

# Successful implementation of a rater training program for medical students to evaluate simulated pediatric emergencies

## Abstract

**Introduction:** Simulation-based training is increasingly used in pediatrics to teach technical skills, teamwork, and team communication, and to improve potential deficiencies in pediatric emergency care. Team performance must be observed, analyzed, and evaluated by trained raters. The structured training of medical students for the assessment of simulated pediatric emergencies has not yet been investigated.

**Methods:** We developed a rater training program for medical students to assess guideline adherence, teamwork, and team communication in simulated pediatric emergencies. Interrater reliability was measured at each training stage using Kendall tau coefficients.

**Results:** In 10 out of 15 pairs of raters interrater reliability was moderate to high ( $\tau > 0.4$ ), whereas it was low in the remaining 5 pairs of raters.

**Discussion:** The interrater reliability showed good agreement between medical students and expert raters at the end of the rater training program. Medical students can be successfully involved in the assessment of guideline adherence as well as teamwork and team communication in simulated pediatric emergencies.

**Keywords:** simulation, medical students, pediatric life support

Nadine Mand<sup>1</sup>  
Tina Stibane<sup>2</sup>  
Helmut Sitter<sup>2</sup>  
Rolf Felix Maier<sup>1</sup>  
Andreas Leonhardt<sup>1</sup>

<sup>1</sup> Philipps University of Marburg, University Hospital Marburg, Department of Pediatrics, Marburg, Germany

<sup>2</sup> Philipps University of Marburg, Dr. Reinfried Pohl Center for Medical Education, Marburg, Germany

## 1. Introduction

Prehospital and intrahospital resuscitation in children occurs less frequently than in adults and is associated with high morbidity and mortality [1], [2], [3], [4]. For the majority of physicians working in pediatrics, routine in these stressful emergencies cannot be achieved. They often feel unconfident and unprepared during resuscitation [5], [6], [7], which has an impact on the delivery of emergency care [8], [9], [10].

Inadequate technical skills [8], [11], lack of teamwork, and communication errors are considered to be reasons for deficiencies in emergency care [12], [13]. Simulation-based training formats address these deficiencies in a targeted manner [14], [15]. However, in addition to structured development and embedding in a meaningful teaching concept [16], [17], effective simulation training requires accurate observation and evaluation of team performance [18]. This is essential for effective debriefing following a simulated or real emergency [19] and for scientific evaluation, quality assurance, and optimization of simulation-based training formats.

In addition to the selection of a valid and reliable assessment tool, specifically trained raters are essential to observe, analyze and evaluate technical skills, teamwork, and team communication in simulated emergencies [18], [20].

Physicians in the early stages of training or clinically inexperienced medical students were involved as raters with varying success [20], [21], [22], [23], [24], [25]. Previously published descriptions of rater training are inconsistent, and details on content are rarely provided [20], [21], [22], [23], [24]. Systematic training of medical students to assess simulated pediatric emergencies has not yet been investigated. This study describes the development, evaluation, and successful implementation of a rater training program for medical students to assess guideline adherence, teamwork, and team communication.

## 2. Methods

### 2.1. Assessment tools

We used the *performance evaluation checklist for pulseless ventricular tachycardia (PEC-PVT)* for evaluating guideline adherence (see attachment 1) [26]. The PEC-PVT evaluates the management of a simulated pediatric emergency with cardiovascular failure due to a shockable heart rhythm. It consists of 31 items, which are divided into three evaluation categories (task not performed; task performed partially, incorrectly, or with delay; and task performed completely). Each item is weighted between 1 and 5 according to its importance to treatment success.

Thus the PEC-PVT illustrates the complexity of a resuscitation situation. The English version has been validated with German pediatric emergency teams [26]. It was translated into German by one of the original authors (EH), followed by a consensus on the final version between EH, the co-authors, and the principal investigator (NM). Based on the phases of the study scenario, individual items were specified, and structured by the principal investigator. To enable the most accurate evaluation of all items, a rater training handbook was developed, thus specifying the rating of each item (see attachment 2). Teamwork and team communication were evaluated using the *Teamwork Emergency Assessment Measure (TEAM)* [27]. Consisting of 11 items that evaluate aspects of team leadership, teamwork, and team communication on a 5-point Likert scale (“never/hardly ever” to “always/nearly always”), as well as a global assessment of teamwork (scale from 1 to 10), this checklist is easy to use. The validity and reliability of the TEAM are high [28], [29]. Since no German version had been published at the beginning of the rater training, the checklist was translated into German using forward-backward translation [30]. Based on the “TEAM Behavioural Markers” published by Cooper [31] and Rall’s Crew Resource Management concept [32], a rater training handbook was designed to provide users with specific and observable “behavioral anchors” that represent the extreme ends of the TEAM observation scale [18].

## 2.2. Participants

Five medical students in intermediate to advanced training (7<sup>th</sup> to 13<sup>th</sup> semester) participated in the rater training. Two of the students had basic simulation knowledge as student tutors in medical education. One student already had experience in emergency medicine as a trained paramedic. All students were beginners in the structured observation and analysis of simulated emergencies.

Each student was trained in the use of one checklist. Training in the application of the specific Performance Evaluation Checklist (PEC-PVT) was conducted for two of the medical students, while three were trained in the application of the Team Emergency Assessment Measure (TEAM).

## 2.3. Rater training

The rater training program was based on previously published framework structures for rater training [18], [33], [34] and included different formats such as self-study and interdisciplinary training. Three core training strategies being essential for the evaluation of observations were incorporated: *Rater Error Training* deals with typical observational errors and their avoidance, *Performance Dimension Training* teaches the recognition of desired behaviors for the item to be examined and in the *Frame of Reference Training*, variations in the quality of a desired behavior is demonstrated [33], [34].

Rater training took place over a period of six weeks and was carried out separately by the principal investigator for each of the two assessment tools.

### Session 1: Conveying expertise

In external courses, the medical students were taught basic knowledge relevant to their assessment tool. Two students attended a two-day Pediatric Advanced Life Support (EPALS) course offered by the European Resuscitation Council, thus learning about the recognition of a critically ill child as well as life support in the event of respiratory and cardiovascular arrest. Three students took part in a one-day Crew Resource Management (CRM) course, learning aspects of team leadership, teamwork, and team communication.

### Session 2: Self-study

The students familiarized themselves with the theoretical aspects of the above-mentioned core training strategies, and typical rater errors to achieve a general understanding of evaluating team performance. They also familiarized themselves with the assessment tools.

### Session 3: First video evaluation

There were two rater trainings: assessing team performance using the PEC-PVT was taught separately from using TEAM. After a structured summary of relevant rater errors followed by a group discussion, two video examples were individually rated by each medical student and the principal investigator. Examples of poor or acceptable team performance were chosen. The team performance was recorded using a paper-and-pencil version of the respective checklists. Subsequently, the individual items were discussed with special attention to those items with the lowest agreement between the medical students and the principal investigator, and reasons for substantial deviation were named.

### Session 4: Rater training handbook, second video evaluation

For the second video session, the above-mentioned two videos with poor and acceptable team performance, as well as three additional videos were evaluated independently using the rater training handbook. The ensuing discussion focused on the individual items with the least agreement. Here, too, training took place separately for the PEC-PVT and the TEAM.

### Session 5: Pilot testing

Over a period of 4 weeks, the medical students and the principal investigator evaluated further nine videos using the checklists and the rater training handbook.

## 2.4. Statistics

In each phase of the rater training, interrater reliability between medical students and the principal investigator was examined pairwise by calculating the Kendall-tau-b correlation coefficient. The Kendall-tau coefficient measures the agreement of raters and thus quantifies the reliability of the observation system. This coefficient is a rank correlation coefficient and takes values between -1 and 1. This measure is suitable for ordinal data such as Likert scales. Only the ranking of the respective values is taken into account. Tau is determined using the individual items of the assessment tools [35]. A weak or low agreement is assumed with a Kendall-tau coefficient of  $\leq 0.4$ , a moderate one with 0.41-0.7, and a high agreement with values  $> 0.7$  [36].

The calculated Kendall-tau correlation coefficients are based on two videos in rater training session 3, five videos in session 4, and nine videos in session 5. The number of videos was restricted due to the complexity of the rater training program. The pairwise matches were determined to evaluate the assessment of a student in comparison to the principal investigator.

Differences in the assessment of individual items were investigated using Mann-Whitney-U tests (for the eleven individual items of TEAM and all items of the PEC-PVT) and t-tests (for the global evaluation of TEAM). Data analysis was carried out using SPSS V24 (Armonk, NY: IBM Corp.).

## 2.5. Ethics

The rater training program was developed as part of the study: "Quality assurance of pediatric emergency care through in-house simulation training at Hessian Children's Hospitals". The Ethics Committee of the Philipps University Marburg approved this study (ref. no. 172/16).

## 3. Results

All five medical students completed the rater training program. Sixteen videos (two videos from rater training session 3, five videos from session 4, and nine videos from session 5) of five pairs of raters each (one student and the principal investigator) were evaluated. With five pairs of observations and three video sessions, 15 Kendall-tau coefficients could be calculated (see table 1).

Five of the 15 Kendall-tau coefficients showed weak agreement between medical students and the principal investigator, five moderate, and five high agreement (see table 1). In three cases, the agreement between the first and second video assessments improved (observation pairs 1, 3, and 5).

The U- and t-tests performed did not show any significant differences between medical students and the principal investigator, i.e. no individual items could be identified

in which the evaluation differed between beginners and expert.

## 4. Discussion

We present a rater training program enabling medical students to evaluate simulated pediatric emergencies. Different training formats and core training strategies were combined [33], [34]. We aimed for evaluating and implementing a rater training program for consecutive studies based on video evaluations of team performance for both experts and trained beginners.

Being usually beginners in assessing team performance, medical students are rarely chosen for evaluating complex simulated emergencies. Freytag and colleagues involved students in the evaluation of teamwork and team communication of simulated emergencies in medical training after a rater training [20]. Interrater reliability was moderate, due to a more lenient assessment by the students as compared to the experienced co-raters [20], [37]. Initial interrater reliability before participation in rater training is not provided. Evans et al. investigated the use of a checklist for the assessment of invasive procedures by students after four hours of rater training and found a good agreement between the ratings of beginners and experts [22]. They concluded that teaching observational skills is possible even without prior background knowledge of the raters [22].

We demonstrated that medical students as beginners in structured behavioral observation can achieve good interrater reliability in the context of a rater training program. This was evident both in the evaluation of complex behaviors such as teamwork and team communication and in the assessment of technical skills such as guideline adherence. After a structured and detailed training of both evaluation strategies and theoretical and practical aspects of pediatric life support (rater training sessions 1 to 3), a good agreement in the evaluation of guideline adherence between beginners and experts was demonstrated, even before the actual video evaluation. Subsequently, the use of rater training handbooks improved the interrater reliability of those evaluating teamwork and team communication (session 4). This *frame of reference* training has been shown in previous studies to be particularly effective in rater training [34]. Extended rater training beyond the evaluation of a few exercise videos (session 5) did not further improve interrater reliability. Medical students can therefore support experts in the observation, analysis, and evaluation of simulated emergencies. This is relevant in the context of limited human resources, as simulation-based training formats are increasingly used in medical education and training. In addition to the training of technical and non-technical skills [38], simulation training can address relevant aspects of patient safety [39], [40] and identifies recurring and often avoidable treatment errors in the management of emergencies [41], [42], [43]. This also requires adequate training of the experts, i.e. teaching physicians.

**Table 1: tau\* for the agreement between students and the principal investigator**

Stage of rater training		First video evaluation (2 videos)	Second video evaluation (5 videos)	Pilot testing (9 videos)
Observation pair	Instrument			
1	TEAM <sup>#</sup>	0,17	0,24	0,36
2	TEAM	0,70	0,57	0,47
3	TEAM	0,14	0,46	0,31
4	PEC-PVT <sup>§</sup>	0,82	0,81	0,79
5	PEC-PVT	0,63	0,90	0,80

\*Kendall-tau  $\tau$ . Assessment of agreement:  $\tau \leq 0.4$  weak,  $0.41 - 0.7$  moderate,  $> 0.7$  high.

<sup>#</sup>TEAM: Team Emergency Assessment Measure. Items 1-11.

<sup>§</sup>PEC-PVT: Performance Evaluation Checklist for Pulseless Ventricular Tachycardia.

In addition to medical expertise, they need to be trained in core evaluation strategies, as well as in typical rater errors.

This study has several limitations.

The Team *Emergency Assessment Measure (TEAM)* was originally developed for experts, not for use by beginners in behavioral observation [27], [44]. Since no pairwise video evaluation was carried out before the start of the rater training program, the construct validity of this checklist is not known for beginners. However, an effect of the rater training can be assumed because 2 out of 3 rater pairs evaluating the TEAM showed improvements in interrater reliability.

The evaluation of the specific performance evaluation checklist (PEC-PVT) showed better interrater reliability in this study as compared to the original publication [26]. While translating the original checklist into German, we detected several ambiguously phrased items. These items were clarified in the German version, which could have improved the interrater reliability.

The rater training presented here is very personnel- and time-consuming and therefore not easily transferable to other studies. Since we could not demonstrate any improvement in interrater reliability in session 5 of the rater training program, training beyond a few exercise videos does not seem to be of any advantage.

## 5. Conclusion

In summary, we were able to demonstrate that medical students can evaluate technical skills, as well as teamwork and team communication in complex simulated emergencies after completing a rater training program. The rater training should be structured and contain established core training strategies. Extended training beyond the use of a few exercise videos does not seem necessary.

## Funding

The Hessian Ministry of Social Affairs and Integration financed pediatric emergency training at Hessian chil-

dren's hospitals, in the context of which the study videos were made.

Open Access funding is provided by the Open Access Publishing Fund of Philipps-University Marburg with the support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

## Competing interests

The authors declare that they have no competing interests.

## Attachments

Available from <https://doi.org/10.3205/zma001629>

1. Attachment\_1.pdf (302 KB)  
PEC-PVT rater training handbook
2. Attachment\_2.pdf (84 KB)  
PEC-PVT

## References

1. Jones H, Wilmshurst SL, Graydon C. Aetiology and outcome of paediatric cardiopulmonary arrest. *Anaesthesia Intensive Care Med.* 2017;18(11):537-540. DOI: 10.1016/j.mpaic.2017.07.003
2. Kendirli T, Erkek N, Köroğlu T, Yıldızdaş D, Bayrakç B, Güzel A, Çtak A, Demirkol D, Ağı H, Arslanköylü AE, Kutlu NO, Tuğgun N, Paksu MS, Anı AB, Kalkan G, Duman M, Dündaröz R, Asloğlu N, Yaman A, Ödek C, Tekin D, Dursan O, Sevketoglu E, Kesici S, Ates C, Gördü Z, Yılmaz HL, Ince E, Karaböcüoğlu M. Cardiopulmonary Resuscitation in Children With In-Hospital and Out-of-Hospital Cardiopulmonary Arrest: Multicenter Study From Turkey. *Pediatr Emerg Care.* 2015;31(11):748-752. DOI: 10.1097/PEC.0000000000000337
3. Holmberg MJ, Ross CE, Fitzmaurice GM, Chan PS, Duval-Arnould J, Grossestreuer AV, Yankama T, Donnino MW, Andersen LW; American Heart Association's Get With The Guidelines-Resuscitation Investigators. Annual Incidence of Adult and Pediatric In-Hospital Cardiac Arrest in the United States. *Circ Cardiovasc Qual Outcomes.* 2019;12(7):e005580.



4. Gräsner JT, Lefering R, Koster RW, Masterson S, Böttiger BW, Herlitz J, Wnent J, Tjelmeland IB, Ortiz FR, Maurer H, Baubin M, Mols P, Hadzibegovic I, Ioannidis M, Skeluc R, Wissenberg M, Salo A, Hubert H, Nikolaou NI, Loczi G, Svavarsdottir H, Semeraro F, Wright PJ, Clarens C, Pijls R, Cebula G, Correia VG, Cimpoesu D, Raffay V, Trenkler S, Markota A, Srömsöe A, Burkart R, Perkins GD, Bossaert LL; EuReCa ONE Collaborators. EuReCa ONE-27 Nations, ONE Europe, ONE Registry: A prospective one month analysis of out-of-hospital cardiac arrest outcomes in 27 countries in Europe. *Resuscitation*. 2016;105:188-195. DOI: 10.1016/j.resuscitation.2016.06.004
5. Lehmann R, Seitz A, Meyburg J, Hoppe B, Hoffmann GF, Tönshoff B, Huwendiek S. Pediatric in-hospital emergencies: real life experiences, previous training and the need for training among physicians and nurses. *BMC Res Notes*. 2019;12(1):19. DOI: 10.1186/s13104-019-4051-4
6. Walsh Ó, Lydon S, O'Connor P. A mixed methods evaluation of paediatric trainee preparedness to manage cardiopulmonary arrests. *Eur J Pediatr*. 2017;176(12):1653-1662. DOI: 10.1007/s00431-017-3017-6
7. Nadel FM, Lavelle JM, Fein JA, Giardino AP, Decker JM, Durbin DR. Teaching Resuscitation to Pediatric Residents: The Effects of an Intervention. *Arch Pediatr Adolesc Med*. 2000;154(10):1049-1054. DOI: 10.1001/archpedi.154.10.1049
8. Hunt EA, Vera K, Diener-West M, Haggerty JA, Nelson KL, Shaffner DH, Pronovost PJ. Delays and errors in cardiopulmonary resuscitation and defibrillation by pediatric residents during simulated cardiopulmonary arrests. *Resuscitation*. 2009;80(7):819-825. DOI: 10.1016/j.resuscitation.2009.03.020
9. Sutton RM, Niles D, French B, Maltese MR, Leffelman J, Eilevstjonn J, Wolfe H, Nishisaki A, Meaney PA, Berg RA, Nadkarni VM. First quantitative analysis of cardiopulmonary resuscitation quality during in-hospital cardiac arrests of young children. *Resuscitation*. 2014;85(1):70-74. DOI: 10.1016/j.resuscitation.2013.08.014
10. Sutton RM, French B, Niles DE, Donoghue A, Topjian AA, Nishisaki A, Leffelman J, Wolfe H, Berg RA, Nadkarni VM, Meaney PA. 2010 American Heart Association recommended compression depths during pediatric in-hospital resuscitations are associated with survival. *Resuscitation*. 2014;85(9):1179-1184. DOI: 10.1016/j.resuscitation.2014.05.007
11. 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. DOI: 10.1097/00006565-200004000-00001
12. Sherman JM, Chang TP, Ziv N, Nager AL. Barriers to Effective Teamwork Relating to Pediatric Resuscitations: Perceptions of Pediatric Emergency Medicine Staff. *Pediatr Emerg Care*. 2020;36(3):e146-e150. DOI: 10.1097/PEC.0000000000001275
13. Risser DT, Rice MM, Salisbury ML, Simon R, Jay GD, Berns SD. The potential for improved teamwork to reduce medical errors in the emergency department. The MedTeams Research Consortium. *Ann Emerg Med*. 1999;34(3):373-383. DOI: 10.1016/s0196-0644(99)70134-4
14. Cheng A, Lang TR, Starr SR, Pusic M, Cook DA. Technology-Enhanced Simulation and Pediatric Education: A Meta-analysis. *Pediatrics*. 2014;133(5):e1313-e1323. DOI: 10.1542/peds.2013-2139
15. Miledler LP, Urlesberger B, Szylid EG, Roehr CC, Schmölzer GM. Simulation-based neonatal and infant resuscitation teaching: a systematic review of randomized controlled trials. *Klin Pediatr*. 2014;226(5):259-267. DOI: 10.1055/s-0034-1372621
16. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach*. 2005;27(1):10-28. DOI: 10.1080/01421590500046924
17. Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simul Healthc*. 2007;2(2):115-125. DOI: 10.1097/SIH.0b013e3180315539
18. Eppich W, Nannicelli AP, Seivert NP, Sohn MW, Rozenfeld R, Woods DM, Holl JL. A Rater Training Protocol to Assess Team Performance. *J Contin Educ Health Prof*. 2015;35(2):83-90. DOI: 10.1002/chp.21270
19. Cheng A, Hunt EA, Donoghue A, Nelson-McMillan K, Nishisaki A, LeFlore J, Eppich W, Moyer M, Brett-Fleegler M, Kleinman M, Anderson J, Adler M, Braga M, Kost S, Stryjewski G, Min S, Podraza J, Lopreiato J, Fidor Hamilton M, Stone K, Reid J, Hopkins J, Manos J, Duff J, Richard M, Nadkarni VM; EXPRESS Investigators. Examining Pediatric Resuscitation Education Using Simulation and Scripted Debriefing: A Multicenter Randomized Trial. *JAMA Pediatrics*. 2013;167(6):528-536. DOI: 10.1001/jamapediatrics.2013.1389
20. Freytag J, Stroben F, Hautz WE, Schaubert SK, Kämmer JE. Rating the quality of teamwork—a comparison of novice and expert ratings using the Team Emergency Assessment Measure (TEAM) in simulated emergencies. *Scand J Trauma Resusc Emerg Med*. 2019;27(1):12. DOI: 10.1186/s13049-019-0591-9
21. Yule S, Rowley D, Flin R, Maran N, Youngson G, Duncan J, Paterson-Brown S. Experience matters: comparing novice and expert ratings of non-technical skills using the NOTSS system. *ANZ J Surg*. 2009;79(3):154-160. DOI: 10.1111/j.1445-2197.2008.04833.x
22. Evans LV, Morse JL, Hamann CJ, Osborne M, Lin Z, D'Onofrio G. The development of an independent rater system to assess residents' competence in invasive procedures. *Acad Med*. 2009;84(8):1135-1143. DOI: 10.1097/ACM.0b013e3181acec7c
23. Möltner A, Lehmann M, Wachter C, Kurczyk S, Schwill S, Loukanova S. Formative assessment of practical skills with peer-assessors: quality features of an OSCE in general medicine at the Heidelberg Medical Faculty. *GMS J Med Educ*. 2020;37(4):Doc42. DOI: 10.3205/zma001335
24. Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. *J Gen Intern Med*. 2009;24(1):74-79. DOI: 10.1007/s11606-008-0842-3
25. Sevdalis N, Lyons M, Healey AN, Undre S, Darzi A, Vincent CA. Observational teamwork assessment for surgery: construct validation with expert versus novice raters. *Ann Surg*. 2009;249(6):1047-1051. DOI: 10.1097/SLA.0b013e3181a50220
26. Schmutz J, Manser T, Keil J, Heimberg E, Hoffmann F. Structured Performance Assessment in Three Pediatric Emergency Scenarios: A Validation Study. *J Pediatr*. 2015;166(6):1498-504. DOI: 10.1016/j.jpeds.2015.03.015
27. Cooper S, Cant R, Porter J, Sellick K, Somers G, Kingsman L, Nestel D. Rating medical emergency teamwork performance: Development of the Team Emergency Assessment Measure (TEAM). *Resuscitation*. 2010;81(4):446-452. DOI: 10.1016/j.resuscitation.2009.11.027
28. Cooper SJ, Cant R. Measuring non-technical skills of medical emergency teams: an update on the validity and reliability of the Team Emergency Assessment Measure (TEAM). *Resuscitation*. 2014;85(1):31-33. DOI: 10.1016/j.resuscitation.2013.08.276
29. Valentine MA, Nembhard IM, Edmondson AC. Measuring teamwork in health care settings: a review of survey instruments. *Med Care*. 2015;53(4):e16-30. DOI: 10.1097/MLR.0b013e31827feef6

30. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P; ISPOR Task Force for Translation and Cultural Adaptation. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*. 2005;8(2):94-104. DOI: 10.1111/j.1524-4733.2005.04054.x
31. Monash University. TEAM Team Emergency Assessment Measure. Victoria: Monash University; 2012. Zugänglich unter/available from: <https://www.monash.edu/search?query=TEAM+Team+Emergency+Assessment+Measure.+2012>
32. Rall M. Human Factors und CRM: Eine Einführung. In: Breuer SP, editor. *Simulation in der Medizin*. Berlin, Heidelberg: Springer-Verlag; 2013. p.136-151. DOI: 10.1007/978-3-642-29436-5\_13
33. Feldman M, Lazzara EH, Vanderbilt AA, DiazGranados D. Rater training to support high-stakes simulation-based assessments. *J Contin Educ Health Prof*. 2012;32(4):279-286. DOI: 10.1002/chp.21156
34. Woehr DJ, Hffcutt AI. Rater training for performance appraisal: a quantitative review. *J Occup Organ Psychol*. 1994;68(3):189-205. DOI: 10.1111/j.2044-8325.1994.tb00562.x
35. Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155-159. DOI: 10.1037//0033-2909.112.1.155
36. Overholser BR, Sowinski KM. Biostatistics primer: part 2. *Nutr Clin Pract*. 2008;23(1):76-84. DOI: 10.1177/011542650802300176
37. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. DOI: 10.1016/j.jcm.2016.02.012
38. Clerihew L, Rowney D, Ker J. Simulation in paediatric training. *Arch Dis Child Educ Pract Ed*. 2016;101(1):8-14. DOI: 10.1136/archdischild-2015-309143
39. Hoffmann N, Kubitz JC, Goetz AE, Beckers SK. Patient safety in undergraduate medical education: Implementation of the topic in the anaesthesiology core curriculum at the University Medical Center Hamburg-Eppendorf. *GMS J Med Educ*. 2019;36(2):Doc12. DOI: 10.3205/zma001220
40. Schildmann J, Salloch S, Peters T, Henking T, Vollmann J. Risks and errors in medicine. Concept and evaluation of an optional study module with integrated teaching of ethical, legal and communicative competencies. *GMS J Med Educ*. 2018;35(3):Doc31. DOI: 10.3205/zma001177
41. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care*. 2008;17(3):216-223. DOI: 10.1136/qshc.2007.023622
42. Ventre KM, Barry JS, Davis D, Baiamonte VL, Wentworth AC, Pietras M, Coughlin L, Barley G. Using in situ simulation to evaluate operational readiness of a children's hospital-based obstetrics unit. *Simul Healthc*. 2014;9(2):102-111. DOI: 10.1097/SIH.0000000000000005
43. Slakey DP, Simms ER, Rennie KV, Garstka ME, Korndorffer JR, Jr. Using simulation to improve root cause analysis of adverse surgical outcomes. *Int J Qual Health Care*. 2014;26(2):144-150. DOI: 10.1093/intqhc/mzu011
44. Cooper S, Cant R, Connell C, Sims L, Porter JE, Symmons M, Nestel D, Liaw SK. Measuring teamwork performance: Validity testing of the Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. *Resuscitation*. 2016;101:97-101. DOI: 10.1016/j.resuscitation.2016.01.026

#### Corresponding author:

Dr. med. Nadine Mand  
Philipps University of Marburg, University Hospital  
Marburg, Department of Pediatrics, Baldingerstraße,  
D-35043 Marburg, Germany  
mandn@staff.uni-marburg.de

#### Please cite as

Mand N, Stibane T, Sitter H, Maier RF, Leonhardt A. Successful implementation of a rater training program for medical students to evaluate simulated pediatric emergencies. *GMS J Med Educ*. 2023;40(4):Doc47. DOI: 10.3205/zma001629, URN: urn:nbn:de:0183-zma0016291

#### This article is freely available from

<https://doi.org/10.3205/zma001629>

**Received:** 2021-07-02

**Revised:** 2023-03-04

**Accepted:** 2023-04-20

**Published:** 2023-06-15

#### Copyright

©2023 Mand et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

# Erfolgreiche Implementierung eines mehrstufigen Beobachtertrainings für Medizinstudierende zur Evaluation simulierter pädiatrischer Notfallsituationen

## Zusammenfassung

**Einleitung:** Auf Simulation basierende Schulungsformate werden zunehmend in der Pädiatrie eingesetzt, um technische Fertigkeiten, Teamarbeit und Teamkommunikation zu trainieren und Defizite in pädiatrischen Notfallversorgungen zu verbessern. Geschulte Beobachter\*innen sind notwendig, um dies im Rahmen simulierter Notfallsituationen zu erfassen, zu analysieren und zu beurteilen. Die strukturierte Schulung von Medizinstudierenden zur Beurteilung pädiatrischer Notfallsituationen im Rahmen von Simulationstrainings wurde bisher nicht untersucht.

**Methoden:** Entwicklung eines mehrstufigen Beobachtertrainings für Medizinstudierende zur Bewertung der Leitlinienadhärenz, sowie der Teamarbeit und -kommunikation. Die Interrater-Reliabilität wurde in jeder Stufe des Trainings durch die Bestimmung von Kendall tau-Koeffizienten ermittelt.

**Ergebnisse:** In 10 von 15 Beobachterpaaren wurde eine moderate bis hohe Interrater-Reliabilität gezeigt ( $\tau > 0,4$ ), in den übrigen 5 Beobachterpaaren eine schwache Übereinstimmung.

**Diskussion:** Die Interrater-Reliabilität zeigte nach Beendigung des Beobachtertrainings eine gute Übereinstimmung zwischen den Beobachter\*innen. Medizinstudierende können demnach erfolgreich in die Evaluation der Leitlinienadhärenz sowie der Teamarbeit und -kommunikation einbezogen werden.

**Schlüsselwörter:** Simulation, Medizinstudierende, pädiatrische Notfallversorgung

Nadine Mand<sup>1</sup>  
Tina Stibane<sup>2</sup>  
Helmut Sitter<sup>2</sup>  
Rolf Felix Maier<sup>1</sup>  
Andreas Leonhardt<sup>1</sup>

1 Philipps Universität Marburg, Universitätsklinikum Marburg, Zentrum für Kinder- und Jugendmedizin, Marburg, Deutschland

2 Philipps Universität Marburg, Dr. Reinfried Pohl-Zentrum für medizinische Lehre, Marburg, Deutschland

## 1. Einleitung

Reanimationen im Kindesalter treten sowohl prä- als auch innerklinisch deutlich seltener auf als bei Erwachsenen und gehen mit einer hohen Mortalität und Morbidität einher [1], [2], [3], [4]. Für die Mehrzahl der in der Kinder- und Jugendmedizin tätigen Ärztinnen und Ärzte ist es deshalb nicht möglich, Routine in diesen stressbehafteten Notfällen zu erlangen. Sie fühlen sich bei Reanimationen häufig unsicher und unvorbereitet [5], [6], [7]. Dies kann zu deutlichen Defiziten in der Notfallversorgung führen [8], [9], [10].

Unzureichende technische Fertigkeiten [8], [11], mangelnde Teamzusammenarbeit und Kommunikationsfehler werden als ursächlich für Defizite im Notfallmanagement angesehen [12], [13]. Um dies gezielt zu trainieren, wurden auch in der Pädiatrie auf Simulation basierende Schulungsformate etabliert [14], [15]. Ein effektives Simulationstraining bedarf jedoch neben einer strukturierten Entwicklung und der Einbettung in ein sinnvolles Lehrkonzept [16], [17] einer möglichst akkuraten Beobachtung und Bewertung der Teamleistung [18]. Dies ist

Voraussetzung sowohl für ein effektives Debriefing im Anschluss an eine simulierte oder reale Notfallsituation [19], als auch für die wissenschaftliche Evaluation, Qualitätssicherung und Optimierung simulationsbasierter Schulungsformate.

Neben der Auswahl eines sinnvollen Bewertungsinstruments sind geschulte Beobachter\*innen notwendig, um valide und reliabel technische Fertigkeiten, Teamzusammenarbeit und Kommunikation im Rahmen simulierter Notfallsituationen zu erfassen, zu analysieren und zu beurteilen [18], [20].

Ärztinnen und Ärzte in frühen Ausbildungsstadien oder klinisch unerfahrene Medizinstudierende wurden als Beobachter\*innen mit wechselndem Erfolg für diesen Zweck eingesetzt [20], [21], [22], [23], [24], [25]. Die publizierten Beschreibungen des Trainings dieser Beobachter\*innen sind uneinheitlich und spezifizieren zumeist nicht Inhalt oder Umfang [20], [21], [22], [23], [24]. Die systematische Schulung von Medizinstudierenden zur Beurteilung pädiatrischer Notfallsituationen im Rahmen von Simulationstrainings wurde bisher nicht untersucht. Die vorliegende Arbeit beschreibt die Entwicklung, Evaluation und erfolgreiche Implementierung eines mehrstufigen Beobachtertrainings für Medizinstudierende zur

Bewertung der Leitlinienadhärenz, sowie der Teamarbeit und -kommunikation.

## 2. Methoden

### 2.1. Auswertungsinstrumente

Für die Auswertung der Leitlinienadhärenz wurde die *Performance Evaluationscheckliste für pulslose ventrikuläre Tachykardien (PEC-PVT)* verwendet (siehe Anhang 1) [25]. Sie evaluiert das Vorgehen eines Teams bei einem simulierten pädiatrischen Notfall mit Herz-Kreislauf-Versagen aufgrund eines defibrillierbaren Herzrhythmus. Bestehend aus 31 Items, die in drei Bewertungskategorien geteilt werden (Aufgabe nicht erfüllt, Aufgabe unvollständig/falsch/verspätet erfüllt, Aufgabe erfüllt), ist sie gut dafür geeignet, komplexe Aufgaben im Rahmen einer Reanimationssituation abzubilden. Zudem wird durch eine Wichtung der einzelnen Items zwischen essenziellen und weniger wichtigen Teilaspekten dieser Notfallsituation unterschieden. Die englische Version wurde an deutschen pädiatrischen Notfallteams validiert [26]. Vor der Anwendung in der vorliegenden Studie erfolgte eine Übersetzung ins Deutsche durch eine der Autorinnen der Originalpublikation und eine nachfolgende Konsensusbildung zwischen den Koautoren der Originalpublikation, sowie der Studienleiterin (NM). Eine Spezifizierung einzelner Items, Homogenisierung und Gliederung anhand der Phasen des Studienszenarios durch die Studienleiterin schloss sich an. Um eine möglichst akkurate Bewertung aller Items zu ermöglichen, wurde zudem ein Anwenderhandbuch durch die Studienleiterin entwickelt (siehe Anhang 2). Darin wurden die Kriterien für die unterschiedliche Bewertung von Maßnahmen während der Simulation spezifiziert.

Die Evaluation der Teamarbeit und -kommunikation erfolgte mittels *Teamwork Emergency Assessment Measure (TEAM)* [27]. Bestehend aus 11 Items, die Aspekte der Teamleitung, der Teamarbeit und der Kommunikation auf einer 5-Punkt Likert-Skala („nie/fast nie“ bis „Immer/fast immer“) bewerten, sowie einer Globalbewertung der Teamarbeit (Skala von 1 bis 10), ist diese Checkliste einfach zu handhaben. Validität und Reliabilität des TEAM sind hoch [28], [29]. Da zum Zeitpunkt des Studienbeginns noch keine deutsche Version publiziert war, erfolgte eine Übersetzung der Checkliste ins Deutsche mittels Forward-Backward-Translation [30]. Auch hier wurde basierend auf den von Cooper veröffentlichten „TEAM Behavioural Markers“ [31] und dem Crew Resource Management Konzept von Rall [32] ein Handbuch konzipiert, um den Anwender\*innen spezifische und beobachtbare „Verhaltensanker“ [18] zur Verfügung zu stellen, die die extremen Enden der vorliegenden TEAM-Beobachtungsskala darstellen.

### 2.2. Teilnehmer\*innen

Es wurden 5 Medizinstudierende im mittleren bis fortgeschrittenen Ausbildungsstand (7. bis 13. Fachsemester) im Beobachtertraining geschult. Zwei der Studierenden wiesen als studentische Tutorinnen Basiskenntnisse in Simulationen auf, eine Studentin hatte bereits Erfahrung in Notfallmedizin als ausgebildete Rettungssanitäterin. Alle Studierenden waren Anfängerinnen in der strukturierten Beobachtung und Analyse simulierter Notfallsituationen.

Jede Studierende wurde in der Anwendung einer Checkliste geschult. Die Schulung in der Anwendung der spezifischen Performance Evaluationscheckliste (PEC-PVT) fand für zwei der Medizinstudierenden statt, während drei in der Anwendung des Team Emergency Assessment Measure (TEAM) trainiert wurden.

### 2.3. Beobachtertraining

Das mehrstufige Schulungskonzept basierte auf publizierten Rahmenstrukturen für Beobachtertraining [18], [33], [34] und beinhaltete unterschiedliche Formate wie Selbststudium und interkollegiales Training, sowie drei für die Bewertung von Beobachtungen essenzielle Trainingsstrategien: im *Rater Error Training* werden typische Beobachtungsfehler und deren Vermeidung thematisiert, das *Performance Dimension Training* lehrt das Erkennen gewünschter Verhalten für das zu untersuchende Item und im *Frame of Reference Training* werden Variationen in der Qualität des gewünschten Verhaltens demonstriert [33], [34].

Die Schulungen fanden über einen Zeitraum von insgesamt sechs Wochen statt und wurden durch die Studienleiterin jeweils getrennt für die beiden Auswertungsinstrumente durchgeführt.

#### Stufe 1: Vermittlung von Fachkenntnissen

Den Medizinstudierenden wurden in externen Kursen für das Auswertungsthema relevante Basiskenntnisse vermittelt. Dazu nahmen zwei Studierende an zweitägigen *Pediatric Advanced Life Support*-Kursen des European Resuscitation Councils teil und drei Studierende an eintägigen *Crew Resource Management*-Kursen. Hier wurde das Erkennen des kritisch kranken Kindes sowie lebenserhaltende Maßnahmen bei Atem- und Herz-Kreislauf-Stillstand, bzw. Aspekte der Teamführung, Teamkooperation und -kommunikation vermittelt.

#### Stufe 2: Selbststudium

Um den Auswertenden ein allgemeines Verständnis für das Bewerten, sowie typische Bewertungsfehler zu vermitteln, wurden theoretische Aspekte o. g. essentieller Trainingsstrategien [33], [34] im Selbststudium erarbeitet. Die Studierenden machten sich darüber hinaus mit den Bewertungsinstrumenten vertraut.



### Stufe 3: Erste Videoauswertung

Nach einer strukturierten Zusammenfassung relevanter Bewertungsfehler mit anschließender Gruppendiskussion erfolgte ein individuelles Bewerten zweier Videobeispiele durch die Medizinstudierenden und die Studienleiterin. Hierzu fanden separate Schulungen für zwei Medizinstudierende in der Anwendung der spezifischen PEC-PVT, sowie für drei Medizinstudierende in der Anwendung des TEAM statt. Es wurden Beispiele für eine schlechte bzw. eine akzeptable Teamleistung gewählt. Die Erfassung der Teamleistung erfolgte auf den jeweiligen Checklisten. Anschließend wurden die Einzelitems mit speziellem Augenmerk auf die Items mit geringster Übereinstimmung zwischen den Medizinstudierenden und der Studienleiterin diskutiert und Gründe für ein substantielles Abweichen benannt.

### Stufe 4: Ausgabe der Beobachterhandbücher, zweite Videoauswertung

Für die zweite Videosession wurden erneut o. g. zwei Videos mit schlechter bzw. akzeptabler Teamleistung, sowie drei zusätzliche Videos anhand der Checklisten unter Nutzung der Beobachterhandbücher durch die Medizinstudierenden und die Studienleiterin unabhängig voneinander bewertet. Die anschließende Diskussion fokussierte auf die Einzelitems mit geringster Übereinstimmung. Auch hier fanden die Schulungen separat für die PEC-PVT und den TEAM statt.

### Stufe 5: Pilottestung

Über einen Zeitraum von ca. 4 Wochen wurden von den Medizinstudierenden und der Studienleiterin weitere neun Videos anhand der Checklisten und unter Nutzung der Handbücher allein ausgewertet.

## 2.4. Statistik

Die Übereinstimmung der Ergebnisse von TEAM bzw. PEC-PVT zwischen den einzelnen Medizinstudierenden einerseits und der Studienleiterin andererseits wurde mittels Berechnung des Kendall-tau-b-Korrelationskoeffizienten paarweise in den unterschiedlichen Phasen des Anwendertrainings überprüft. Der Kendall-tau-Koeffizient misst die Übereinstimmung der Beobachter und quantifiziert damit die Reliabilität des Beobachtungssystems. Dieser Koeffizient ist ein Rangkorrelationskoeffizient und nimmt Werte zwischen -1 und 1 an. Dieses Maß eignet sich für ordinale Daten, also zum Beispiel für Likert-Skalen. Dabei wird nur die Rangordnung der jeweiligen Werte berücksichtigt. Die Bestimmung von tau erfolgt mittels aller Einzelitems der Fragebögen [35]. Eine schwache bzw. geringe Übereinstimmung wird bei einem Kendall-tau-Koeffizienten von  $\leq 0,4$ , eine moderate bei  $0,41-0,7$  und eine hohe Übereinstimmung bei Werten  $>0,7$  angenommen [36].

Den berechneten Kendall-tau-Korrelationskoeffizienten liegen in Stufe 3 zwei Videos, in Stufe 4 fünf Videos und in Stufe 5 neun Videos zu Grunde. Die Festlegung der Anzahl der Videos ist in erster Linie dem großen Aufwand für jedes einzelne Video geschuldet. Es erfolgte die Bestimmung der paarweisen Übereinstimmungen, um jeweils die Einschätzung einer Studierenden im Vergleich zur Studienleiterin zu evaluieren.

Um mögliche Unterschiede zwischen den Bewertungen der Studierenden und der Studienleiterin auf Itemebene darzustellen, erfolgten nach Beendigung des Beobachtertrainings für den TEAM Mann-Whitney-U Tests (für die elf Einzelitems) und t-Tests (für die Globalbewertung). Analog dazu wurden Mann-Whitney-U-Tests für die Auswertung mittels PEC-PVT durchgeführt.

Die Datenanalyse erfolgte mit SPSS V24 (Armonk, NY: IBM Corp.).

## 2.5. Ethikvotum

Die Entwicklung des Beobachtertrainings erfolgte im Rahmen der Studie „Qualitätssicherung pädiatrischer Notfallversorgung durch Inhouse Simulationstraining an hessischen Kinderkliniken.“ Es liegt ein Ethikvotum der Ethikkommission der Philipps Universität Marburg vor (AZ: 172/16).

## 3. Ergebnisse

Alle fünf Medizinstudierenden absolvierten das Beobachtertraining vollständig. Es wurden 16 Videos (zwei Videos aus Stufe 3, fünf Videos aus Stufe 4 und neun Videos aus Stufe 5 des Beobachtertrainings) von jeweils 5 Beobachtungspaaren (jeweils eine Studentin und die Studienleiterin) ausgewertet. Bei 5 Beobachtungspaaren und 3 Zeitpunkten konnten 15 Kendall-tau-Koeffizienten berechnet werden (siehe Tabelle 1).

Fünf der 15 tau zeigten eine schwache Übereinstimmung zwischen Medizinstudierenden und der Studienleitung, fünf eine moderate und fünf eine hohe Übereinstimmung (siehe Tabelle 1). Bei drei Studierenden verbesserte sich die Übereinstimmung zwischen erster und zweiter Videoauswertung (Beobachtungspaar 1, 3 und 5).

Die durchgeführten U- und t-Tests zeigten keine signifikanten Unterschiede zwischen Studentinnen und der Studienleiterin auf Itemebene, d. h. es konnten keine einzelnen Items identifiziert werden, in denen die Auswertung von Anfängerinnen und Expertin differierten.

## 4. Diskussion

In dieser Arbeit wird ein mehrstufiges Beobachtertraining vorgestellt, das Medizinstudierende ermöglicht, simulierte pädiatrische Notfallsituationen strukturiert zu evaluieren. Es wurden dazu unterschiedliche Schulungsformate und gut evaluierte Trainingsstrategien kombiniert [33], [34]. Ziel war es, für konsekutive Studien, die auf Video-

Tabelle 1: tau\* für die Übereinstimmung zwischen Studierenden und der Studienleiterin

Auswertungszeitpunkt		Erste Videoauswertung (2 Videos)	Zweite Videoauswertung (5 Videos)	Pilottestung (9 Videos)
Beobachtungs-paar	Instrument			
1	TEAM <sup>#</sup>	0,17	0,24	0,36
2	TEAM	0,70	0,57	0,47
3	TEAM	0,14	0,46	0,31
4	PEC-PVT <sup>§</sup>	0,82	0,81	0,79
5	PEC-PVT	0,63	0,90	0,80

\*Kendall-tau  $\tau$ . Beurteilung der Übereinstimmung:  $\tau \leq 0,4$  schwach,  $0,41 - 0,7$  moderat,  $> 0,7$  hoch.

<sup>#</sup>TEAM: Team Emergency Assessment Measure. Untersuchung der Einzelitems 1-11.

<sup>§</sup>PEC-PVT: Performance Evaluationscheckliste für pulslose ventrikuläre Tachykardie.

auswertungen von Teamleistungen sowohl durch Expertinnen und Experten als auch durch trainierte Anfänger\*innen beruhen, das Beobachtertraining zu evaluieren und die Übereinstimmung der Bewertung zu überprüfen.

Medizinstudierende sind in der Regel Anfänger\*innen in der Bewertung von komplexen simulierten Notfallsituationen, für die sie deshalb bisher kaum herangezogen wurden. Freytag und Kollegen integrierten Studierende in die Bewertung von Teamzusammenarbeit und Kommunikation studentischer Notfallsimulationen nach vorherigem Beobachtertraining [20]. Es zeigte sich eine nachsichtiger Bewertung im Vergleich zu den erfahrenen Mitbeobachtern, was zu einer moderaten Interrater-Reliabilität führte [20], [37]. Initiale Interrater-Reliabilitäten vor der Teilnahme am Beobachtertraining wurden nicht bestimmt. Evans et al. untersuchten die Anwendung einer Checkliste zur Beurteilung von invasiven Prozeduren durch Studierende nach vierstündigem Beobachtertraining und fanden eine gute Übereinstimmung zwischen den Beobachtungen von Anfänger\*innen und Expertinnen bzw. Experten [22]. Sie schlossen daraus, dass die Vermittlung solcher Beobachtungsfähigkeiten auch ohne vorheriges Hintergrundwissen der Bewertenden möglich ist [22].

In unserer Arbeit konnten wir zeigen, dass Medizinstudierende als Anfänger\*innen in der strukturierten Verhaltensbeobachtung im Rahmen eines Beobachtertrainings gute Interrater-Reliabilitäten erreichen können. Dies zeigte sich sowohl in der Evaluation komplexer Verhaltensweisen wie Teamarbeit und -kommunikation als auch in der Bewertung technischer Fertigkeiten wie der Leitlinienadhärenz. Bereits nach der Vermittlung von Basiskonzepten (Stufen 1 bis 3 des Trainings) wurde eine gute Übereinstimmung in der Bewertung der Leitlinienadhärenz zwischen Anfängerinnen und Expertin nachgewiesen. Wir führen dies auf die strukturierte und ausführliche Vermittlung von Kenntnissen des *Pediatric Life Supports* noch vor der eigentlichen Videoauswertung zurück. Die Nutzung von Beobachterhandbüchern verbesserte im Folgenden v. a. die Interrater-Reliabilität der die Teamarbeit und Teamkommunikation Bewertenden (Stufe 4). Dieses *Frame of Reference Training* hat sich in früheren Studien als besonders effektiv in Bewerbertrainings gezeigt [34]. Zuvor strukturiert vermittelte Grundkenntnisse des Crew

Ressource Managements hatten nicht zu einer adäquaten Interrater-Reliabilität geführt. Ein erweitertes Beobachtertraining über die Auswertung von wenigen Übungsvideos hinaus (Stufe 5) scheint dagegen die Interrater-Reliabilität nicht weiter zu verbessern.

Medizinstudierende können demnach Expertinnen und Experten in der Beobachtung, Analyse und Bewertung von simulierten Notfallsituationen unterstützen. Dies ist im Rahmen eingeschränkter Personalressourcen relevant, da simulationsbasierte Schulungsformate in der medizinischen Aus- und Weiterbildung zunehmend genutzt werden. Neben dem Training technischer und nicht-technischer Fertigkeiten [38] erfolgen Simulationstrainings mit dem Ziel, Patientensicherheit relevante Aspekte zu thematisieren [39], [40] sowie wiederkehrende und häufig vermeidbare Behandlungsfehler im Notfallmanagement zu identifizieren [41], [42], [43]. Hierzu ist auch eine adäquate Ausbildung nicht-studentischer Dozent\*innen, d. h. der lehrenden Ärztinnen und Ärzte erforderlich. Neben den medizinischen Fachkompetenzen sollte eine strukturierte Vermittlung von Grundkenntnissen in der Bewertung, sowie in typischen Bewertungsfehlern in die Ausbildung von Dozent\*innen integriert werden.

Diese Studie hat mehrere Limitationen.

Der *Team Emergency Assessment Measure* wurde ursprünglich für Expertinnen und Experten, nicht für die Anwendung durch Anfänger\*innen in der Verhaltensbeobachtung entwickelt und zeigte hier eine gute Validität und Reliabilität [27], [44]. Da vor Beginn des eigentlichen Beobachtertrainings keine paarweise Videoauswertung erfolgte, ist die Konstruktvalidität dieser Checkliste bei Anfänger\*innen nicht bekannt. Es zeigten sich jedoch bei 2 von 3 die TEAM-Checkliste auswertenden Beobachtungspaaren Verbesserungen in der Interrater-Reliabilität, so dass wir von einem Effekt des Beobachtertrainings ausgehen.

Die Evaluation der spezifischen Performance Evaluationscheckliste (PEC-PVT) ergab in dieser Arbeit im Vergleich zur Originalpublikation bessere Interrater-Reliabilitäten [26]. Bei der Übersetzung der Originalcheckliste ins Deutsche zeigten sich einzelne Items als nicht eindeutig formuliert. Diese wurden in der hier verwendeten deut-

schen Version präzisiert, was allein bereits zu einer verbesserten Interrater-Reliabilität geführt haben könnte. Das hier vorgestellte Beobachtertraining ist sehr personal- und zeitintensiv und damit nicht uneingeschränkt auf andere Studien übertragbar. Da wir keine Verbesserung der Beobachterübereinstimmung in Stufe 5 des Beobachtertrainings nachweisen konnten, sehen wir keinen Vorteil von über wenige Übungsvideos hinausgehende Schulungen. Die Form des optimalen Schulungsmodells bleibt damit offen.

## 5. Schlussfolgerung

Zusammenfassend konnten wir zeigen, dass Medizinstudierende nach Durchlaufen eines Beobachtertrainings anhand von Videos technische Fertigkeiten, sowie Teamarbeit und -kommunikation von interprofessionellen Notfallteams in komplexen simulierten Notfallsituationen evaluieren können. Das Beobachtertraining sollte strukturiert erstellt sein und etablierte Schulungskonzepte enthalten. Ein erweitertes Training über die Verwendung von wenigen Übungsvideos hinaus, scheint nicht notwendig.

## Förderung

Das Hessische Ministerium für Soziales und Integration finanzierte die Durchführung von pädiatrischen Notfalltrainings an hessischen Kinderkliniken, im Rahmen welcher Studienvideos erfolgten.

Open Access Förderung durch den Open Access Publikationsfonds der Philipps-Universität Marburg mit Unterstützung der Deutschen Forschungsgemeinschaft (DFG).

## Interessenkonflikt

Die Autor\*innen erklären, dass sie keinen Interessenkonflikt im Zusammenhang mit diesem Artikel haben.

## Anhänge

Verfügbar unter <https://doi.org/10.3205/zma001629>

1. Anhang\_1.pdf (302 KB)  
PEC-PVT Anwenderhandbuch
2. Anhang\_2.pdf (84 KB)  
PEC-PVT

## Literatur

1. Jones H, Wilmshurst SL, Graydon C. Aetiology and outcome of paediatric cardiopulmonary arrest. *Anaesthesia Intensive Care Med.* 2017;18(11):537-540. DOI: 10.1016/j.mpaic.2017.07.003
2. Kendirli T, Erkek N, Köroğlu T, Yıldızdaş D, Bayrakç B, Güzel A, Çtak A, Demirkol D, Ağın H, Arslanköylü AE, Kutlu NO, Tuysun N, Pakso MS, Anl AB, Kalkan G, Duman M, Dündaröz R, Asloğlu N, Yaman A, Ödek C, Tekin D, Dursan O, Sevketoglu E, Kesici S, Ates C, Gördü Z, Yılmaz HL, Ince E, Karaböcüoğlu M. Cardiopulmonary Resuscitation in Children With In-Hospital and Out-of-Hospital Cardiopulmonary Arrest: Multicenter Study From Turkey. *Pediatr Emerg Care.* 2015;31(11):748-752. DOI: 10.1097/PEC.0000000000000337
3. Holmberg MJ, Ross CE, Fitzmaurice GM, Chan PS, Duval-Arnould J, Grossestreuer AV, Yankama T, Donnino MW, Andersen LW; American Heart Association's Get With The Guidelines-Resuscitation Investigators. Annual Incidence of Adult and Pediatric In-Hospital Cardiac Arrest in the United States. *Circ Cardiovasc Qual Outcomes.* 2019;12(7):e005580.
4. Gräsner JT, Lefering R, Koster RW, Masterson S, Böttiger BW, Herlitz J, Wnent J, Tjelmeland IB, Ortiz FR, Maurer H, Baubin M, Mols P, Hadzibegovic I, Ioannides M, Skeluc R, Wissenberg M, Salo A, Hubert H, Nikolaou NI, Loczi G, Svavarsdottir H, Semeraro F, Wright PJ, Clarens C, Pijls R, Cebula G, Correia VG, Cimpoesu D, Raffay V, Trenkler S, Markota A, Srömsöe A, Burkart R, Perkins GD, Bossaert LL; EuReCa ONE Collaborators. EuReCa ONE-27 Nations, ONE Europe, ONE Registry: A prospective one month analysis of out-of-hospital cardiac arrest outcomes in 27 countries in Europe. *Resuscitation.* 2016;105:188-195. DOI: 10.1016/j.resuscitation.2016.06.004
5. Lehmann R, Seitz A, Meyburg J, Hoppe B, Hoffmann GF, Tönshoff B, Huwendiek S. Pediatric in-hospital emergencies: real life experiences, previous training and the need for training among physicians and nurses. *BMC Res Notes.* 2019;12(1):19. DOI: 10.1186/s13104-019-4051-4
6. Walsh Ó, Lydon S, O'Connor P. A mixed methods evaluation of paediatric trainee preparedness to manage cardiopulmonary arrests. *Eur J Pediatr.* 2017;176(12):1653-1662. DOI: 10.1007/s00431-017-3017-6
7. Nadel FM, Lavelle JM, Fein JA, Giardino AP, Decker JM, Durbin DR. Teaching Resuscitation to Pediatric Residents: The Effects of an Intervention. *Arch Pediatr Adolesc Med.* 2000;154(10):1049-1054. DOI: 10.1001/archpedi.154.10.1049
8. Hunt EA, Vera K, Diener-West M, Haggerty JA, Nelson KL, Shaffner DH, Pronovost PJ. Delays and errors in cardiopulmonary resuscitation and defibrillation by pediatric residents during simulated cardiopulmonary arrests. *Resuscitation.* 2009;80(7):819-825. DOI: 10.1016/j.resuscitation.2009.03.020
9. Sutton RM, Niles D, French B, Maltese MR, Leffelman J, Eilevstjonn J, Wolfe H, Nishisaki A, Meaney PA, Berg RA, Nadkarni VM. First quantitative analysis of cardiopulmonary resuscitation quality during in-hospital cardiac arrests of young children. *Resuscitation.* 2014;85(1):70-74. DOI: 10.1016/j.resuscitation.2013.08.014
10. Sutton RM, French B, Niles DE, Donoghue A, Topjian AA, Nishisaki A, Leffelman J, Wolfe H, Berg RA, Nadkarni VM, Meaney PA. 2010 American Heart Association recommended compression depths during pediatric in-hospital resuscitations are associated with survival. *Resuscitation.* 2014;85(9):1179-1184. DOI: 10.1016/j.resuscitation.2014.05.007
11. 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. DOI: 10.1097/00006565-200004000-00001
12. Sherman JM, Chang TP, Ziv N, Nager AL. Barriers to Effective Teamwork Relating to Pediatric Resuscitations: Perceptions of Pediatric Emergency Medicine Staff. *Pediatr Emerg Care.* 2020;36(3):e146-e150. DOI: 10.1097/PEC.0000000000001275

13. Risser DT, Rice MM, Salisbury ML, Simon R, Jay GD, Berns SD. The potential for improved teamwork to reduce medical errors in the emergency department. The MedTeams Research Consortium. *Ann Emerg Med.* 1999;34(3):373-383. DOI: 10.1016/s0196-0644(99)70134-4
14. Cheng A, Lang TR, Starr SR, Pusic M, Cook DA. Technology-Enhanced Simulation and Pediatric Education: A Meta-analysis. *Pediatrics.* 2014;133(5):e1313-e1323. DOI: 10.1542/peds.2013-2139
15. Miledler LP, Urlesberger B, Szyld EG, Roehr CC, Schmölzer GM. Simulation-based neonatal and infant resuscitation teaching: a systematic review of randomized controlled trials. *Klin Padiatr.* 2014;226(5):259-267. DOI: 10.1055/s-0034-1372621
16. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach.* 2005;27(1):10-28. DOI: 10.1080/01421590500046924
17. Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simul Healthc.* 2007;2(2):115-125. DOI: 10.1097/SIH.0b013e3180315539
18. Eppich W, Nannicelli AP, Seivert NP, Sohn MW, Rozenfeld R, Woods DM, Holl JL. A Rater Training Protocol to Assess Team Performance. *J Contin Educ Health Prof.* 2015;35(2):83-90. DOI: 10.1002/chp.21270
19. Cheng A, Hunt EA, Donoghue A, Nelson-McMillan K, Nishisaki A, LeFlore J, Eppich W, Moyer M, Brett-Fleegler M, Kleinman M, Anderson J, Adler M, Braga M, Kost S, Stryjewski G, Min S, Podraza J, Lopreiato J, Fidor Hamilton M, Stone K, Reid J, Hopkins J, Manos J, Duff J, Richard M, Nadkarni VM; EXPRESS Investigators. Examining Pediatric Resuscitation Education Using Simulation and Scripted Debriefing: A Multicenter Randomized Trial. *JAMA Pediatrics.* 2013;167(6):528-536. DOI: 10.1001/jamapediatrics.2013.1389
20. Freytag J, Stroben F, Hautz WE, Schaubert SK, Kämmer JE. Rating the quality of teamwork—a comparison of novice and expert ratings using the Team Emergency Assessment Measure (TEAM) in simulated emergencies. *Scand J Trauma Resusc Emerg Med.* 2019;27(1):12. DOI: 10.1186/s13049-019-0591-9
21. Yule S, Rowley D, Flin R, Maran N, Youngson G, Duncan J, Paterson-Brown S. Experience matters: comparing novice and expert ratings of non-technical skills using the NOTSS system. *ANZ J Surg.* 2009;79(3):154-160. DOI: 10.1111/j.1445-2197.2008.04833.x
22. Evans LV, Morse JL, Hamann CJ, Osborne M, Lin Z, D'Onofrio G. The development of an independent rater system to assess residents' competence in invasive procedures. *Acad Med.* 2009;84(8):1135-1143. DOI: 10.1097/ACM.0b013e3181acec7c
23. Möltner A, Lehmann M, Wachter C, Kurczyk S, Schwill S, Loukanova S. Formative assessment of practical skills with peer-assessors: quality features of an OSCE in general medicine at the Heidelberg Medical Faculty. *GMS J Med Educ.* 2020;37(4):Doc42. DOI: 10.3205/zma001335
24. Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. *J Gen Intern Med.* 2009;24(1):74-79. DOI: 10.1007/s11606-008-0842-3
25. Sevdalis N, Lyons M, Healey AN, Undre S, Darzi A, Vincent CA. Observational teamwork assessment for surgery: construct validation with expert versus novice raters. *Ann Surg.* 2009;249(6):1047-1051. DOI: 10.1097/SLA.0b013e3181a50220
26. Schmutz J, Manser T, Keil J, Heimberg E, Hoffmann F. Structured Performance Assessment in Three Pediatric Emergency Scenarios: A Validation Study. *J Pediatr.* 2015;166(6):1498-504. DOI: 10.1016/j.jpeds.2015.03.015
27. Cooper S, Cant R, Porter J, Sellick K, Somers G, Kingsman L, Nestel D. Rating medical emergency teamwork performance: Development of the Team Emergency Assessment Measure (TEAM). *Resuscitation.* 2010;81(4):446-452. DOI: 10.1016/j.resuscitation.2009.11.027
28. Cooper SJ, Cant R. Measuring non-technical skills of medical emergency teams: an update on the validity and reliability of the Team Emergency Assessment Measure (TEAM). *Resuscitation.* 2014;85(1):31-33. DOI: 10.1016/j.resuscitation.2013.08.276
29. Valentine MA, Nembhard IM, Edmondson AC. Measuring teamwork in health care settings: a review of survey instruments. *Med Care.* 2015;53(4):e16-30. DOI: 10.1097/MLR.0b013e31827feef6
30. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P; ISPOR Task Force for Translation and Cultural Adaptation. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health.* 2005;8(2):94-104. DOI: 10.1111/j.1524-4733.2005.04054.x
31. Monash University. TEAM Team Emergency Assessment Measure. Victoria: Monash University; 2012. Zugänglich unter/available from: <https://www.monash.edu/search?query=TEAM+Team+Emergency+Assessment+Measure.+2012>
32. Rall M. Human Factors und CRM: Eine Einführung. In: Breuer SP, editor. *Simulation in der Medizin.* Berlin, Heidelberg: Springer-Verlag; 2013. p.136-151. DOI: 10.1007/978-3-642-29436-5\_13
33. Feldman M, Lazzara EH, Vanderbilt AA, DiazGranados D. Rater training to support high-stakes simulation-based assessments. *J Contin Educ Health Prof.* 2012;32(4):279-286. DOI: 10.1002/chp.21156
34. Woehr DJ, Hffcutt AI. Rater training for performance appraisal: a quantitative review. *J Occup Organ Psychol.* 1994;68(3):189-205. DOI: 10.1111/j.2044-8325.1994.tb00562.x
35. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155-159. DOI: 10.1037//0033-2909.112.1.155
36. Overholser BR, Sowinski KM. Biostatistics primer: part 2. *Nutr Clin Pract.* 2008;23(1):76-84. DOI: 10.1177/011542650802300176
37. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med.* 2016;15(2):155-163. DOI: 10.1016/j.jcm.2016.02.012
38. Clerihew L, Rowney D, Ker J. Simulation in paediatric training. *Arch Dis Child Educ Pract Ed.* 2016;101(1):8-14. DOI: 10.1136/archdischild-2015-309143
39. Hoffmann N, Kubitz JC, Goetz AE, Beckers SK. Patient safety in undergraduate medical education: Implementation of the topic in the anaesthesiology core curriculum at the University Medical Center Hamburg-Eppendorf. *GMS J Med Educ.* 2019;36(2):Doc12. DOI: 10.3205/zma001220
40. Schildmann J, Salloch S, Peters T, Henking T, Vollmann J. Risks and errors in medicine. Concept and evaluation of an optional study module with integrated teaching of ethical, legal and communicative competencies. *GMS J Med Educ.* 2018;35(3):Doc31. DOI: 10.3205/zma001177
41. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care.* 2008;17(3):216-223. DOI: 10.1136/qshc.2007.023622
42. Ventre KM, Barry JS, Davis D, Baiamonte VL, Wentworth AC, Pietras M, Coughlin L, Barley G. Using in situ simulation to evaluate operational readiness of a children's hospital-based obstetrics unit. *Simul Healthc.* 2014;9(2):102-111. DOI: 10.1097/SIH.000000000000005



43. Slakey DP, Simms ER, Rennie KV, Garstka ME, Korndorffer JR, Jr. Using simulation to improve root cause analysis of adverse surgical outcomes. *Int J Qual Health Care*. 2014;26(2):144-150. DOI: 10.1093/intqhc/mzu011
44. Cooper S, Cant R, Connell C, Sims L, Porter JE, Symmons M, Nestel D, Liaw SK. Measuring teamwork performance: Validity testing of the Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. *Resuscitation*. 2016;101:97-101. DOI: 10.1016/j.resuscitation.2016.01.026

**Korrespondenzadresse:**

Dr. med. Nadine Mand  
Philipps Universität Marburg, Universitätsklinikum  
Marburg, Zentrum für Kinder- und Jugendmedizin,  
Baldingerstraße, 35043 Marburg, Deutschland  
mandn@staff.uni-marburg.de

**Bitte zitieren als**

Mand N, Stibane T, Sitter H, Maier RF, Leonhardt A. Successful implementation of a rater training program for medical students to evaluate simulated pediatric emergencies. *GMS J Med Educ*. 2023;40(4):Doc47.  
DOI: 10.3205/zma001629, URN: urn:nbn:de:0183-zma0016291

**Artikel online frei zugänglich unter**

<https://doi.org/10.3205/zma001629>

**Eingereicht:** 02.07.2021

**Überarbeitet:** 04.03.2023

**Angenommen:** 20.04.2023

**Veröffentlicht:** 15.06.2023

**Copyright**

©2023 Mand et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.