Teaching a Hypothesis-driven Physical Diagnosis Curriculum to Pulmonary Fellows Improves Performance of First-Year Medical Students

Bashar S. Staitieh, Ramin Saghafi, Jordan A. Kempker, and David A. Schulman

Department of Medicine, Division of Pulmonary, Allergy, Critical Care Medicine, and Sleep Medicine, Emory University School of Medicine, Atlanta, Georgia

Abstract

Rationale: Hypothesis-driven physical examination emphasizes the role of bedside examination in the refinement of differential diagnoses and improves diagnostic acumen. This approach has not yet been investigated as a tool to improve the ability of higher-level trainees to teach medical students.

Objectives: To assess the effect of teaching hypothesis-driven physical diagnosis to pulmonary fellows on their ability to improve the pulmonary examination skills of first-year medical students.

Methods: Fellows and students were assessed on teaching and diagnostic skills by self-rating on a Likert scale. One group of fellows received the hypothesis-driven teaching curriculum (the “intervention” group) and another received instruction on head-to-toe examination. Both groups subsequently taught physical diagnosis

to a group of first-year medical students. An oral examination was administered to all students after completion of the course.

Measurements and Main Results: Fellows were comfortable teaching physical diagnosis to students. Students in both groups reported a lack of comfort with the pulmonary examination at the beginning of the course and improvement in their comfort by the end. Students trained by intervention group fellows outperformed students trained by control group fellows in the interpretation of physical findings ($P < 0.05$).

Conclusions: Teaching hypothesis-driven physical examination to higher-level trainees who teach medical students improves the ability of students to interpret physical findings. This benefit should be confirmed using validated testing tools.

Keywords: medical education; hypothesis-driven physical examination

(Received in original form May 24, 2015; accepted in final form January 4, 2016)

Supported by National Institutes of Health training grant T32 HL 076118-09 (B.S.S.).

Author Contributions: conception and design—B.S.S., R.S., and D.A.S.; data acquisition and analysis—B.S.S., R.S., J.A.K., and D.A.S.; drafting the work for important intellectual content—B.S.S., R.S., J.A.K., and D.A.S.; all authors approved the submitted manuscript.

Correspondence and requests for reprints should be addressed to Bashar Staitieh, M.D., 615 Michael Street, Suite 205, Atlanta, GA 30322. E-mail: bashar.staitieh@emory.edu

Ann Am Thorac Soc Vol 13, No 4, pp 489–494, Apr 2016
Copyright © 2016 by the American Thoracic Society
DOI: 10.1513/AnnalsATS.201505-297OC
Internet address: www.atsjournals.org

Despite continued advances in diagnostic technologies, the physical examination remains a key component of clinical assessment. In-patient rounds provide essential time for medical students to learn skills of physical diagnosis from attending physicians and upper-level trainees. However, studies have demonstrated that bedside teaching during attending rounds has declined since the 1970s (1), that residents are spending less time at the

bedside with their students (2), and that residents are no better at physical diagnosis than the students they are charged with teaching (3). These findings call into question the prevalence and effectiveness of bedside teaching. Although the downstream impact of this trend is difficult to assess, the need for improved physical diagnosis education clearly exists.

In the past decade, the hypothesis-driven physical examination has emerged as

a way of conducting the physical examination that enhances its usefulness as a diagnostic modality (4). As opposed to the traditional head-to-toe (HTT) examination, hypothesis-driven physical examination uses a patient's specific complaint to direct the physical exam and contextualize its findings (5–7). Whereas the traditional examination is frequently taught as a checklist of findings that are to be elicited in every patient, hypothesis-driven physical examination

teaches students to think about each physical exam technique as a diagnostic tool for discovering the etiology of a specific symptom (Figure 1). By emphasizing techniques that would be most valuable in determining a diagnosis, hypothesis-driven physical examination has proven its value for learners who are just starting to develop their skills in diagnostic reasoning (4).

Thus far, hypothesis-driven physical examination has been studied predominately as a tool for teaching physical diagnosis to medical students (8), although it has been tested in the setting of resident education as well (9). In most instances, hypothesis-driven physical examination is taught in a formal setting, particularly during general medicine rotations. To date, only one study has focused specifically on the utility of hypothesis-driven physical examination in teaching the pulmonary physical examination (10). No published study has examined the effects of hypothesis-driven physical examination on the ability of higher-level trainees to teach their students, leaving an important knowledge gap on the effects that “teaching teachers” have on medical students.

During the first-year medical student pulmonary module at the Emory University School of Medicine (Atlanta, Georgia), course

material introducing the fundamentals of pulmonary medicine is complemented by bedside physical diagnosis sessions conducted by pulmonary fellows without formal training in teaching. In this study, we evaluated whether a single session teaching hypothesis-driven physical examination to pulmonary fellows improved the diagnostic abilities of their medical students, as assessed by an interactive oral examination.

Methods

This was a prospective cohort study of a single class of first-year medical students at Emory University during their pulmonary module in February 2014. During this module, all medical students received a standard curriculum, which included a faculty-led small group session on the pulmonary history and physical examination, as well as lectures on auscultation and the approach to the pulmonary patient. Small group attendance was mandatory; lecture attendance was not. In addition, during this month, groups of approximately eight students participated in a single 2-hour bedside teaching session with a pulmonary fellow. A list of in-patients with prominent pulmonary findings was

generated on the day of each session and handed out to pulmonary fellows immediately before the session.

At 2 weeks before the first bedside teaching session, 10 pulmonary fellows volunteered, based on schedule availability, to serve as facilitators on dates stipulated by the medical school. The fellows available on a given date were then randomly selected for placement in either the control or intervention group. The intervention group received a 30-minute didactic session, developed in coordination with Emory University School of Medicine administration and designed on the principles of hypothesis-driven physical examination.

The goal of the curriculum was to train fellows to help their students think through differential diagnoses during the physical examination and perform maneuvers that would best help them narrow down the lists that they generated. The curriculum was discussed in person and distributed electronically to intervention group fellows, who were given explicit instructions to refrain from sending it to their students or fellows in the control group. Fellows who did not receive the hypothesis-driven physical examination curriculum were instructed for a comparable amount of time in the general HTT exam that has been practiced in prior years during these sessions.

The students were randomized to attend the pulmonary bedside teaching session from a fellow who had received hypothesis-driven physical examination training or by one who had not. Randomization was performed at the level of students’ pre-existing societies, which are learning communities of approximately 35 students each into which students are placed upon enrollment to the medical school. Society placement attempts to balance a number of factors, including prior academic performance, to obtain four equal groups. After the physical diagnosis session, a small sample of students was asked to describe the content of the bedside activity to ensure that it reflected the intended group assignment (i.e., that the fellows adhered to the appropriate teaching methodology).

Survey Data

All fellows completed an anonymous self-assessment survey to assess their previous training in and comfort with physical diagnosis at the beginning of the hypothesis-driven physical examination course. The first-year medical students also completed

Figure 1. A representative example of the different strategies used in teaching the pulmonary exam by the traditional head-to-toe (HTT) method and the hypothesis-driven physical examination method. Sections shown were culled from curriculum materials used to teach fellows in the control (HTT) and intervention (hypothesis-driven physical examination) groups.

self-assessment surveys on their proficiency with pulmonary physical diagnosis before their teaching session with the fellows and at the end of the pulmonary course before their oral examinations. The surveys used five-point Likert scales (1–5, with 1 being “strongly disagree,” 2 being “disagree,” 3 being “neutral,” 4 being “agree,” and 5 being “strongly agree”). Because students were surveyed anonymously before their group assignments, aggregate survey data were analyzed descriptively.

Training Session

Each medical student participated in a single physical diagnosis session as a part of a group composed of approximately eight students. Sessions were led by a fellow who had received either the hypothesis-driven physical examination curriculum (the “intervention” group) or the traditional HTT curriculum (the “control” group). Each session lasted for 2 hours and incorporated patient visits at one of three large academic hospitals. Patients were prescreened and noted to have relevant pulmonary physical findings. Patients with similar diagnoses and physical findings were selected at each site to minimize variability in student experience. Students

were unaware of their fellow’s group placement.

Oral Examinations

In preparation for the oral examination, a list of eight potential diagnoses with characteristic history and physical findings was provided to all students: asthma, emphysema, interstitial lung disease, pleural effusion, pneumonia, pneumothorax, pulmonary edema, and pulmonary embolism. Instructions detailing the format of the exam were also provided. From the list above, four different exams designed by individuals unaware of the content of the hypothesis-driven physical examination curriculum were generated to test the students on their abilities in several different domains, including the ability to elicit historical data and physical findings, interpret those findings, synthesize a diagnosis, and order appropriate additional diagnostic testing.

Each metric was assessed using a scripted question followed by scoring of the student’s answers without further prompting by the examiner (an example can be found in Figure 2). Exams were developed in coordination with the pulmonary module director (D.A.S.) and

tested by two independent examiners on volunteer fourth-year medical students to ensure that the grading rubrics accurately reflected the students’ intentions.

Oral examinations were administered by both control and intervention fellows, all of whom were instructed in its administration. Students were randomly assigned to examiners by the medical school administration. Examiners were unaware of the students’ prior group assignment. Each student was assigned a “fail,” “pass,” or “honors” rating based on their performance relative to that expected of a first-year medical student. Specific examples were provided to the examiners in a pretest didactic session with a question and answer period to minimize interrater variability. Two test organizers (B.S.S. and R.S.) circulated among the testing rooms to observe and ensure appropriate and uniform exam administration.

Statistical Analysis

Because surveys were anonymous, paired pre- and postintervention statistical analysis was not available; only descriptive aggregate data were assessed; *t* tests were used to compare oral exam scores between the control and intervention groups. For

Exam Maneuver	Response	Scoring—check if asked
Inspection	State: “The patient’s work of breathing is mildly increased.”	
Auscultation	State: “Auscultation reveals diminished breath sounds on both sides.”	
	Ask: “Name three diagnoses from your list that can present with diminished breath sounds.” (Answers: emphysema, asthma, PTX, PNA, effusion)	
Percussion	State: “Percussion reveals increased resonance on both sides.”	
	Ask: “Name two diagnoses from your list that can present with hyperresonant percussion.” (Answers: asthma, emphysema, PTX)	
Tactile Fremitus	State: “Fremitus symmetrically decreased throughout”	
	Ask: “Name two diagnoses from your list that can present with a decrease in tactile fremitus.” (Answer: pleural effusion, emphysema, asthma, pneumothorax)	
Forced expiratory time	State: “Forced expiratory time > 9 seconds”	
	Ask: “Name two diagnoses from your list that present with increased forced expiratory time.” (Answers: emphysema, asthma)	

☐ Score (unshaded): ____/3
(max 3)

☐ Score (shaded): ____/3
(max 3)

Figure 2. Sample oral exam question from chronic obstructive pulmonary disease case. The examiner reads from the script, as outlined, and scores the student’s responses up to a maximum of three points in each category. A list of eight diagnoses was provided to students and examiners before the testing session. PNA = pneumonia; PTX = pneumothorax.

comparisons of overall assessment by the oral exam graders (honors versus passing), a chi-square test was used.

To assess the validity of the oral examination, student scores were compared with their scores on the written examination taken at the end of their pulmonary module. This examination consists of questions similar in format to those on United States Medical Licensing Examination step 1 that cover the breadth of pulmonary medicine. Because the questions are vetted and validated every year, they serve as useful discriminators of student performance.

Pearson correlation analysis was used to determine correlations between oral and written exam scores. ANOVA was performed to assess both interexaminer and interexamination variability in scoring. A *P* value less than 0.05 was considered statistically significant for all tests. Students who were tested by fellows with whom they had worked during the physical examination sessions were left in their original groups for analysis.

This study was granted an educational waiver for review by the Institutional Review Board at Emory University (Institutional Review Board no. 00073321).

Results

A total of 141 students were divided into groups of 72 control students and 69 intervention students based on society

assignments. Average student grades for both groups before the pulmonary module were provided by the medical school administration to assess for confounding by academic performance. There was no significant difference in prior average test scores between control and intervention group students ($82.3 \pm 5.4\%$ versus $82.6 \pm 5.1\%$). Of the 10 participating pulmonary fellows, four first-year fellows and one third-year fellow constituted the intervention group. Three first-year, one second-year, and one third-year fellow constituted the control group.

Survey Data

A total of 12 surveys were collected from pulmonary fellows (Figure 3). In the aggregate, fellows endorsed a lack of formal instruction in the pulmonary exam during their training (average score = 1.7 ± 0.2), but neither agreed nor disagreed with the statement that they were aware of the evidence supporting physical diagnostic maneuvers (average score = 2.8 ± 0.3). They generally agreed with the statement that they had received informal training on exam techniques during the course of their training (average score = 3.6 ± 0.2), and were also comfortable performing and teaching the pulmonary examination (average scores = 4.3 ± 0.2 and 4.1 ± 0.2 , respectively).

In the preintervention period, 113 (80.0%) medical student surveys were returned; 123 (87.0%) were returned in the

postintervention period. Students in both groups were not comfortable with their skills in physical diagnosis at the start of the course, but improved after their fellow-led sessions (from average scores of 2.5 to 4.0 on the five-point Likert scale). The two groups demonstrated similar average changes before and after their fellow-led sessions.

Oral Examinations

Of the 141 medical students, only three were graded by fellows who had trained them in physical diagnosis (i.e., an unblinded assessment), two from the intervention group and one from the control group. Overall, oral exam scores showed a significant, but weak, linear correlation with written exam scores ($\rho = 0.2$, $P = 0.03$).

There was a trend in improved oral examination scores in those students taught by fellows trained in hypothesis-driven physical examination (88 versus 86%, $P = 0.09$). All students passed the examination, but a significantly higher portion of the students taught by intervention group fellows received "honors" (16 versus 7%, $P < 0.01$).

As shown in Figure 4, students taught by intervention fellows scored higher in the interpretation of physical findings (2.4 versus 2.2 points on a three-point scale, $P < 0.05$), but not in the other three examination sections (eliciting historical data and physical findings, synthesizing a diagnosis, and ordering appropriate additional diagnostic testing). In addition, no significant interexamination or interexaminer variability in scoring was noted by ANOVA.

Discussion

This novel study reports the results of a fellow-initiated education project to improve the teaching of the pulmonary physical examination to first-year medical students. We found that teaching senior-level trainees how to teach their medical students hypothesis-driven physical examination significantly improved the ability of students to interpret physical examination findings.

Given the significant amount of time senior trainees spend with their junior colleagues, understanding and improving the educational dynamics between these

Figure 3. Pulmonary fellows reported high levels of comfort performing and teaching the pulmonary examination to students. A total of 12 fellows completed an anonymous five-question survey. The questions, shown on the y axis, related to their degree of agreement with statements on teaching and understanding pulmonary physical diagnosis. Averages of Likert scores are presented along the x axis ($\pm$ SEM).

Figure 4. Medical students taught by hypothesis-driven physical examination–trained fellows (the “intervention” group) outperformed their colleagues taught by head-to-toe (HTT)–trained fellows (the “control” group) at the interpretation of physical findings on their oral examination. No significant differences were found between groups taught by hypothesis-driven physical examination–trained fellows and those taught by HTT-trained fellows by oral examination in the domains of history taking, physical diagnosis technique, or final diagnosis. Students taught by hypothesis-driven physical examination–trained fellows scored significantly higher on the section of the examination relating to the interpretation of physical findings (2.36 versus 2.15 points on a three-point scale; * $P < 0.05$).

groups is essential. Furthermore, despite the clear importance of the physical examination in both clinical practice and diagnostic reasoning, our data and that of others (discussed in the background section) suggest that it is not taught beyond the early years of training. In that regard, our intervention offers a means to improve both student and fellow education.

Our assessment of subjective variables by Likert scales and more objective variables by oral examination allowed assessment of trainee comfort as well as the performance of the intervention. We were not surprised that students were equally comfortable with the physical examination at the end of their pulmonary course, regardless of whether or not they worked with a fellow specifically trained to teach them. However, we note that the focus of the surveys was on student comfort level, and not on fellow teaching skill, which leaves us unable to comment directly on student perception of their teachers.

Students in both groups received lectures and a small group session on physical diagnosis, which likely explains the excellent overall performance on the oral examination. The fact that students taught by intervention group fellows still showed a significant improvement in performance, despite the very good baseline score, adds further credence to the argument

that our intervention is beneficial. Our fellows only received one half-hour training session with the intervention curriculum; each student attended only one 2-hour session with those fellows. An educational benefit, despite such limited exposure, argues for the conceptual power of hypothesis-driven physical examination and highlights the key benefits in teaching higher-level trainees how to teach their students.

To develop a reliable oral examination, we conducted a pretest session in which the same volunteer fourth-year medical students were tested by different examiners. The students then reviewed the grading sheets with the examiners to ensure that the sheets adequately reflected their intentions. Before the examination, we conducted a session to teach the participating fellows how to administer the examination. Finally, two of the study authors (R.S. and B.S.S.) assessed examiner performance in real time during the oral examinations to ensure that scoring was conducted in a uniform manner. Posttest, we performed ANOVA to assess variability between scores for particular cases to ensure that cases were of relatively equal difficulty, and found no significant differences. In addition, we analyzed interexaminer scoring variability on a question-by-question basis and overall to ensure that no outliers existed among the

examiners. No significant variability was found in that assessment.

Of note, the oral examination was developed with local expertise (provided by the module director and medical school administration) to ensure that the test measured what we were attempting to capture and that the questions would be a reasonable assessment of a first-year student’s knowledge, both in content and in style.

We validated the oral examinations as a measure of student knowledge by comparison to the students’ scores with their written examination scores, an imperfect validation metric, given that the two tests may well be evaluating different skills, but the best we had available. Despite that, the statistically significant correlation, weak though it is, suggests that the oral examination was indeed useful in measuring student performance.

We recognize that the slight difference in scores seen in the interpretation of physical findings may be statistically significant, but may not actually be educationally important. Given that the metric included a total of only three items, the difference represents an improvement of 7% in intervention group scores, which we believe to be of considerable benefit. Of note, the minor amount of unblinding that occurred during the oral examinations did not alter the results when we

excluded those students from the analysis. Importantly, the significant improvement in the number of students earning “honors” on the examination, although perhaps a more subjective measure than the scores themselves, still suggests that the students taught by fellows trained in hypothesis-driven physical examination performed at a higher level. However, the teaching ability of the fellows was not directly assessed, and, despite their comparable training level and random group placement, one group may have benefited from stronger teachers overall.

Although we had expected that the technique of hypothesis-driven physical examination would enable the students to improve their history-taking skills, no difference was seen in this regard, either as a result of the aforementioned ceiling effect or the possibility that our intervention offers no such advantage. We note, however, that discussion of hypothesis-driven physical examination in the literature typically argues for a more formal structure (e.g., anticipation of findings, execution of maneuvers, identification and interpretation of findings, justification of diagnosis) than what was used in this study (11), which could also explain why the intervention did not perform as well in other areas. In the future, a more discriminating examination

technique, such as an objective structured clinical examination, would potentially expose a difference between the groups. For this first attempt at using hypothesis-driven physical examination to improve the diagnostic skills of first-year students, we believe it was important to assess reasoning independent of actual exam technique, as the latter could significantly confound the former.

Given the significant improvement in student performance achieved with a relatively small investment of time and personnel, we believe that this model could be expanded to great effect. Specifically, we plan to train future pulmonary fellows in hypothesis-driven physical examination for the purpose of teaching their students; we believe this model could be useful in other domains of advanced-level graduate medical education as well. Furthermore, we suspect that relying on the principles of hypothesis-driven physical examination when teaching students on bedside rounds during their clerkship rotations will help them better understand the utility of the physical examination in clinical care. More studies will be necessary to determine the impact of this style of intervention on more formal aspects of medical student performance (e.g., clerkships, United States Medical Licensing Examination, objective

structured clinical examination), as well as the effects on the reasoning and teaching skills of the advanced-level trainees.

Conclusions

Our study shows that the commonplace habit of asking advanced trainees to teach first-year medical students physical diagnosis may be insufficient to guarantee the students an optimal learning experience. This study supports the use of the hypothesis-driven physical examination as a tool, not only to teach medical students how to approach physical diagnosis, but also to improve the ability of higher-level trainees to teach students. This hypothesis should be tested in other settings using established and discriminating assessment instruments, such as the objective structured clinical examination. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank Dr. Hugh Stoddard of the Emory School of Medicine (Atlanta, GA) for his critical reading of the manuscript, as well as the 2013–2014 fellows of the Emory University Division of Pulmonary, Allergy, and Critical Care Medicine for their time in participating in the study.

References

- 1 LaCombe MA. On bedside teaching. *Ann Intern Med* 1997;126: 217–220.
- 2 Smith MA, Gertler T, Freeman K. Medical students' perceptions of their housestaffs' ability to teach physical examination skills. *Acad Med* 2003;78:80–83.
- 3 Mangione S, Nieman LZ. Pulmonary auscultatory skills during training in internal medicine and family practice. *Am J Respir Crit Care Med* 1999;159:1119–1124.
- 4 Yudkowsky R, Otaki J, Lowenstein T, Riddle J, Nishigori H, Bordage G. A hypothesis-driven physical examination learning and assessment procedure for medical students: initial validity evidence. *Med Educ* 2009;43:729–740.
- 5 Kassirer JP. Teaching clinical medicine by iterative hypothesis testing: let's preach what we practice. *N Engl J Med* 1983;309:921–923.
- 6 Benbassat J, Bauml R, Heyman SN, Brezis M. Viewpoint: suggestions for a shift in teaching clinical skills to medical students: the reflective clinical examination. *Acad Med* 2005;80: 1121–1126.
- 7 Yudkowsky R, Downing S, Klamen D, Valaski M, Eulenberg B, Popa M. Assessing the head-to-toe physical examination skills of medical students. *Med Teach* 2004;26:415–419.
- 8 Ramani S, Ring BN, Lowe R, Hunter D. A pilot study assessing knowledge of clinical signs and physical examination skills in incoming medicine residents. *J Grad Med Educ* 2010;2:232–235.
- 9 Yudkowsky R, Bordage G, Lowenstein T, Riddle J. Residents anticipating, eliciting and interpreting physical findings. *Med Educ* 2006;40:1141–1142.
- 10 Benbassat J, Bauml R. Narrative review: should teaching of the respiratory physical examination be restricted only to signs with proven reliability and validity? *J Gen Intern Med* 2010;25: 865–872.
- 11 Nishigori H, Masuda K, Kikukawa M, Kawashima A, Yudkowsky R, Bordage G, Otaki J. A model teaching session for the hypothesis-driven physical examination. *Med Teach* 2011;33:410–417.