

ORIGINAL ARTICLE

## ASSESSING RESIDENT DIAGNOSTIC SKILLS USING A MODIFIED BRONCHIOLITIS SCORE

Andrea Rivera-Sepulveda<sup>1,2</sup>, Muguette Isona<sup>3</sup>.

<sup>1</sup>Pediatrics, Emergency Medicine, Nemours Children's Hospital, Orlando, FL, United States,

<sup>2</sup>University of Puerto Rico Medical Sciences Campus, School of Health Professions and School of Medicine, San Juan, Puerto Rico,

<sup>3</sup>Emergency Department, San Juan City Hospital, San Juan, Puerto Rico.

### ABSTRACT

**Background:** Resident milestones are objective instruments that assess the resident's growth, progression in knowledge, and clinical diagnostic reasoning; but they rely on the subjective appraisal of the supervising attending. Little is known about the use of standardized instruments that may complement the evaluation of resident diagnostic skills in the academic setting.

**Objectives:** Evaluate a modified bronchiolitis severity assessment tool by appraising the inter-rater variability and reliability between pediatric attendings and pediatric residents.

**Methods:** Cross-sectional study of children under 24 months of age who presented to a Community Hospital's Emergency Department with bronchiolitis between January-June 2014. A paired pediatric attending and resident evaluated each patient. Evaluation included age-based respiratory rate (RR), retractions, peripheral saturation, and auscultation. Cohen's kappa (K) measured inter-rater agreement. Inter-rater reliability (IRR) was assessed using a one-way random, average measures intra-class correlation (ICC) to evaluate the degree of consistency and magnitude of disagreement between inter-raters. Value of >0.6 was considered substantial for kappa and good internal consistency for ICC.

**Results:** Twenty patients were evaluated. Analysis showed fair agreement for the presence of retractions (K=0.31), auscultation (K=0.33), and total score (K=0.3). The RR (ICC=0.97), SpO<sub>2</sub> (ICC=1.0), auscultation (ICC=0.77), and total score (ICC=0.84) were scored similarly across both raters, indicating excellent IRR. Identification of retractions had the least agreement across all statistical analysis.

**Conclusion:** The use of a standardized instrument, in conjunction with a trained resident-teaching staff, can help identify deficiencies in clinical competencies among residents and facilitate the learning process for the identification of pertinent clinical findings.

### Introduction

Appropriate use of patient history and physical examination are essential to clinical practice. Literature suggests that current medical students and medical residents are deficient in some aspects of physical diagnosis; auscultation being the most prominent.<sup>1</sup> For residents-in-training, the challenge lies in learning to perform physical examinations particular to their specialty; while the challenge with the resident-teaching staff rests in identifying both subtle and

obvious mistakes made by residents in history and physical examinations.<sup>2</sup> In addition to history-taking and physical examination skills, one needs to consider medical literature evidence of accuracy in exam maneuvers, including sensitivity, specificity, and clinical diagnostic reasoning.

Pediatric residents' growth and progression in knowledge and clinical diagnostic reasoning is usually measured by the combination of resident milestones from the Accreditation Council for Graduate Medical Education (ACGME), and the American Academy of Pediatrics (AAP). These general and pediatric milestones are useful in focusing teaching and the assessment of clinical diagnostic reasoning in residency training by clarifying performance expectations for both residents and teaching faculty.<sup>3</sup> Although resident milestones are a set of objective assessment instruments, they rely on the subjective appraisal of the attending. Therefore,

### ARTICLE HISTORY

Received 30 October 2020

Accepted 3 January 2021

### KEYWORDS

resident, pediatric; diagnostic skills; bronchiolitis, inter-rater, score, board-certified, pediatrician, pediatric emergency department.

### ABBREVIATIONS

IRR: Inter-rater reliability  
ICC: Intra-class correlation  
ACGME: Accreditation Council for Graduate Medical Education  
AAP: American Academy of Pediatrics  
PED: Pediatric Emergency Department  
RR: Respiratory rate  
SpO<sub>2</sub>: Oxygen saturation  
K: Cohen's kappa

**Address for Correspondence:** Andrea Rivera-Sepulveda MD, MSc, FAAP; Nemours Children's Hospital, Department of Pediatrics/ Division of Pediatric Emergency Medicine, 6535 Nemours Parkway, Orlando, Florida 32827, USA.

**Email:** [rivera.andreav@gmail.com](mailto:rivera.andreav@gmail.com)

©2021 Pediatric Oncall

the use of a standardized evaluation instrument may complement the evaluation of resident diagnostic skills in the academic setting. The objective of this study is to evaluate a modified bronchiolitis severity assessment tool by appraising the inter-rater variability and reliability between pediatric attendings and pediatric residents.

### Methods & Materials

**Study design and setting:** This is a cross-sectional study performed in a community teaching hospital. The mainly urban population under study consisted of a convenience sample of children under 24 months of age who presented to the Pediatric Emergency Department from January 01, 2014 to June 30, 2014. We included children with a primary or secondary diagnosis of clinical bronchiolitis. The principal investigators identified patients with potential for recruitment. We recruited children with bronchiolitis whose parents consented to participate and provided a signed parental informed consent. Bronchiolitis was defined as clinical evidence of lower respiratory tract involvement such as wheezing, rhonchi, crackles or chest wall retractions with or without upper respiratory tract infection. We excluded children who required immediate therapeutic management or intubation per the physician's clinical criteria; children known to have another reason for respiratory distress, such as prematurity, chronic lung disease, bronchopulmonary dysplasia, bronchiectasis, gastroenteritis, liver function impairment and/or congenital heart disease; and patients with a diagnosis of pneumonia by chest radiography.

A pediatric resident and either a non-board-certified general pediatrician or board-certified general pediatric attending, known as raters, evaluated each recruited patient. Rater pairs were formed based on the availability of another physician. We excluded patients when there was only one physician available or the other physician was not of a different level of clinical background. Both physicians were instructed to perform a complete physical examination on the pediatric patient while awake and at rest. Each member of the rater pairs performed a patient evaluation simultaneously but independently from each other, and before initiation of therapeutic intervention. The physicians were instructed to record the findings of the respiratory examination in individual bronchiolitis score sheets that had the same identification number. We also recorded the patient's age in months and the level of clinical background of the rater (i.e., non-board-certified general pediatrician, board-certified general pediatrician, or pediatric resident). The raters were blinded to each other's assessment and the meaning of the total score. The bronchiolitis score was not used in treatment decisions of patients.

**Measurements:** The bronchiolitis score included standard respiratory parameters. The clinical evaluation tool from Goebel et al<sup>4</sup> was modified to include an age-based respiratory rate.<sup>5,6,7</sup> The modified bronchiolitis severity assessment tool included four sub-scores: 1) age-based respiratory rate (RR) (score of 1-3); 2) anatomic location of retractions (score 0-3); 3) peripheral oxygen saturation (score 0-3); and 4) quality of wheezes (score 0-3) (Table 1). Peripheral

capillary oxygen saturation (SpO<sub>2</sub>) was recorded as the first read after 30 seconds of stable signal during spot check, while the child was breathing room air. Total score ranged from 1 to 12 points, with higher scores indicating greater respiratory distress. The sum of the sub-scores determined mild (1-6 points), moderate (7-9 points) or severe bronchiolitis (10-12 points). Prior to the implementation of the modified bronchiolitis assessment tool, the instrument was validated between the non-board-certified general pediatrician and board-certified general pediatrician and found to be a reliable instrument with almost perfect agreement and excellent internal consistency.

**Outcome measures:** The primary outcome measure was to evaluate the reliability of the bronchiolitis severity assessment tool among pairs of pediatric attendings and pediatric residents.

**Statistical Analysis:** Data was analyzed using descriptive statistics, including dispersion measures such as standard deviation and frequency distribution. Categorical variables were analyzed using frequency and percentages; continuous variables were analyzed using means and standard deviation if normally distributed, or median (interquartile range) and percentages, if not normally distributed. The effect of training level was assessed by comparing agreement for overall severity of illness classification and the numeric value of the model score between faculty and residents. Pearson correlation was used to assess correlation between patient's age and the inter-rater level of agreement by overall score and diagnosis. Inter-rater level of agreement was measured with Cohen's kappa (K) for discrete variables and Spearman's Rho for ordinal variables. Inter-rater level of agreement is considered to be slight if kappa ranges below 0.2, fair if kappa ranges from 0.21 to 0.4, moderate if kappa ranges from 0.41 to 0.6, substantial if kappa ranges from 0.61 to 0.8, and almost perfect if kappa is greater than 0.81.<sup>8</sup> Because the raters were paired from a convenience sample of pediatric resident physicians within the Pediatric Emergency Department, we performed an inter-rater reliability (IRR) test using a one-way random, average measures intra-class correlation coefficient (ICC). This prevents the ICC from accounting for systemic deviations due to specific raters or two-way coder-subject interaction by the newly paired raters for each subject. The ICC was used to assess the degree of consistency and the magnitude of disagreement between inter-rater scores. Intra-class correlation is considered to have poor internal consistency if the value is below 0.4, fair internal consistency if the value is 0.4 to 0.59, good internal consistency if the value is 0.6 to 0.74, and excellent internal consistency if the value is over 0.75. A value of >0.6 was considered as substantial agreement for kappa, and as good internal consistency for ICC.<sup>8</sup> A sample size of twenty was estimated to detect a correlation coefficient of 0.5 at an alpha error of 5% and beta error of 20%. A p value <0.05 was considered as statistically significant. Statistical analyses were made with SPSS 21 for Mac OS X Windows (IBM Corp., Armonk, NY, USA). This study was approved by the San Juan City Hospital Institutional Review Board from San Juan, Puerto Rico.

**Results**

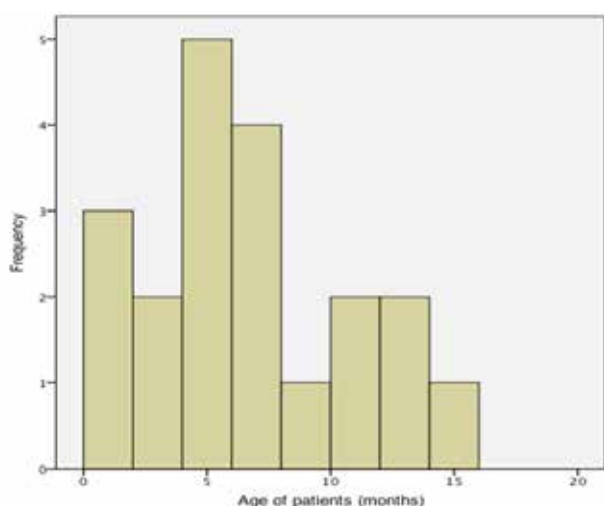
Twelve providers (10 pediatric residents, 1 non-board-certified general pediatrician and 1 board-certified general pediatrician) participated in the completed assessment of 20 patients for a total of 40 clinical assessments during the 6-month study period. Patients' ages ranged from 1 to 15 months, with a mean age of 6 + 4 months (Figure 1). The inter-rater's total scores had a wide distribution that ranged from 1 to 7, with a median score of 3 for both raters (Figure 2).

Sub-score analysis showed almost perfect agreement in RR (K=0.9) and SpO2 (K=1.0) (Table 2). Inter-raters had fair agreement for the presence of retractions (K=0.31), auscultation (K=0.33), and total severity score (K=0.3); but had high-moderate correlation between the ranked variables among inter-raters. Inter-item correlation was high-moderate for auscultation

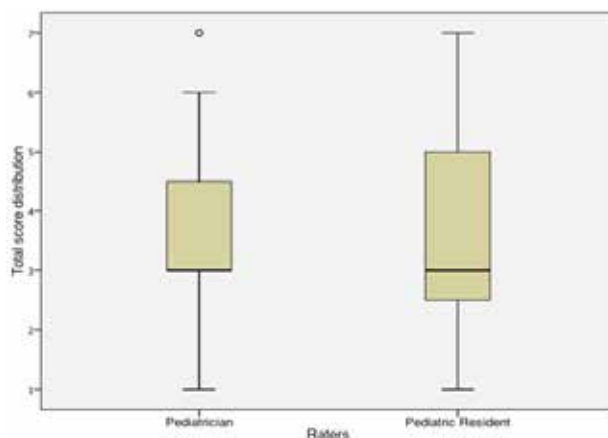
(R=0.61; p=0.004) and total severity score (R=0.72; p=0.001) and showed near perfect correlation with RR (R=0.95; p=0.001). The sub-score for RR (ICC=0.97), SpO2 (ICC=1.0), auscultation (ICC=0.77), and total severity score (ICC=0.84) were scored similarly across both raters, indicating excellent internal consistency and IRR. The variable that showed the least agreement across all statistical analyses was the presence of retractions.

Total score analysis in relation to the distribution of bronchiolitis severity was 90% for mild bronchiolitis and 10% for moderate bronchiolitis. A subgroup analysis controlling for age group (1-3, 4-6, 7-9, 10-12, 13-15 months) showed high-moderate correlation with inter-rater agreement based on total severity score (R=0.69; p=0.001); but low-moderate correlation based on diagnosis of severity (R=0.45; p=0.052).

**Figure 1.** Patient distribution based on age.



**Figure 2.** Total modified bronchiolitis score distribution among raters.



**Table 1.** Modified bronchiolitis score<sup>a</sup>

Variable	0 point	1 point	2 points	3 points
Respiratory rate				
Age ≤2 months		≤60	61-69	≥70
Age 2-12 months		≤50	51-59	≥60
Age 12-24 months		≤40	41-44	≥45
Flaring/Retractions	None	Subcostal or intercostal	2 of the following: subcostal, intercostal, substernal OR nasal flaring	3 of the following: subcostal, intercostal, substernal, suprasternal, supraclavicular OR nasal flaring/head bobbing
Site of infection	Sinus antero-lateral thigh	Port site (cholecystectomy)	Disseminated - inguinal nodes, multiple sinuses	Disseminated - Skin, blood, central venous catheter site, lung
Oxygen saturation (% at room air)	>95	90-94	85-89	<85
Auscultation	Normal breath sounds, no wheezing	End-expiratory wheezes ONLY	Full expiratory wheeze	Inspiratory and expiratory wheeze OR diminished breath sounds OR both

<sup>a</sup>Adapted from Goebel J et al.<sup>4</sup>

**Table 2.** Inter-rater reliability and internal consistency by sub-score

	<b>Cohen's kappa</b>	<b>Spearman's Rho</b>	<b>Pearson's correlation</b>	<b>Intra-class correlation (95% CI)</b>
Respiratory rate	0.90*	0.92*	0.95*	0.97 (0.93-0.99)
Retractions	0.31*	0.48*	0.39	0.57 (-0.10-0.83)
Oxygen saturation	1.0	1.0	1.0	1.0
Auscultation	0.33*	0.63*	0.61*	0.77 (0.41-0.91)
Total score	0.30*	0.72*	0.72*	0.84 (0.59-0.94)
Diagnosis	0.44*	0.44*	0.44*	0.63 (0.07-0.86)

\*P value <0.05 as statistically significant.

### Discussion

The modified bronchiolitis score showed significant reliability across paired raters in their assessment of respiratory distress in children with bronchiolitis. There were higher levels of agreement for observed quantitative parameters (i.e. oxygen saturation and respiratory rate) than for subjective parameters (i.e. retractions and wheezing). This may be because subjective parameters such as the assessment of lung sounds rely heavily on the experience of the clinician, the acuity of the clinician's hearing, and his or her personal interpretation of what is heard.<sup>9,10,11</sup> Inter-rater comparison showed that pediatric residents had a wider range of total scores and a 50% reduced ability to identify the presence of 2 or more retractions versus the pediatric attending. This inconsistency in the physical examination can lead to inaccurate assessment of respiratory distress, which in turn may ultimately affect appropriate critical management of airway compromise. An analysis controlling for age group seemed to show an effect on the level of agreement for the variation seen in total score and diagnosis of severity. These findings differ from Gajdos et al<sup>5</sup> who determined that the use of a respiratory score between a physician, nurse and respiratory therapist for assessment of respiratory status of children hospitalized with bronchiolitis showed no differences in weighted kappa estimates in accordance to age group. In our study, the age distribution was skewed to the right, with younger children being more frequently affected with bronchiolitis. The disconnect between high-moderate correlation for total severity score and low-moderate correlation for diagnosis of severity may be explained by the respiratory assessment tool's range in points required for a diagnosis to be made for mild bronchiolitis when compared to moderate bronchiolitis severity.

The use of a respiratory assessment tool can be easily implemented bedside to evaluate the patients' clinical signs. In previous studies, certain clinical parameters, such as retractions, wheezing, respiratory rate and oximetry have been included in respiratory scores, because they have been shown to be strongly related to respiratory distress.<sup>5,12-23</sup> Respiratory assessment is considered an integral part of the clinical reasoning process for physicians-in-training.<sup>9,24</sup> Auscultation is used to assess changes in lung sounds that may be associated with certain respiratory pathologies or dysfunction.<sup>25</sup> However, the presence or absence of

retractions is more telling about the degree of respiratory distress than auscultatory findings<sup>26</sup>, given that the use of accessory muscles is a depiction of the chest cavity maintaining adequate ventilation, and is therefore a representation of work of breathing.<sup>27</sup> Respiratory rate has been associated with lower respiratory tract infection<sup>28,29</sup> and has been stressed as a predictive value in the assessment of respiratory distress in bronchiolitis.<sup>30</sup> Furthermore, pulse oximetry is an objective and easily reproducible parameter, which may not require inter-observer assessment as evidenced by its almost perfect agreement and excellent internal consistency. However, new AAP recommendations on the diagnosis and management of bronchiolitis discourage the continuous use of pulse oximetry due to its limitations on the evaluation of increasing respiratory distress.<sup>31</sup>

Studies have evaluated respiratory scores for inter-observer variability and reproducibility based on kappa value between providers.<sup>5,16,17,21,23</sup> However, there is limited literature showing the use of standardized scoring tools for the assessment of resident skills. An example of a standardized test used to evaluate physicians-in-training is the objective structured clinical examination (OSCE). The OSCE is an assessment tool that aims to evaluate the competence of medical students and physicians-in-training through their performance in simulated cases.<sup>32</sup> This affords the examiner a theoretical and systematic evaluation of a standardized patient, albeit somewhat limited in the assessment of the critically ill patient. Studies have shown that trainees often enter residencies with significant deficiencies in clinical skills<sup>33,34,35</sup>, with a lack of proficiency in physical examination skills.<sup>36,37</sup> Therefore, it is the role of the core faculty to assess, evaluate and directly observe trainees to evaluate resident competence and milestones.<sup>38</sup> This is achieved by making direct observations of students and residents, while they take histories and conduct physical examinations<sup>37,39,40,41</sup>; after which faculty members could give the students and residents constructive feedback on the appropriateness and accuracy of their history-taking and physical-examination techniques and of their interpretation of the findings.<sup>42</sup>

Evaluating resident competence includes an in-depth evaluation of mastery of knowledge, demonstration of observed behaviors, representation of characteristics and behaviors with numbers, mindful practice through

reflection and self-assessment, and demonstration of standardized outcomes for knowledge, skills and behaviors.<sup>43</sup> Utilization of evaluation tools allows for individual assessments to provide learner feedback (assessment for learning), followed by aggregated assessment data used for higher stake decisions (assessment of learning).<sup>44</sup> Lastly, the implementation of an assessment scoring tool provides education benefits by standardizing resident exposure to evidence-based medicine, such as the assessment of respiratory distress, and improvement of skills at communicating respiratory status<sup>45</sup>; remembering that it is the users of the tools, not the tools, which determine the validity of the assessment.<sup>46</sup>

### Limitations

As part of the study, we did not collect identifiable information about the resident. This prevented us from evaluating whether the resident's clinical deficiencies were persistent or deviated from patient to patient. We also lacked information about year of training, and therefore were unable to perform a secondary analysis on inter-rater agreement stratified by the level of training. We did not provide follow up on the residents that showed deficiency in their clinical skills as shown by higher degree of disagreement between raters to show improvement in clinical skills after remediation using the same modified clinical scoring tool. Patients' low disease severity, given that 90% of patients had mild bronchiolitis, may have played a role in the residents' ability to assess subtle clinical differences between mild and moderate disease. We did not take into consideration confounding variables that may have affected the physical evaluation of the patient; such as tachypnea secondary to dehydration, or as a result of undiagnosed pneumonia.

### Conclusion

Clinical diagnostic reasoning is an essential skill for practicing physicians, especially physicians-in-training. Assessing increased work of breathing is a challenging process that improves with practice and experience. The use of a modified bronchiolitis score would enhance resident teaching and assessment of clinical diagnostic reasoning through early identification of challenges in resident learners. This allows for more effective remediation outcomes and promotes a more standardized teaching and evaluation process.

### Acknowledgement

Research reported is supported in part by the National Institute of Minority Health and Health Disparities of the National Institutes of Health Award Number R25MD007607. The content is solely the responsibility of the authors and does not necessarily represent the views of The National Institute of Health.

### Compliance with Ethical Standards

Funding None

Conflict of Interest None

### References:

1. Durning SJ, Artino AR, Schuwirth L, van der Vleuten C. Clarifying assumptions to enhance our understanding and assessment of clinical reasoning. *Acad Med* 2013;88:442-448.

2. Pinsky, LE, & Wipf, JE. Learning and Teaching at the Bedside. University of Washington, Department of Medicine. <http://depts.washington.edu/phsdx/gettingstarted.html>. Accessed November 1, 2018.
3. Hicks PJ, Englander R, Schumacher DJ, Burke A, Benson BJ, Guralnick S, et al. Pediatrics milestone project: next steps toward meaningful outcomes assessment. *J of Grad Med Educ* 2010;2:577-584.
4. Goebel J, Estrada B, Quinonez J, Noorkarim N, Sanford D, Boerth RC. Prednisolone plus albuterol versus albuterol alone in mild to moderate bronchiolitis. *Clin Pediatr* 2000;39:213-220.
5. Gajdos V, Beydon N, Bommenel L, Pellegrino B, de Pontual L, Bailleux S, et al. Inter-observer agreement between physicians, nurses, and respiratory therapists for respiratory clinical evaluation in bronchiolitis. *Pediatr Pulmonol.* 2009;44:754-762.
6. Berg MD, Schexnayder SM, Chameides L, Terry M, Donoghue A, Hickey RW, et al. Part 13: pediatric basic life support: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circ* 2010;122:S862-S875.
7. Fleming S, Thompson M, Stevens R, Heneghan C, Plüddemann A, Maconochie I, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *Lancet.* 2011;377:1011-1018.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.
9. Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. *N Engl J Med* 2014;370(8):744-751.
10. Alsmadi S, Kahya YP. Design of a DSP-based instrument for real-time classification of pulmonary sounds. *Comput Biol Med* 2008; 38(1):53-61.
11. Pasterkamp H, Kraman SS, Wodicka GR. Respiratory sounds: advances beyond the stethoscope. *Am J Respir Crit Care Med* 1997; 156(3):974-987.
12. Berger I, Argaman Z, Schwartz SB, Segal E, Kiderman A, Branski D, Kerem E. Efficacy of corticosteroids in acute bronchiolitis: short-term and long-term follow-up. *Pediatr Pulmonol* 1998;26:162-166.
13. Dabbous IA, Tkachyk JS, Stamm SJ. A double blind study on the effects of corticosteroids in the treatment of bronchiolitis. *Pediatr* 1966;37:477-484.
14. Gadomski AM, Lichenstein R, Horton L, King J, Keane V, Permutt T. Efficacy of albuterol in the management of bronchiolitis. *Pediatr* 1994;93:907-912.
15. Kristjansson S, Lodrup Carlsen KC, Wennergren G, Strannegard IL, Carlsen KH. Nebulised racemic adrenaline in the treatment of acute bronchiolitis in infants and toddlers. *Arch Dis Child* 1993; 69:650-654.
16. Liu LL, Gallaher MM, Davis RL, Rutter CM, Lewis TC, Marcuse EK. Use of a respiratory clinical score among different providers. *Pediatr Pulmonol* 2004;37:243-248.
17. Lowell DI, Lister G, Von Koss H, McCarthy P. Wheezing in infants: the response to epinephrine. *Pediatr* 1987;79:939-945.
18. Richter H, Seddon P. Early nebulized budesonide in the treatment of bronchiolitis and the prevention of postbronchiolitic wheezing. *J Pediatr* 1998;132:849-853.
19. Tal A, Bavilski C, Yohai D, Bearman JE, Gorodischer



- R, Moses SW. Dexamethasone and salbutamol in the treatment of acute wheezing in infants. *Pediatr* 1983;71:13-18.
20. Schuh S, Canny G, Reisman JJ, Kerem E, Bentur L, Petric M, Levison H. Nebulized albuterol in acute bronchiolitis. *J Pediatr* 1990;117:633-637.
21. Wang EE, Milner RA, Navas L, Maj H. Observer agreement for respiratory signs and oximetry in infants hospitalized with lower respiratory infections. *Am Rev Respir Dis* 1992;145:106-109.
22. Wainwright C, Altamirano L, Cheney M, Cheney J, Barber S, Price D, Moloney S, et al. A multicenter, randomized, double-blind, controlled trial of nebulized epinephrine in infants with acute bronchiolitis. *N Engl J Med* 2003;349:27-35.
23. Walsh P, Gonzales A, Satar A, Rothenberg SJ. The interrater reliability of a validated bronchiolitis severity assessment tool. *Pediatr Emerg Care* 2006;22:316-320.
24. Higgs J, Jones M, Loftus S, Christensen N. *Clinical reasoning in the health professions*, 4th edition. Amsterdam: Elsevier Health Sciences; 2018:3-12.
25. Sole ML, Bennett M. Comparison of airway management practices between registered nurses and respiratory care practitioners. *Am J Crit Care* 2014;23(3):191-199.
26. Ahmed A, Graber MA. Evaluation of the adult with dyspnea in the emergency department. In Hockberger RS (Ed.), *UpToDate*. <http://www.uptodate.com/contents/evaluation-of-the-adult-with-dyspnea-in-the-emergency-department>. Accessed June 30, 2018.
27. Mulholland EK, Olinsky A, Shann FA. Clinical findings and severity of acute bronchiolitis. *Lanc* 1990;335:1259-1261.
28. Mahabee-Gittens EM, Grupp-Phelan J, Brody AS, Donnelly LF, Bracey SE, Duma EM, Mallory ML, Slap GB. Identifying children with pneumonia in the emergency department. *Clin Pediatr (Phila)* 2005;44:427-435.
29. Margolis P, Gadomski A. The rational clinical examination. Does this infant have pneumonia? *JAMA* 1998;279:308-313.
30. American Academy of Pediatrics. Subcommittee on Diagnosis and Management of Bronchiolitis. *Diagnosis and Management of Bronchiolitis*. *Pediatr* 2006;118(4):1774-1793.
31. Ralston SL, Lieberthal AS, Meissner HC, et al. Clinical Practice Guideline: The Diagnosis, Management, and Prevention of Bronchiolitis. *Pediatr* 2014;134(5):e1474-e1502.
32. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *J Med Educ* 2011;45(12):1181-1189.
33. Wray NP, Friedland JA. Detection and correction of house staff error in physical diagnosis. *JAMA* 1983;249:1035-1037.
34. Woolliscroft JO, Stross JK, Silva J Jr. Clinical competence certification: a critical appraisal. *J Med Educ* 1984;59:799-805.
35. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med* 1995; 123:795-9.
36. Fletcher SW, O'Malley MS, Bunce LA. Physicians' abilities to detect lumps in silicone breast models. *JAMA* 1985;253:2224-8.
37. Mangione S, Nieman LZ. Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency. *JAMA* 1997;278:717-722.
38. Govaerts, MJ, Van de Wiel MW, Schuwirth LW, Van der Vleuten CP, Muijtjens AM. Workplace-based assessment: raters' performance theories and constructs. *Adv. In Health Sci. Educ* 2013;18(1):375-396.
39. Mangione S, Torre DM. Teaching of pulmonary auscultation in pediatrics: a nationwide survey of all U.S. accredited residencies. *Pediatr Pulmonol* 2003;35:472-476.
40. Mangione S, Burdick WP, Peitzman SJ. Physical diagnosis skills of physicians in training: a focused assessment. *Acad Emerg Med* 1995;2:622-629.
41. Mangione S. Cardiac auscultatory skills of physicians-in-training: a comparison of three English speaking countries. *Am J of Med* 2001;110:210-216.
42. Bordage G. Where are the history and physical? *Can Med Assoc J* 1995;152:1595-1598.
43. Hodges, BD. *The shifting discourses of competence. The Question of Competence: Reconsidering Medical Education in the Twenty-First Century*. Ithaca: Cornell University Press; 2012:14-41.
44. van der Vleuten, CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach* 2012;34(3):205-214.
45. Talib HJ, Lax Y, Reznik M. The Impact of a Clinical Asthma Pathway on Resident Education. *BioMed Research Int* 2018.
46. Moonen-van Loon, JM, Overeem K, Donkers HH, Van der Vleuten, CP, Driessen EW. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Adv Health Sci Educ* 2013;18(5):1087-1102.