Original Publication

OPEN ACCESS

Self-Directed Rater Training for Pediatric History and Physical Exam Evaluation (P-HAPEE) Rubric, a Validated Written H&P Assessment Tool

Marta A. King, MD*, Carrie A. Phillipi, MD, PhD, Paula M. Buchanan, PhD, Linda O. Lewin, MD

*Corresponding author: martak78@gmail.com

Citation: King MA, Phillipi CA, Buchanan PM, Lewin LO. Self-directed rater training for Pediatric History and Physical Exam Evaluation (P-HAPEE) rubric, a validated written H&P assessment tool. *MedEdPORTAL*. 2017;13:10603.
https://doi.org/10.15766/mep_2374-8265.10603

Copyright: © 2017 King et al. This is an open-access publication distributed under the terms of the Creative Commons Attribution-NonCommercial license.

Abstract

Introduction: We developed, revised, and implemented self-directed rater training materials in the course of a validity study for a written Pediatric History and Physical Exam Evaluation (P-HAPEE) rubric. **Methods:** Core training materials consist of a single-page instruction sheet, sample written history and physical (H&P), and detailed answer key. We iteratively revised the materials based on reviewer comments and pilot testing. Eighteen attending physicians and five senior residents underwent self-directed training, scored 10 H&Ps, and completed a rubric utility survey in the course of the validity study. We have since implemented the P-HAPEE rubric and self-directed rater training in a pediatric clerkship. Based on input from reviewers, study raters, faculty members, and medical student users, we have also developed and implemented additional optional supplemental training materials. **Results:** Pilot testing indicated that training takes approximately 1 hour. While reviewers endorsed the training format, several suggested having optional supplemental materials available. Nineteen out of 23 volunteer study raters completed the rubric utility survey. All described the rubric as good or very good and indicated strong to very strong interest in continued use. **Discussion:** The P-HAPEE rubric offers a novel, practical, reliable, and valid method for supervising physicians to assess pediatric written H&Ps and can be implemented using brief, self-directed rater training.

Keywords

Faculty Development, Assessment, Clinical Reasoning, Medical Documentation, History and Physical, Written Communication, Self-Directed Training

Appendices

- A. P-HAPEE Rater Instructions and Scoring Rubric.docx
- B. Sample HP1 With Detailed Scoring Key.docx
- C. Sample HP2 With Detailed Scoring Key.docx
- D. Rater Training Booklet.docx
- E. Rubric Introduction.pptx

All appendices are peer reviewed as integral parts of the Original Publication.

Educational Objectives

By the end of this session, learners will be able to:

1. Implement the Pediatric History and Physical Exam Evaluation rubric self-directed rater training.
2. Reliably assess the quality of pediatric written histories and physicals (H&Ps).
3. Provide trainees with specific goals for H&P improvement.

Introduction

The written history and physical exam (H&P) reflects a writer's competency in numerous core entrustable professional activities, including gathering a history and performing a physical exam, prioritizing a differential, recommending and interpreting common diagnostic and screening tests, and documenting a clinical encounter.¹

While the expectation of obtaining and documenting an H&P is ubiquitous in medical training^{2,3} and a requirement of the USMLE Step 2 CS,⁴ structured formative or summative evaluation is rare. Although a majority of Canadian medical students in McCleod's study² described note writing as a valuable educational activity and reported the impact of note feedback on subsequent documentation as being significant, they felt written documentation assessment did not reflect clinical abilities. McCleod linked this discrepancy to student and program director concerns regarding lack of standardized evaluation criteria, considerable interrater variability, and feedback delays.² In a 2016 survey of pediatric clerkship directors,

varying expectations and lack of faculty time were commonly cited barriers to written H&P assessment.⁵ Missing, inconsistent, and unreliable assessment could result in poor understanding of expectations and diminished value of the written H&P as a teaching and evaluation tool.

The Pediatric History and Physical Exam Evaluation (P-HAPEE) rubric is a written pediatric H&P assessment tool with supporting content, internal structure, and response process validity.⁶ We developed, revised, and implemented brief self-directed rater training during our rubric development and validity study.

We deliberately chose self-directed training over the structured training described by others⁷ as it allows for asynchronous completion at the individual rater's convenience. We hypothesized this training would be easier to implement with a large group of busy physicians who would be scoring H&Ps at different times and more practical for institutions interested in adopting P-HAPEE.

Methods

Rater Training Material Development

Core rater training materials consist of the P-HAPEE rubric and a single-page instruction cover sheet (Appendix A). Sample H&Ps with detailed P-HAPEE rubric answer keys are also used for training (Appendices B & C).

We drafted the rubric and instructions (Appendix A) based on the literature as well as written H&P guidelines solicited from a number of medical schools. Although our validity study focused on H&Ps written by third-year medical students, our goal was to create a rubric that could be used to review and provide feedback for any H&P regardless of author training level. In the rubric instructions, we clarified the fact that the rubric is criterion referenced, with a score of 5 representing a theoretical ideal H&P for a given patient written by an experienced attending without time restraints. We made a deliberate decision to keep rater instructions limited to a single rubric cover sheet to allow raters to move quickly to rubric use. We selected a typical general pediatric patient (an infant with bronchiolitis) and an average-quality student H&P (Appendix B) from among those submitted by students as a routine part of the pediatric clerkship. Three of us then used the P-HAPEE rubric to score this H&P, drafting a detailed answer key via consensus discussion (Appendix B). In the course of the validation study, we iteratively revised the rubric and training materials based on critical review by over 50 clinicians, educators, residents, and medical students.⁶

Rater Training Material Pilot Testing and Revision

We asked a national group of 15 undergraduate and graduate medical educator volunteers to review the rubric and instruction sheet (Appendix A), score the sample H&P (Appendix B), and discuss their experience during one of two facilitated hour-long conference-call focus groups. We asked reviewers to comment on both the rubric and instructions and to discuss their sample H&P scores. We then pilot tested the rubric and training materials with a group of general pediatric inpatient attendings at Saint Louis University.

We made further edits to the rubric and training materials based on reviewer and user feedback. This iterative process is detailed in the paper by King, Phillipi, Buchanan, and Lewin.⁶ We included clarification for the additional history (past medical, family, social history, and review of systems item) as raters often commented on not wanting to penalize a student who did a stellar job for one element but a poor job with another. Similarly, in addition to an early term change from *problem list* to *problem identification*, we clarified that credit could be given to students identifying problems in a more extended discussion without an enumerated problem list. We included specific instruction regarding common questions, such as awarding a maximum of 3 points for a patient with only a single hospital problem, and clarified that a student could potentially score a 5 for his or her plan even if his or her problem identification was inaccurate provided the student appropriately addressed the identified problems. We included a specific

recommendation about maintaining a 48-hour turnaround time for feedback and offered practical advice about reviewing the student's assessment first as it puts the rest of the H&P into perspective and allows for a clearer idea as to whether the writer's H&P is hypothesis driven. In response to early rater requests, we also included specific examples of assessments with scores of 1, 3, and 5 based on H&Ps written for the same patient.

Rater training implementation and evaluation—validity study participants: We asked the 18 attending physician and five senior resident volunteer raters in our validity study to complete self-directed training prior to scoring a selection of 10 H&Ps. As in our pilot testing, training consisted of reviewing the P-HAPEE rubric with instruction cover sheet (Appendix A), using the P-HAPEE rubric to score the sample H&P (Appendix B), reviewing the detailed scoring key (Appendix B), and asking any clarifying questions. Raters also had the option of requesting access to additional scored H&Ps. We asked raters to fill out a seven-item anonymous rubric utility survey after completing H&P scoring. Questions inquired about the easiest and hardest items to score, as well as missing and unnecessary items, and solicited additional suggestions and comments. Some items used a 5-point Likert scale to inquire about ease of use and interest in using the rubric for documentation assessment.

Supplemental Rater Training Materials Development

Supplemental rater training materials consist of an “exceeds expectations” version of the H&P for the infant with bronchiolitis with a detailed P-HAPEE answer key (Appendix C) and a comprehensive rater instruction booklet (Appendix D). We developed these materials based on input from focus group participants, study raters, faculty, and medical student users. Marta A. King revised the original sample H&P based on the consensus answer key (Appendix B) to create the “exceeds expectations” version. Three of us reviewed this detailed answer key again, coming to consensus regarding score and comments. This finalized “exceeds expectations” sample is located in Appendix C.

While keeping the rubric instruction to a brief cover sheet format (Appendix A), we have developed a more comprehensive instructional booklet (Appendix D) with detailed explanation of each rubric item as an additional optional resource for raters to review and reference.

Rater training—Saint Louis University faculty: We continued to use the self-directed rater training during rubric implementation in the pediatric clerkship at Saint Louis University. Ward attendings received the rubric with instruction cover sheet (Appendix A), then were asked to score an H&P. We asked those new to H&P scoring to score the sample H&P, review the answer key (Appendix B), and ask any clarifying questions. Additional optional training materials (Appendices C & D) were also made available. We also included contact information in case of questions or comments regarding the rubric, H&P scoring process, and/or access to additional scored H&Ps.

Training material access: At the Saint Louis University School of Medicine, training materials (Appendices A-C) were stored in a virtual folder accessible by link from the instruction cover sheet and from the clerkship web-based syllabus. Based on the preference of several attendings, we also distributed printed copies of the P-HAPEE rubric and cover sheet (Appendix A) to students during clerkship orientation and asked them to offer a physical document copy of the rubric as well as their H&P to the reviewing attendings. We also turned the rubric into a Google form, allowing immediate web-based access.

Rubric Introduction

Use by faculty: We developed a PowerPoint presentation to introduce the rubric and its validity evidence to faculty (Appendix E). We have presented this information at several national conferences and at our local faculty meetings. At the Saint Louis University School of Medicine, H&P evaluation is a clerkship expectation. We initially introduced the rubric during a department research colloquium. We sought rubric

review and feedback from the general academic pediatric faculty early in development, with several faculty members participating in pilot testing. We then shared the rubric and instructions with subspecialty attendings, soliciting additional feedback.

Attendings are expected to share and discuss the P-HAPEE rubric with the student who created the H&P. If the H&P is found to be below expectations, the student might be allowed to revise and resubmit the initial H&P or, alternatively, submit another one. If the H&P is found to meet or exceed expectations, the attending signs off on that clerkship requirement. Failure to fulfill this or any other clerkship requirement would negatively impact the student's clerkship grade in the professionalism domain.

Use by students: Saint Louis University third-year clerkship students are required to document an H&P for each patient they admit and need to have at least two evaluated and found acceptable for level of training by the attending of record for the patient. We shared the rubric, instruction cover sheet (Appendix A), and sample H&Ps with detailed scoring sheets (Appendices B & C) with the students at the start of the clerkship. In lieu of the detailed rater training booklet (Appendix D), we provided them with a detailed "How to Write a Great H&P" guide. We also included an instruction session on using written and oral case presentations to practice and demonstrate clinical reasoning during clerkship orientation and an early clerkship lecture. Students at Oregon Health & Science University received a copy of the rubric at the start of the clerkship. As with attendings, we included contact information in case of questions or comments regarding the rubric, H&P scoring process, and/or access to additional scored H&Ps.

Results

Pilot Testing and Validity Study

Focus group participants estimated training to take approximately 1 hour. While both focus group participants and volunteer raters endorsed this training format and quickly moved on to actual scoring, several suggested the option of additional supplemental materials (Appendices C & D). None of the 23 validity study raters requested access to additional scored sample H&Ps after completion of the initial rater training. Nineteen out of 23 study volunteer raters completed the anonymous rubric utility survey. All described the rubric as good or very good, and all indicated strong to very strong interest in continued use. The overall interrater reliability (IRR) was .85. Individual P-HAPEE section IRRs were .81 for history, .83 for physical exam and diagnostic studies, and .89 for information synthesis and clinical reasoning. Further validity evidence is indicated in the following section.

Validity Evidence⁶

Content:

- Literature review:
 - Competencies outlined by the Association of American Medical Colleges, the Accreditation Council on Graduate Medical Education, and core pediatric hospital medicine.
 - Previously published written H&P evaluation tools, a consultation letter evaluation tool, and an oral case presentation evaluation tool.
 - Written H&P instructions and sample H&Ps solicited from a number of medical schools.
- Incorporation of medical education frameworks:
 - The early formation of the problem representation and hypothesis-driven nature of information gathering by expert clinicians is reflected in the information-gathering anchors.
 - Diagnostic study anchors reflect the RIME (reporter, interpreter, manager, educator) framework progression from novice to expert clinicians.
 - Information-synthesis anchors incorporate the concept of illness scripts with inclusion of critical information and exclusion of extraneous data, encapsulation with synthesis of lower-level details into a smaller number of higher-level abstract terms proximal to the diagnosis, and the use of semantic qualifiers as a marker of expert clinical reasoning.

- Iterative process and extensive multisource review:
 - Critical review by more than 50 local and national medical educators and clinicians.
 - Fifteen undergraduate and graduate medical educators underwent self-directed rater training, reviewed and scored a sample H&P, and discussed the rubric during one of two facilitated hour-long focus groups.
 - Rubric pilot testing and modification based on feedback from students and raters.

Internal structure:

- IRR:
 - Whole rubric: .85.
 - History: .81.
 - Physical exam and diagnostic studies: .83.
 - Information synthesis and clinical reasoning: .89.
 - Global assessment (below, meets, exceeds expectations): .89.
- Item correlation:
 - Cronbach's α = .93.
 - Highest correlations between an individual item and total score were for assessment (.84) and differential diagnosis (.82).
 - Pairwise correlation between items ranged from .78 between assessment and differential diagnosis to .39 between patient introduction and vital signs/growth parameters.
 - Strong correlation (Spearman's ρ = .867, p < .001) between the calculated mean score and rater's global assessment.
- Factor analysis:
 - An unrotated principle components analysis indicated that there was a single-factor solution with an eigenvalue of 6.044, accounting for 60.4% of the variance.

Response process:

- Time to review (and score):
 - All: 16.3 (6.2) minutes.
 - Attending physicians: 15.1 (6.2) minutes.
 - Senior residents: 19.3 (6.0) minutes; p < .001.
- Rubric utility survey:
 - No missing or unnecessary rubric items identified.
 - All described the P-HAPEE rubric utility as good or very good and endorsed strong or very strong interest in continued use.

Clerkship Use

We implemented the P-HAPEE rubric and rater instructions in the Saint Louis University pediatric clerkship in 2014. As of fall 2016, we have used the rubric for over 450 students and 900 written H&Ps at our institution. All general and subspecialty floor team attendings (~50) participate in H&P scoring as part of their teaching role.

Since implementing the rubric, from students we have received only a handful of questions, all dealing with clerkship practicalities (e.g., "Do I need to turn in the actual H&P to the clerkship office?", "What should I do if I have not heard back from the attending in a certain period of time?", "What if the attending has gone on vacation?", etc.). Questions from attendings have been rare and also have focused on clerkship practicalities (e.g., how to respond to a student who is very late turning in an H&P, the appropriateness of asking attendings to score an H&P of a patient that they are not taking care of, etc.). There have been no requests for access to additional scored H&Ps from either attendings or students and no instances of students questioning or appealing their written H&P assessment with the clerkship

directors. Compliance has not been a problem, and students have submitted the required two H&Ps by the end of the clerkship.

Discussion

We developed, revised, and implemented self-directed rater training materials in the course of a validity study for P-HAPEE, a rubric for assessing pediatric written H&Ps. We have successfully used the same training format and materials throughout rubric pilot testing, validity study, and clerkship implementation.

Challenges and Potential Solutions

Rater variability: We have intentionally not monitored or enforced rater training completion during pilot testing, validity study, or rubric implementation. Study raters who were asked to complete the same self-directed training achieved very good IRR (.85).⁶ Given our current use of the rubric as a low-stakes formative learning exercise (with no cases to date of students questioning or appealing their written H&P assessment), we have not evaluated the IRR or reviewed any grading outliers found in practical use.

Those wishing to use H&P scoring for higher-stakes evaluation could consider having raters discuss scored H&Ps as a group, with detailed explanation provided by the facilitators. The raters could then continue working together and review each other's scores until agreement is reached. Additional options might include monitoring and offering additional training for grading outliers, as well as periodically having two faculty members score the same H&P to monitor for IRR.

Timeliness of feedback: We included specific recommendations about turnaround time of less than 48 hours. Unfortunately, as we did not systemically gather information about timeliness of written H&P feedback before or after P-HAPEE implementation, we cannot compare pre-/postimplementation outcomes. Comparison of pre-/postrubric implementation outcomes might constitute a fruitful area for further inquiry for centers interested in P-HAPEE rubric implementation.

As a start, in addition to making the timeliness and its rationale clear to both students and raters, we suggest using the H&Ps that students document in patients' physical or electronic charts as opposed to a separate H&P created just for the purpose of evaluation, with the expectation of the H&P being documented within 24 hours of admission. Although this does not allow students as much time and flexibility to polish the H&P, it reflects real-life patient care and documentation expectations of physicians.

Competing clinical demands: At Saint Louis University, the attending responsible for H&P evaluation is the patient's attending of record. At times, reconciling patient care and other team teaching responsibilities with H&P evaluation and feedback has been difficult. As the rubric has been shown to be reliable with senior residents acting as raters,⁶ an option might be the team's senior resident and attending sharing assessment duties. Another option might be separation of clinical and assessment duties. In our validity study, raters were able to reliably score H&Ps without being involved in a given patient's care.⁶ This method is currently employed by the Oregon Health & Science University, where the clerkship directors are responsible for H&P evaluation, and by the University of Maryland, where a designated teaching attending is available. An additional option might include having a teaching resident, chief resident, inpatient rotation director, or faculty coach/mentor who would be responsible for these evaluations be assigned to each student for the rotation.

Rubric Implementation Suggestions

We suggest that persons in charge of H&P assessment review all of the main and supplemental education materials and adapt rater instructions and rubric introduction to their settings as needed. We suggest those interested in rubric implementation consider presenting the rubric and validity evidence PowerPoint presentation (Appendix E) during a faculty meeting and/or a faculty development session. Those in charge of H&P assessment should also house materials in a format accessible to the raters and include a

mechanism for raters to contact them in case of questions and to provide feedback regarding the rating process.

Future Directions

Future steps include implementing a student H&P peer review process, extending the use of the rubric to advanced learners, and determining the minimum number of written H&P P-HAPEE assessments needed to reliably establish competency. We are also interested in evaluating the convergent and discriminant correlational relationship between P-HAPEE scores and other factors. Future directions also include assessing the reliability of self and peer scoring and determining whether structured H&P feedback utilizing the P-HAPEE rubric results in improvement in the quality of subsequent H&Ps.

Marta A. King, MD: Associate Professor, Department of Pediatrics, Saint Louis University School of Medicine

Carrie A. Phillipi, MD, PhD: Professor, Department of Pediatrics, Oregon Health & Science University School of Medicine

Paula M. Buchanan, PhD: Associate Professor, Center for Outcomes Research, Saint Louis University

Linda O. Lewin, MD: Associate Professor, Department of Pediatrics, University of Maryland School of Medicine

Acknowledgments

The authors wish to thank the Council on Medical Student Education in Pediatrics Richard T. Sarkin Foundation as well as Dr. Robert Wilmott and the Saint Louis University Department of Pediatrics for grant support for this project. They also want to acknowledge the many medical educators and clinicians who reviewed and provided feedback during Pediatric History and Physical Exam Evaluation rubric development as well as the attending physician and senior resident volunteer H&P raters.

Disclosures

Dr. Phillipi reports limited duties as an associate editor and assessment lead for the Computer-Assisted Learning in Pediatric Program for MedU. She received an annual payment of \$2,500 last year for this work.

Funding/Support

None to report.

Prior Presentations

King MA, Phillipi CA, Buchanan PM, Lewin LO. Developing validity evidence for the written Pediatric History and Physical Exam Evaluation rubric. *Acad Pediatr*. 2017;17(1):68-73. <https://doi.org/10.1016/j.acap.2016.08.001>

King MA, Phillipi CA, Slavin S, Buchanan P, Lewin L. Development and multicenter validation of a written Pediatric History and Physical Exam Evaluation (P-HAPEE) rubric. Platform presentation at: Pediatric Academic Societies Annual Meeting; April 26, 2015; San Diego, CA.

King MA, Phillipi CA, Slavin S, Buchanan P, Lewin L. Development and multicenter validation of a written Pediatric History and Physical Exam Evaluation (P-HAPEE) rubric. Research plenary presentation at: Pediatric Hospital Medicine Annual Meeting; July 23-26, 2015; San Antonio, TX.

King MA, Phillipi CA, Slavin S, Buchanan P, Lewin L. Development and multicenter validation of a written Pediatric History and Physical Exam Evaluation (P-HAPEE) rubric. Poster presented at: Council on Medical Student Education in Pediatrics Annual Meeting; April 6-9, 2016; St. Louis, MO.

King MA, Holmes A, Ottolini M, et al. Teach and assess clinical reasoning using practical, validated written and oral case presentation rubrics: P-HAPEE and PBEAR. Workshop presented at: Council on Medical Student Education in Pediatrics Annual Meeting; April 6-9, 2016; St. Louis, MO.

King MA, Holmes A, Ottolini M, et al. Teach and assess clinical reasoning using practical, validated written and oral case presentation rubrics: P-HAPEE and PBEAR. Workshop presented at: Pediatric Academic Societies Annual Meeting; April 30-May 3, 2016; Baltimore, MD.

King MA, Holmes A, Ottolini M, et al. Teach and assess clinical reasoning using practical, validated written and oral case presentation rubrics: P-HAPEE and PBEAR. Workshop presented at: Council on Medical Student Education in Pediatrics Annual Meeting; March 29-April 1, 2017; Portland, OR.

King MA, Holmes A, Ottolini M, et al. Teach and assess clinical reasoning using practical, validated written and oral case presentation rubrics: P-HAPEE and PBEAR. Workshop presented at: Pediatric Academic Societies Annual Meeting; May 6-9, 2017; San Francisco, CA.

Ethical Approval

This publication contains data obtained from human subjects and received ethical approval.

References

1. *Core Entrustable Professional Activities for Entering Residency: Curriculum Developer's Guide*. Washington, DC: Association of American Medical Colleges; 2014.
 2. McLeod PJ. Assessing the value of student case write-ups and write-up evaluations. *Acad Med*. 1989;64(5):273-274. <https://doi.org/10.1097/00001888-198905000-00016>
 3. Ratcliffe TA, Hanson JL, Hemmer PA, Hauer KE, Papp KK, Denton GD. The required written history and physical is alive, but not entirely well, in internal medicine clerkships. *Teach Learn Med*. 2013;25(1):10-14. <https://doi.org/10.1080/10401334.2012.741538>
 4. *Step 2 Clinical Skills (CS): Content Description and General Information*. <http://www.usmle.org/pdfs/step-2-cs/cs-info-manual.pdf>. Published 2003. Accessed March 6, 2014.
 5. King MA, Parikh P, Buchanan PM, Li STT. Medical student pediatric written history and physical exam instruction, expectations, assessment, and perceived barriers: a survey of North American pediatric clerkship directors. Abstract presented at: Pediatric Academic Societies Annual Meeting; May 6-9, 2017; San Francisco, CA.
 6. King MA, Phillipi CA, Buchanan PM, Lewin LO. Developing validity evidence for the written Pediatric History and Physical Exam Evaluation rubric. *Acad Pediatr*. 2017;17(1):68-73. <https://doi.org/10.1016/j.acap.2016.08.001>
 7. Baker EA, Ledford CH, Fogg L, Way DP, Park YS. The IDEA assessment tool: assessing the reporting, diagnostic reasoning, and decision-making skills demonstrated in medical students' hospital admission notes. *Teach Learn Med*. 2015;27(2):163-173. <https://doi.org/10.1080/10401334.2015.1011654>
-

Received: December 21, 2016 | **Accepted:** May 15, 2017 | **Published:** July 21, 2017