (0.54 and 0.42, $p < 0.025$).

Discussion

We developed an instrument to measure pediatric residents' self-efficacy in CRM skills and found validity evidence for the scores from our instrument through content validity, internal structure, and relationship to other variables including known group comparison and comparison with performance of CRM skills. We will review our findings in light of other work on CRM skills, self-efficacy and performance, consider possible reasons why only two of our self-efficacy factors significantly correlated with performance, and reflect on where our work fits within the research on self-assessment.

Table 4 Correlation between residents' self-efficacy scores and observed performance scores

Self-efficacy score	ANTS score		Ottawa GRS score	
	Correlation	<i>p</i> -value	Correlation	<i>p</i> -value
Situation awareness	0.52	0.004	0.54	0.002
Team management	0.25	0.185	0.28	0.133
Environment management	0.44	0.016	0.42	0.022
Decision making	0.33	0.076	0.35	0.058

We drew heavily from previous work in the field of crisis resource management and personal experience of the principal investigators and other experts when developing our instrument. In spite of this attempt to ensure content validity, two of our items, “follow Pediatric Advanced Life Support algorithm” and “consider a variety of explanations for the symptoms”, were eliminated during factor analysis likely because they refer more to medical knowledge than CRM skills. The reasons for the poor performance of the other two eliminated items, “stay calm yourself” and “elicit suggestions from other team members”, are less clear however in a subsequent study by our group, residents were found to be particularly unable to self-assess their level of calmness and their degree of interaction with the team while leading simulated resuscitations (Plant et al. 2010). Our residents may lack insight into their ability and performance of these two specific skills. Factor analysis in this study indicated the remaining 20 items fell with good internal consistency reliability into four distinct areas of CRM skills: situation awareness, team management, environment management, and decision making. In spite of the fact that we referenced both the ANTS (Fletcher et al. 2003) and Ottawa GRS (Kim et al. 2006) instruments when developing ours, two of our factors (team management and environment management) are not shared with either of these instruments. This discrepancy is likely because of the complexity of the construct of CRM skills and possibly because our methods and target audiences differ from those of Fletcher et al. and Kim et al. Interestingly, in our study, the inter-item correlations for observers’ scores on the ANTS and Ottawa GRS instruments were so high that we created composite observer scores for each. It is unclear whether our high degree of inter-item correlations was a function of the instruments or the “halo effect” (Nunnally 1978).

We found significant positive correlations between two factors from our self-efficacy instrument and performance as measured by the composite observer scores on the ANTS and Ottawa GRS instruments. Since published guidelines suggest that correlations with an absolute value between 0.3 and 0.5 are considered moderate (Cohen et al. 2003), our findings give some support the assertion that self-efficacy is related to performance of the associated skills, at least in the context of resuscitation (Bandura 1977; Maibach et al. 1996). Maibach et al. (1996, p. 95) have discussed the theoretical importance of self-efficacy in resuscitation training, stating, “it is likely to influence the development of and real-time access to other cognitive, affective, psychomotor, and social aspects of resuscitation proficiency”. Studies have provided conflicting evidence regarding this hypothesis. In a study of internal medicine residents’ ability to follow ACLS algorithms during simulated resuscitations, there was no correlation between self-efficacy and performance (Wayne et al. 2006). In another study, a large majority of pediatric residents expressed confidence in technical skills such as endotracheal intubation, whereas only a minority performed those skills properly (Nadel et al. 2000a). A more recent study found positive correlations between self-efficacy in technical skills and the initiation, but not the successful performance, of those behaviors during simulated resuscitations. This same study showed a moderate correlation ($r = 0.48$) between self-efficacy in general resuscitation skills and observers’ assessment of their global performance (Turner et al. 2009). While their study provides the first positive evidence for a link between self-efficacy and performance during resuscitations, the study did not allow for the demonstration and measurement of CRM skills, since the simulated resuscitations were not performed in a team context. Our study is unique in that we examined the relationship between self-efficacy and performance of CRM skills during interdisciplinary simulated resuscitations.

We found correlation with performance for self-efficacy in situation awareness and environment management, but not for team management and decision making. A possible

explanation for this finding is that level of experience may affect ability to predict skill level. “Personal performance mastery experiences” are thought to be among the most direct and powerful factors affecting self-efficacy (Bandura 1997). It is assumed that successful practice of skills increases self-efficacy whereas unsuccessful practice decreases self-efficacy. A lack of experience with skills may, therefore, limit an individual’s ability to accurately assess self-efficacy, a concept called the “dual burden” by Kruger and Dunning (1999). In addition, there is some evidence that exposure to benchmarking examples and feedback improves an individual’s ability to self-assess accurately (Lane and Gottlieb 2004; Martin et al. 1998). Pediatric residents have considerable experience with patient assessment and resource acquisition, activities corresponding to situation awareness and environment management. While engaged in these patient care activities, they are exposed to benchmarks as they observe their peers and attending physicians demonstrate the related skills, receive feedback regarding their own performance and are likely to reflect on their performance in light of this feedback. Accordingly, due to their experiences requiring situation awareness and environment management skills, residents in this study may have been better able to assess their abilities. In contrast, pediatric residents have less experience, especially early in their training, with independent decision-making and team management, the other two factors on our self-efficacy instrument and, as a result, may be unable to accurately assess these abilities.

Our results support that there is a role for assessment of self-efficacy in crisis resource management training. In light of our findings and the fact that self-efficacy is a form of self-assessment, the conclusion that “physicians are inaccurate self-assessors” may be premature and an oversimplification. Many of the studies on which this conclusion is based have been criticized for their suboptimal quality (Colthart et al. 2008; Davis et al. 2006). In this study, we specifically addressed the issue of measurement and used psychometrically robust approaches to measuring both self-efficacy and performance. We were able to find some correlation between self-efficacy and performance suggesting that the ability to self-assess may be task and context specific.

Our study has several limitations. First, since this study was incorporated into a resident resuscitation curriculum and a variety of scenarios were used to maximize the educational experience, the scenarios were not identical for all residents. The complexity and therefore difficulty of scenarios may have varied, however, the residents were not evaluated on their medical knowledge or technical skills and should have been able to demonstrate the same level of CRM skills in each scenario. Second, we were unable to find a validated rating instrument for CRM skill performance that was developed specifically for use in pediatrics. Since CRM skills are generic behavioral skills, generalizable across fields of medicine and we were able to achieve good inter-rater reliability with instruments developed for adult practitioners, the ANTS and Ottawa GRS seem to have been appropriate for our setting. Finally, our study through its quantitative nature was not designed to elucidate the underlying reasons for our pattern of findings. Self-assessment is a qualitative process and qualitative inquiry into an individual’s approach to self-assessment may shed further light on the factors that determine its accuracy (Colliver et al. 2005; Colthart et al. 2008; Ward et al. 2002).

Conclusion

With this study, we add to the evidence that self-efficacy correlates with performance of resuscitation skills, at least in the domains of situation awareness and environment

management. Our findings highlight the need for more in-depth research into the determinants of self-assessment. When applied to specific domains and in a well-defined context with adequate feedback and benchmarks, self-assessment may accurately inform self-directed learning and evaluation.

Acknowledgments The authors of this article would like to thank Drs. Natalie Cvijanovich, Sanjeev Datar, and Maurice Zwass for their review of the self-efficacy instrument. This study was supported in part by a UCSF Faculty Education Research Grant. Dr. Plant was supported by a NICHD Institutional Training Grant for Pediatric Critical Care T32HD049303-03.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: H-Freeman and Company.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In T. Urdan & F. Pajares (Eds.), *Self-efficacy beliefs of adolescents*. Greenwich, CT: Information Age Publishing.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Colliver, J. A., Verhulst, S. J., & Barrows, H. S. (2005). Self-assessment in medical practice: A further concern about the conventional research paradigm. *Teaching and Learning in Medicine*, 17, 200–201.
- Colthart, I., Bagnall, G., Evans, A., Allbutt, H., Haig, A., Illing, J., et al. (2008). The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. *Medical Teacher*, 30, 124–145.
- Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *JAMA*, 296, 1094–1102.
- Downing, S. M. (2003). Validity: On meaningful interpretation of assessment data. *Medical Education*, 37, 830–837.
- Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N., & Patey, R. (2003). Anaesthetists' Non-Technical Skills (ANTS): Evaluation of a behavioural marker system. *British Journal of Anaesthesia*, 90, 580–588.
- Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N., & Patey, R. (2004). Rating non-technical skills: Developing a behavioural marker system for use in anaesthesia. *Cognition, Technology & Work*, 6, 165–171.
- Fletcher, G. C., McGeorge, P., Flin, R. H., Glavin, R. J., & Maran, N. J. (2002). The role of non-technical skills in anaesthesia: A review of current literature. *British Journal of Anaesthesia*, 88, 418–429.
- Gaba, D., Fish, K., & Howard, S. (1994). *Crisis management in anesthesiology*. New York: Churchill Livingstone.
- Holmes, W. C., & Shea, J. A. (1998). A new HIV/AIDS-targeted quality of life (HAT-QoL) instrument: Development, reliability, and validity. *Medical Care*, 36, 138–154.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Kane, M. T. (1992). The assessment of professional competence. *Evaluation & The Health Professions*, 15, 163–182.
- Kim, J., Neilipovitz, 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, 2167–2174.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121–1134.

- Landis, J. R., & Kock, C. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.
- Lane, J. L., & Gottlieb, R. P. (2004). Improving the interviewing and self-assessment skills of medical students: Is it time to readopt videotaping as an educational tool? *Ambulatory Pediatrics*, *4*, 244–248.
- Maibach, E. W., Schieber, R. A., & Carroll, M. F. (1996). Self-efficacy in pediatric resuscitation: Implications for education and performance. *Pediatrics*, *97*, 94–99.
- Martin, D., Regehr, G., Hodges, B., & McNaughton, N. (1998). Using videotaped benchmarks to improve the self-assessment ability of family practice residents. *Academic Medicine*, *73*, 1201–1206.
- Nadel, F. M., Lavelle, J. M., Fein, J. A., Giardino, A. P., Decker, J. M., & Durbin, D. R. (2000a). Assessing pediatric senior residents' training in resuscitation: Fund of knowledge, technical skills, and perception of confidence. *Pediatric Emergency Care*, *16*, 73–76.
- Nadel, F. M., Lavelle, J. M., Fein, J. A., Giardino, A. P., Decker, J. M., & Durbin, D. R. (2000b). Teaching resuscitation to pediatric residents: The effects of an intervention. *Archives of Pediatrics and Adolescent Medicine*, *154*, 1049–1054.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill, Inc.
- Plant, J., Corden, M., Mourad, M., O'Brien, B., & van Schaik, S. (2010). *Understanding the self-assessment Process: Preliminary results of a study evaluating residents' performance of crisis resource management skills*. Paper presented at the AAMC Western Group on Educational Affairs Annual Meeting, Monterey, CA.
- Reader, T., Flin, R., Lauche, K., & Cuthbertson, B. H. (2006). Non-technical skills in the intensive care unit. *British Journal of Anaesthesia*, *96*, 551–559.
- Reznek, M., Smith-Coggins, R., Howard, S., Kiran, K., Harter, P., Sowb, Y., et al. (2003). Emergency medicine crisis resource management (EMCRM): Pilot study of a simulation-based crisis management course for emergency medicine. *Academic Emergency Medicine*, *10*, 386–389.
- Shea, J. A., & Fortna, G. F. (2002). Psychometric methods. In G. R. Norman, C. P. M. van der Vleuten, & D. I. NEWBLE (Eds.), *International Handbook of Research in Medical Education: Part 1*. Boston: Kluwer Academic Publishers.
- Stevens, J. P. (2002) *Applied multivariate statistics for the social sciences*. Mahwah: Lawrence Erlbaum Associates.
- Streiner, D. L. (1994). Figuring out factors: The use and misuse of factor analysis. *Canadian Journal of Psychiatry*, *39*, 135–140.
- Turner, N. M., Lukkassen, I., Bakker, N., Draaisma, J., & ten Cate, O. T. (2009). The effect of the APLS-course on self-efficacy and its relationship to behavioural decisions in paediatric resuscitation. *Resuscitation*, *80*, 913–918.
- van Schaik, S. M., von Kohorn, I., & O'Sullivan, P. (2008). Pediatric resident confidence in resuscitation skills relates to mock code experience. *Clinical pediatrics (Philadelphia)*, *47*, 777–783.
- Wang, Y.-W., Davidson, M. M., Yakuchko, O. F., Savoy, H. B., Tan, J. A., & Bleier, J. K. (2003). The scale of ethnocultural empathy: Development, validation, and reliability. *Journal of Counseling Psychology*, *50*, 221–234.
- Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education: Theory and Practice*, *7*, 63–80.
- Wayne, D. B., Butter, J., Siddall, V. J., Fudala, M. J., Wade, L. D., Feinglass, J., et al. (2006). Graduating internal medicine residents' self-assessment and performance of advanced cardiac life support skills. *Medical Teacher*, *28*, 365–369.
- Yee, B., Naik, V. N., Joo, H. S., Savoldelli, G. L., Chung, D. Y., Houston, P. L., et al. (2005). Nontechnical skills in anesthesia crisis management with repeated exposure to simulation-based education. *Anesthesiology*, *103*, 241–248.
- Yudkowsky, R., Downing, S. M., & Sandlow, L. J. (2006). Developing an institution-based assessment of resident communication and interpersonal skills. *Academic Medicine*, *81*, 1115–1122.