Residents-as-Teachers Curricula: A Critical Review
Robert E. Post, MD, R. Glen Quattlebaum, MD, MPH, and Joseph J. Benich, III, MD

Abstract

Purpose
Residents serve as medical students’ primary teachers for practical clinical skills. The purpose of this study is to provide an updated systematic review of the literature on residents-as-teachers curricula to determine the most evidence-based curricula and evaluation strategy.

Method
In 2008, the authors performed a systematic review of the literature with PubMed using the MESH terms “internship and residency” and “teaching,” as well as a key word search of the term “residents as teachers.” The search was limited to publications in English from 1975 to 2008.

Results
A total of 24 studies met inclusion criteria. Eleven (45.8%) were uncontrolled studies, seven (29.2%) were randomized control trials, and six (25%) were nonrandomized controlled trials. The mean sample size of all studies was 39.6. Evaluation was performed by a variety of means including objective structured teaching exams (5; 20.8%), videotape evaluations (6; 25.0%), learner evaluations (11; 45.8%), and self-questionnaires (7; 29.2%). The mean intervention length was 7.6 hours, and the most common intervention was based on the One-Minute Preceptor.

Conclusions
Research on residents-as-teachers curricula is limited by both the number of studies and their methodology. Despite this, the results demonstrate that residents-as-teachers curricula can significantly improve residents’ teaching skills. In addition, the studies’ methodologies have improved over time. Using these data, the authors recommend an evidence-based intervention and evaluation, which would include a three-hours-or-longer intervention (and, if possible, periodic reinforcement) based on the One-Minute Preceptor. The evaluation should be a randomized controlled trial using an objective structured teaching examination.


Residents are a crucial link in teaching future generations of physicians. Previous research shows that residents serve as medical students’ primary teachers for practical clinical skills, that one third of a medical student’s fund of knowledge was directly attributable to residents, and that 20% of a resident’s time was spent on teaching activities—a higher percentage than that of any other teaching faculty.

Residents also benefit from teaching medical students. Weiss and Needman showed that teaching leads to better knowledge acquisition for the teacher than do self-study or lecture attendance. Other research reveals that residents’ job satisfaction is augmented by teaching duties.

Despite their critical role as teachers, until recently few residency programs provided formal education in teaching methods. However, Morrison et al. showed that, by 2000, 55% of U.S. residencies provided instruction on teaching. But of those that provided such instruction, they provided it for only an average of 11 hours per resident for their entire residency.

Although several books have been written on the subject, only a 2004 review has systematically reviewed the medical research to determine whether residents-as-teachers programs were effective. This review found only 13 studies of various experimental designs, most with methodological limitations. While describing the various training methods for residents, the authors of the 2004 review did not recommend the most optimal teaching strategy. We carried out the present study to provide an updated review of the literature on residents-as-teachers curricula. Also, we hoped that our findings would allow us to formulate recommendations about the most evidence-based curricula and evaluation strategy, to assist residency faculty in assessing the benefits of each residents-as-teachers curriculum so they can adopt one best suited for their goals.

Method
In 2008, we performed a systematic review of the literature with PubMed using the MESH terms “internship and residency” and “teaching,” as well as a key word search of the term “residents as teachers.” The search was limited to publications in English from 1975 to 2008.

We evaluated each citation found in both searches and marked 47 articles for further evaluation. The abstracts of those articles were reviewed for relevance to the subject of residents-as-teachers, study design, and outcome measures. Articles that were deemed pertinent were those that included an intervention that attempted to improve residents’ teaching skills. The categories for study design were randomized controlled studies, nonrandomized controlled studies, and uncontrolled studies. Studies that reported outcomes using self-report methods, learner evaluations, or direct observation methods such as videotaped evaluations and objective structured teaching examinations (OSTEs) were included in the analysis. An OSTE is a

Dr. Post is a resident, Department of Family Medicine, Medical University of South Carolina, Charleston, South Carolina.

Dr. Quattlebaum is a resident, Department of Family Medicine, Medical University of South Carolina, Charleston, South Carolina.

Dr. Benich is a resident, Department of Family Medicine, Medical University of South Carolina, Charleston, South Carolina.

Correspondence should be addressed to Dr. Post, 190 Calhoun St., Charleston, SC 29425; telephone: (843) 792-2383; fax: (843) 792-3598; e-mail: (post@must.edu).
direct observation of a research participant's teaching ability as judged by predetermined, observable criteria. Studies pertaining to training fellows or faculty members were excluded. Descriptive studies and pilot studies of later published works were also excluded. In total, 24 studies met the criteria. The references of these 24 articles were evaluated to obtain further studies; however, none that met the inclusion criteria were discovered.

Each of us independently reviewed the 24 articles to determine the number of participants, a description of the intervention, teaching methods, outcome measures, method of measurement, randomization, and the presence of a control group. The mean and the median of the sample sizes and intervention lengths were calculated. We compared and agreed on our interpretations of the studies. We organized these studies from the perspective of the program evaluators—first by outcome measure and then by study type.

Results
Descriptive statistics
Overall, 24 studies met our criteria (see Table 1). The majority of these (11; 45.8%) were uncontrolled studies, followed by randomized controlled trials (7; 29.2%) and nonrandomized controlled trials (6; 25.0%). Residents were studied from a variety of disciplines: internal medicine (42.9% of studies), obstetrics–gynecology (28.6%), pediatrics (23.8%), surgery (19.0%), and family medicine (14.3%). Individuals in the studies were evaluated by a variety of means including direct observation methods (OSTEs) (5; 20.8%), videotape evaluations (6; 25.0%), learner evaluations (11; 45.8%), and self-questionnaires (7; 29.2%). Five studies (20.8%) used a combination of outcome measures.

The mean sample size of all studies was 39.6, but sizes ranged from 6 to 145 participants. The median sample size was 25. Uncontrolled studies had the lowest number of participants (mean 28.3), and two did not report populations. Nonrandomized controlled studies had the highest mean number of participants (52.4), and randomized controlled studies had a mean of 45.0 participants.

The interventions were based on various theoretical models, including the One-Minute Preceptor (also referred to as the Five Microskills of Clinical Teaching), the Teaching Improvement Project Systems, problem-based learning modules, and unspecified workshops and retreats. The One-Minute Preceptor was used by four of the studies. The mean length of time for the interventions was 7.6 hours, with a range of 1 to 15 hours. The median intervention length was eight hours. Studies that either did not report intervention length at all or reported length in days or weeks without including hours per day or hours per week were not included in these calculations.

In looking at the studies performed on residents' teaching skills, one key aspect was the method by which residents' teaching skill improvements were assessed. The techniques used for these evaluations varied significantly in their objectivity and rigor. We present the findings in the following paragraphs, from the most rigorous evaluation methods (direct evaluation by OSTEs and videotape) to the least rigorous (learner evaluations and resident self-evaluations).

Direct evaluation
Five studies used direct evaluation by OSTEs as the preferred method of evaluation. Two of these were randomized controlled studies. Morrison et al.11 evaluated 62 second-year internal medicine residents (33 intervention, 29 control). The intervention group participated in a 13-hour teaching curriculum. Residents were evaluated before and after the intervention with an OSTE. OSTE scores of the experimental group after the intervention were significantly higher ($P < .001$) than the scores of the controls. Dunnington and DaRosa12 evaluated 62 residents (intervention versus controls not specified) at two surgery training programs. The intervention group participated in a two-day, 10.5-hour teaching skills course. Reminders of course content were given to participants at two and four weeks after completion of the course. All participants were evaluated with an OSTE six to seven months later. Results were mixed, as some skills showed a significant difference between the two groups and some skills did not, with no consistency across the two residency programs.

Only one nonrandomized controlled trial used an OSTE. Gaba et al.13 evaluated 24 obstetrics–gynecology residents (13 intervention, 11 control). The intervention residents participated in a 10.5-hour teaching workshop and then were evaluated with an OSTE. Overall, the intervention group scored significantly higher than the control group ($P = .001$) on the OSTE.

Two uncontrolled studies used OSTEs. Zabar et al.14 used an OSTE at their program as the tool to evaluate residents. Sixty-five residents from the first-, second-, and third-year classes were compared. The results showed the highest scores in overall teaching performance and communication skills were in the third-year class, followed by the second-year class, and then the first-year class. The third-year class's residents rated significantly higher ($P = .05$) than the interns. Residents were given feedback after taking the OSTE each year, and improvement was demonstrated with experience. However, we could not attribute the improvement simply to improved teaching skills. Some of the improvement could have been from an evolution of knowledge and professionalism as well. White et al.15 studied 21 first-, second-, and third-year pediatrics residents. The intervention was a 3.5-hour teaching workshop. Teaching encounters were observed by faculty members in an OSTE before and after the intervention, and residents improved in all teaching skills that were taught in the workshop (no $P$ values reported).

Videotape evaluations
Overall, six studies used evaluations of videotaped teaching sessions as the evaluation method, two randomized controlled trials and four uncontrolled studies.

In the first randomized study, D'eon16 examined 16 residents (8 intervention, 8 control) in various medicine and surgery specialties. The intervention was a two-day teaching workshop. All participants were videotaped giving a presentation before and after the workshop. The videotapes were analyzed for various teaching skills by two undergraduate education students. The intervention group was rated significantly higher after the course in two key teaching areas: the opening ($P < .05$) and the use of instructional objectives ($P = .1$).
### Table 1
Residents-as-Teachers Studies that Met the Authors’ Criteria, 2008*

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Site</th>
<th>Participants</th>
<th>Specialty</th>
<th>Length</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morrison et al</td>
<td>2004</td>
<td>RC, DO</td>
<td>62</td>
<td>Internal medicine</td>
<td>13 hours</td>
<td>Yes (P &lt; .01)</td>
</tr>
<tr>
<td>Dunnignton et al</td>
<td>1998</td>
<td>RC, DO</td>
<td>62</td>
<td>Surgery</td>
<td>10.5 hours</td>
<td>No</td>
</tr>
<tr>
<td>Gaba et al</td>
<td>2007</td>
<td>NC, SA</td>
<td>24</td>
<td>Obstetrics–gynecology</td>
<td>10.5 hours</td>
<td>Yes (P = .001)</td>
</tr>
<tr>
<td>Zabro et al</td>
<td>2004</td>
<td>UC, DO</td>
<td>65</td>
<td>Obstetrics–gynecology</td>
<td>10.5 hours</td>
<td>No (P no value reported)</td>
</tr>
<tr>
<td>White et al</td>
<td>1997</td>
<td>UC, DO</td>
<td>21</td>
<td>Pediatrics</td>
<td>3.5 hours</td>
<td>Yes (P no value reported)</td>
</tr>
<tr>
<td>D’eon et al</td>
<td>2004</td>
<td>RC, VE</td>
<td>16</td>
<td>Surgery</td>
<td>2 days (hours unspecified)</td>
<td>Yes (P = .1)</td>
</tr>
<tr>
<td>Edwards et al</td>
<td>1988</td>
<td>RC, VE</td>
<td>22</td>
<td>Obstetrics–gynecology/family medicine/ Internal medicine</td>
<td>Unspecified</td>
<td>Yes (P &lt; .01), but not sustained at six months</td>
</tr>
<tr>
<td>Barth et al</td>
<td>1997</td>
<td>UC, VE</td>
<td>6</td>
<td>Surgery</td>
<td>Unspecified</td>
<td>Yes (P = .002)</td>
</tr>
<tr>
<td>Roberts et al</td>
<td>1994</td>
<td>UC, VE</td>
<td>Unspecified</td>
<td>Pediatrics</td>
<td>8 hours</td>
<td>Yes (P no value reported)</td>
</tr>
<tr>
<td>Bing-You et al</td>
<td>1990</td>
<td>UC, VE</td>
<td>26</td>
<td>Internal medicine</td>
<td>8 hours</td>
<td>Yes (P = .05)</td>
</tr>
<tr>
<td>Lawson et al</td>
<td>1980</td>
<td>UC, VE</td>
<td>20</td>
<td>Family medicine/ Internal medicine</td>
<td>13 hours</td>
<td>Yes (P &lt; .001)</td>
</tr>
<tr>
<td>Furrer et al</td>
<td>2001</td>
<td>NC, LE</td>
<td>57</td>
<td>Internal medicine</td>
<td>1 hour</td>
<td>No</td>
</tr>
<tr>
<td>Pavone et al</td>
<td>1998</td>
<td>NC, LE</td>
<td>57</td>
<td>Internal medicine</td>
<td>1 hour</td>
<td>Yes (P &lt; .05)</td>
</tr>
<tr>
<td>Batchett et al</td>
<td>1982</td>
<td>UC, LE</td>
<td>53</td>
<td>Pediatrics</td>
<td>6 hours</td>
<td>Yes (P &lt; .01)</td>
</tr>
<tr>
<td>Busch et al</td>
<td>2006</td>
<td>NC, LE</td>
<td>27</td>
<td>Pediatrics and obstetrics–gynecology</td>
<td>Unspecified</td>
<td>Yes (P = .02)</td>
</tr>
<tr>
<td>Hammod et al</td>
<td>2004</td>
<td>NC, LE</td>
<td>Unspecified</td>
<td>Obstetrics–gynecology</td>
<td>1 day (hours unspecified)</td>
<td>Yes (P = .05)</td>
</tr>
<tr>
<td>Synnott et al</td>
<td>1996</td>
<td>NC, LE</td>
<td>44</td>
<td>Internal medicine</td>
<td>3 hours</td>
<td>Yes (P &lt; .03)</td>
</tr>
<tr>
<td>Edwards et al</td>
<td>1988</td>
<td>NC, LE</td>
<td>145</td>
<td>13 different specialties</td>
<td>1/2 day (hours unspecified)</td>
<td>Yes (P &lt; .05)</td>
</tr>
<tr>
<td>Jafari et al</td>
<td>2007</td>
<td>UC, LE</td>
<td>10</td>
<td>Gastroenterology</td>
<td>Unspecified</td>
<td>Faculty more than residents (P &lt; .05)</td>
</tr>
<tr>
<td>Fraterreri and Kasuda</td>
<td>2003</td>
<td>UC, LE</td>
<td>17</td>
<td>Obstetrics–gynecology</td>
<td>4.5 hours</td>
<td>Yes (P &lt; .05)</td>
</tr>
<tr>
<td>Wipf et al</td>
<td>1999</td>
<td>UC, LE</td>
<td>17</td>
<td>Obstetrics–gynecology</td>
<td>4.5 hours</td>
<td>Yes (P &lt; .05)</td>
</tr>
<tr>
<td>Litteman et al</td>
<td>1994</td>
<td>UC, LE</td>
<td>72</td>
<td>Internal medicine</td>
<td>6 hours</td>
<td>Yes (P &lt; .001)</td>
</tr>
<tr>
<td>Edwards et al</td>
<td>1986</td>
<td>UC, SA</td>
<td>18</td>
<td>Unspecified</td>
<td>Unspecified</td>
<td>Yes (P &lt; .001)</td>
</tr>
</tbody>
</table>

*The authors performed a systematic review of the literature with PubMed using the MESH terms “Internship and residency” and “teaching,” as well as a keyword search of the term “residents as teachers.” The search was limited to publications in English from 1972 to 2008. The authors evaluated each citation found in both searches, and marked 47 articles for further evaluation. The abstracts of those articles were reviewed for pertinence to the subjects of residents-as-teachers, study design, and outcome measures. Articles that were deemed pertinent were those that included an intervention that attempted to improve residents’ teaching skills. Studies pertaining to training fellows or faculty members were excluded. Descriptive studies and pilot studies of later published works were also excluded. In total, the 24 studies described in the table met the criteria.

*RC, randomized controlled study; NC, nonrandomized controlled study; UC, uncontrolled study; DO, direct observation study; VE, videotaped evaluation study; LE, learner evaluation study; SA, self-assessment study.

Edwards et al studied 22 obstetrics–gynecology residents (13 randomized to an intervention group with 9 controls). The intervention group received critiques, instruction, and feedback about their teaching skills, whereas the control group received no feedback or instruction. Videotaped teaching sessions were evaluated by two psychology graduate students before the intervention and six months after. Overall teaching quality scores were significantly increased (P < .017) in the intervention group during instruction. No skills in the control group were significantly increased. In comparing the two groups at six months postintervention, the intervention group scored significantly higher (P < .017) than the control group only in the skill of “communicating objectives.” At six months...
postintervention, the intervention group's overall teaching quality score decreased, showing a need for periodic reinforcement of skills.

Four uncontrolled studies used videotaping as the preferred evaluation method. Barth et al. studied six senior residents in a surgery residency. The residents were videotaped performing a teaching session as a baseline. A second teaching session was videotaped without intervention. Before the third session, the residents were given a lecture and reviewed their own teaching videos. Between the third and final teaching sessions, they were given a feedback session by a teaching consultant. Each teaching session was evaluated and scored by an independent review panel. The only significant improvement in scores ($P = .002$) was between the third and final teaching sessions, where self-review was not effective and one-on-one training was the most effective.

Roberts et al. studied residents in all years of a pediatrics residency. Their intervention was two 4-hour sessions as part of a residents-as-teachers retreat. Videotaping was used in the pre- and posttest and also in a test of defining educational terms. The ability to define educational terms increased from 26% preintervention to 94% postintervention.

Lawson and Harvill asked 20 family medicine and internal medicine residents to prepare a lecture, which was videotaped. This was evaluated using a skills evaluation instrument, which served as the pretest. These residents then attended a 13-week teaching program. After the intervention, they were asked to revise and reteach the pretest lecture, which was again videotaped and evaluated. Scores on the evaluation instrument significantly increased ($P < .001$) after the intervention.

Bing-You evaluated 26 internal medicine residents who attended an eight-hour workshop. Videotapes were evaluated for various teaching skills before the course and anywhere from 2 to 11 months after the course. The impacts on various skills were mostly improvements ($P < .05$), but they varied depending on the residents' training levels. All residents had declining organizational skills after the intervention ($P < .05$).

**Learner evaluations**

The most commonly used method of evaluation was learner-completed questionnaires, used in 11 of 24 studies. Two randomized controlled studies used this method. Hewitt et al. evaluated 53 pediatric residents spread throughout all training years (27 intervention, 26 control). The intervention group received a teaching program consisting of two 3-hour workshops with feedback sessions. The intervention group had increased confidence ($P < .05$) and received better feedback from students, faculty, and fellow residents than did the control group ($P < .01$).

Furney et al. studied 57 internal medicine residents (28 intervention, 29 control) spread across all training years. The intervention was a one-hour lecture and role-play session in which the residents taught the One-Minute Preceptor. Learners (interns and medical students) rated significant improvements in the intervention group in all skills except "teaching general rules." However, learner ratings of overall teaching effectiveness were not statistically significant ($P < .05$) between the intervention and control groups.

Five of the six nonrandomized controlled trials that we found in any category were learner evaluations. Edwards et al. studied 145 first-year residents across various specialties. The experimental group participated in a half-day teaching skills course. Residents who participated in the course were rated by medical students to be significantly better ($P < .05$) than controls in four skills: knowledge, organization, skills demonstration, and overall teaching.

Hammond et al. studied obstetrics-gynecology residents; the residents at one teaching site participated in an intervention, and residents at three other teaching sites were used as controls. The intervention was a one-day teaching skills workshop. Medical students evaluated residents at the end of each rotation and scored the residents at the intervention site significantly higher ($P = .05$) than the controls.

Pandachuck et al. studied faculty and residents who participated in a two-day teaching-enhancement workshop versus controls who did not. Medical students evaluated all participants. Students' mean ratings of the 22 instructors (a combination of faculty and residents) of the intervention group were significantly increased ($P < .0012$) after the workshop. Ratings for controls (their number was unspecified) were unchanged.

Busari et al. evaluated 27 pediatrics and obstetrics-gynecology residents (14 intervention, 13 control). Medical students evaluated the teaching abilities of the residents before and after the intervention. The intervention was a 2-hour teaching workshop. Medical students rated the intervention group to have significantly better teaching skills ($P = .02$) after the workshop than did the controls.

Spickard et al. evaluated 44 second- and third-year internal medicine residents (22 intervention, 22 controls). The intervention group participated in a three-hour teaching skills workshop. The participants were evaluated by students as well as by themselves. Residents' self-assessments ($P < .01$) and students' assessments of those residents ($P < .03$) were significantly improved in the intervention group but not in the control group. However, teaching skills were not significantly changed.

Four uncontrolled studies used learner evaluations. Litzelman et al. used a clinical teaching retreat as their intervention. Directly after the retreat and again in six months, 72 first-year internal medicine residents were evaluated by third-year medical students, who rated the residents as significantly better teachers ($P < .05$) on both occasions.

Wipf et al. performed a review of 446 evaluations of internal medicine residents in the three years before and after a six-hour course on residents' teaching began. Data were obtained from teaching assessment forms that were filled out by students and interns about their senior residents. Mean scores significantly increased ($P < .001$) each year after the intervention.

Frattarelli and Kasuya used medical students to evaluate 17 obstetrics-gynecology residents after a 4.5-hour training program on how to teach. The students' ratings indicated that the residents did not have improved teaching skills after the course.
Jafri et al. compared five gastrointestinal residents with five gastrointestinal faculty after both groups had brief training on teaching principles. Both groups then taught a number of problem-based learning sessions. The groups were evaluated by the students who attended the sessions. Overall, the faculty scored significantly higher ($P < .05$) than the residents on the student questionnaires.

Self-evaluations
Self-questionnaires were used by one randomized controlled, one nonrandomized controlled, and five uncontrolled studies. In most situations, self-evaluation was used as a secondary outcome measure. The only study in which this was the sole method of evaluation was in a study by Edwards et al. Eighteen residents from multiple specialties were asked about their teaching skills before and after a course on clinical teaching. The residents rated their skills significantly higher ($P < .001$) after the course than before the course. The authors attempted to have students evaluate these residents as well, but not enough responses were obtained.

The remaining six studies used self-evaluations as a secondary outcome measure. Gaba et al. had residents complete self-assessment questionnaires before and after the teaching workshop (no $P$ values reported). Overall, the intervention residents had higher self-assessment scores after the intervention. Litzelman et al. asked residents to evaluate themselves on clinical teaching skills immediately after the retreat and again in six months. The residents significantly rated themselves higher ($P < .01$) both times. Residents also had significantly improved ($P < .05$) self-ratings on a number of teaching skills after the intervention in the study described by Pratizzarelli and Kasuya. Participants in the study by Bing-You investigated to themselves as more effective teachers after the intervention ($P < .05$). Roberts et al. assessed a teaching inventory of behaviors and attitudes toward teaching, which showed increased scores (no $P$ values reported). Furney et al. showed that on self-assessment, the residents in the intervention group reported statistically significant improvements in all behaviors ($P < .05$). In all of these seven self-assessment evaluations, residents rated themselves as better teachers after the intervention.

Other studies on residents’ teaching and learning
One study did not fit into any of the categories previously described. Weiss and Needleman evaluated 43 pediatrics residents (18 intervention, 25 controls) in a randomized controlled study. These residents were given a pretest on a topic and then were randomly assigned to teach the topic or listen to a lecture about it. Six to eight weeks later, they were given a posttest, and the “listeners” were found to have significantly higher knowledge acquisition ($P < .01$) than the “listeners.”

Discussion
General analysis of the literature
Our review demonstrates that the current research on residents-as-teachers curricula is limited both by the number of studies and by their methodologies. Half of the articles on this topic did not qualify for this review because they were descriptive in nature. This could be because relatively few journals are available to publish this type of research or because of too little interest in funding such research, or both. In addition, only 3 of the 24 studies that met our criteria did not find statistically significant improvements in the teaching abilities of the study participants. This suggests that there may be a publication bias against studies with negative results.

In addition to a paucity of this type of study in the literature, the methodologies of published studies on this topic had clear limitations. Most published studies were nonrandomized or uncontrolled. Only seven of these were randomized controlled trials. Half of the studies relied on self-report surveys and learner surveys instead of OSTE or evaluations of videotaped teaching sessions, and none of the studies focused on the most important objective outcome, learners’ knowledge acquisition. It has been shown that OSTEs are reliable and valid methods for determining the teaching competencies of residents and should be the standard method for evaluating resident’s teaching in the future. Although Ibry developed a reliable and valid self-assessment, that method is still relatively flawed compared with independent evaluation methods.

The sample sizes were fairly small, and the range of sample sizes was wide. Despite this fact, significant improvements were found almost universally. There were large differences in the interventions themselves. The programs varied in length from 1 hour to 13 weeks. The studies also varied by theoretical basis, some using the One-Minute Preceptor (Five Microskills for Clinical Teaching), others using self-created teaching curricula, and some with unspecified workshop methods. All of these methods increased the participants’ teaching ability or confidence, but none compared curricula. Finally, the studies varied by the participants—some using interns and some using upper-level residents. One study showed interns performed poorly compared with third-year residents. The variation in training years between investigations makes analyzing effectiveness among these studies difficult.

The types of primary outcomes also varied. Some aimed to increase teaching skills for the benefit of medical students, others for residents. Buchel and Edwards showed that, depending on the level of education of the learner, there is disagreement in what skills are necessary to make the most effective clinical teacher. This suggests that, for example, programs aiming to teach medical students may not be as effective as those showing them how to teach other residents. Further study in this area is needed so that future interventions can be evidence based.

The findings of the majority of the randomized controlled trials studied demonstrate a statistical improvement in the OSTE scores, videotape evaluations, and learner evaluations. The nonrandomized uncontrolled studies also consistently showed increases in residents’ teaching confidence, ability, and interns’ and students’ evaluations. Although a small number of studies suggested that teaching skills decrease over time, some research suggests that job satisfaction and teaching skills may linger at some level.

The findings of our review also demonstrate that, over time, the studies on this subject are becoming more sophisticated. The Wamsley et al. review in 2004 revealed that a majority of the studies were uncontrolled, nonrandomized studies, but since then only one has been uncontrolled and
Our recommended intervention and how to study it

The intervention. Despite the limitations of current data, there seems to be sufficient research for us to recommend a type of intervention that represents the most evidence-based curricula and evaluation strategy. Based on the research presented here, our recommended intervention to help residents become better teachers would be based on the approach of the One-Minute Preceptor (also known as the Five Microskills for Clinical Teaching). The One-Minute Preceptor is a teaching method that involves five distinct skills. First, the preceptor discovers what the learner thinks is happening in a clinical situation; then, the preceptor probes the learner to understand his or her reasoning. The preceptor then teaches some general concepts about the topic, reinforces the correct information that the learner stated, and, finally, corrects the mistakes that the learner made in a nonjudgmental fashion. Curricula of this type are available and have already been shown to increase teaching ability. In addition, using the same intervention would allow comparison with existing research. The intervention could be effective even if it took place for only three hours, but longer programs and periodic reinforcement would almost certainly be more effective.

How to study the intervention. We suggest the following ways to study our recommended type of intervention adequately. Using a study population of 40 residents would give a study enough power to show a significant effect of the intervention. The sample should include participants from all training years to analyze the effect of experience on the efficacy of the intervention, which may be a confounder in current studies. Ideally, residents from multiple specialties should be used. A randomized controlled study would be ideal, with an OSTE performed before the intervention, and another after the intervention. A repeat OSTE at six months to one year would also be important to investigate whether the treatment effect persists.

Summing up
In summary, our review suggests that education using the One-Minute Preceptor model with periodic reinforcement would be the best intervention. Using OSTE performance as a pre- and postintervention outcome measure would be best to assess the utility of the intervention. However, because of the complexity and cost of an intervention of this nature, it may not be feasible for all programs to implement. In that case, the next-best approach would use videotaped evaluations and a shorter intervention. A customized, individually designed curriculum without reinforcement could also be used, but this would likely be time consuming to create, and it would not be a validated measure. Also, using this less intense approach without a previously validated measure could compromise the benefits of educating the residents to be effective teachers. If videotaped evaluations are not feasible, then medical student or intern evaluations of their resident teachers would be the next best approach because these are still a third-party method of evaluation. Self-assessments should be avoided because results will be obscured by self-bias.

Future research
Several questions remain unanswered about residents-as-teachers curricula. The ideal length of each program is unknown, as well as whether a dose–response relationship exists. There is no consistent theoretical basis of the educational programs and no comparison among the programs. As stated above, to make the intervention evidence based, research is needed to determine what skills are necessary to be an effective clinical teacher. Finally, no studies have investigated the link between residents’ teaching skills and a standardized comparator such as learner self-exams and scores on the United States Medical Licensing Examination.

Acknowledgments
This research was supported in part by grant D55HP05150 from the Health Resources and Services Administration.

References
21 Bing-You RG. Differences in teaching skills and attitudes among residents after their formal instruction in teaching skills. Acad Med. 1990;65:483–484.

Did You Know?

In 2003, researchers at Johns Hopkins University School of Medicine discovered, through the results of exercise tests, that the prognosis for heart disease in women is different from the prognosis for men.

For other important milestones in medical knowledge and practice credited to academic medical centers, visit the “Discoveries and Innovations in Patient Care and Research Database” at (www.aamc.org/innovations).