An Educational Intervention for Contextualizing Patient Care and Medical Students’ Abilities to Probe for Contextual Issues in Simulated Patients

Alan Schwartz, PhD  
Saul J. Weiner, MD  
Ilene B. Harris, PhD  
Amy Binns-Calvey, BA

Clinical Decision Making Requires 2 distinct skills: classifying patients’ conditions into diagnostic and management categories that permit the application of best-evidence guidelines, and individualizing or contextualizing care for patients when their circumstances and needs require variations from the standard approach to care. Contextualization is the process of identifying individual patient circumstances (their context) and, if necessary, modifying the plan of care to accommodate those circumstances.

An error occurs when a physician recommends an inappropriate plan of care. A biomedical error occurs when a physician does not identify physiological processes in the patient that require an alternative approach to appropriately care for the patient; a contextual error occurs when a physician does not identify contextual factors, such as access to care, that must be addressed to appropriately plan care. Research has demonstrated that contextual errors can be identified using standardized patients.

Objective To evaluate an educational intervention designed to increase physicians’ skills in incorporating the patient’s context in assessment and management of care and to thereby decrease the rate of contextual errors.

Design, Setting, and Participants Quasi-randomized controlled trial, with assessments by blinded observers. Fourth-year medical students (n = 124) in internal medicine subinternships at the University of Illinois at Chicago or Jesse Brown Veterans Administration Medical Center between July 2008 and April 2009 and between August 2009 and April 2010 participated and were assessed.

Intervention A 4-hour course on contextualization.

Main Outcome Measures Probing for contextual issues in an encounter, probing for medical issues in an encounter, and developing an appropriate treatment plan. Outcomes were assessed using 4 previously validated standardized patient encounters performed by each participant and were adjusted for subinternship site, academic year, time of year, and case scenario.

Results Students who participated in the contextualization workshops were significantly more likely to probe for contextual issues in the standardized patient encounters than students who did not (90% [95% confidence interval (CI), 87%-94%] vs 62% [95% CI, 54%-69%], respectively) and significantly more likely to develop appropriate treatment plans for standardized patients with contextual issues (69% [95% CI, 57%-81%] vs 22% [95% CI, 12%-32%]). There was no difference between the groups in the rate of probing for medical issues (80% [95% CI, 75%-85%] vs 81% [95% CI, 76%-86%]) or developing appropriate treatment plans for standardized patients with medical issues (54% [95% CI, 42%-67%] vs 66% [95% CI, 53%-79%]).

Conclusion Medical students who underwent an educational intervention were more likely to contextualize care for individual standardized patients.

JAMA. 2010;304(11):1191-1197  
www.jama.com

Context A contextual error occurs when a physician does not identify elements of a patient’s environment or behavior, such as access to care, that must be addressed to appropriately plan care. Research has demonstrated that contextual errors can be identified using standardized patients.
be observed in the medical record and are likely to be missed in studies of error that rely on chart review.7-10

It has been shown that contextual errors can be measured by using standardized patients and that resident physicians are more likely to make contextual than biomedical errors.11 A study with almost 400 unannounced standardized patient visits in situ demonstrated that attending internists were more likely to make contextual errors than biomedical errors when treating actors they believed were real patients.12 The objective of this study was to evaluate an educational intervention designed to increase physicians’ skills in identifying patient context and to decrease the rate of contextual errors.

**METHODS**

**Setting and Participants**

All fourth-year medical students at the University of Illinois at Chicago (UIC) College of Medicine who were enrolled in internal medicine subinternships at one of 2 sites (UIC or Jesse Brown Veterans Administration Medical Center) between July 2008 and April 2009 or between August 2009 and April 2010 were eligible to participate (May and June were excluded because few students do subinternships at the end of the year). Medical students were targeted because changing established habits of practitioners is difficult; fourth-year medical students were targeted because students are unlikely to be adequately comfortable with clinical encounters in general until their clinical years. The clerkships at the 2 sites share a common daily didactic session and a weekly grand rounds session, but other teaching activities and all clinical activities are specific to sites.

Students were invited to participate at the start of their subinternship; they were told that the purpose of the study was to validate new standardized patient assessments (without discussion of the content of the assessments or study hypotheses). Although the study was determined to be exempt by the UIC institutional review board, students were asked to provide written informed consent for inclusion of their data in the research project. Standardized patient recruitment, training, and testing were carried out at a dedicated assessment and simulation clinical performance center.

Demographic characteristics were self-reported by students on medical school applications. To assess comparability of groups, race/ethnicity was self-reported using Office of Management and Budget 1997 categories14 and recoded for this study as white vs nonwhite.

**Intervention**

The intervention consisted of 4 weekly case-based 1-hour sessions designed to help the students develop knowledge and skills in contextualizing patient care. The objectives of the sessions were for students to apply the following concepts and related skills in patient assessment and management:1-12: (1) 4 components of clinical expertise (clinical state, research evidence, patient context, and patient preferences); (2) the role of patient context in clinical expertise; (3) domains of patient context (cognitive abilities, emotional state, cultural and spiritual beliefs, access to care, social support, caretaker responsibilities, attitudes to illness, relationship with clinicians, and economic situation); (4) contextual red flags (anything about a patient situation suggesting that unaddressed contextual factors may be contributing to problems in his or her care); (5) contextual assumptions (assumptions about a patient’s context that may be correct or incorrect); and (6) contextual errors (errors in management caused by not taking into account patient context).

Students were taught to apply these concepts to contextualize patient care by listening to the narrative surrounding the patient’s complaints; identifying contextual red flags and, when identified, formulating a “contextual differential” from 10 contextual domains; exploring contextual assumptions for each domain in their differential; asking questions to narrow the differential (contextual assessment); characterizing the relevant patient context; and merging contextual assessment findings with information about clinical state, research evidence, and patient preferences to formulate a contextually appropriate plan of care.

All sessions were conducted by one of the authors (S.J.W.), a medicine-pediatrics physician; the first 2 sessions were co-taught with another author (I.B.H.) with expertise in medical education and qualitative methods. During the first session, students discussed and applied the basic concepts to a written case describing contextual issues. During the second session, students applied the basic concepts to another written case and then described and applied the concepts to their own patients on service. During the final 2 sessions, students applied and developed their knowledge and skills by interviewing patients with potential contextual issues at the bedside, with one of the authors (S.J.W.) serving as a guide and role model. Neither the standardized patients nor the cases used in the assessment procedure were components of the intervention.

**Procedure**

Students at UIC are assigned by the medical school to 1 of 2 internal medicine subinternship sites, with approximately 6 to 8 students per site per month, using a computerized lottery program. Students are randomly ordered and sequentially have their preferences for fourth-year schedules (order and site of clerkships) applied, subject to constraints on the number of students in each clerkship at each site. Each month, students at one site served as the intervention group and students at the other site as the control group; assignment of site to intervention or control was alternated to reduce the risk of contamination by students discussing their subinternship experiences with the subsequent cohort at the same site and to control effects of site.

Students in both groups who had agreed to participate in the study un-
derwent standardized patient assessment at the end of the month at the clinical performance center (3-10 days following the final workshop for the intervention group). The assessment consisted of 4 standardized patient encounters (cases A, B, C, and D) for each student, with each standardized patient presenting a different variant (of 4 variants) of each case. Subinterns each month were randomly assigned to 1 of the 16 possible permutations of case and variant. For example, a student could receive the baseline variant of case A, the biomedical variant of case B, the contextual variant of case C, and the biomedical/contextual variant of case D, but would always participate in 4 total encounters: 1 baseline, 1 biomedical, 1 contextual, and 1 biomedical/contextual variant.

In all cases and all variants, the standardized patient presented a typical clinical scenario but mentioned 2 red flags that could indicate a biomedically or contextually atypical diagnosis. In baseline variants, patients presented no symptoms of the atypical diagnosis if students probed these red flags. In biomedical variants, student probing of the biomedical red flag led the patient to present further symptoms of the medically atypical diagnosis; similarly, in the contextual variants, probing the contextual red flag led the patient to present further evidence of the contextually atypical situation. In the biomedical/contextual variants, each red flag, if probed, led to presentation of the combination of evidence for both the biomedically and the contextually atypical situation. The development and validation of the cases and variants used a panel of expert internists to ensure that not identifying the atypical diagnosis would lead to a plan of care that would be unequivocally inappropriate for the patient. Case details are presented elsewhere.

An example of a case was that of a 42-year-old man who reports worsening asthma; his current prescriptions include a daily brand-name steroid inhaler and albuterol as needed. During the encounter, the patient mentions that he wakes up wheezing and coughing at night and that things have been tough since he lost his job. Nighttime coughing is a biomedical red flag that, if probed in the biomedical or biomedical/contextual variant, will reveal several symptoms of gastroesophageal reflux disease and if probed in the other variants will result in denial of reflux disease symptoms. That things have been tough since he lost his job is a contextual red flag that, if probed in the contextual or biomedical/contextual variant, will reveal several indications that the patient uses his daily inhaler only intermittently because of the cost of the prescription and if probed in the other variants will result in denial of such financial difficulties. A physician who does not probe either red flag, or who does not modify the treatment plan to take the resulting information into account, will provide inappropriate care; both the biomedical error and the contextual error have serious implications for the patient’s health.

During each encounter, the student performed a history and focused physical examination of the standardized patient and recorded a diagnosis and management plan using the clinical performance center’s Web-based postencounter data collection tool. Encounters were digitally videotaped. Each student performed the 4 encounters sequentially on a single day, with only short breaks.

The combination of cases and variants presented was counterbalanced to subinternship months through a permuted block design with a block size of 4 months, using a computer-generated schedule (eTable 1, available at http://www.jama.com). Standardized patients were blinded to the assignment of students to groups. A standardized patient trainer (A.B.-C.), also blinded to group assignment, filled out an 11- to 16-item (depending on case and variant) quality assessment checklist on each actor’s portrayal after each encounter and provided feedback to the actor if any items were missed.

Outcomes

Primary outcomes were whether the student probed for the biomedical red flag, whether the student probed for the contextual red flag, and whether the student’s management plan addressed the problem(s) presented in the case and variant. Each encounter’s recording was reviewed by a standardized patient trainer (A.B.-C.) and coded for whether the student probed the biomedical red flags, contextual red flags, or both; the same trainer reviewed all recordings in the study and was blinded to the assignment of students to intervention or control group. Management plans were retrieved and coded by an investigator (A.S.) who had no other contact with students, using predefined checklists to determine whether students addressed the problem(s) presented in the case and variant; this coding was also performed blinded to the assignment of students to intervention or control group and without knowledge of whether the student had probed either red flag (accordingly, students who planned appropriate care were scored as successful, even if they had done so without probing relevant red flags). Outcomes were coded dichotomously.

Hypotheses and Data Analysis

We hypothesized that participants in the intervention group would probe contextual red flags more often than control group participants and also write appropriate management plans for the contextual and biomedical/contextual variants more often than control group participants. We did not hypothesize any difference between the groups on probing of biomedical red flags or appropriate baseline or biomedical variant management plans.

These hypotheses were tested by fitting logistic mixed models to the primary outcome variables, using as predictors the trial group (intervention or control), case variant (baseline, biomedical, contextual, or biomedical/contextual), interaction between trial group and case variant, case scenario (A, B, C, or D), academic year (2008 or 2009), early vs late portion of the year (July through October coded as “early” and November through April coded as “late”), and subinternship site (UIC or Jesse Brown Veterans Administration Medical Center). To account for...
clustering of encounters within students, we used the method of generalized estimating equations,\textsuperscript{15,16} with a compound symmetry covariance matrix. We predicted a significant main effect of trial group for contextual probing but not for biomedical probing. For management plan, we predicted a significant interaction between trial group and case variant, such that intervention participants write better management plans in variants including contextual qualifiers.

In a secondary analysis, probing of contextual and biomedical red flags (nested within the presence of contextual or biomedical qualifiers in the variant, respectively) was assessed as a predictor for writing an appropriate treatment plan. Data analysis was conducted using the GENLIN procedure in SPSS version 16 (SPSS, Chicago, Illinois).

**Sensitivity Analysis**

To determine whether the results were robust in the face of some unmeasured bias among participants lost to follow-up, worst-case sensitivity analysis was conducted by including in the data set those intervention students who did not appear for their standardized patient assessments but attributing to each the observed mean probability of success (in probing and treatment planning) of control students who faced the same standardized patient encounters (case and variants) for which the intervention participant was scheduled. This imputation assumes that the intervention participants who did not appear for standardized patient assessment received no benefit from the intervention.

Because these dependent variables were no longer strictly dichotomous, the GENMOD procedure (with generalized estimating equations) in SAS version 9.2 (SAS Institute Inc, Cary, North Carolina) was used to fit probit models to the completers-only data set and to the sensitivity analysis data set using the same predictors as in the logistic regressions.

**Sample Size**

Based on past research and pilot studies,\textsuperscript{11,12} it was estimated that approximately 50% of students in the control group would plan an appropriate treatment in the contextual variants and that the intervention would be considered successful if 75% of students in the intervention group planned appropriate treatment in these variants. A power calculation suggested that 108 students would be sufficient to detect such an unadjusted difference, with 80% power using a 1-tailed Fisher exact test at an a priori significance level of \( P < .05 \), which we considered to be a conservative sample size given the use of relatively more powerful statistical methods. However, we report 2-tailed statistical tests and confidence intervals (CIs).

**RESULTS**

A total of 189 students (of 230 approached) agreed to participate over the study period. Seventeen withdrew before outcome assessment (owing to conflicts with residency interviews or to family or health problems); 48 did not present for assessments, also owing to conflicts with residency interviews or other responsibilities, and we were unable to schedule make-up assessments within the following month. As a result, 65 students in total participated in the intervention group (across sites), and 59 participated in the control group (across sites) (FIGURE). Student performance was always analyzed in the group to which they were assigned.

**TABLE 1** displays the participant characteristics. The rate of completion for consenting participants did not differ significantly by subinternship site \( (\chi^2=0.007, P=.93) \) or by assignment to intervention vs control group \( (\chi^2=0.107, P=.74) \).

Within each group, there was no differ-

---

**Figure. Participant Flow**

230 Eligible participants identified
118 University of Illinois at Chicago Medical Center (UIC)
112 Jesse Brown Veterans Affairs Medical Center (JBVA)

41 Declined to participate
19 UIC
22 JBVA

189 Provided written informed consent
99 UIC
90 JBVA

94 Assigned to receive educational intervention based on site assignment\textsuperscript{a}
51 UIC
43 JBVA

95 Assigned to undergo observation only based on site assignment\textsuperscript{a}
48 UIC
47 JBVA

94 Received educational intervention as assigned

65 Assessed
26 Attended 4 teaching sessions
25 Attended 3 teaching sessions
12 Attended 2 teaching sessions
11 Attended 1 teaching session
11 Withdrawed before assessment
7 UIC
4 JBVA
18 Did not present for assessment
8 JBVA

65 Included in primary analysis

59 Assessed
6 Withdrawed before assessment
2 UIC
4 JBVA
30 Did not present for assessment
13 UIC
17 JBVA

59 Included in primary analysis

---

\textsuperscript{a} Each month, students at one site served as the intervention group and students at the other site as the control group; assignment of site to intervention or control was alternated to reduce the risk of contamination by students discussing their subinternship experiences with the subsequent cohort at the same site and to control effects of site. Student performance was always analyzed in the group to which students were assigned.
ience in completion by sex, race/ethnicity, age, or United States Medical Licensing Examination Step 2 Clinical Knowledge score. Students in the intervention group did not always attend all 4 teaching sessions; the Figure reports the number of students in the intervention group completing 4, 3, 2, or 1 sessions.

The 124 students completed a total of 494 encounters; 2 students completed only 3 rather than 4 planned encounters because the actors portraying their missing encounters were ill. A review of quality-assessment checklists indicated that the standardized patients provided perfect portrayals in 96% to 100% of encounters, depending on case and variant. In 1 encounter, a standardized patient portrayed the baseline variant of a single case instead of the intended contextual variant; this encounter was scored as a baseline variant encounter.

**Probing**

Students in the control group probed contextual red flags in 61% (95% CI, 55%-67%) of encounters, while students in the intervention group probed contextual red flags in 86% (95% CI, 82%-90%) of encounters (adjusted odds ratio, 3.75 [95% CI, 1.59-8.77]). Students in the control group probed biomedical red flags in 77% (95% CI, 72%-82%) of encounters, while students in the intervention group probed biomedical red flags in 86% (95% CI, 82%-90%) of encounters, and adjusted rates of probing are shown in Table 2. Case was a significant predictor of probing of both biomedical (P < .001) and contextual (P < .001) red flags. The biomedical red flag was more likely to be elicited in case A than in other cases, and the contextual red flag was more likely to be elicited in case C than in other cases. There was no effect of subinternship site, academic year, or early vs late month on probing of either red flag.

### Treatment Plans

Raw and adjusted proportions of encounters in which students wrote appropriate treatment plans are shown in Table 2, by variant. Overall, appropriate plans were written most often in the baseline variant, less often when either biomedical or contextual qualifiers were present, and least often when both were present in the variant. There was a significant effect of case scenario (P < .001; case C was significantly easier than cases A, B, and D), but no effect of subinternship site, academic year, or early vs late month.

Students who participated in the intervention group were much more likely to write an appropriate treatment plan in the contextual variant than students in the control group (67% of encounters [95% CI, 55%-79%] vs 24% [95% CI, 13%-35%]). There were no significant differences between intervention and control students in likelihood of an appropriate management plan in the baseline, biomedical, or biomedical/contextual variants.

### Relationship of Probing to Treatment Planning

Probing was associated with a statistically significant higher probability of an appropriate treatment plan. In context

---

**Table 1.** Characteristics of Students Enrolled in the Study and Assessed by Standardized Patients

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No. (%)</th>
<th>Assigned (n = 189)</th>
<th>Assessed (n = 124)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (n = 95)</td>
<td>Intervention (n = 94)</td>
<td>Control (n = 59)</td>
</tr>
<tr>
<td>White race</td>
<td>Men 54 (57)</td>
<td>45 (48)</td>
<td>31 (53)</td>
</tr>
<tr>
<td></td>
<td>White race 28.8 (3.6)</td>
<td>28.3 (3.6)</td>
<td>28.7 (2.5)</td>
</tr>
<tr>
<td></td>
<td>USMLE step 2 clinical knowledge score, mean (SD) 226 (21.6)</td>
<td>226 (21.9)</td>
<td>224 (19.9)</td>
</tr>
<tr>
<td>Academic year 2009-2010</td>
<td>54 (57)</td>
<td>52 (55)</td>
<td>38 (64)</td>
</tr>
</tbody>
</table>

**Table 2.** Rates of Probing and Appropriate Treatment Plans in Control and Intervention Groups

<table>
<thead>
<tr>
<th>Probing</th>
<th>Unadjusted Proportion (95% CI)</th>
<th>Adjusted Proportion (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (236 Encounters)</td>
<td>Intervention (258 Encounters)</td>
</tr>
<tr>
<td></td>
<td>P Value&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Biomedical red flag</td>
<td>0.77 (0.72-0.82)</td>
<td>0.77 (0.72-0.82)</td>
</tr>
<tr>
<td></td>
<td>P Value&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Contextual red flag</td>
<td>0.61 (0.55-0.67)</td>
<td>0.86 (0.82-0.90)</td>
</tr>
<tr>
<td></td>
<td>0.62 (0.54-0.69)</td>
<td>0.90 (0.87-0.94)</td>
</tr>
<tr>
<td>Appropriate treatment plan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline variant</td>
<td>0.97 (0.93-1.00)</td>
<td>0.97 (0.93-1.00)</td>
</tr>
<tr>
<td></td>
<td>0.96 (0.92-1.00)</td>
<td>0.98 (0.96-1.00)</td>
</tr>
<tr>
<td>Biomedical variant</td>
<td>0.58 (0.45-0.71)</td>
<td>0.53 (0.41-0.65)</td>
</tr>
<tr>
<td></td>
<td>0.66 (0.53-0.79)</td>
<td>0.54 (0.42-0.67)</td>
</tr>
<tr>
<td>Contextual variant</td>
<td>0.24 (0.13-0.35)</td>
<td>0.67 (0.55-0.79)</td>
</tr>
<tr>
<td></td>
<td>0.22 (0.12-0.32)</td>
<td>0.69 (0.57-0.81)</td>
</tr>
<tr>
<td>Biomedical/contextual variant</td>
<td>0.21 (0.10-0.32)</td>
<td>0.27 (0.16-0.38)</td>
</tr>
<tr>
<td></td>
<td>0.16 (0.06-0.25)</td>
<td>0.27 (0.16-0.38)</td>
</tr>
</tbody>
</table>

**Abbreviation:** CI, confidence interval.

<sup>a</sup>For comparisons of adjusted rates based on tests of model effects in the logistic regression analysis and adjusted for case scenario (A, B, C, or D), subinternship site (site 1 or site 2), academic year (2008 or 2009), and early vs late month in year.

<sup>b</sup>For comparisons of unadjusted rates based on Χ<sup>2</sup> tests with 1 df.

<sup>c</sup>Rates per encounter, across 124 participants.

©2010 American Medical Association. All rights reserved.
tual variants, appropriate plans were written 4% (95% CI, 0%-21%) of the time (across trial groups and cases) when the contextual red flag was not probed and 57% (95% CI, 46%-67%) of the time when it was probed (35% [95% CI, 21%-52%] in the control group and 71% [95% CI, 58%-82%] in the intervention group); in biomedica
tual variants, appropriate plans were written 15% (95% CI, 4%-35%) of the time when the biomedical red flag was not probed and 67% (95% CI, 56%-76%) of the time when it was probed (68% [95% CI, 52%-81%] in the control group and 65% [95% CI, 50%-78%] in the intervention group). However, the association with intervention persisted when controlling for pro
ing (P = .008), suggesting that the in
tervention was directly associated with treatment planning independent of probing.

**Sensitivity Analysis of Students Not Completing the Assessment**

Even though the students who did not present for the standardized patient as
cessments were similar in number and characteristics across the control and intervention groups, sensitivity analyses were conducted using multiple im
pulation with the conservative assump
tion that the intervention students who did not appear for standardized pa
tient assessment received no benefit from the intervention. Each analysis produced the same pattern of findings (eTable 2). Although the worst case re
clected a smaller effect size for pro
ing the contextual red flag (60% vs 82% rather than 62% vs 90%) and for plan
ning appropriate treatment in the con
textual variant (22% vs 55% rather than 22% vs 69%), the effect of intervention remained statistically significant (P < .001) in the worst-case analysis.

**COMMENT**

These results suggest that an educational intervention can be associated with better performance in the ability of medical students to contextualize care for individual patients. Students who participated in contextualization workshops were significantly more likely to probe for contextual issues in the standardized patient encounters than students who did not and significantly more likely to develop appropriate treatment plans for standardized patients with contextual issues. There was no difference between the groups in the rate of probing for bi
omedical issues or developing appropriate treatment plans for standardized pa
tients with biomedical issues.

When students did not probe a contextual red flag that was indicative of a contextually atypical condition, they rarely planned appropriate care. Even when students in the intervention group successfully probed the red flag, they planned appropriate care only 71% of the time. These rates of treatment planning in contextual variants are similar to those for attending physicians presented with the same cases by unannounced standardized patients in their office settings. We speculate that not only is it more difficult to provide care for (biomedically or contextually) atypical conditions that are not identified but that contextualization of care (even when contextually atypical conditions are identified) is more difficult than providing evidence-based care for typical conditions and is also less likely to be acquired by physicians as part of medical training. The difficulty might reflect that contextualization is less likely to leverage the brain’s strengths in pattern recognition.

To develop an appropriate treatment plan in the biomedical/contextual variants, a student must identify both biomedical and contextual issues and incorporate each into the plan. This is considerably more difficult than iden
tifying and incorporating a single red flag, both because there are now 2 is
sues to address and because the discovery of the first complexity may lead the student to premature closure around the diagnosis and reduce the chance of iden
tifying and incorporating the second complexity. This possibility could not be tested, because we did not present students with cases that included 2 bi
omedical or 2 contextual issues.

**Limitations**

This study is limited by the number of cases and variants and the nature of the sample population. Although participants were assigned to sites using a quasi-random process independent of this trial, assignment of site to inter
vention or control was alternated. Al
though this reduced the risk of cont
amination, it may increase the risk that participants could predict their alloca
tion to trial group. If some contamina
tion between sites did occur each month, that would be expected to have biased the results toward the null hy
pothesis. Because the design is quasi-randomized, strong causal inferences about the effect of the intervention should not be drawn.

A substantial number of students consented to participate in the study but did not undergo the standardized pa
tient assessment outcome measure. Many informed us that given their resi
dency interviews and requirements of other clerkships, they were unavail
able for any of the times for which the standardized patients and testing cen
ter were available in the following month. Although the sensitivity anal
ysis suggests that this loss to assess
ment was unlikely to change the find
ings and that this loss would not be expected to be biased, it remains a limi
tation of this study.

Participants were assessed on per
formance in 4 standardized patient en
counters based on 4 previously vali
dated cases that varied the presence or absence of contextual and biomedical qualifiers in a factorial design. Al
though this design allowed examining associations with the intervention across a minimal number of encour
ers, additional encounters with con
textual qualifiers (in a randomized de
sign) could permit within-participant measurement of the generalizability of the effect of the intervention and individual student performance and might detect effects of the intervention in the biomedical/contextual encounter.

Participants were fourth-year medi
cal students, and it is not known whether the associations would be as
strong in more experienced residents or attending physicians. It is also unknown how long the new skills inculcated by the educational intervention will be retained. A study at our institution introduced third-year medical students to a different set of cognitive skills (the hypothesis-driven physical examination) using an educational intervention of intensity similar to that of this one and measured retention a year later.

Some skills (anticipation, elicitation, and documentation) demonstrated decay, while another (discrimination) was enhanced (presumably as a result of greater clinical experience); as a result, diagnostic accuracy did not decline after 1 year. The authors suggested that live teaching and learning encounters with standardized patients, particularly when feedback is provided, may enhance retention.

**CONCLUSIONS**

Medical students are typically trained to identify biomedical red flags that may alter their diagnosis and management of patients but are rarely trained to identify contextual red flags that may be equally vital in providing appropriate care. Similarly, practicing physicians are tracked for adherence to quality measures, such as the Healthcare Effectiveness Data and Information Set, that do not incorporate contextual issues; hence, deficits are unlikely to be addressed.

Yet contextualization of care is an important skill for physicians. Not considering patient context in management plans may result in harms of a magnitude equal to not appreciating a biomedical finding. Moreover, several contextual factors, such as access to care, religion, and socioeconomic status, are associated with health disparities, and not identifying and integrating patient context in clinical decision making may worsen these disparities. The skills required for contextual probing and contextualization in treatment planning are teachable, but students may not acquire them through current medical school curricula. Curricula and activities that emphasize contextualization may be warranted.

**Author Contributions:** Dr Schwartz had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** Schwartz, Weiner, Harris, Binns-Calvey.

**Acquisition of data:** Weiner, Binns-Calvey.

**Analysis and interpretation of data:** Schwartz.

**Drafting of the manuscript:** Schwartz.

**Critical revision of the manuscript for important intellectual content:** Weiner, Harris, Binns-Calvey.

**Statistical analysis:** Schwartz.

**Obtained funding:** Schwartz, Weiner, Harris.

**Administrative, technical, or material support:** Weiner, Harris, Binns-Calvey.

**Study supervision:** Schwartz.

**Financial Disclosures:** Drs Schwartz and Weiner reported being owners of a company that provides management consulting services to clinicians and institutions interested in collecting customer service and performance data using unannounced standardized patient methods. They have not to date received any consulting fees, honoraria, contracts, or other payments. No other authors reported disclosures.

**Funding/Support:** This project was funded in part by a National Board of Medical Examiners (NBME) Edward J. Stemmler, MD, Medical Education Research Fund grant. There were no other external sponsors.

**Role of the Sponsor:** The NBME had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

**Disclaimer:** This project does not necessarily reflect NBME policy, and NBME support provides no official endorsement.

**Online-Only Material:** eTable 1 and eTable 2 are available at http://www.jama.com.

**Additional Contributions:** We gratefully acknowledge the assistance of the Allan L. and Mary Graham Clinical Performance Center at the University of Illinois at Chicago College of Medicine and Naomi Ashley-Benedict, BA, University of Illinois at Chicago, in scheduling medical student encounters and assisting in data collection. Ms Benedict was employed by the Department of Medicine during this study as a visiting research specialist.

**REFERENCES**


