Teaching Empathy to Medical Students: An Updated, Systematic Review
Samantha A. Batt-Rawden, MBChB, Margaret S. Chisolm, MD, Blair Anton, and Tabor E. Flickinger, MD, MPH

Abstract

Purpose
Some research shows that empathy declines during medical school. The authors conducted an updated, systematic review of the literature on empathy-enhancing educational interventions in undergraduate medical education.

Method
The authors searched PubMed, EMBASE, PsycINFO, CINAHL, Scopus, and Web of Science (January 1, 2004 through March 19, 2012) using key words related to undergraduate medical education and empathy. They independently selected and reviewed all English-language articles that described an educational intervention designed to promote empathy in medical students, assessing the quality of the quantitative studies using the Medical Education Research Study Quality Instrument (MERSQI).

Results
The authors identified and reviewed the full texts of 18 articles (15 quantitative and 3 qualitative studies). Included interventions used one or more of the following—patient narrative and creative arts (n = 7), writing (n = 3), drama (n = 1), communication skills training (n = 4), problem-based learning (n = 1), interprofessional skills training (n = 1), patient interviews (n = 4), experiential learning (n = 2), and empathy-focused training (n = 1). Fifteen articles reported significant increases in empathy. Mean effect size was 0.23. Mean MERSQI score was 10.13 (range 6.5–14).

Conclusions
These findings suggest that educational interventions can be effective in maintaining and enhancing empathy in undergraduate medical students. In addition, they highlight the need for multicenter, randomized controlled trials, reporting long-term data to evaluate the longevity of intervention effects. Defining empathy remains problematic, and the authors call for conceptual clarity to aid future research.

A growing body of research has established the importance of empathy in several key aspects of medicine. Physician empathy leads to improved patient satisfaction,1,2 greater adherence to therapy,3,4 better clinical outcomes,4,7 and lower malpractice liability.9 Empathy, however, is difficult to define. One commonly accepted definition in patient care situations is “a cognitive attribute that involves an ability to understand the patient’s inner experiences and perspective and a capability to communicate this understanding.”9,9 Clinical empathy, as it relates to the patient–doctor relationship, is often divided into two dimensions. The first, affective, describes the passive emotional response of one individual to the emotions of another.9,10 The second, cognitive, is an active skill that may be acquired and is amenable to nurturing. This aspect of empathy has been described as “detached concern,” or the ability of one individual to understand the experiences of another without invoking a personal emotional response.10

Two major systematic reviews of studies of empathy measures in medicine were recently published.11,12 These reviews highlight the variety of instruments available to measure empathy and present evidence regarding the reliability and validity of these instruments. Examples from these reviews of reliable and valid self-report questionnaires (completed by trainees and/or medical professionals) include the Jefferson Scale of Physician Empathy (JSPE),9,13,14 the Questionnaire Measure of Emotional Empathy,15 from which the Balanced Emotional Empathy Scale (BEES)16 was developed, the Interpersonal Reactivity Index,17 and the Empathy Construct Rating Scale (ECRS),18 which is used primarily in nursing research. Examples from these reviews of reliable and valid observed measures of empathy include the Consultation and Relational Empathy (CARE)19 measure and the Four Habits Coding Scheme.20,21 Although correlations between self-reported and observed empathy have been shown, disagreement remains regarding the validity of self-report questionnaires as an accurate measure of empathy outcomes.22

Researchers also disagree about the outcomes of tracking trends of empathy among students and professionals. Another recent systematic review suggested that a significant decline in empathy occurs during medical school and residency.23 This decline in empathy is generally viewed as a valid finding,24–26 although it has not been without challenge.27 Researchers agree that empathy levels vary according to certain characteristics of trainees, such as female gender,28–30 younger age,28 and specialty choice.14,30,32,33 Researchers also agree that training, particularly exposure to patients and the clinical environment, can affect empathy.24,35 The following characteristics specifically can lead to...
empathy decline: psychological factors, the “hidden curriculum,” unsuitable learning environments, cynicism/loss of idealism, and the perceived need for detachment.33–47

The aim of this review was to update and systematically expand the most recently published review of the literature on empathy-enhancing educational interventions for undergraduate medical education (by Stepien and Baerinstein39). This review sought to answer the following three questions: (1) What new interventions have been implemented (since 2003) to promote empathy in medical students? (2) How has the effectiveness of these interventions been measured? and (3) What is the quality of evidence for these interventions?

Method

Literature search and eligibility criteria

We created search strategies for six bibliographic electronic databases (PubMed, EMBASE, PsychINFO, CINAHL, Scopus, and Web of Science) to capture English-language, peer-reviewed literature published between January 1, 2004 and March 19, 2012. We defined the concepts of undergraduate medical education and empathy using a combination of controlled vocabulary terms applicable to each database and key word terms and phrases. We constructed our base search strategy in PubMed and then translated it for the other databases. One author (B.A.) ran all searches on March 19, 2012, imported citations into a citation management system, and removed duplicates.

We ensured that our search strategies captured the previously published review46 and all relevant studies included in that review. We applied a date limit to focus our review on studies published since 2003.

For our review, we included studies published in English and in peer-reviewed journals, describing an intervention to promote empathy among medical students. We excluded articles if they did not meet these criteria or if the full text was unavailable. To identify additional studies, we conducted a hand search of the reference lists of those studies that we included in full-text review.

Article selection

Two authors (M.S.C. and T.E.F.) identified the relevant articles for full-text review by reviewing the titles and abstracts and reaching a consensus together.

Data extraction and synthesis

Two authors (M.S.C. and S.A.B.R.) independently reviewed the full texts of the relevant articles, using a predetermined data extraction form developed for this review. Data extraction fields included authors, year of publication, study design, participants, intervention, outcome measures, and key findings. Discussion with a third author (T.E.F.) resolved any differences in data extraction. We calculated effect sizes, where possible, using the Cohen d measure.

Quality assessment

Two of the authors (two of S.A.B.R., M.S.C., and T.E.F.) independently assessed the quality of the included quantitative articles using the Medical Education Research Study Quality Instrument (MERSQI), a tool specifically developed to evaluate educational studies.49 This 10-item scale assesses the domains of study design, sampling, type of data, validity of the evaluation instrument, data analysis, and outcomes. Intraclass correlation coefficients for the MERSQI have been reported at 0.72 to 0.98 for interrater and 0.78 to 0.99 for intrarater reliability. Criterion validity has been assessed by expert quality rating, citation rates, and publication impact factors.49,50

Results

Our initial search identified 3,212 titles, from which we selected 1,486 for abstract review and 54 for full-text review (see Figure 1). We selected 18 articles for final review. Appendix 1 summarizes the 15 quantitative studies we reviewed, and Appendix 2 summarizes the 3 qualitative studies.

Study characteristics

Of the 18 articles we selected for final review, 7 were single-group pre–post comparisons, 3 single-group posttest evaluations only, 4 nonrandomized two-group studies, and 3 modified cohort controlled studies. Only 1 was a fully randomized controlled trial.

Empathy measures

Of the 15 quantitative studies, 12 used validated outcome measures in various combinations. Nine studies employed self-report questionnaires—7 studies used the JSPE,33,34,41 1 study used the Empathy Tendency Scale (ETS)31,52 and the Empathic Skill Scale (ESS),31,52 and 1 study used the BEES46 and the ECRS.18 Two studies employed observed measures—1 used CARE53 from the point of view of first-person patient assessors, and another used the Staff–Patient Interaction Rating Scale (SPIR)53 from the point of view of third-person assessors. The remaining 4 studies used nonvalidated, self-report measures developed by the study investigators.

Study quality

We found the mean MERSQI score (possible range 5–15.5) for the 15 included quantitative studies to be 10.13. The lowest score was 6.5, and the highest was 14. We found scores to be limited by common methodological flaws—lack of a control group, nonrandomized design, conducted at a single institution, lack of preintervention or baseline measurement, and measurement of attitudes rather than skills or patient outcomes.

Types of interventions

Patient narrative and creative arts interventions. Seven studies reported on interventions based around the patient narrative and the creative arts, including creative writing, blogging, drama, poetry, fiction, and film. Such interventions fit primarily into the affective domain of empathy. Of these studies, two reported on small-group sessions and facilitated discussions of poetry54 and/or short stories55 concerning the doctor–patient relationship, allowing a student to appreciate the patient’s point of view. Muszkat and colleagues4 reported a significant increase in empathy scores using their author-developed survey. Intriguingly, Shapiro and colleagues55 reported a significant increase in scores on the BEES but not on the ECRS. In response, the authors proposed that the ECRS items measure the cognitive domain of empathy and the BEES items measure the affective domain.

Writing interventions. Three studies used varying methods of writing to enhance empathy. DasGupta and Charon56

1172

Academic Medicine, Vol. 88, No. 8 / August 2013
required students to compose reflective essays on personal experiences of illness, suggesting that “explicit awareness of … [one’s] own feelings and experiences deepens the capacity to respond empathically.” Shapiro and colleagues described an intervention in which students were trained in point-of-view writing and subsequently required to compose essays from the patient’s point of view. The authors hypothesized that this practice might foster an understanding of the “patient’s emotional and social perspective about his or her illness and its consequences.” Both studies reported a significant increase in the empathic content of these essays.

In the third such study, Rosenthal and colleagues required each student to post at least one blog entry per clerkship during the third year. These entries then were used as triggers for discussion. The authors reported no significant change in empathy, but they argued that such a finding in itself was positive, as one would expect empathy to decline in the third year. The authors also performed a subgroup analysis on students selected from their peers for the Gold Humanism Honor Society (GHHS). Interestingly, although GHHS students had significantly higher posttest empathy scores than their classmates, their scores did not differ pretest. The authors posited that students’ awareness of their GHHS selection could have provided positive reinforcement of their tendency toward humanistic values and behaviors, which had been perceived by their peers. In addition, this knowledge may have encouraged them to develop these qualities further during their clerkship experiences. In the subsequent cohort, students were not informed of GHHS selection until after the administration of the posttest JSPE. The authors found that these students scored significantly lower than the baseline on the posttest. However, their scores rose significantly, returning to the baseline, after they were informed of their GHHS selection. Therefore, such programs, which validate humanistic behavior, may contribute to medical students’ maintenance of empathy.

Drama interventions. Lim and colleagues used drama to teach empathy, coaching students “how to act-in-role.” Akin to methods employed by communication workshops targeting the cognitive domain, the exercises in this study focused on building participants’ acting skills to enhance their ability to portray empathy and were successful in significantly increasing empathy. Van Winkle and colleagues also used drama to enhance empathy, which included a student portraying the challenges of aging, followed by small-group discussions. These authors reported significant increases in empathy on the posttest immediately after the workshop, but the improvement was not sustained on the 7-day or 26-day posttests.

Communication skills training interventions. Four studies used communication skills training as an intervention, reflecting the authors’ preference for the cognitive definition of empathy. This component of empathy traditionally has been targeted for modification because it can be considered a skill, whereas affective empathy is regarded as a personal trait, which lies beyond the scope of teaching. Bombeke and colleagues implemented a communication skills training consisting of small-group interactive teaching and role-play. Although differences between cohorts were not significant, the authors identified a trend that JSPE scores had increased in the intervention group but had declined in the control group. Fernández-Olano and colleagues described an intervention in which students participated in communication skills workshops, which included role-playing, coaching on formulating empathic phrases, and conveying empathy verbally and nonverbally. The authors found a significant postintervention increase in empathy, which suggests that empathy is a skill that can be modified by educational strategies.
Interpersonal skills training intervention. Turiniami and colleagues developed an interpersonal skills training course that included a lecture followed by role-playing; these simulations were videotaped and subsequently analyzed. The authors primarily investigated the impact of the training course on participants’ communication skills, but they also included an “empathy and reflection” item on their survey. Participants’ scores on this item increased significantly post intervention.

Patient interview interventions. Four studies reported on interventions that included interviewing a patient for his or her perspective. Both Mullen and colleagues and Yuen and colleagues reported positive posttest outcomes following students’ visits to chronically ill patients in their own homes. Kommalage, on the other hand, showed students videos of interviews in which patients and relatives described their hematological diseases and the resultant socioeconomic problems. Pre–post intervention increases in empathy were significant only for female participants. Shapiro and colleagues offered first-year medical students the opportunity to meet with patients one-to-one for four months. Although the authors found that this practice increased empathy scores on the SPIR, students’ confidence in their communication skills decreased.

Experiential learning interventions. Two studies simulated a patient experience, and both reported a significant increase in empathy. Both sets of authors hypothesized that students might develop empathy toward certain patient groups after personally experiencing their inherent challenges and symptoms. In a randomized controlled trial, Bunn and Terpstra exposed participants to a 40-minute simulated auditory hallucination via individual headphones. They asked participants to follow written directions, complete a Mini-Mental State Examination, and attempt to interact with peers while experiencing the simulation. Varkey and colleagues conducted an “aging game,” simulating the loss of vision, hearing, manual dexterity, mobility, continence, independence, and dignity that occurs with advancing age.

Empathy intervention. One study developed specific curricula targeted exclusively toward enhancing empathy. Bayne developed a program consisting of didactic and experiential content, including communication skills training and role-play, in an attempt to address both domains of empathy. Facilitators acknowledged the external characteristics of the decline of empathy, working with students to develop strategies to overcome perceived barriers to empathy in practice. This intervention had the highest effect size in our review. It uniquely employed the CARE measure as a patient-reported outcome measure, which suggests that participants were able to translate the empathy they cultivated during this intervention into practice.

Discussion

Our findings suggest that educational interventions can successfully cultivate empathy in undergraduate medical students and that such interventions are well received by participants. Many of the studies we included corroborated findings from the broader literature—that empathy declines during medical school and is associated with gender, ethnicity, and specialty choice. In addition, several studies reported that higher baseline empathy scores were associated with greater pre–post differences in empathy, suggesting that certain interventions should target certain groups. Further research, then, might focus on the specific attributes that predispose a student to being more susceptible to different interventions.

We also found that the included studies were limited by common methodological flaws, including lack of control groups, small sample sizes, single institutions, lack of preintervention or baseline measurements, and lack of long-term follow-up. Our findings highlight the need for multicenter, randomized controlled trials, reporting long-term data to evaluate the longevity of intervention effects.

Still, an operational definition of empathy remains elusive, and only three articles included in our review addressed this issue, clarifying which domains of empathy the authors were targeting with their interventions. The problematic nature of defining empathy is evident in two studies that reported interventions that failed to translate increased empathy scores into practice, using objective structured clinical examinations with standardized patients to test participants’ empathy in practice. In addition, the studies included in our review produced conflicting results pertaining to participants’ self-confidence in their communications skills post intervention, with two studies reporting a decrease in self-confidence and two reporting an increase. For example, posited that “unprofessional students can ‘fake’ professional behavior” in exam situations. In addition, the studies included in our review produced conflicting results pertaining to participants’ self-confidence in their communications skills post intervention, with two studies reporting a decrease in self-confidence and two reporting an increase. For example, posited that “unprofessional students can ‘fake’ professional behavior” in exam situations. The crux of the matter, however, is captured by Shapiro and colleagues, who postulated that empathic skills developed through point-of-view writing may not translate into behavior, a theory reiterated by many authors when discussing their respective interventions. This paradox raises two questions. The first question is how useful current validated, self-report questionnaires are in predicting perceived empathy in practice. To investigate this question further, one might consider employing the use of second-person and third-person assessments in addition to self-report questionnaires. The second question is whether domains of empathy as influenced by such teaching methods are valued by assessors in examination situations and, more important, by patients. Further research must inform medical educators of the clinical impact of educational interventions to increase empathy.

Conclusions

The findings of our systematic review suggest that educational interventions can be successful in maintaining and
enhancing empathy in undergraduate medical students. Although continuing to implement such strategies will further clarify best practices, more rigorous research, especially large, properly controlled longitudinal studies, is needed to inform recommendations for medical education. Moving forward, educational scholars and researchers should consider addressing the widely reported characteristics of the decline in empathy, including psychological factors such as stress and fatigue, the “hidden curriculum,” unstable learning environments, loss of idealism, and the perceived need for detachment. In his study, Bayne highlighted the importance of role models and the reciprocal nature of empathy development in training, suggesting that “[j]indeed, perhaps students need to receive more empathy from faculty, other physicians, and even their patients before they can truly understand how to establish empathic connections.” Thus, educators should consider using the practice of relationship-centered care as the fundamental building block for their educational interventions to teach empathy.

Funding/Support: Dr. Batt-Rawden was awarded the Junior Association for the Study of Medical Education (JASME) medical elective bursary in 2012, the Gilchrist Educational Grant for Academically Outstanding Individuals, and the Doubleday Fund award, and would like to thank these organizations which supported the elective at Johns Hopkins University where this research was conducted.

Other disclosures: None.

Ethical approval: Not applicable.

References
### Appendix 1

**Characteristics of 15 Quantitative Studies of Empathy-Enhancing Interventions, Identified in a Systematic Review of the Literature Published Between January 1, 2004 and March 19, 2012**

<table>
<thead>
<tr>
<th>First author, year</th>
<th>Design and participants</th>
<th>Intervention</th>
<th>Outcome measure</th>
<th>Significant increase in empathy reported?</th>
<th>Effect size (Cohen d)</th>
<th>MERSQI score (5–15.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Winkle, 2012</td>
<td>Pre–post comparison of 187 preclinical pharmacy and 183 medical students</td>
<td>Workshop with performance and discussion</td>
<td>JSE (HPS/S-version)</td>
<td>Yes</td>
<td>0.19</td>
<td>11.5</td>
</tr>
<tr>
<td>Bayne, 2011</td>
<td>Pre–post comparison of 22 clinical students</td>
<td>Didactic and role-play workshops during six-week clerkship</td>
<td>CARE</td>
<td>Yes</td>
<td>0.45</td>
<td>12</td>
</tr>
<tr>
<td>Bombeke, 2011</td>
<td>Nonrandomized two-group cohort</td>
<td>Communication skills training across five-year curriculum, four-hour small-group sessions including role-play</td>
<td>JSPE</td>
<td>No</td>
<td>N/A</td>
<td>11</td>
</tr>
<tr>
<td>Karaoglu, 2011</td>
<td>Pretest–posttest comparison of 137 preclinical students and 66 medical residents</td>
<td>Three problem-based learning discussions around scenarios from different character viewpoints, one focusing on communication and empathy</td>
<td>ETS, ESS</td>
<td>No</td>
<td>N/A</td>
<td>10.5</td>
</tr>
</tbody>
</table>

(Appendix Continues)
### Appendix 1, Continued

<table>
<thead>
<tr>
<th>First author, year</th>
<th>Design and participants</th>
<th>Intervention</th>
<th>Outcome measure</th>
<th>Significant increase in empathy reported?</th>
<th>Effect size (Cohen d)</th>
<th>MERSQI score (5–15.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kommalage, 2011&lt;sup&gt;69&lt;/sup&gt;</td>
<td>Pre–post comparison of 144 preclinical students</td>
<td>Two videos on patient perspective during hematology modules</td>
<td>JSPE</td>
<td>Yes</td>
<td>0.14</td>
<td>9.5</td>
</tr>
<tr>
<td>Lim, 2011&lt;sup&gt;69&lt;/sup&gt;</td>
<td>Nonrandomized two-group pre–post of 149 clinical students</td>
<td>Module including drama, role-play, motivational interviewing</td>
<td>JSPE</td>
<td>Yes</td>
<td>Not enough data to calculate</td>
<td>10</td>
</tr>
<tr>
<td>Rosenthal, 2011&lt;sup&gt;69&lt;/sup&gt;</td>
<td>Nonrandomized two-group, pre–post comparison of 209 clinical students</td>
<td>Component of curriculum including blogging, discussing articles, film, and fiction</td>
<td>JSPE-S</td>
<td>Yes</td>
<td>0.08</td>
<td>11</td>
</tr>
<tr>
<td>Tiuraniemi, 2011&lt;sup&gt;69&lt;/sup&gt;</td>
<td>Pre–post comparison of 126 clinical students and 183 psychology students</td>
<td>Interpersonal skills training lecture and small-group simulations and role-play</td>
<td>Author's own</td>
<td>Yes</td>
<td>0.30</td>
<td>8</td>
</tr>
<tr>
<td>Mullen, 2010&lt;sup&gt;57&lt;/sup&gt;</td>
<td>Posttest of 240 preclinical students</td>
<td>Interviews with chronically ill patient at home and his or her caregiver</td>
<td>Author's own</td>
<td>Yes</td>
<td>Not enough data to calculate</td>
<td>7</td>
</tr>
<tr>
<td>Muszkat, 2010&lt;sup&gt;54&lt;/sup&gt;</td>
<td>Posttest of 44 clinical students</td>
<td>Five 1-hour meetings involving discussion of poetry</td>
<td>Author's own</td>
<td>Yes</td>
<td>Not enough data to calculate</td>
<td>6</td>
</tr>
<tr>
<td>Bunn, 2009&lt;sup&gt;71&lt;/sup&gt;</td>
<td>Randomized controlled trial of 150 clinical students</td>
<td>40-minute auditory hallucination simulation</td>
<td>JSPE</td>
<td>Yes</td>
<td>Not enough data to calculate</td>
<td>14</td>
</tr>
<tr>
<td>Shapiro, 2009&lt;sup&gt;55&lt;/sup&gt;</td>
<td>Modified cohort controlled study of 79 preclinical students</td>
<td>Weekly group meetings and exploratory discussion with patients</td>
<td>SPIR</td>
<td>Yes</td>
<td>0.14</td>
<td>12.5</td>
</tr>
<tr>
<td>Fernández-Olano, 2008&lt;sup&gt;64&lt;/sup&gt;</td>
<td>Pre–post comparison of 127 preclinical students and 66 medical residents</td>
<td>25-hour workshop on communication skills and expression of empathy</td>
<td>JSPE</td>
<td>Yes</td>
<td>0.23</td>
<td>10.5</td>
</tr>
<tr>
<td>Varkey, 2006&lt;sup&gt;63&lt;/sup&gt;</td>
<td>Pre–post comparison of 84 preclinical students</td>
<td>Three-hour practical simulation of aging</td>
<td>Author's own</td>
<td>Yes</td>
<td>0.24</td>
<td>6.5</td>
</tr>
<tr>
<td>Shapiro, 2004&lt;sup&gt;55&lt;/sup&gt;</td>
<td>Modified cohort controlled study of 22 preclinical students</td>
<td>Eight 1-hour small-group sessions on fiction and poetry addressing the doctor–patient relationship</td>
<td>BEES, ECRS</td>
<td>Yes, No</td>
<td>0.30, N/A</td>
<td>12</td>
</tr>
</tbody>
</table>

*MERSQI indicates Medical Education Research Study Quality Instrument; JSPE, Jefferson Scale of Physician Empathy; JSPE-S, Jefferson Scale of Physician Empathy, Student Version; CARE, Consultation and Relational Empathy Scale; ETS, Empathy Tendency Scale; ESS, Empathic Skill Scale; SPIR, Staff–Patient Interaction Rating scale; BEES, Balanced Emotional Empathy Scale; and ECRS, Empathy Construct Rating Scale.

### Appendix 2

**Characteristics of Three Qualitative Studies of Empathy-Enhancing Interventions, Identified in a Systematic Review of the Literature Published Between January 1, 2004 and March 19, 2012**

<table>
<thead>
<tr>
<th>First author, year</th>
<th>Design and participants</th>
<th>Intervention</th>
<th>Assessment technique</th>
<th>Increase in empathy reported?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yuen, 2006&lt;sup&gt;68&lt;/sup&gt;</td>
<td>Nonrandomized, two-group pre–post of 50 clinical and former clinical students</td>
<td>Visiting chronically ill patients at home</td>
<td>Interviews</td>
<td>Yes</td>
</tr>
<tr>
<td>Shapiro, 2006&lt;sup&gt;67&lt;/sup&gt;</td>
<td>Cohort controlled study (randomized at group level) of 92 preclinical students</td>
<td>Small-group sessions on poetry and narratives; participants required to write essays from patient's point of view</td>
<td>Analysis of essays</td>
<td>Yes</td>
</tr>
<tr>
<td>DasGupta, 2004&lt;sup&gt;64&lt;/sup&gt;</td>
<td>Posttest of 11 preclinical students</td>
<td>Six-week reflective writing seminar</td>
<td>Open-ended student questionnaire</td>
<td>Yes</td>
</tr>
</tbody>
</table>