Title

Physical Examination Education in Graduate Medical Education—A Systematic Review of the Literature

Journal

Journal of General Internal Medicine, 28(8)

ISSN

0884-8734

Authors

Mookherjee, Somnath
Pheatt, Lara
Ranji, Sumant R
et al.

Publication Date

2013-08-01

DOI

10.1007/s11606-013-2380-x

Peer reviewed

Physical Examination Education in Graduate Medical Education—A Systematic Review of the Literature

Somnath Mookherjee, MD[1], Lara Pheatt, MA[2], Sumant R. Ranji, MD[3], and Calvin L. Chou, MD, PhD[4]

[1]Department of Medicine, Division of General Internal Medicine, University of Washington, Seattle, WA, USA; [2]Columbia University, New York, USA; [3]Department of Medicine, Division of Hospital Medicine, University of California, San Francisco, USA; [4]Department of Medicine, Division of General Internal Medicine, University of California, San Francisco, USA.

OBJECTIVES: There is widespread recognition that physical examination (PE) should be taught in Graduate Medical Education (GME), but little is known regarding how to best teach PE to residents. Deliberate practice fosters expertise in other fields, but its utility in teaching PE is unknown. We systematically reviewed the literature to determine the effectiveness of methods to teach PE in GME, with attention to usage of deliberate practice.

DATA SOURCES: We searched PubMed, ERIC, and EMBASE for English language studies regarding PE education in GME published between January 1951 and December 2012.

STUDY ELIGIBILITY CRITERIA: Seven eligibility criteria were applied to studies of PE education: (1) English language; (2) subjects in GME; (3) description of study population; (4) description of intervention; (5) assessment of efficacy; (6) inclusion of control group; and (7) report of data analysis.

STUDY APPRAISAL AND SYNTHESIS METHODS: We extracted data regarding study quality, type of PE, study population, curricular features, use of deliberate practice, outcomes and assessment methods. Tabulated summaries of studies were reviewed for narrative synthesis.

RESULTS: Fourteen studies met inclusion criteria. The mean Medical Education Research Study Quality Instrument (MERSQI) score was 9.0 out of 18. Most studies (n=8) included internal medicine residents. Half of the studies used resident interaction with a human examinee as the primary means of teaching PE. Three studies "definitely" and four studies "possibly" used deliberate practice; all but one of these studies demonstrated improved educational outcomes.

LIMITATIONS: We used a non-validated deliberate practice assessment. Given the heterogeneity of assessment modalities, we did not perform a meta-analysis.

CONCLUSIONS AND IMPLICATIONS OF KEY FINDINGS: No single strategy for teaching PE in GME is clearly superior to another. Following the principles of deliberate practice and interaction with human examinees may be beneficial in teaching PE; controlled studies including these educational features should be performed to investigate these exploratory findings.

Published online April 9, 2013

KEY WORDS: physical examination; deliberate practice; graduate medical education.
J Gen Intern Med 28(8):1090–9
DOI: 10.1007/s11606-013-2380-x

BACKGROUND

Physical examination (PE) is an essential aspect of the patient–physician encounter. PE provides information that is critical to accurate diagnosis and can potentially save costs through decreased testing.[1–4] However, published reports have lamented the inadequacy of the PE skills of practicing clinicians,[5–7] and the use of advanced diagnostic tests continues to rise on an annual basis.[8] Although training in PE is a core element of undergraduate medical education (UME), it is irregularly taught in graduate medical education (GME). Despite decades of pleas to reprioritize PE,[9] GME training in aspects of PE is underemphasized whenever formally assessed.[10–14] Furthermore, residents' examination skills in many contexts, including cardiovascular, pulmonary, breast, rheumatologic, and genitourinary systems, among others,[15–34] are suboptimal.

Deficiencies in PE skills are not limited to the performance of examination maneuvers or the identification of abnormalities. Residents struggle to understand the significance of PE findings,[35,36] tend to have low self-confidence in their PE skills,[37] and may spend only a short time actually performing a PE.[38] Undergoing residency training in itself cannot be expected to resolve these deficiencies.[39] Thus, there is a growing consensus that PE education in GME must be improved.[40–42] Reflecting this, new competency-based milestones for internal medicine require that residents accurately perform PE, track important changes, and identify findings that influence clinical decision making.[43]

It is not known what educational strategies will succeed in providing the skills needed by soon-to-be practicing physicians. Deliberate practice is an educational strategy that is known to be an effective means of achieving expertise[44] and has been shown to be effective in teaching clinical skills in medical education.[45–47] In this method of training, the learner repetitively practices skills and un-

dergoes assessment with feedback, resulting in observed improvement of skill performance.[44,45] Deliberate practice would seem to be an ideal means of teaching PE, but this has not yet been systematically examined.

PE teaching methods used in GME are best studied, understood, and optimized separately from UME because the educational context of resident physicians is unique. Residents have little protected time to attend didactics, are focused on experiential learning in the course of their patient care responsibilities, work long hours, and are under constant pressure to adhere to duty hour requirements.

There are few previously published reviews synthesizing the literature on best practices on teaching PE, in particular with regard to GME. A recent non-systematic literature review of models for teaching PE did not investigate the efficacy of these, but offered potential themes for further investigation.[48] A recent systematic review of the efficacy of simulation based medical education (SBME) to teach cardiac auscultation found that SBME appears to be an effective instructional approach.[49] A systematic review of musculoskeletal clinical skills teaching found that maximizing engagement and realistic context for learners may be effective.[50] Finally, a systematic review of the efficacy of the use of patients in teaching and assessing intimate examination skills suggested that patient involvement is beneficial, but was also limited in scope in terms of the type of examination investigated.[51] Given the lack of a literature synthesis on effective means of teaching PE in GME, we systematically reviewed the literature to determine which teaching methods are likely to be effective.

METHODS

Literature Search

We searched for all studies regarding PE education in GME published between January 1951 and December 2012. In consultation with a research librarian, we developed a search strategy designed to be as sensitive as possible. Search strings used for PubMed, Elsevier Biomedical and Pharmacological Database (EMBASE) and Education Resources Information Center (ERIC) databases are listed in Figure 1. Further studies were added based on knowledge of the field by the authors, articles suggested by experts in medical education, review of references from the included papers, and review of PubMed "related articles."

Eligibility Criteria

We used a standard definition of PE skills: "the process of evaluating objective anatomic findings through the use of observation, palpation, percussion, and auscultation. The information obtained must be thoughtfully integrated with the patient's history and pathophysiology."[52] This definition encompasses the PE skills required of clinicians during typical patient encounters, but excludes several domains related to PE separate from our research question, such as studies primarily focused on advanced technologies (such as hand-held ultrasound), procedural skills, or communication skills.

We defined studies of PE education as those which had the intent of improving PE skills. Studies assessing existing skills or attitudes towards PE were excluded. Studies were additionally required to meet seven inclusion criteria: (1) English language; (2) subjects enrolled in GME (studies with multiple levels of learners were eligible if both intervention and control groups included GME subjects and were analyzed separately for educational outcomes); (3) description of the study population, including number of participants and level of training; (4) description of an educational intervention to improve PE skills; (5) assessment of efficacy of the intervention; (6) contemporaneous comparison group, or a historical control group with the same level of training; and (7) report of data analysis (descriptions of outcomes without statistical analysis were not included).

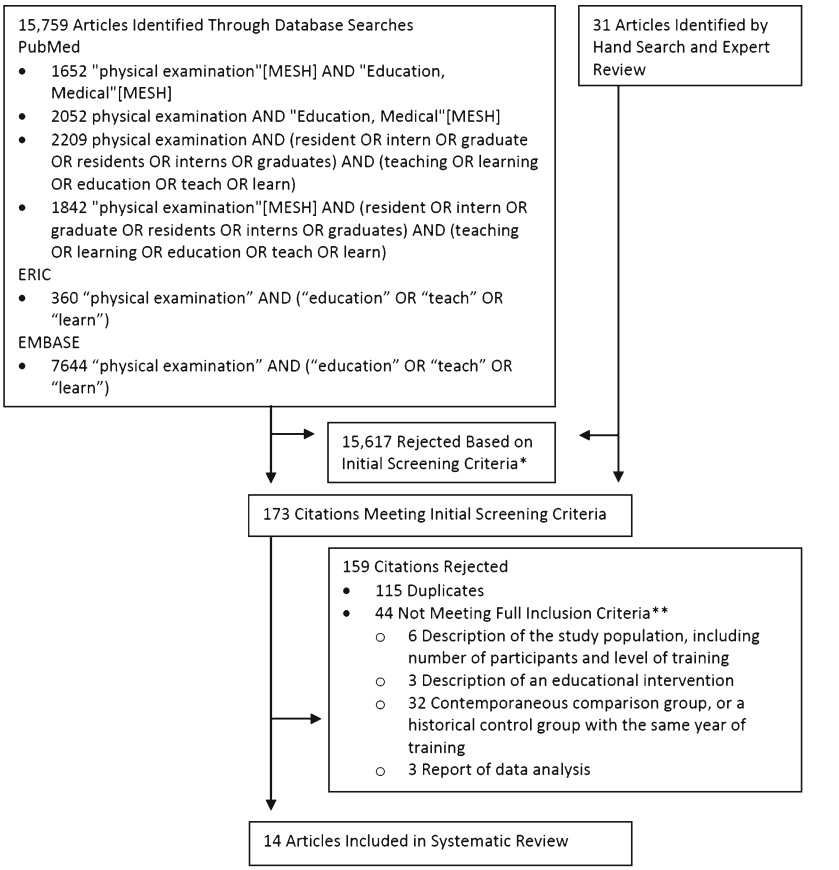
Title and Abstract Review

Retrieved citations were screened by title and abstract review by one of two authors (SM and CLC), to determine if the subject matter of the study was PE education and met inclusion criteria (1) and (2) as listed above. Next, each of these studies was independently reviewed by at least two authors (SM and either CLC or LP) to determine if they met eligibility criteria (3) through (7) listed above. Full text articles were retrieved if necessary to determine eligibility. Conference abstracts were eligible if they met the seven criteria listed above.

Study Review

Studies meeting inclusion criteria were independently reviewed by at least two authors (SM and either CLC or LP). Three quantitative tools were used for assessment of study quality, data extraction, and assessment of use of deliberate practice. To analyze the quality of the studies, we used the medical education research study quality instrument (MERSQI), which has been shown to be associated with study quality.[53] We followed the methods described by the authors of the MERSQI: for each of ten items, each study was scored at the highest possible level. Individual raters' scores were used to determine inter-rater reliability.

For data extraction, we developed a tool based on the Best Evidence Medical Education Collaboration protocol.[54] Extracted data included type of PE, nation where study was performed, amount of teaching time, level and number of learners in intervention and control groups. Educational outcomes were classified using Kirkpatrick's hierarchy[55]:



* Studies regarding physical examination education meeting the first two of seven inclusion criteria: (1) English language; (2) subjects enrolled in GME.
** The remaining five inclusion criteria: (3) description of the study population, including number of participants and level of training; (4) description of an educational intervention to improve PE skills; (5) assessment of the efficacy of the intervention; (6) contemporaneous comparison group, or a historical control group with the same year of training); and (7) report of data analysis. Categorized by the first inclusion criteria not met.

Figure 1. Search and selection of included articles.

Kirkpatrick level 0 (KL0) = no assessment of impact; Kirkpatrick level 1 (KL1) = assessment of reaction; Kirkpatrick level 2a (KL2a) = assessment of attitudes or perceptions; Kirkpatrick level 2b (KL2b) = assessment of knowledge or skills; and Kirkpatrick level 3 (KL3) = assessment of changes in behavior. No studies reached Kirkpatrick level 4. Other than for KL1 (in which a control group is not applicable), outcomes were examined only if compared to the control group. We classified outcomes at each Kirkpatrick level as "X" = not measured, or not compared to a control group; "0" = not better than control group; or "1" = beneficial (intervention group with significantly better outcome than control at $p \leq 0.05$).

We developed a four component rubric to analyze interventions for the use of deliberate practice techniques; there was no existing instrument for this purpose. The rubric was based on previously defined salient elements of deliberate practice,[45,56] which were modified for use in the rubric. The four elements are: (1) repetitive performance of skills by the learner; (2) assessment of skills by the teacher; (3) specific feedback to the learner by the teacher; and (4) observation of improved performance in a controlled setting. To analyze each paper, we used a two-step process. First, we used the rubric to ascertain the presence of each element: papers were scored for the presence of these by two authors (SM and CLC), where 0= not reported, 1=reported. This step served to help train the authors to recognize the key principles of deliberate practice, determine if they were present, and to standardize our assessments. Although deliberate practice can be broken down into these discrete components for analytical purposes, it is an effective combination of these elements that validates that an intervention is anchored in deliberate practice. Thus, using the first step as background, we then performed a global assessment for the use of deliberate practice and assigned each paper a global deliberate practice score, where: 0=no use of deliberate practice or unable to determine; 1=possible use of deliberate practice; or 2=definite use of deliberate practice.

We determined whether the studies used interaction between human examinees and the learners both for the teaching intervention and in the assessment of educational outcomes. "Learner interaction with human examinees" was defined as direct learner contact with a human as part of the

teaching (in contrast to observation of videos, audio recordings, or other electronic methods).

For each of the quantitative scales, disagreements between raters were resolved by discussion to determine final consensus ratings. This review was conducted according to the principles of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[57]

Statistical Analysis and Data Synthesis

Inter-rater reliability was determined for the elements of the MERSQI scores, individual elements of deliberate practice and the global deliberate practice score by calculating kappa. We used the consensus mean and median MERSQI scores with standard deviations to describe the overall quality of included studies. We did not perform meta-analysis of effect sizes, given the heterogeneity of outcome measurements; instead, we integrated findings by narrative synthesis by means of group review of tabulated summaries of studies. All analyses were performed in SAS (version 9.2, Cary, NC).

RESULTS

Characteristics of Eligible Studies

Of the 15,790 citations retrieved, 14 studies met inclusion criteria (Fig. 1). Thirteen were performed in the United States[58–70] and one in Denmark.[71] The cardiac examination was most frequently studied,[59,62,64,67,69,71] followed by the pelvic examination[61,65,68] and the musculoskeletal examination.[58,70] Four studies include randomization of the intervention.[58,61,67,71] Participants were predominantly internal medicine residents;[58–61,63,65–67,69,70] other participants included family practice residents,[59,62] pediatrics residents,[64,65,68] and Danish "house officers" of unknown post graduate year.[71] The majority of studies (n=12) made assessments at KL2b (knowledge and skill).[58,59,61–67,69–71] Two studies made assessments at KL 3 (changes in behavior).[60,65]

Methodological Quality

The mean consensus MERSQI score was 9.0, with a standard deviation of 1.1 and a median score of 9.0, indicating fair overall study quality. Total consensus MERSQI scores for each paper are shown in Table 1. There was perfect inter-rater agreement [kappa=1.0 (95 % CI= 1.0, 1.0)] for all but two items in the ten point MERSQI scale. These items were "sampling" [kappa=0.44 (95 % CI= −0.16, 1.0)] and "content validity" [kappa=0 (95 % CI= −0.52, 0.52)].

Features of Curricula

Table 1 summarizes major features of the curricula. Half of the studies (n=7) used resident interaction with a human examinee as the primary means of teaching PE.[58,60,61,63,65,68,69] Four of these studies used actual patients in the clinical context: genital exams in continuity clinic,[65] bedside rounds with internal medicine inpatients,[69] breast exams in a breast care clinic,[60] and pelvic exams in a pediatric clinic.[68] Two of the studies used "teaching associates," in which the examinee is primarily responsible for teaching the learner. Most of the remaining studies (n=7) targeted the cardiac examination, and the primary educational intervention did not include interaction with patients.[59,62–64,67,70]

There was a wide range in the amount of time spent teaching PE. Five studies relied on a single teaching intervention[58,61,63,64,71] that ranged in length from 3 min[64] to 4 hours.[71] The remaining studies consisted of multiple sessions,[59,60,62,65–70] ranging from ongoing teaching in an elective[68] to three 45-min sessions.[62]

Use of Deliberate Practice

None of the studies explicitly stated that they used deliberate practice. We scored three studies as utilizing "definite use of deliberate practice,"[60,63,69] four with "possible use of deliberate practice,"[58,59,65,68] and seven with "no use of deliberate practice or unable to determine."[61,62,64,66,67,70,71] There was good inter-rater reliability in our assessment of components of deliberate practice as well as the global deliberate practice score: repetitive performance of skills by the learner [kappa=0.86 (95 % CI=0.59, 1.0)]; assessment of skills by the teacher [kappa= 0.72 (95 % CI=0.38, 1.0)]; specific feedback to the learner by the teacher [kappa=0.71 (95 % CI=0.33, 1.0)]; observation of improved performance in a controlled setting [kappa=0.81 (95 % CI=0.46, 1.0)]; and global deliberate practice score [kappa=0.76 (95 % CI=0.46, 1.0)].

Evaluation Design and Assessment Tools

Only three studies used a previously described assessment tool. These were an arthritis examination checklist using an arthritis educator,[58] a multi-modal cardiac examination test,[59] and a checklist to assess pelvic examination skills.[61]

We evaluated each study for the use of human interaction with residents as part of the assessment of the intervention. The majority of studies (n=8) used some form of patient examination and interaction as part of the assessment, even if interaction with a human examinee was not part of the intervention. These assessments used "actual" patient volunteers,[67,71] patient educators,[58,61] and objective structured clinical examinations (OSCEs).[63,66,69,70] For studies not using human examination as part of the assessment (n=

Table 1. Summary of Included Studies

First Author and Year Published	Examination type	MERSQI Score	DP Score	Learner Characteristics[a]	Summary of Curricula	Assessment Strategies	Educational Outcomes[b]
Branch 1999[58]	MSK	11	1	IM Residents. Intervention: 16 Control: 11	Demonstration with patient educator. One session, length of time not reported. Interactive demonstration of MSK exam by arthritis educator.	Standardized arthritis educators used previously validated 85-point checklist.	KL 2b: Skill assessment=1
Criley 2008[59]	Cardiac	9	1	IM and FM Residents. Intervention: 34 Control: 25	Interactive web-based cardiac examination tutorial with virtual patients. Sessions with instructor for orientation, review of materials. Average 5.2 +/− 2 hours on teaching, including the 3 hours with the instructor.	Interactive multimedia computer program to assess knowledge of physiology, audio skills, visual skills, and integration of audio and video skills; survey to assess learner confidence.	KL 2a: Self-assessed confidence=0 KL 2b: Interactive multimedia test=1
Freund 1998[60]	Breast	9.5	2	IM Residents. Intervention: 15 Control: 13	Supervised examination of patients.Three to four 3.5-hour sessions. Technique observed and feedback given. Syllabus given, mammograms reviewed.	Chart review: frequency of screening breast examinations.	KL 3: Chart review=1
Herbers 2003[61]	Pelvic	10	0	IM Residents. Intervention: 39 Control: 33	Small group session with trained patient educator. Unknown length of single session. Four to six interns with groups of two trainers with discussion of perceptions of exam, demonstration by trainers, guided examination by each intern.	Patient educator assessment.	KL 2b: Skill assessment=1
Horiszny 2001[62]	Cardiac	9	0	FM residents. Intervention: 11 Control: 4	Multimedia classroom lecture. Three 45-min sessions. Reviewed unknowns from pre-test, discussed extra diastolic and systolic sounds and murmurs. Played heart sounds, instructed on how to identify, informally tested learners.	Identification of pre-recorded heart sounds.	KL 2b: Skill assessment=1
Houck 2002[63]	Thyroid	8	2	IM residents. Intervention: 19 Control: 20	Multimedia workshop including patient demonstration and practice. Sixty-minute session with slide presentation, video demonstration, demonstration on patient, precepted exam on real patient with verification of exam.	Survey to assess reaction to curriculum; OSCE.	KL 1: 1 KL 2b: OSCE – findings=1 OSCE – technique=0
Iversen 2006[71]	Cardiac	10	0	72 house officers. 4×4 design with 18 residents in each group.	Multimedia classroom lecture and high-quality stethoscope. Four-hour course with theoretical introduction to auscultation and technique. Included 3 hours of training with recorded murmurs.	Actual patient volunteers to assess accuracy of diagnosis.	KL 2b: Patient diagnoses= 0 (both groups)
Keren 2005[64]	Cardiac	9	0	Pediatrics and Pediatrics Emergency Medicine Residents. Intervention: 24 Control: 28	Brief audio tutorial. Three min session with audio teaching of heart sounds based system for grading murmurs.	Assessment to grade accuracy of intensity of pre-recorded heart sounds.	KL 2b: Skill assessment=0
Leder 2005[65]	Pelvic	8.5	1	Pediatrics and IM residents. Intervention: 44 Control: 49	Precepted examination in actual clinical setting. Goal of six exams in 6 months. Preceptors trained by authors. Preceptors reviewed genital anatomy and exam, observed exam, gave feedback, reviewed documentation.	Chart review: documentation of appropriate pediatric genital examination; written tests of knowledge and confidence.	KL 2a: Self-assessed confidence=1 Self-assessed competence=0 KL 2b: Knowledge test=0 KL 3: Documentation review=0 (data summarized and not shown)

(*continued on next page*)

Table 1. (continued)

First Author and Year Published	Examination type	MERSQI Score	DP Score	Learner Characteristics[a]	Summary of Curricula	Assessment Strategies	Educational Outcomes[b]
Mangione 1994[66]	General	8	0	IM residents. Intervention: 56 Control: 11	Multimedia classroom lectures. Twelve hour-long sessions. Lectures on emergency PE and validated findings. Used visual aids, video, audio, remote stethophone, synthesizer.	OSCE; survey to assess general knowledge exam and attitudes.	KL 1: 1 KL 2a: Attitude towards PE=0 KL 2b: Written knowledge=0 Standardized patient=0
Oddone 1993[67]	Cardiac	10	0	IM and FM residents. Intervention: 29 Control: 27	Cardiology simulator. Eight hour-long sessions. Simulator used to demonstrate common findings and proper techniques.	Actual patient volunteers to assess cardiac examination, in combination with simulator.	KL 2b: Simulator for findings and diagnoses=0 Patient diagnoses=0
Rabinovitz 1987[68]	Pelvic	6.5	1	Pediatrics residents. Intervention: 14 Control: 6	Adolescent medicine elective. Emphasis throughout the elective on how to do pelvic exam. RN and MD model behavior and provide immediate feedback.	Survey for self-assessment of confidence and competence.	KL 2a: Self-assessed comfort=1 Self-assessed confidence=1
Smith 2006[69]	Cardiac	8	2	IM Residents. Intervention Group 1: 26 Intervention Group 2: 24 Control: 28 residents	Two types of bedside teaching. Three 2-hour sessions. 1) Demonstration and practice—residents examined, reported to the group. Teacher demonstrated, explained, confirmed or corrected findings. Residents practiced until objectives met. 2) Collaborative discovery—residents examine, report observations and techniques used. Teacher suggests explicit criteria. All re-examine until consensus reached and objectives met.	OSCE.	KL 2b: OSCE technique=1 (both groups) OSCE key findings=1 (Collaborative discovery group)
Smith 2005[70]	MSK	9	0	IM Residents. Two interventions with separate controls. Group 1: Intervention: 26 Control: 10 Group 2: Intervention: 7 Control: 8	Classroom, small group sessions. Three 50-min educational sessions on painful knee or painful shoulder. Reviewed anatomy, common complaints and disorders, discussed cases.	OSCE, survey to assess reaction to curriculum.	KL 1: 1 KL 2b: OSCE – exam and diagnostic skills=1 both groups

[a]*For studies in which there were non-graduate medical education (GME) participants, only GME participation information is summarized here*
[b]*Only outcomes compared to a control group are listed (except for KL1 level assessments). For KL1, 1=positive reaction to the curriculum. For all other Kirkpatrick levels, 1=intervention group with significantly better educational outcome than control group, and 0=intervention group not with significantly better outcome than control group*
MERSQI medical education research study quality instrument; DP deliberate practice; MSK musculoskeletal; IM Internal Medicine; FM Family Medicine; PE physical examination; OSCE Observed Structured Clinical Examination

6), major assessment modalities included the use of recorded or multimedia assessments,[59,62,64] chart reviews,[60,65] and survey.[68]

We categorized studies by the types of assessment used at each Kirkpatrick level. At KL1[63,66,70] and KL2a,[59,65,66,68] surveys were the primary means of assessment. Several assessment strategies were used to assess efficacy of curricula at KL2b.[58,59,61–67,69–71] At KL3, two studies used chart review to assess for screening for breast cancer[60] and for appropriate documentation of genital anatomy in cases of suspected sexual abuse.[65]

Educational Outcomes

We tabulated studies by the use of deliberate practice, use of human examinees as part of the educational intervention, and efficacy of the intervention at the highest Kirkpatrick level examined (Table 2). The three studies that were rated with definite use of deliberate practice all had beneficial educational outcomes. Of the seven studies with deliberate practice scores of 1 or 2, only one study did not have beneficial educational outcomes.[65] Of seven studies for which there was unclear or no evidence of deliberate practice, three showed favorable educational outcomes.

Of the seven studies that included learner interaction with human examinees as part of the curriculum, only one did not result in beneficial educational outcomes—this was the same study of precepting pediatric genital examinations as noted above.[65]

All three studies that assessed reaction (KL1) to their curriculum reported favorable results at this level. For higher Kirkpatrick levels, improved outcomes were seen in multiple types of curricula, but these outcomes were often mixed. For example, precepted pediatric genital examination resulted in increased confidence[65] (KL2a), but no improvement in knowledge (KL2b), self-assessed competence (KL2a) or behavior (KL3). Length of time spent on an educational intervention did not clearly affect whether there were favorable educational outcomes.[66,68] In fact, single session interventions could be effective.[58,61]

COMMENTS

To our knowledge, this is the first systematic review of PE education in GME. In the studies examined, a variety of PE skills were taught, and educational outcomes were assessed using various measures. Therefore, we have limited ability to directly compare educational outcomes between curricula. However, two themes emerge as the most salient for effective PE teaching. First, we propose that the use of deliberate practice may be suited for the effective teaching of PE. Despite the diversity of interventions reviewed for this study, all but one of the interventions rated as having possible or definite use of deliberate practice had significant educational benefit. Interestingly, a single teaching session that is anchored in deliberate practice[63] may be more effective than multiple classroom sessions where deliberate practice is not used.[66] On this basis, we suggest that attention to deliberate practice when designing PE curricula may be of benefit.

Secondly, interaction with human examinees may be useful to residents. Technologically rooted approaches such as simulation and repetitive practice with recordings can be effective in achieving specific learning objectives. However, a practitioner must ultimately be facile with the identification and interpretation of findings in the actual clinical context. For example, it is not surprising that learning the cardiac examination with simulation and audio recordings does not necessarily result in the translation of these skills to the clinical context.[67] We agree that simulation has an important place in PE education, particularly in the context of repetition and pattern recognition. Nevertheless, we submit that interaction with human examinees during PE teaching is more likely to improve the actual practice of PE at the bedside.

We cannot conclude that one PE teaching method is superior to another. This finding should strike a cautionary

Table 2. Categorization of Studies by Use of Deliberate Outcome, Use of Learner Interaction with Human Examinees and Efficacy of Educational Intervention

	Intervention group with better educational outcomes than control at highest Kirkpatrick level assessed	Intervention group and control group with same educational outcomes at highest Kirkpatrick level assessed
Possible or definite use of deliberate practice (deliberate practice score=1 or 2)	**Branch 1999**[58] Criley 2008[59] **Freund 1998**[60a] **Houck 2002**[63a] **Rabinovitz 1987**[68] **Smith 2006**[69a]	**Leder 2005**[65]
No use of deliberate practice or unable to determine (deliberate practice score=0)	**Herbers 2003**[61] Horiszny 2001[62] Smith 2005[70]	Iversen 2006[71] Keren 2005[64] Mangione 1994[66] Oddone 1993[67]

[a]*Studies scored as "definite use of deliberate practice"*
Studies that used learner interaction with human examinees as part of the educational intervention are in bold

note. While medical educators may be inclined to take advantage of current technologically rooted trends to teach PE, such as high-fidelity simulation, we found little convincing evidence in the GME literature that simulation will be more effective in teaching PE than other techniques.[67] Similarly, while it is tempting for educators to advocate for "going back to bedside" as the solution for how to improve residents' sub-optimal PE skills, it is not obvious that simply spending more time with patients will result in better trained residents.[65]

Our study has several limitations. First, the initial title and abstract screen was performed by a single author, and was limited to the three databases that were felt to be the highest yielding on this topic: PubMed, ERIC and EMBASE. The quality of the included studies based on the MERSQI score was lower than average;[53] this is a limitation of work in the field rather than our review. We used a non-validated means of assessing studies for the use of deliberate practice because no other instrument was available; however, we had good inter-rater reliability with our assessment, and consensus was reached between authors for all conclusions regarding deliberate practice. In terms of the educational outcomes, we did not calculate or meta-analyze effect sizes. Given the heterogeneity of assessment modalities, we do not believe that this type of analysis is of practical utility to medical educators. Rather, we used dichotomized summaries of whether the intervention was effective at each Kirkpatrick level studied. This method of literature synthesis ("vote-counting") is known to be susceptible to bias due to under-consideration of study features;[72] nevertheless, we feel that this was a useful tool for our exploratory analysis given the small number of studies that met eligibility criteria. Furthermore, our results may be subject to publication bias, as unsuccessful studies of interventions to improve PE skills may go unpublished. We attempted to minimize this possibility by searching multiple databases, including ERIC and EMBASE, which include conference proceedings and other "grey literature." Finally, we did not evaluate studies on the basis of their incorporation of aspects of diagnostic accuracy; the hypothesis-driven PE, which requires that practitioners use clinical reasoning to forecast, elicit, and make sense of PE findings, is a critical approach that residents must develop.[73]

Despite these limitations, we can make several recommendations for ongoing work in GME PE education. First, we recommend that the two approaches that we have suggested—the use of deliberate practice and of human examinees—be subjected to rigorous evaluation. Secondly, this review highlights the fact that the optimal methods of teaching PE in GME remain unknown—we encourage greater use of true randomized, controlled study designs in future work. Finally, and most ambitiously, we recommend the development of a national GME blueprint outlining competence in PE for practicing physicians in every specialty. The heterogeneity in assessment modalities, inconsistency in Kirkpatrick levels assessed and highly variable definitions of what competence entails make it evident that there is no consensus on how competence in various realms of PE may be manifested or how to best measure it. The ACGME internal medicine milestones for competency in PE also support the need for such a blueprint.[43] A comprehensive blueprint detailing which PE skills graduating residents should have, what competence in these skills entails, and how this is best taught and evaluated is a necessity. Our study begins to address the last of these domains. A PE competency blueprint will permit standardized assessments of PE competency across institutions, will allow for meaningful comparison of PE educational interventions, and will highlight the integral role of PE in physicianship.

Acknowledgements:
Contributors: *The authors wish to thank Judy Maselli, MSPH for her assistance in the statistical analyses.*

Funders: *The Society of Directors of Research in Medical Education provided an honorarium to Dr. Mookherjee and Dr. Chou to support this work.*

Prior presentations: *This work was presented at the Society of Directors of Research in Medical Education (SDMRE) Summer Meeting: Annapolis, MD, June 17, 2012. Prior versions of this work were presented at the University of California at San Francisco Medical Education Day, April 25, 2011; Society of General Internal Medicine (SGIM) National Conference: Phoenix, AZ, May 4–7, 2011; and Western Group of Educational Affairs (WGEA) Regional Conference, Stanford School of Medicine, CA, April 30–May 3, 2011.*

Conflict of Interest: *The authors declare that they do not have a conflict of interest.*

Corresponding Author: *Somnath Mookherjee, MD; Department of Medicine, Division of General Internal Medicine, University of Washington, Magnuson Health Sciences Center, Room B-503, Box 356429, Seattle, WA 98195-6429, USA (e-mail: smookh@u.washington.edu).*

REFERENCES

1. **Fitzgerald FT.** Physical diagnosis versus modern technology. A review. West J Med. 1990;152(4):377–382.
2. **Flegel KM.** Does the physical examination have a future? CMAJ. 1999;161(9):1117–1118.
3. **Fletcher RH, Fletcher SW.** Has medicine outgrown physical diagnosis? Ann Intern Med. 1992;117(9):786–787.
4. **Verghese A, Horwitz RI.** In praise of the physical examination. BMJ. 2009;339:b5448.
5. **Ende J, Fosnocht KM.** Clinical examination: still a tool for our times? Trans Am Clin Climatol Assoc. 2002;113:137–150.
6. **Peixoto AJ.** Birth, death, and resurrection of the physical examination: clinical and academic perspectives on bedside diagnosis. Yale J Biol Med. 2001;74(4):221–228.
7. **Reilly BM.** Physical examination in the care of medical inpatients: an observational study. Lancet. 2003;362(9390):1100–1105.
8. **Smith-Bindman R, Miglioretti DL, Larson EB.** Rising Use of diagnostic medical imaging in a large integrated health system. Heal Aff. 2008;27(6):1491–1502.
9. **Zoneraich S, Spodick DH.** Bedside science reduces laboratory art. Appropriate use of physical findings to reduce reliance on sophisticated and expensive methods. Circulation. 1995;91(7):2089–2092.

10. **Goldenberg DL, DeHoratius RJ, Kaplan SR, Mason J, Meenan R, Perlman SG, et al.** Rheumatology training at internal medicine and family practice residency programs. Arthritis Rheum. 1985;28(4):471–476.
11. **Mangione S.** The teaching of cardiac auscultation during internal medicine and family medicine training–a nationwide comparison. Acad Med J Assoc Am Med Coll. 1998;73(10 Suppl):S10–S12.
12. **Mangione S.** The teaching of chest auscultation in U.S. Internal medicine and family practice medicine residencies. Acad Med: J Assoc Am Med Coll. 1999;74(10 Suppl):S90–S92.
13. **Mangione S, Duffy FD.** The teaching of chest auscultation during primary care training: has anything changed in the 1990s? Chest. 2003;124(4):1430–1436.
14. **Mangione S, Nieman LZ, Gracely E, Kaye D.** The teaching and practice of cardiac auscultation during internal medicine and cardiology training. A nationwide survey. Ann Intern Med. 1993;119(1):47–54.
15. **Chalabian J, Formenti S, Russell C, Pearce J, Dunnington G.** Comprehensive needs assessment of clinical breast evaluation skills of primary care residents. Ann Surg Oncol. 1998;5(2):166–172.
16. **Dhuper S, Vashist S, Shah N, Sokal M.** Improvement of cardiac auscultation skills in pediatric residents with training. Clin Pediatr. 2007;46(3):236–240.
17. **Dugoff L, Everett MR, Vontver L, Barley GE.** Evaluation of pelvic and breast examination skills of interns in obstetrics and gynecology and internal medicine. Am J Obstet Gynecol. 2003;189(3):655–658.
18. **Endean ED, Sloan DA, Veldenz HC, Donnelly MB, Schwarcz TH.** Performance of the vascular physical examination by residents and medical students. J Vasc Surg. 1994;19(1):149—54. discussion 55–6.
19. **Gaskin PR, Owens SE, Talner NS, Sanders SP, Li JS.** Clinical auscultation skills in pediatric residents. Pediatrics. 2000;105(6):1184–1187.
20. **Heiligman RM.** Proficiency of PG-1 internal medicine residents in performing the pelvic examination. Acad Med J Assoc Am Med Coll. 1998;73(3):347.
21. **Johnson JE, Carpenter JL.** Medical house staff performance in physical examination. Arch Intern Med. 1986;146(5):937–941.
22. **Jones JS, Hunt SJ, Carlson SA, Seamon JP.** Assessing bedside cardiologic examination skills using "Harvey," a cardiology patient simulator. Acad Emerg Med Off J Soc Acad Emerg Med. 1997;4(10):980–985.
23. **Lang F.** Resident behaviors during observed pelvic examinations. Fam Med. 1990;22(2):153–155.
24. **Li JT.** Assessment of basic physical examination skills of internal medicine residents. Acad Med J Assoc Am Med Coll. 1994;69(4):296–299.
25. **Mangione S.** Cardiac auscultatory skills of physicians-in-training: a comparison of three English-speaking countries. Am J Med. 2001;110(3):210–216.
26. **Mangione S, Burdick WP, Peitzman SJ.** Physical diagnosis skills of physicians in training: a focused assessment. Acad Emerg Med Off J Soc Acad Emerg Med. 1995;2(7):622–629.
27. **Mangione S, Nieman LZ.** Cardiac auscultatory skills of internal medicine and family practice trainees. A comparison of diagnostic proficiency. JAMA The Journal of the American Medical Association. 1997;278(9):717–722.
28. **Mangione S, Nieman LZ.** Pulmonary auscultatory skills during training in internal medicine and family practice. Am J Respir Crit Care Med. 1999;159(4 Pt 1):1119–1124.
29. **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(2):232–235.
30. **St Clair EW, Oddone EZ, Waugh RA, Corey GR, Feussner JR.** Assessing housestaff diagnostic skills using a cardiology patient simulator. Ann Intern Med. 1992;117(9):751–756.
31. **Vukanovic-Criley JM, Criley S, Warde CM, Boker JR, Guevara-Matheus L, Churchill WH, et al.** Competency in cardiac examination skills in medical students, trainees, physicians, and faculty: a multicenter study. Arch Intern Med. 2006;166(6):610–616.
32. **Vukanovic-Criley JM, Hovanesyan A, Criley SR, Ryan TJ, Plotnick G, Mankowitz K, et al.** Confidential testing of cardiac examination competency in cardiology and noncardiology faculty and trainees: a multicenter study. Clin Cardiol. 2010;33(12):738–745.
33. **Wilt TJ, Cutler AF.** Physician performance and patient perceptions during the rectal examination. J Gen Intern Med. 1991;6(6):514–517.
34. **Wray NP, Friedland JA.** Detection and correction of house staff error in physical diagnosis. JAMA The Journal of the American Medical Association. 1983;249(8):1035–1037.
35. **Herrle SR, Corbett EC Jr, Fagan MJ, Moore CG, Elnicki DM.** Bayes' Theorem and the physical examination: probability assessment and diagnostic decision making. Academic Medicine: Journal of the Association of American Medical Colleges. 2011;86(5):618–627.
36. **Yudkowsky R, Bordage G, Lowenstein T, Riddle J.** Residents anticipating, eliciting and interpreting physical findings. Medical Education. 2006;40(11):1141–1142.
37. **Wu EH, Fagan MJ, Reinert SE, Diaz JA.** Self-confidence in and perceived utility of the physical examination: a comparison of medical students, residents, and faculty internists. Journal of General Internal Medicine. 2007;22(12):1725–1730.
38. **Sharma S.** A single-blinded, direct observational study of PGY-1 interns and PGY-2 residents in evaluating their history-taking and physical-examination skills. The Permanente Journal. 2011;15(4):23–29.
39. **Willett LL, Estrada CA, Castiglioni A, Massie FS, Heudebert GR, Jennings MS, et al.** Does residency training improve performance of physical examination skills? The American Journal of the Medical Sciences. 2007;333(2):74–77.
40. **Kugler J, Verghese A.** The physical exam and other forms of fiction. Journal of General Internal Medicine. 2010;25(8):756–757.
41. **Horwitz RI, Kassirer JP, Holmboe ES, Humphrey HJ, Verghese A, Croft C, et al.** Internal medicine residency redesign: proposal of the internal medicine working group. Am J Med. 2011;124(9):806–812.
42. **Feddock CA.** The lost art of clinical skills. Am J Med. 2007;120(4):374–378.
43. **Green ML, Aagaard EM, Caverzagie KJ, Chick DA, Holmboe E, Kane G, et al.** Charting the road to competence: developmental milestones for internal medicine residency training. Journal of Graduate Medical Education. 2009;1(1):5–20.
44. **Ericsson KA.** Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Academic Medicine: Journal of the Association of American Medical Colleges. 2004;79(10 Suppl):S70–S81.
45. **Duvivier RJ, van Dalen J, Muijtjens AM, Moulaert VR, Van der Vleuten CP, Scherpbier AJ.** The role of deliberate practice in the acquisition of clinical skills. BMC Medical Education. 2011;11(1):101.
46. **McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB.** Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? a meta-analytic comparative review of the evidence. Academic Medicine: Journal of the Association of American Medical Colleges. 2011;86(6):706–711.
47. **McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB.** Medical education featuring mastery learning with deliberate practice can lead to better health for individuals and populations. Academic Medicine: Journal of the Association of American Medical Colleges. 2011;86(11):e8–e9.
48. **Easton G, Stratford-Martin J, Atherton H.** An appraisal of the literature on teaching physical examination skills. Education for Primary Care: An Official Publication of the Association of Course Organisers, National Association of GP Tutors, World Organisation of Family Doctors. 2012;23(4):246–254.
49. **McKinney J, Cook DA, Wood D, Hatala R.** Simulation-Based Training for Cardiac Auscultation Skills: Systematic Review and Meta-Analysis. Journal of General Internal Medicine. 2012.
50. **O'Dunn-Orto A, Hartling L, Campbell S, Oswald AE.** Teaching musculoskeletal clinical skills to medical trainees and physicians: a best evidence in medical education systematic review of strategies and their effectiveness: BEME guide No. 18. Medical Teacher. 2012;34(2):93–102.
51. **Jha V, Setna Z, Al-Hity A, Quinton ND, Roberts TE.** Patient involvement in teaching and assessing intimate examination skills: a systematic review. Medical Education. 2010;44(4):347–357.
52. **Campbell EJ, Lynn C.** The physical examination. In: **Walker H, Hall W, Hurst J, eds.** Clinical methods: the history, physical, and laboratory examinations. Third Editionth ed. Boston: Butterworths; 1990.
53. **Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM.** Association between funding and quality of published medical education research. JAMA: the Journal of the American Medical Association. 2007;298(9):1002–1009.
54. Best Evidence Medical and Health Professional Education. Available at: http://www.bemecollaboration.org/. Accessed January 30, 2013.
55. **Kirkpatrick D.** Evaluation of training. In: **Craig RMI, ed.** Training and development handbook. New York: McGraw-Hill; 1967:87–112.
56. **Wayne DB, Butter J, Siddall VJ, Fudala MJ, Wade LD, Feinglass J, et al.** Mastery learning of advanced cardiac life support skills by internal medicine residents using simulation technology and deliberate practice. Journal of General Internal Medicine. 2006;21(3):251–256.

57. **Moher D, Liberati A, Tetzlaff J, Altman DG.** Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Medicine. 2009;6(7):e1000097.
58. **Branch VK, Graves G, Hanczyc M, Lipsky PE.** The utility of trained arthritis patient educators in the evaluation and improvement of musculoskeletal examination skills of physicians in training. Arthritis Care and Research: The Official Journal of the Arthritis Health Professions Association. 1999;12(1):61–69.
59. **Criley JM, Keiner J, Boker JR, Criley SR, Warde CM.** Innovative web-based multimedia curriculum improves cardiac examination competency of residents. J Hos Med Off Publ Soc Hos Med. 2008;3(2):124–133.
60. **Freund KM, Burns RB, Antab L.** Improving residents' performances of clinical breast examination. J Cancer Educ Off J Am Assoc Cancer Educ. 1998;13(1):20–25.
61. **Herbers JE Jr, Wessel L, El-Bayoumi J, Hassan SN, St Onge JE.** Pelvic examination training for interns: a randomized controlled trial. Acad Med: J Assoc Am Med Coll. 2003;78(11):1164–1169.
62. **Horiszny JA.** Teaching cardiac auscultation using simulated heart sounds and small-group discussion. Fam Med. 2001;33(1):39–44.
63. **Houck WA, Soares-Welch CV, Montori VM, Li JT.** Learning the thyroid examination–a multimodality intervention for internal medicine residents. Teach Learn Med. 2002;14(1):24–28.
64. **Keren R, Tereschuk M, Luan X.** Evaluation of a novel method for grading heart murmur intensity. Arch Pediatr & Adolesc Med. 2005;159(4):329–334.
65. **Leder MR, French GM.** Precepting genital exams: challenges in implementing a complex educational intervention in the continuity clinic. Ambul Pediatr: The Off J Ambul Pediatr Assoc. 2005;5(2):112–116.
66. **Mangione S, Peitzman SJ, Gracely E, Nieman LZ.** Creation and assessment of a structured review course in physical diagnosis for medical residents. J Gen Intern Med. 1994;9(4):213–218.
67. **Oddone EZ, Waugh RA, Samsa G, Corey R, Feussner JR.** Teaching cardiovascular examination skills: results from a randomized controlled trial. Am J Med. 1993;95(4):389–396.
68. **Rabinovitz S, Neinstein LS, Shapiro J.** Effect of an adolescent medicine rotation on pelvic examination skills of paediatric residents. Med Educ. 1987;21(3):219–226.
69. **Smith CA, Hart AS, Sadowski LS, Riddle J, Evans AT, Clarke PM, et al.** Teaching cardiac examination skills. A controlled trial of two methods. J Gen Intern Med. 2006;21(1):7–12.
70. **Smith CC, Newman L, Davis RB, Yang J, Ramanan R.** A comprehensive new curriculum to teach and assess resident knowledge and diagnostic evaluation of musculoskeletal complaints. Med Teach. 2005;27(6):553–558.
71. **Iversen K, Sogaard Teisner A, Dalsgaard M, Greibe R, Timm HB, Skovgaard LT, et al.** Effect of teaching and type of stethoscope on cardiac auscultatory performance. Am Heart J. 2006;152(1):85 e1–7.
72. **Nurius PS, Yeaton WH.** Research synthesis reviews: an illustrated critique of "hidden" judgments, choices, and compromises. Clin Psychol Rev. 1987;7(6):695–714.
73. **Yudkowsky R, Bordage G, Lowenstein T, Riddle J.** Residents anticipating, eliciting and interpreting physical findings. Med Educ. 2006;40(11):1141–1142.