

Effects of Using an Abdominal Simulator to Develop Palpatory Competencies in 3rd Year Medical Students.

Robert M. Hamm (✉ Robert-hamm@ouhsc.edu)

Department of Family and Preventive Medicine, University of Oklahoma Health Sciences Center

David M. Kelley

Department of Family and Preventive Medicine, University of Oklahoma Health Sciences Center

Jose A Medina

Physician Associate Program, University of Oklahoma Health Sciences Center

Noreen S. Syed

Department of Family and Preventive Medicine, University of Oklahoma Health Sciences Center

Geraint A. Harris

Great Plains Family Medicine Residency Program, Oklahoma City, OK

Frank J. Papa

Texas College of Osteopathic Medicine, University of North Texas, Fort Worth TX

Research Article

Keywords: Abdominal examination, diagnosis, medical students, palpation, physical examination, simulation

DOI: <https://doi.org/10.21203/rs.3.rs-256815/v1>

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

[Read Full License](#)

Abstract

Background. US medical students seldom practice enough on patients to master the palpation skills of the physical examination of the abdomen. We evaluated the impact of extended practice with an abdominal simulator which provides concurrent and summative feedback on the appropriate depth of palpation and coverage of relevant areas.

Methods. All third-year medical students were given the opportunity to study with the AbSim simulator during the family medicine rotation. The competence of those who studied with the simulator was measured by its sensors, before and after a training session that included visual feedback regarding the depth and coverage of the student's manual pressure. Additionally, they reported their confidence in their abdominal examination skill at the beginning and end of the rotation.

Results. 119 (86.9%) of 137 students filled out the initial questionnaire, and 73 (61.3%) studied the tutorial. Pre-training competence was predictable by gender, but not by month of third year nor by previous surgery or internal medicine rotations. There was little relation between students' confidence in their abdominal examination skills and their measured competence. The simulator training had a highly significant effect, improving overall competence (4 measures, all p 's < 0.001).

Conclusions. A tutorial using an abdominal simulator increased medical students' competency examining the abdomen, both the depth of palpation and the thoroughness of coverage. Interpretation of changes in confidence are uncertain, because confidence was unrelated to objectively measured competence, although low initial confidence did influence student choice to study with the abdominal simulator.

Introduction

We report a study assessing the use of an abdominal simulator to give 3rd year medical students practice experience with the palpation of the abdomen, an essential skill for accurately diagnosing abdominal pain (Ferguson, 1990; Medford-Davis et al., 2016). The ability to physically examine the abdomen is an essential tool in the rapid identification of likely causes of abdominal pain. The competence builds on a basic understanding of abdominal anatomy and pathology, which can be acquired from text and graphics, yet it also has sensory-motor components. As it involves sensory motor skills, we may assume it is best learned through actually doing the physical movements, rather than simply watching (Clark & Mayer, 2008), and through repeated practice with appropriate feedback (Ericsson, 2007). As a review of 32 studies showed, the more hours spent on a clinical skill the more it is mastered (McGaghie, Issenberg, Petrusa, & Scalese, 2006). However, palpation of patients' abdomens evokes pain, even when executed by skilled clinicians, and so patients and teachers are hesitant to permit repeated exploration. One's fellow students and hired "standardized patient" actors (Sachdeva et al., 1997) can tolerate students' repeated palpations, but their abdomens do not accurately demonstrate how various abnormalities feel, and their acting may not accurately reflect what patients with specific sources and types of pain do. Well-designed

abdominal simulators could potentially fill the gap (Okuda et al., 2009), as they can approximately represent the feel and location of various abnormalities and some aspects of the behavior of patients with sensitive abdomens, without actually hurting people.

There has been a rapid growth in the number and variety of simulators used in medical education. These can range from very simple devices that give practice in specific tasks (e.g., listening to lungs and hearts, stitching wounds, making surgical incisions, or performing injections) to highly sophisticated computer controlled manikins that can provide individuals or teams with realistic experiences, extended through time, performing procedures or managing medical emergencies (Jones, Passos-Neto, & Braghiroli, 2015; Munshi, Lababidi, & Alyousef, 2015). Compared with traditional clinical medical education, deliberate practice using simulation can be superior (McGaghie, Issenberg, Cohen, Barsuk, & Wayne, 2011). Twelve simulator features and best practices for integrating simulation into medical education were identified (McGaghie, Issenberg, Petrusa, & Scalese, 2010). A simulator that focusses on palpation was described by Anders et al (2010), whose device measures manual pressure intensity and rate of applying it for the chiropractor's clinical task of identifying and treating myofascial pain.

There are many reports at conferences and in the literature describing abdominal simulators. The range of technical sophistication varies, from a manikin where the teacher sets up for the student by inserting physical organs into the cavity (Hyde, Erolin, & Ker, 2012; Mahaboob et al., 2010) to ones with computerized control of abnormalities and concurrent and summative feedback of student performance (ACDET's ABSIM system, used in the current research). While companies around the world (e.g., Laerdal in Norway, General Doctor in China, Oniko in the Ukraine) advertise abdominal palpation manikins, their functioning has not been described in the literature. Further, there has been little research evaluating simulation's contribution to mastery of the physical examination of the abdomen. The only mention of learning to examine the abdomen in a large review (Okuda et al., 2009) was as a control condition. As of 2012, there had yet to be any comparative effectiveness studies comparing an abdominal simulator with a specified alternative teaching method (Cook et al., 2012). Ribeiro (Ribeiro, Lederman, Elias, & Nunes, 2016) reviewed the theory and practice of constructing devices that provide accurate simulation of the haptic features needed for training of palpation or surgical procedures, noting that "most of the studies consider only one point of contact, which can limit the simulation realism (p. 1)," but did not include any assessments of devices designed specifically for the palpation of the abdomen.

Despite exhortations that the evaluation of educational interventions should use good measurements (Cheng et al., 2014; McNulty, Halama, & Espiritu, 2004; Zendejas, Brydges, Hamstra, & Cook, 2013) that are related to relevant theories of the acquisition of knowledge and skill (Ericsson, 2007; Jones et al., 2015), it has been observed that much research assessing methods of medical education is methodologically weak (Bewley & O'Neil, 2013). A 2011 review of simulation in medical education found only 14 papers comparing simulation-based education with traditional methods (McGaghie et al., 2011).

In the present study, the educational intervention focusses on the execution of one particular component of the skill of diagnosing abdominal illness, the palpation of the abdomen. It is measured with the lowest

two of Kirkpatrick's 4 levels (Kirkpatrick & Kirkpatrick, 2006): students' self-assessed evaluation of the learning experience (level 1), and objective measures of performance (level 2). Thus this study does not address students' clinical behavior nor patient outcomes, measurements which would be infeasible for routine medical school instruction and inappropriate for training that is focused only on a particular subcomponent of a skill.

The abdominal simulator used in this study was designed to train the basic elements of the physical examination of the abdomen. It has a realistic-feeling torso with the capability to simulate a number of abnormalities and patient responses to palpation (vocal, abdominal wall guarding), and to provide students with feedback regarding their technique (depth and location of palpation with respect to likely location of organs). In the present study we made use of its ability to provide students with feedback regarding the depth and location of their palpation of a realistic feeling abdomen, without any exposure to the feel of abdominal abnormalities or to patient behavior during examination of a tender organ.

Methods

Participants

All 3rd year medical students at this public allopathic medical school are required to take a 4-week rotation in family medicine. They were asked to volunteer for the study during class on the first day of the rotation.

Abdominal Simulator

The study used the AbSim abdominal simulator, which is a frame approximately 56 cm long, 35 cm wide, and 27 cm tall, covered with a rubbery silicon skin, with visible and palpable ribs and pelvis defining the bounds of the abdomen, and an umbilicus in the center. Underneath the visible layer are a layer of spongy padding, a sensor pad, and a taut rubber sheet that supports the sensor pad. The attached computer has a program which can display a schematic outline of an abdomen showing the location of pressure detected by the sensor pad, and visually highlights the locations of key organs.

The simulator provided training directed at developing sensory and motor skills representing appropriate: 1) depth and 2) thoroughness of the student's performance of an abdominal examination. The simulator's various multi-modal formative feedback mechanisms (tactile, visual and vocal feedback) were used both to reinforce appropriate palpatory depth and thoroughness, and to call attention to the need for remediation.

Three forms of feedback to the student were utilized in this study: 1) display of the pressure the student is currently exerting wherever it is sensed, coded as light (gray), deep (blue), or too deep (red); (2) cumulative display of maximum pressure detected at each location during an exam, superimposed on a cartoon that includes the outline of organ locations (Fig. 1-a); (3) list of organs with check marks indicating whether light, deep, or too-deep palpation was detected over the organ (Fig. 1-b). The

boundaries between no, light, deep, and too-deep palpation were calibrated based on clinical experience, as checked by multiple physicians both locally and at various conferences (see Appendix 2 in Supplementary Materials).

Questionnaires

At the beginning and end of the 4-week clerkship, all participating students filled out a questionnaire addressing their self confidence regarding examining the abdomen and their preferences for various modes of instruction. The questionnaire at the end of the clerkship also asked 5 questions regarding their opinion about studying with an abdominal simulator, how many abdomens they had palpated during the month, and whether they had previously done their 8-week clerkships in surgery and internal medicine, in which the abdominal exam is also commonly taught and used (see Supplement, Appendix 1).

Measurement

The simulator's software provided simulation-based palpatory measurements. Both prior to and following training, it stored screenshots documenting each student's 1) depth of palpatory efforts in terms of light, deep or too deep, and 2) the specific areas of the abdomen palpated and not palpated. These measurements were taken twice on each occasion, once with instructions to demonstrate one's light palpatory technique, and again with instructions to explore areas deeply. Additionally, during the simulator's training sequence the software presented the student with a "baseline" summary of the student's performance on another "light, and then deep" examination, and subsequent to the training that process was repeated as a "comparison" summary. These summaries are not stored by the computer, but while the student reviewed them the researcher saved a screenshot. All six of these displays were analyzed by calculating the extent to which the student explored the 11 organs at the requested depth. The baseline and comparison reports also showed the thoroughness of the subjects' palpatory efforts (i.e., the percentage of the abdomen that was palpated or not palpated) and the percent of the total abdomen palpated at each depth. The initial depth scores are used to assess the impact of students' previous clerkship experiences as the third year passed. The comparison of the depth scores from before and after the training identifies palpatory changes (in both thoroughness and depth) resulting from their training via the simulator.

Procedure:

During the clerkship's first day, a researcher explained the study, sought informed consent for participation, administered the initial questionnaire, and offered signup sheets for students to schedule individual sessions studying the abdominal simulator with a researcher. It was stated that even if they did not plan to study with the simulator, their questionnaires were of interest. Reminders were sent using the campus email and calendar system (Microsoft Outlook). For participants who did not schedule a session because they did not yet know their clinic schedule, one email message was sent offering open slots once they had a better idea of their availability. Effort was made to schedule the session during the first week, though occasionally it was necessary to schedule during the second week.

During the individual sessions (see Appendix 2), the researcher oriented the student to the simulator, and explained that there would be assessment, then training, then another assessment. During the pre-training assessment, the subjects were requested to perform three separate palpatory examinations of the abdomen; first palpating only lightly, then palpating only deeply, and third palpating both lightly and deeply. During these assessments the students were given no visual feedback regarding the depth or location of their palpatory efforts via the simulator's monitor, until after the third examination, when the monitor was turned so they could see its record, as the baseline for the upcoming training.

The training sequence consisted of four phases. 1) Inspection of the baseline summary (e.g., Figs. 1a and 1b) which shows a graph of the location and depth of their palpations, and a list of the areas of the abdomen where key organs are located (appendix, gallbladder, ovaries, etc.) and the maximum depth with which they had palpated each area. 2) Training about the appropriate depth of palpation. The student palpated the abdomen while observing the schematic torso outline without organ locations, as the current depth and location of their palpation was displayed on the screen (colored dots indicating light, deep, and too deep palpation). This training was intended to enable the students to calibrate the depth of their palpatory efforts to those criteria established by the investigators, enabling them to recognize when their palpatory efforts are light, deep and too deep. 3) Training about the appropriate thoroughness of palpation. This training consisted of the student palpating the abdomen while observing a schematic torso on the monitor, which this time displayed the specific locations of the various organs likely to present as common and/or important abdominal disorders. The monitor provided real time, visual feedback revealing which areas of the abdomen (and the underlying organs) they had palpated so far. This cumulative display of the deepest pressure they had explored with at each location was intended to reinforce the performance of a thorough palpatory examination of the abdomen. 4) Performance of a comparison palpation, both light and deep, with monitor screen turned away, and subsequently viewing their coverage and depth, comparing before versus after the training, and discussing what was palpated and what was missed. Students were allowed to take as long as they wished on each phase of the training. The training time was not recorded, but most of the students completed the entire session within a half hour.

After the training, for assessment the student repeated the separate light and deep abdominal examinations, with each recorded separately and with no feedback. Finally, the student was reminded of the importance of filling out the questionnaire at the end of the clerkship.

During the fourth and final week, participating students were given the second questionnaire by the clerkship administrative assistant (with the statement that completing it was important even if one had not been able to schedule a session with the simulator). Participants who had not filled it out were sent a copy (Microsoft Word document) by email with encouragement to complete it and send it back.

Data Analysis:

Data were gathered at 3 time points during the 4-week clerkship (first day; first week; last week), plus demographics (from the final questionnaire). Analysis addresses the prediction of measures at each time point from preexisting measures. Thus, we use demographics to predict initial confidence in one's abdominal examination skills; demographics and initial confidence to predict participation with the simulator and pre-training competence (palpation depth and coverage); demographics, confidence, and initial competence to predict post-training competence; and all the above to predict end of rotation confidence and attitudes regarding training with an abdominal simulator. Because not all students used the simulator, two separate analyses are required for some of these models, one with all students but excluding competence data, and one including competence data that includes only students who did the tutorial.

For each prediction three levels of relation were done; pairwise relations (correlations, differences) between each predictor and the dependent variable; multivariate analyses (linear and logistic regression); and structural equation modeling. Summary results are presented in the main paper, organized by interest and importance; the remaining results are in the Supplementary Appendices, which are organized according to the framework described above.

The questions addressed include: Does previous 3rd year training improve students' initial confidence and competence, and whether they are interested in spending time training with an abdominal simulator? Does the training improve competence and confidence? Do the students think that the training was worth the time?

To measure the quality of students' abdominal examination when instructed to palpate *lightly*, we calculated summary scores from the visual report of the light exam for each of 11 organs, for each region of the abdomen, and overall. This score gives most credit for a palpation measured as light, less for deep palpation, even less for too-deep palpation, and 0 for no palpation. A similar procedure applied to the report of the *deep* exam yields a score of how well calibrated the student is when instructed to palpate deeply. Formulas are in Figs. 4a and 4b. The deep exam scoring is also applied to the baseline and comparison records produced as part of the tutorial, which are based on both the student's light palpation and deep palpation, but the computer reports only the deepest palpation sensed during the entire exam. Finally, the simulator's tutorial record also reports the proportion of the area of the abdomen (including areas outside the 11 specified organ locations) that were palpated lightly, deeply, too deeply, or not at all. All these measures include both calibration of palpation depth and adequacy of palpation coverage.

Results

Participants: 137 third year medical students were enrolled in the twelve family medicine clerkship blocks over the academic year (July 2017 to June 2018); 119 (86.9%) of these agreed to participate in the study and filled out the initial questionnaire. Of these participants, 73 (61.3%) completed a session using the abdominal simulator (data for one of these was lost), and 105 (88.2%) filled out the final questionnaire. There was no trend over the year in the proportion of students enrolling in the study to complete the initial

questionnaire, $r = -0.05$ (excluding the final month, June, when the students knew no simulation training would be provided due to researcher absence), and a nonsignificant decrease in the proportion of participants who completed the final questionnaire (correlation with clerkship month, $r = -0.31$). However, the proportion of participants who arranged to study with the abdominal simulator declined greatly over time: the 3-month averages were 81% for July - September, then 73%, 61%, and finally 30% for April - May, with a -0.70 correlation between participation rate and month ($p < .01$). (See also Supplement, Appendix 3.)

Competence: Overall changes in calibration and coverage in students' exploration of abdomen. The baseline and comparison reports provided by the simulator in its tutorial mode included an estimate of the proportion of the abdominal area that had been palpated lightly, deeply, too deeply, or not at all (Fig. 1a). As the student had been instructed to examine the abdomen using both a light and a deep touch, a deep palpation should have been registered at each location. After training, the average proportion deeply palpated increased ($t = 3.7$, $df = 60$, $p < 0.001$), though it was still less than 50% of the abdominal area (Fig. 2). The proportion lightly palpated also increased ($t = 6.8$), while there was much less "too deep" palpation ($t = 6.7$). The students learned to perform more thorough exams, because the proportion not palpated decreased by almost a third ($t = 9.7$).

Changes in the palpation of various organs. From the record of the depth of palpation over each organ (Fig. 1-b) we computed the student's average light and deep scores in each zone, as defined in Fig. 3. At every level of aggregation, there was general improvement in the students' ability to demonstrate an appropriate (i.e., expert-calibrated) level of light palpation in each (Fig. 4-a; each t-test significant with $p < 0.001$, $N = 72$ students). The ability to demonstrate an appropriate deep level of palpation also improved (Fig. 4-b; each t-test significant at $p < 0.001$, $N = 72$, except for the organs located in the left lower abdomen (left ovary, colon), $p = 0.04$, and the upper middle abdomen (gastric/epigastric region), $p = 0.34$).

Use of the simulator's recording capabilities to assess student performance. Presumably third year medical students are improving their physical exam skills as they rotate through various clerkships. The simulator's pre-training metrics should reveal changes in the depth and thoroughness of their palpatory abdominal examination over time, as influenced by previous rotations, independent of any simulator training. There was no statistically significant change in the proportion of the abdomen palpated at each depth (correlations with month ranged from $r = -0.222$ for % not palpated ($p = 0.08$) to $r = 0.183$ for deep palpation ($p = 0.16$). There was, however, a suggestion that previous experience helped the students take advantage of the simulator training, because the post-training deep palpation score increased significantly (correlation with month, $r = 0.24$, $p = 0.04$), as did the proportion of abdomen palpated ($r = 0.28$, $p = 0.26$). It should be remembered the number of participants fell off as the year progressed.

Besides family medicine, the surgery and internal medicine clerkships probably offer students the most instruction and experience with the physical examination of the abdomen, with less in the obstetrics & gynecology and pediatrics clerkships. As the year progressed, it became more likely the students had experienced each specific clerkship when they appeared for the abdominal simulator training session.

Having completed surgery or internal medicine produced slight improvements on the various abdominal simulator measures prior to this study's training, but the differences were not statistically significant.

Gender differences in depth of palpation and thoroughness. Though not planned, we looked at possible differences in the depth of the palpatory efforts of males (N = 35) and females (N = 27). Before training, males tended to demonstrate appropriate (i.e., expert calibrated) levels of both light and deep palpation more than females (Fig. 5: light ideal: difference = 9%, $p = 0.06$; deep ideal: difference = 13%, $p = 0.006$; deep ideal on training Baseline, difference = 13%, $p = 0.007$). These gender differences were somewhat reduced but still significant after training (light ideal: difference = -1%, $p = 0.72$; deep ideal: difference = 5%, $p = 0.04$; deep ideal on post training Comparison, difference = 11%, $p = 0.002$). There was a similar pattern, of initial gender differences reduced by simulator training, in the thoroughness of the abdominal exams. On the baseline examination before training, on average the men palpated a greater percentage of the abdomen than the women (61% vs 55%, $F = 3.99$, $p = 0.05$). On the post-training comparison exam, the difference (72% vs 70%) was not statistically significant.

Relations among initial confidence, studying with the abdominal simulator, and final competence. The multivariate analyses suggest that student self selection into the group that trained with the simulator confounds the comparison between the confidence of those who did and did not train. As the third year passed, students had more confidence in their abdominal examination skills at the beginning of the family medicine clerkship and fewer chose to study with the simulator. Within each month, those who studied with the simulator reported having less confidence. Studying with the simulator increased students' confidence, yet this served only to catch them up with the confidence of those who did not study with the simulator. The multivariate analyses do not satisfactorily address all these variables simultaneously, but one of the structural equation models reported in Supplemental Appendix 7 shows that all three links delineated here are statistically significant, consistent with the suggestion. High initial confidence and abdominal simulator training each predict higher confidence at the end of the rotation, and higher initial confidence predicts that the students will NOT study with the simulator.

Student evaluation of simulator training. Do students who have studied with the abdominal simulator state a more positive evaluation and attitude about it than students who did not study with it, on the final questionnaire? Among the students who worked with the abdominal simulator, most felt they had learned from it, though they qualified the statement. Thus, 19 (31%) said they learned a lot; 19 (31%) said they learned a little but would need to use it more to really benefit; 20 (33%) said they learned a little, and it was about all they would ever learn from such a device; and 3 (5%) said they learned nothing new. This question was not relevant for those who did not study with the abdominal simulator.

Asked whether they'd want to study with an abdominal simulator without an instructor, 93.4% of 61 students who had studied with the AbSim said "yes," compared to 84.4% of 32 students who did not study with it (NS, Chi-Square ($df = 1$) = 1.974, $p = 0.16$). Asked whether classroom demonstrations with an abdominal simulator, without the opportunity to palpate it, would be useful, 50.8% of those who had used the AbSim and 59.4% of those who had not used it said "yes" (NS). Asked whether they would want to

have guidance using an abdominal simulator, or to use it independently, 63.9% of the AbSim studiers and 65.5% of the non-studiers said they'd prefer guidance (NS). 77.0% of the AbSim studiers and 63.3% of the non-studiers would recommend integrating the abdominal simulator into the Family Medicine clerkship curriculum (NS).

Students were asked to rank order 9 modalities for learning about examining the abdomen at both the beginning and the end of the clerkship. There was little change in the relative preference for studying individually, or as a group, with an abdominal simulator, either among those who studied with the AbSim or those who did not. The most preferred mode was a standardized patient, then a real patient, and then individual study with an abdominal simulator was tied with palpating a fellow student. The less preferred modalities were instructor demonstration, video tutorial, lecture with powerpoint, and finally reading.

Discussion

The study of the impact of using an abdominal simulator to train 3rd year medical students in the palpatory aspects of the physical examination of the abdomen showed that the training produced significant improvements in the calibration of the depth of the medical students' abdominal palpation, and in the thoroughness of their examination of the abdominal organs. This improvement was larger than produced by the training and clinical experiences the students had had on their previous rotations. This study makes a contribution because no study has previously been published showing the effects of abdominal simulator training on these two skills.

This study involved both main types of simulation-based research described by Cheng et al (2014): "1) studies that assess the efficacy of simulation as a training methodology and (2) studies where simulation is used as an investigative methodology" (p. 1091). The AbSim simulator provided the training, and its measurements before and after are the source of the objective measurement of changes in competence.

The AbSim abdominal simulator used here has unique features: a) it simulates the feel of the abdomen; b) it measures the student's palpation, both position and depth; and c) it provides these measurements as both concurrent and summary feedback for the student. To the best of our knowledge, it is the only commercially available abdominal simulation trainer that has these useful features, and that has software specifically designed to take the learner through a sequence of experiences that build up competence. Still it is not completely realistic. It is full of air, not liquid, so it does not make informative sounds when tapped. The resistance to a probe is provided by a taut rubber sheet, not by variously shaped organs suspended in liquid. Its "skin" is that of a slender person, so it does not provide experience palpating through various thicknesses of fat.

The training procedure includes 5 of the 12 features that McGaghie et al (McGaghie et al., 2010) identified as promoting learning: "(i) feedback; (ii) deliberate practice; (iii) curriculum integration [in that the students may well need it in their current clinical placement]; ... (v) simulation fidelity; ... (xi) instructor [researcher] training." It can be viewed as a way to support the development of higher levels of

performance through supervised practice of each part of a skill in isolation and subsequent reintegration with the execution of the whole skill (Ericsson, 2004, 2015).

To address Bewley and O'Neil's (2013) observation that much research assessing methods of medical education is methodologically weak, we aimed for: precise definition of the goals of the training in terms of particular subskills of competence; accurate and relevant measurements of the skill; effective training that provides the opportunity to practice the skill with concurrent as well as summative feedback; the best feasible sample; and an adequate design for evaluation of the effect of studying with the simulator.

The study's sample was good in that over 80% of the students in the third-year medical school class participated by filling out the questionnaires at the beginning and end of the family medicine clerkship month; and over 60% of these participants actually took the training, although the proportion doing so declined during the year. The observations were taken over the course of 11 months, so the results are applicable to students at all points in the educational calendar. However, practical considerations – the rights of research participants, and an already overcommitted curriculum – precluded requiring all students to participate. Hence participation was voluntary, as was whether participants studied with the simulator. We did not randomize students to studying or not studying with the abdominal simulator, and we had no palpation competence measures for the nonusers of the simulator. Further, to minimize burden (which could have reduced participation), no measurement using the simulator was obtained at the end of the month; so the only competence data were obtained in one hour-long session at the beginning of the month, before and immediately after the training.

For each individual studying with the simulator we had multiple measures before and after the training, and the measures were taken in fine grain (dots at locations) and summarized in multiple ways (over all, and for particular regions and particular organs, and with separate assessments of light and deep palpation technique). While we found highly statistically significant improvements in the quality of the palpation examination immediately after training – including the thoroughness of the students' coverage, and the calibration of their light and deep palpation, overall and at most regions of the abdomen – we do not know how long these effects would last.

An incidental observation was that on the average female students started off palpating more lightly than men (on the other hand men more often palpated "too deeply"), perhaps because on average they weigh less or are more careful about not hurting people. The differences between men and women were somewhat reduced by the training.

One particular location that almost all students failed to palpate, even after training, was the epigastric area, i.e., the esophageal/gastric junction in the upper middle abdomen. Perhaps this is exceptionally difficult to explore adequately in real patients as well, suggesting that instructors should give special attention to this area. However, it may be a feature of this particular simulator: perhaps the sensors are not as sensitive in that region, which is near the edge of the sensor pad.

In addition to the simulator's objective measurements, the students provided their assessments of the training (those who chose to do it) and of the general concept of simulator training (all participants). Despite the substantial improvements in objectively measured competence, there was no difference between end-of-clerkship evaluations by the participants who did and did not experience the abdominal simulator training. We had expected a difference. One interpretation is that participants who did not get simulation training chose not to because of difficulty in scheduling it (e.g., from their off-campus clinical site) rather than because of disinterest. An alternative interpretation would be that the training made little impression three weeks later.

Use of abdominal simulator to measure student performance. In addition to studying the effect of the simulator's tutorial, the other main focus of the study was the utility and validity of the simulator's measurements. We asked whether the students' previous clerkship rotations, especially surgery and internal medicine, produced better performance as measured by the abdominal simulator prior to the students' training. Our results indicate that the initial measure of palpation competence was essentially level throughout the year; neither rotating through specialties in which abdominal palpation is necessary, nor the cumulative total of all previous clerkships, produced much improvement. We had expected that prior experience would improve student competence and that it would be reflected in the simulator's measures.

Assuming the measures are valid, what does this finding imply? Perhaps students acquire accurate palpatory calibration on each clerkship and then rapidly lose it when it is not daily reinforced. Perhaps this particular aspect of the skill is little emphasized elsewhere, because it is not viewed as the primary skill needed in the physical examination of real patient abdomens. Third year medical students may hear about this element of the physical examination of the abdomen on third year rotations, but not have much opportunity to practice. There was a large range in the amount of previous experience students anecdotally reported to the researcher, depending on the particular clinics and physicians they shadowed. On the final questionnaire in this study they also reported a wide range of chances to palpate patient abdomens on the family medicine clerkship. A few students who exhibited clear competence, as observed by the researcher and measured by the simulator, explained that they had previously worked as emergency medical technicians or nurses. This emphasizes the important role of having the opportunity to practice on numerous patients, presumably with feedback.

What is proven by the increase in student performance after training, as measured by the abdominal simulator's measurements, which naturally are linked tightly to the particular calibration standards as embodied on the simulator? At minimum, it indicates the students learned to temporarily adjust their palpation of the simulator in accord with the instructions – indeed it may be a very accurate measure of this particular subcomponent of abdominal examination competence – but it does not prove they have learned this skill permanently, nor does it prove competence at examining the wide range of body types, nor the broad range of examination skills taught by other physicians.

However, the fact that the very experienced students did well on the simulator's measures suggests it captures an essential part of an integrated competence. This implies that most medical students are not mastering the integrated skill of the physical exploration of the abdomen during their third-year rotations. Many of the students reported that they had not previously been guided to focus in such detail upon their palpatory technique, and so they paid close attention to the simulator's feedback on how they were doing.

Is there value in having the simulator train students to perform the palpatory exam to a very particular standard – to calibrate to the level set by one expert – when patient body types and complaints may require a different range of pressures and techniques? We'd argue that there is benefit learning to execute a skill in a particular way, because one becomes aware of the range of possibilities in doing so. It does not stop the physician from recalibrating when the situation demands it; rather it facilitates the ability to do so.

In conclusion, though we don't know for sure if the simulator's assessment of competence is generally valid, other than for immediate confirmation that the students adjust their calibration as the simulator tells them to do, there is plenty of indirect evidence to think it is. Presumably they learn how it is possible to calibrate, and they are made aware of depth and coverage of palpation as important elements of their palpatory exam, whether or not the particular depth and coverage to which they were trained here is the best one in general.

Limitations. An ideal study for assessing training with a simulator would be able to tell us the cost per long term stable improvement in competence, as compared to the costs and improvements of alternative training methods. Costs of training as done in this study would include buying or leasing the simulator and the time of the researcher who guided the student through the procedure. We did not test whether unsupervised students following only the instructions of the tutorial program would have the motivation to repeat the examination multiple times sufficient to improve their skills.

Although we showed the training produced improvement between the pre and immediate post training measures of the particular subskills – depth and coverage – it is just part of the total skill of the physical examination of the abdomen, and other subskills merit supervised practice as well.

A major weakness, as we have acknowledged, is the non-randomized nature of the study, and the fact that even among study participants we had measures only from those who chose to arrange a session with the abdominal simulator. Thus we have competence measures only for those who were interesting in getting the training. We cannot compare the competence of participants who studied using the simulator and those who did not, nor with non-participants (who did not even fill out the questionnaires). Hence a possibility that our non-randomized non-comparative design cannot discount is that only the least competent students chose to study, as self-remediation. The fact that fewer participants opted to spend the hour with the simulator as the year went by is consistent with that alternative theory. It is less of a commitment for a program to provide a simulator as remediation, than as a requirement for all students.

Our study design also could not tell us how long the effects of the training last. We don't have competence measures at the end of the month, so we cannot compare the effects of the simulator to the effects of the clinical experiences offered by the family medicine, internal medicine, and surgery clerkships. We did not investigate the difference between the effect of expensive, guided individual use of the simulator versus less costly group instruction, or individual sessions using the simulator guided only by its computer program. We did not quantify the cost of the simulator per student; as this study focused on individual measurement, it used researchers not teachers, and did not compare individual instruction with team or group exercises with a simulator.

The AbSim simulator was used both to provide the training and to assess student performance on the trained skill. To allay concern about simply "training for the test," it would be better to have an independent assessment of change in competence. And to assess the accuracy of the simulator's assessments, it could be used before and after a different form of training, and/or compared with an alternative valid measurement of competence.

Conclusion

The study used a novel abdominal simulator to both measure individual students' competence at palpatory exploration of the abdomen, and to train them using immediate and summative feedback to improve their competence.

On multiple measures, the participating students showed significant improvements. Additional studies with alternative designs and measurements could assess the robustness of these gains in skill.

Declarations

Funding. No external funding was received for the conduct of the study. The Department of Family and Preventive Medicine, and its Research Division, provided salary support for researchers and the secretaries who did data entry. ACDEC, Inc., provided the AbSim abdominal simulators.

Conflicts of interest/competing interests. Papa developed the AbSim simulator, provided it for the study without charge, and owns ACDEC, Inc., the company which sells it. He had no role in the data analysis. No other authors have any conflict of interest with respect to the area of study.

Ethics approval. All procedures performed in studies involving human participants were in accordance with the ethical standards as set forth by US federal regulations, Oklahoma state law, and University of Oklahoma policies and procedures. The study was approved by the Institutional Review Board of the University of Oklahoma Health Sciences Center (IRB #6997 and #9253).

Consent. Informed consent was obtained from all individual participants included in the study.

Data and Code availability. Data are available upon request from Robert Hamm, Robert-hamm@ouhsc.edu or robertmhamm@gmail.com.

Authors' contributions. Harris wrote some of the questions in the questionnaires and participated in preparation for preliminary studies. Medina and Hamm developed the procedure for assessing light and deep palpation. Hamm, Papa, and Kelley developed the study procedure and collaborated planning the study. Syed and Hamm conducted the individual study sessions with the students. Hamm did the statistical analysis. Kelley provided access to the students as clerkship director. Hamm and Kelley did the writing.

Consent for publication. All authors have reviewed the manuscript and approved it.

Acknowledgements. Jessica Brockhaus coordinated the researchers' access to the students to recruit them. Stacy Wigley and Dawn Kent entered data. ACDET, Inc. (Frank Papa, Steve Fogoros, Nick Papa, Mike Oglesby) provided the abdominal simulators, calibration, and maintenance.

References

1. Anders, H. L., Corrie, M., Jan, H., Cuno, R., Marianne, H., Kristian, M., & Per, A. (2010). Standardized simulated palpation training—development of a palpation trainer and assessment of palpatory skills in experienced and inexperienced clinicians. *Man Ther*, *15*(3), 254–260. doi:10.1016/j.math.2010.01.003
2. Bewley, W. L., & O'Neil, H. F. (2013). Evaluation of medical simulations. *Mil Med*, *178*(10 Suppl), 64–75. doi:10.7205/MILMED-D-13-00255
3. Cheng, A., Auerbach, M., Hunt, E. A., Chang, T. P., Pusic, M., Nadkarni, V., & Kessler, D. (2014). Designing and conducting simulation-based research. *Pediatrics*, *133*(6), 1091–1101. doi:10.1542/peds.2013-3267
4. Clark, R., & Mayer, R. (2008). Learning by viewing vs learning by doing: Evidence based guidelines for principled learning environments. *Performance Improvement*, *47*, 5–13.
5. Cook, D. A., Brydges, R., Hamstra, S. J., Zendejas, B., Szostek, J. H., Wang, A. T.,... . Hatala, R. (2012). Comparative effectiveness of technology-enhanced simulation versus other instructional methods: a systematic review and meta-analysis. *Simul Healthc*, *7*(5), 308–320. doi:10.1097/SIH.0b013e3182614f95
6. Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med*, *79*(10 Suppl), S70-81.
7. Ericsson, K. A. (2007). An expert-performance perspective of research on medical expertise: the study of clinical performance. *Med Educ*, *41*(12), 1124–1130. doi:10.1111/j.1365-2923.2007.02946.x
8. Ericsson, K. A. (2015). Acquisition and maintenance of medical expertise: a perspective from the expert-performance approach with deliberate practice. *Acad Med*, *90*(11), 1471–1486. doi:10.1097/ACM.0000000000000939

9. Ferguson, C. M. (1990). Inspection, auscultation, palpation, and percussion of the abdomen. In H. K. Walker, W. D. Hall, & J. W. Hurst (Eds.), *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd Edition ed., pp. 473–477). Boston, MA: Butterworths.
10. Hyde, L., Erolin, C., & Ker, J. (2012). Creation of abdominal palpation model prototype for training of medical students in detection and diagnosis of liver disease. *J Vis Commun Med, 35*(3), 104–114. doi:10.3109/17453054.2012.713855
11. Jones, F., Passos-Neto, C. E., & Braghiroli, O. F. M. (2015). Simulation in medical education: Brief history and methodology. *Principles and Practice of Clinical Research, 1*(2), 56–63.
12. Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating Training Programs: The Four Levels* (3 ed.). San Francisco: Berrett-Koehler.
13. Mahaboob, S., Lim, L. K., Ng, C. L., Ho, Q. Y., Leow, M. E., & Lim, E. C. (2010). Developing the "NUS Tummy Dummy", a low-cost simulator to teach medical students to perform the abdominal examination. *Ann Acad Med Singapore, 39*(2), 150–151.
14. McGaghie, W. C., Issenberg, S. B., Cohen, E. R., Barsuk, J. H., & Wayne, D. B. (2011). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med, 86*(6), 706–711. doi:10.1097/ACM.0b013e318217e119
15. McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2006). Effect of practice on standardised learning outcomes in simulation-based medical education. *Med Educ, 40*(8), 792–797. doi:10.1111/j.1365-2929.2006.02528.x
16. McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Med Educ, 44*(1), 50–63. doi:10.1111/j.1365-2923.2009.03547.x
17. McNulty, J. A., Halama, J., & Espiritu, B. (2004). Evaluation of computer-aided instruction in the medical gross anatomy curriculum. *Clin Anat, 17*(1), 73–78. doi:10.1002/ca.10188
18. Medford-Davis, L., Park, E., Shlamovitz, G., Suliburk, J., Meyer, A. N., & Singh, H. (2016). Diagnostic errors related to acute abdominal pain in the emergency department. *Emerg Med J, 33*(4), 253–259. doi:10.1136/emered-2015-204754
19. Munshi, F., Lababidi, H., & Alyousef, S. (2015). Low- versus high-fidelity simulations in teaching and assessing clinical skills. *Journal of Taibah University Medical Sciences, 10*(1), 12–15.
20. Okuda, Y., Bryson, E. O., DeMaria, S., Jr., Jacobson, L., Quinones, J., Shen, B., & Levine, A. I. (2009). The utility of simulation in medical education: what is the evidence? *Mt Sinai J Med, 76*(4), 330–343. doi:10.1002/msj.20127
21. Ribeiro, M. L., Lederman, H. M., Elias, S., & Nunes, F. L. S. (2016). Techniques and Devices Used in Palpation Simulation with Haptic Feedback. *Acm Computing Surveys, 49*(3). doi: 10.1145/2962723
22. Sachdeva, A. K., Wolfson, P. J., Blair, P. G., Gillum, D. R., Gracely, E. J., & Friedman, M. (1997). Impact of a standardized patient intervention to teach breast and abdominal examination skills to third-year

medical students at two institutions. *Am J Surg*, 173(4), 320–325. doi:10.1016/S0002-9610(96)00391-1

23. Zendejas, B., Brydges, R., Hamstra, S. J., & Cook, D. A. (2013). State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. *Ann Surg*, 257(4), 586–593. doi:10.1097/SLA.0b013e318288c40b