

Assessing Resident Diagnostic Skills Using a Modified Bronchiolitis Score

Andrea V Rivera-Sepulveda (✉ rivera.andreav@gmail.com)

Saint Louis University School of Medicine <https://orcid.org/0000-0002-8423-3981>

Muguette Isona

San Juan City Hospital

Research note

Keywords: resident, pediatric, diagnostic skills, bronchiolitis, inter-rater, score, board-certified, pediatrician, pediatric emergency department

Posted Date: December 19th, 2019

DOI: <https://doi.org/10.21203/rs.2.19264/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Pediatric Oncall on January 1st, 2021. See the published version at <https://doi.org/10.7199/ped.oncall.2021.10>.

Abstract

Objective: The purpose of this study is to evaluate resident's assessment of respiratory distress compared to attending's assessment in the academic setting.

Results: We evaluated resident's assessment of respiratory distress in children through inter-rater reliability of a bronchiolitis severity assessment tool among pediatric attendings and pediatric residents. 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. 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), peripheral saturation (ICC=1.0), auscultation (ICC=0.77), and total score (ICC=0.84) were scored similarly across both raters, indicating excellent IRR. The identification of retractions had the least level of agreement.

Introduction

Patient history and physical examination are essential to clinical practice. Literature suggests that medical students and residents are deficient in aspects of physical diagnosis; auscultation being the most prominent¹. The challenge with the resident-teaching faculty rests in identifying subtle and obvious mistakes made by residents in history and physical examinations²; in conjunction with medical evidence of accuracy in exam maneuvers, including sensitivity, specificity, and clinical diagnostic reasoning.

Pediatric residents' growth, acquisition of knowledge, and clinical diagnostic reasoning is usually measured by the descriptive performance expectations from the combination of resident milestones from the Accreditation Council for Graduate Medical Education (ACGME), and the American Academy of Pediatrics (AAP)³. Although resident milestones are objective assessment instruments, they rely on the subjective appraisal of the attending. Therefore, the use of a standardized instrument may complement the evaluation of resident diagnostic skills in the academic setting. The objective of this study is to evaluate resident's assessment of respiratory distress through inter-rater reliability of a bronchiolitis severity assessment tool among pediatric attendings and pediatric residents.

Methods

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 (PED) 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. 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, prematurity, chronic lung disease, bronchopulmonary dysplasia, bronchiectasis, gastroenteritis, liver function impairment, 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.⁴ was modified to include an age-based respiratory rate⁵⁻⁷. 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 saturation (score 0-3); and 4) quality of wheezes (score 0-3) (Table 1). Peripheral capillary oxygen saturation (SpO₂) 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.

Table 1
Modified bronchiolitis score^a

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
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
^a Adapted from Goebel J et al. ⁴				

Outcome measures

The primary outcome measure was to evaluate resident's assessment of respiratory distress through inter-rater 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 comparison between level of clinical backgrounds was evaluated by level of agreement for overall severity of illness classification (i.e., mild, moderate or severe) and the numeric value of the modified bronchiolitis severity assessment score. 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 from 0.21 to 0.4, moderate from 0.41

to 0.6, substantial from 0.61 to 0.8, and almost perfect if kappa is greater than 0.81⁸. Because the raters were paired from a convenience sample of pediatric resident physicians within the PED, 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 0.39 or below, fair from 0.4 to 0.59, from 0.6 to 0.74, and excellent if the value is over 0.75. A value > 0.6 was considered as substantial agreement for kappa, and as good internal consistency for ICC⁸. A sample size of twenty patients per pair 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 statistically significant. Statistical analyses were made with SPSS 21 for Mac OS X Windows (IBM Corp., Armonk, NY, USA).

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 a 6-month period. Patients' ages ranged from 1 to 15 months, with a mean age of 6 months (SD 4). The inter-raters' total scores had a wide distribution that ranged from 1 to 7, with a median score of 3 for both raters (Fig. 1).

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 presented, however, high-moderate correlation between ranked variables. 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 presence of retractions showed the least agreement across all statistical analyses.

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

	Cohen's kappa	Spearman's Rho	Pearson's correlation	Intra-class correlation (95% CI)
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.				

Total score analysis in relation to the distribution of bronchiolitis severity was 90% for mild bronchiolitis and 10% for moderate bronchiolitis. An 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 ($R = 0.45$; $p = 0.052$).

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 the assessment of auscultatory findings relies on the clinician's experience, acuity of hearing, and personal interpretation⁹. 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 affect the management of airway compromise. An analysis controlling for age group showed 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.⁵ 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.

Respiratory assessment tools can be easily implemented bedside given its strong relationship with respiratory distress, and reproducibility^{5,10-12}. Respiratory assessment is considered an integral part of the clinical reasoning process for physicians-in-training¹³. Auscultation is used to assess lung sounds that may be associated with respiratory pathologies or dysfunction¹⁴. However, the presence or absence of retractions is more telling about the degree of respiratory distress than auscultatory findings¹⁵, 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 breathing¹⁶. Respiratory rate has been associated with lower respiratory tract infection^{17,18}, and has been stressed as a predictive value in the assessment of respiratory distress in bronchiolitis¹⁹. 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 consistency²⁰.

There are standardized tests that evaluate physicians-in-training through medical school, but not in residency²¹. There is limited literature on the implementation of standardized scoring tools for the assessment of resident skills. Studies have shown that trainees often enter residencies with significant deficiencies in clinical skills²²⁻²⁴, with a lack of proficiency in physical examination^{25,26}. It is the role of the faculty to assess, and directly observe resident learners to evaluate competence and milestone achievement on the appropriateness and accuracy of history-taking, and physical examination techniques, as well as the interpretation of their findings²⁷⁻³². Therefore, 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 status³³⁻²⁵.

Conclusions

The use of a scoring tool would enhance resident teaching and evaluation of clinical diagnostic reasoning through early identification of diagnostic challenges. This allows for more effective resident remediation, improved patient and resident outcomes, and promotes a more standardized teaching and evaluation process.

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 consider 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.

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

SpO₂: Oxygen saturation

K: Cohen's kappa

SD: Standard deviation

Declarations

Ethics approval and consent to participate: This manuscript adheres to the appropriate reporting guidelines and community standards for data management. It adheres to the highest of ethical standards and rigorous methodology. All participants' parents or guardians provided written informed consent prior to enrolment in the study. This study was approved by the San Juan City Hospital Institutional Review Board (approval no. B0270214).

Consent to publish: All participants' parents or guardians provided written informed consent prior to enrolment in the study, which included the publication of the results.

Availability of data and materials: The data that supports the findings of this study was used under license for the current study, and is not in the public domain.

Competing interests: The authors declare that they have no competing interests.

Funding: The authors received no specific funding for this work.

Authors' contributions: Dr. Rivera-Sepulveda conceptualized and designed the study, acquired and interpreted data. Dr. Isona made substantial contributions to conception and design, analysis and interpretation of data. Both authors were involved in drafting, critically revising and providing final approval of the manuscript.

Acknowledgements: 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.

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(4):442–448.
2. Pinsky, LE, & Wipf, JE. Learning and Teaching at the Bedside. University of Washington, Department of Medicine. <http://depts.washington.edu/physdx/gettingstarted.html>. Accessed November 1, 2018.
3. Hicks PJ, Englander R, Schumacher DJ, Burke A, Benson BJ, Guralnick S, Ludwig S, Carraccio C. Pediatrics milestone project: next steps toward meaningful outcomes assessment. *J of Grad Med Educ* 2010;2(4):577-584.
4. Goebel J, Estrada B, Quinonez J, Noorkarim N et al. Prednisolone plus albuterol versus albuterol alone in mild to moderate bronchiolitis. *Clin Pediatr* 2000;39(4):213-220.
5. Gajdos V, Beydon N, Bommenel L, Pellegrino B, de Pontual L, Bailleux S, Labrune P, Bouyer J. Inter-observer agreement between physicians, nurses, and respiratory therapists for respiratory clinical evaluation in bronchiolitis. *Pediatr Pulmonol.* 2009;44(8):754-762.
6. Berg MD, Schexnayder SM, Chameides L, Terry M, Donoghue A, Hickey RW, Berg RA, Sutton RM, Hazinski MF. **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. *Lanc* 2011;377(9770): 1011-1018.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
9. 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.
10. 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. ☒
11. 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. ☒
12. 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.
13. Higgs J, Jones M, Loftus S, Christensen N. *Clinical reasoning in the health professions*, 4th edition. Amsterdam: Elsevier Health Sciences; 2018:3-12.

14. 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.
15. 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.
16. Mulholland EK, Olinsky A, Shann FA. Clinical findings and severity of acute bronchiolitis. *Lanc* 1990;335:1259–1261.
17. 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.
18. Margolis P, Gadomski A. The rational clinical examination. Does this infant have pneumonia? *JAMA* 1998;279:308–313.
19. American Academy of Pediatrics. Subcommittee on Diagnosis and Management of Bronchiolitis. Diagnosis and Management of Bronchiolitis. *Pediatr* 2006;118(4):1774-1793.
20. Ralston SL, Lieberthal AS, Meissner HC, et al. Clinical Practice Guideline: The Diagnosis, Management, and Prevention of Bronchiolitis. *Pediatr* 2014;134(5):e1474-e1502.
21. 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.
22. Wray NP, Friedland JA. Detection and correction of house staff error in physical diagnosis. *JAMA* 1983;249:1035-1037.
23. Woolliscroft JO, Stross JK, Silva J Jr. Clinical competence certification: a critical appraisal. *J Med Educ* 1984;59:799-805.
24. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med* 1995; 123:795-9.
25. Fletcher SW, O'Malley MS, Bunce LA. Physicians' abilities to detect lumps in silicone breast models. *JAMA* 1985;253:2224-8.
26. Mangione S, Nieman LZ. Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency. *JAMA* 1997;278:717-722.
27. 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.
28. 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.
29. Mangione S, Burdick WP, Peitzman SJ. Physical diagnosis skills of physicians in training: a focused assessment. *Acad Emer Med* 1995;2:622-629.
30. Mangione S. Cardiac auscultatory skills of physicians-in-training: a comparison of three English speaking countries. *Am J of Med* 2001;110:210-216.

31. Bordage G. Where are the history and physical? *Can Med Assoc J* 1995;152:1595- 1598.
32. 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.
33. 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.
34. Talib HJ, Lax Y, Reznik M. The Impact of a Clinical Asthma Pathway on Resident Education. *BioMed Research Int* 2018.
35. 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.

Figures

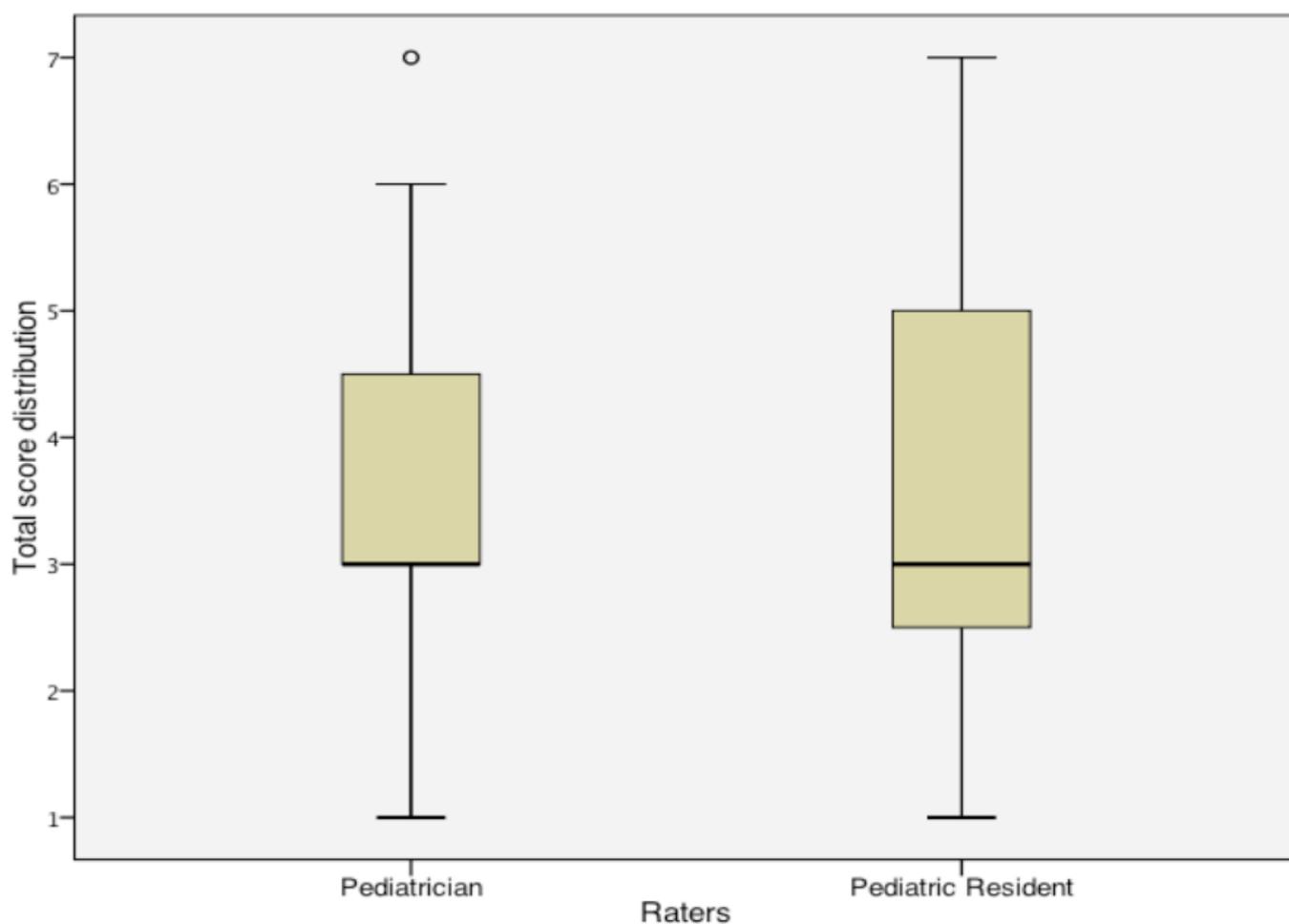


Figure 1

Total modified bronchiolitis score distribution among raters.