An educational intervention for contextualizing patient care and medical students' abilities to probe for contextual issues in simulated patients

Alan Schwartz, PhD[1,2], Saul J. Weiner, MD[2,3,4], Ilene B. Harris, PhD[1,5], Amy Binns-Calvey[1]

[1]Department of Medical Education, University of Illinois at Chicago

[2]Department of Pediatrics, University of Illinois at Chicago

[3]Department of Medicine, University of Illinois at Chicago

[4]Jesse Brown VA Medical Center

[5]Department of Pathology, University of Illinois at Chicago

Revision date: June 12, 2010

Word count: 4096

Corresponding Author:

Alan Schwartz

Department of Medical Education (mc 591)

808 S. Wood St, 986 CME

Chicago, IL 60612

Email: alansz@uic.edu

Phone: 312-996-2070

Fax: 312-413-2048

Abstract

Context: A contextual error occurs when a physician fails to identify elements of a patient's environment or behavior, such as access to care, that must be addressed in order 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 VA Medical Center from July 2008 – April 2009 and August 2009 – 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%]), 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 patient with medical issues (54% [95% CI, 42-67%] vs 66% [95% CI, 53-79%]).

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

Introduction

Clinical decision making requires two distinct skills: (1) classifying patients' conditions into diagnostic and management categories that permit the application of best evidence guidelines, and (2) individualizing or *contextualizing* care for patients when their circumstances and needs require variations from the standard approach to care.[1-2] 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.[3] A *biomedical error* occurs when a physician fails to identify physiological processes in the patient that require an alternative approach to care; a *contextual error* occurs when a physician fails to identify contextual factors, such as access to care, that similarly demand an alternative approach in order to appropriately care for the patient.[4] Although the failure of physicians to attend to contextual clues[5] and the frequency of some contextual clues of substantial health importance (eg, medication nonadherence[6]) have been studied, the frequency of contextual error – inappropriate care due to failure to identify or incorporate contextual information – is unknown, because such errors can not 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 (SPs) and that resident physicians were 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

Participants and Setting

All fourth-year (M4) medical students at the University of Illinois at Chicago College of Medicine (UIC) who were enrolled in internal medicine subinternships at one of two sites (UIC or Jesse Brown VA Medical Center) between July 2008 – April 2009 or August 2009 – 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;[13] 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 two 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 (CPC).

Demographic characteristics were self-reported by students on medical school applications. Race/ethnicity was self-reported using OMB 1997 categories[17] and recoded for this study as white vs non-white for comparisons of participants assigned to groups.

Intervention

The intervention consisted of 4 weekly case-based one-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, 14] (1) four 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 health care providers, and economic situation); (4) contextual *red flags* (anything about a patient situation suggesting that unaddressed contextual factors may be contributing to problems in their 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 a failure to take 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; 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 incorporating contextual assessment 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 (SW), a medicine-pediatrics physician; the first two sessions were co-taught with another author (IH) 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 SW serving as a guide and role model. Neither 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 one of two 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 underwent SP assessment at the end of the month at the CPC (3-10 days following the final workshop for the intervention group). The assessment consisted of four SP encounters for each student (cases A, B, C, and D), with each SP presenting a different one of 4 variants of each case. Subinternships 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 four total encounters: one baseline, one biomedical, one contextual, and one biomedical/contextual variant.

In all cases and all variants, the SP presented a typical clinical scenario but mentioned two 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 contextually atypical situation. The development and validation of the cases and variants used a panel of expert internists to ensure that failure to identify the atypical diagnosis would lead to a plan of care that would be unequivocally inappropriate for the patient.[11] Case details are presented elsewhere.[12]

An example of a case was a 42 year old man who presents with complaints of 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. Night-time coughing is a biomedical red flag that, if probed in the biomedical or biomedical/contextual variant, will reveal several symptoms of GERD, and if probed in the other variants will result in denial of symptoms of GERD. 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 due to the cost of the prescription, and that if probed in the other variants will result in denial of such financial difficulties. A physician who fails to probe either red flag, or fails to 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 SP, and recorded a diagnosis and management plan using the CPC's web-based post-encounter data collection tool. Encounters were digitally videorecorded. Each student performed the four 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 four months, using a computer-generated schedule (see eTable 1). Standardized patients were blinded to the assignment of students to groups. A standardized patient trainer (AB), also blinded, filled out an 11-16 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/variant. Each encounter's recording was reviewed by a standardized patient trainer (AB) and coded for whether the student probed the biomedical and/or contextual red flags; 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 who had no other contact with students (AS) using predefined checklists to determine whether they addressed the problem(s) presented in the case/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 was 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 (1) probe contextual red flags more often than control group participants, and (2) write correct 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 correct 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 arm (intervention or control), case variant (baseline, biomedical, contextual, or biomedical/contextual), interaction between trial arm and case variant, case scenario (A, B, C, or D), academic year (2008 or 2009), early vs late portion of the year (July – October coded as "early" and November – April coded as "late") and subinternship site (UIC or Jesse Brown VA Medical Center). To account for clustering of encounters within students, we used the method of generalized estimating equations,[15-16] with a compound symmetry covariance matrix. For contextual probing, we predicted a significant main effect of trial arm; for biomedical probing, we did not predict a significant main effect of trial arm. For management plan, we predicted a significant interaction between trial arm 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 a correct treatment plan. Data analysis was conducted using SPSS 16 (GENLIN procedure).

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 SP assessments, but attributing to each the observed mean probability of success (in probing and treatment planning) of control students who faced the same SP encounters (case and variants) for which the intervention participant was scheduled. This imputation assumes that the intervention participants who did not appear for SP assessment received no benefit from the intervention.

Because these dependent variables were no longer strictly dichotomous, probit models were fitted using SAS 9.2 (PROC GENMOD, with generalized estimating equations) 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,[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 one-tailed Fisher's 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.

Results

A total of 189 students (of 230 approached) agreed to participate over the study period. Seventeen withdrew before outcome assessment (due to conflicts with residency interviews or family or health problems); 48 failed to show up for assessments, also due 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 students participated in the control group (across sites). Student performance was always analyzed in the group to which they were assigned.

Figure 1 illustrates the flow of participants in the study. Table 1 displays the participant characteristics. The rate of completion of consented participants did not differ significantly by subinternship site ($\chi^2(1)=0.007$, $p=0.93$), or by assignment to intervention vs control group ($\chi^2(1)=0.107$, $p=0.74$). Within each group, there was no difference in completion by sex, race, age, or USMLE Step 2 Clinical Knowledge score. Students in the intervention group did not always attend all four teaching sessions; Figure 1 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; two 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 SPs provided perfect portrayals in 96-100% of encounters, depending on case and variant. In one encounter, an SP portrayed the baseline variant of one 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% confidence interval [CI], 55 – 67%) of encounters, while students in the intervention group probed contextual red flags in 86% of encounters (95% CI, 82 – 90%); adjusted odds ratio (aOR), 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, and students in the intervention group probed biomedical red flags in 77% (95% CI, 72% – 82%) of encounters (aOR, 0.75; 95% CI, 0.33 – 1.68). Raw and adjusted rates of probing are shown in Table 2. Case was a significant predictor of probing of both biomedical ($p<.001$) and contextual red flags ($p<.001$). The biomedical red flag was more likely to be elicited in case A than other cases, and the contextual red flag was more likely to be elicited in case C than 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 correct treatment plans are shown in Table 2, by variant. Overall, correct 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 arm were much more likely to write a correct treatment plan in the contextual variant than students in the control arm (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 a correct 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 a correct treatment plan. In contextual variants, correct plans were written 4% (95% CI, 0% - 21%) of the time (across trial arms 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 biomedical variants, correct 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). The association with intervention persisted when controlling for probing, however (p=.008), suggesting that the intervention was directly associated with treatment planning independent of probing.

Sensitivity analysis of students not completing the assessment

Even though the students who failed to present for the SP assessments were similar in number and characteristics across the control and intervention groups, sensitivity analyses were conducted using multiple imputation with the conservative assumption that the intervention students who did not appear for SP assessment received no benefit from the intervention. Each analysis produced the same pattern of findings. (eTable 2) Although the worst case reflected a smaller effect size for probing the contextual red flag (60% vs 82% rather than 62% vs 90%) and for planning appropriate treatment in the contextual 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 SP encounters than students who did not, and significantly more likely to develop appropriate treatment plans for SPs with contextual issues. There was no difference between the groups in the rate of probing for biomedical issues or developing appropriate treatment plans for SPs with biomedical issues.

When students failed to 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 in attending physicians presented with the same cases by unannounced standardized patients in their office settings.[12] 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 harder than identifying and incorporating a single red flag, both because there are now two issues 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 identifying and incorporating the second complexity. This possibility could not be tested because we did not present students with cases that included two biomedical or two 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, sites were alternated between serving as intervention and control. Although this reduced the risk of contamination, it may increase the risk that participants could predict their allocation to trial arm. If some contamination between sites did occur each month, that would be expected to have biased the results toward the null hypothesis. 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 SP assessment outcome measure. Many informed us that given their residency interviews and requirements of other clerkships, they were unavailable for any of the times for which the standardized patients and testing center were available in the following month. Although the sensitivity analysis suggests that this loss-to-assessment was unlikely to change the findings, and this loss would not be expected to be biased, it remains a limitation of this study.

Participants were assessed on performance in four SP encounters based on four previously validated cases that varied the presence or absence of contextual and biomedical qualifiers in a factorial design. Although this design allowed examining associations with the intervention across a minimal number of encounters, additional encounters with contextual qualifiers (in a randomized design) 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 medical 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 similar intensity to this one and measured retention a year later.[18] 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.[18]

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 (HEDIS), that do not incorporate contextual issues; hence, deficits are unlikely to be addressed.

And yet, contextualization of care is an important skill for physicians. Failure to consider patient context in management plans may result in harms of a magnitude equal to failure to appreciate a biomedical finding. Moreover, several contextual factors, such as access to care, religion, and socioeconomic status, are associated with health disparities, and the failure to identify and integrate 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.

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 disclosures were reported.

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

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

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

Additional contributions: The authors 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

1. Weiner SJ. From research evidence to context: the challenge of individualizing care. *ACP J Club.* 2004;141:A11-A12.
2. Safford M, Allison J, Kiefe C. Patient Complexity: More Than Comorbidity. The Vector Model of Complexity. *J Gen Intern Med.* 2007;22(0):382-390.
3. Kohn LT, Corrigan J, Donaldson MS. *To err is human : building a safer health system*. Washington, D.C.: National Academy Press; 2000.
4. Weiner SJ. Contextual error. In: Kattan MW, ed. *Encyclopedia of Medical Decision Making*: SAGE; 2008:198-202.
5. Levinson W, Gorawara-Bhat R, Lamb J. A study of patient clues and physician responses in primary care and surgical settings. *JAMA.* 2000;284(8):1021.
6. DiMatteo MR. Variations in Patients' Adherence to Medical Recommendations: A Quantitative Review of 50 Years of Research. *Med Care.* 2004;42(3):200-209.
7. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med.* Feb 7 1991;324(6):370-376.
8. Studdert DM, Mello MM, Gawande AA, et al. Claims, errors, and compensation payments in medical malpractice litigation. *N Engl J Med.* May 11 2006;354(19):2024-2033.
9. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care.* Mar 2000;38(3):261-271.
10. Weiner SJ. Medical Error. In: Mullner R, ed. *Encyclopedia of Health Services Research.* Vol Vol II. SAGE Publications; 2009:730-735.
11. Weiner SJ, Schwartz A, Yudkowsky R, et al. Evaluating physician performance at individualizing care: a pilot study tracking contextual errors in medical decision making. *Med Decis Making.* Nov-Dec 2007;27(6):726-734.
12. Weiner SJ, Schwartz A, Weaver FM, et al. Contextual errors and failures in individualizing patient care: A multicenter study. *Ann Intern Med.* In press.
13. Grimshaw JM, Shirran L, Thomas R, et al. Changing Provider Behavior: An Overview of Systematic Reviews of Interventions. *Med Care.* 2001;39(8):II2-II45.
14. Weiner SJ. Contextualizing medical decisions to individualize care: lessons from the qualitative sciences. *J Gen Intern Med.* 2004;19:281-285.
15. Diggle P, Heagerty P, Liang K, Zeger S. *Analysis of longitudinal data*: Oxford University Press, USA; 2002.
16. Zeger S, Liang K. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986;42(1):121-130.
17. Office of Management and Budget. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. *Fed Regist.* 1997;62:58781-58790.
18. Yudkowsky R, Otaki J, Lowenstein T, Riddle J, Nishigori H, Bordage G. A hypothesis-driven physical examination learning and assessment procedure for medical students: initial validity evidence. *Med Educ.* Aug 2009;43(8):729-740.

Figure Legends

Figure 1. Participant flow through the trial.

Tables

Table 1: Characteristics of students enrolled in the study and assessed by standardized patients

	Assigned (n=189)		Assessed (n=124)	
	Control (n=93)	Intervention (n=95)	Control (n=59)	Intervention (n=65)
Male sex (%)	54 (58%)	45 (47%)	31 (53%)	29 (45%)
White race (%)	34 (37%)	30 (32%)	22 (37%)	21 (32%)
Age (mean ± SD)	28.8 ± 3.6	28.8 ± 3.6	28.7 ± 2.5	28.8 ± 3.1
USMLE 2CK (mean ± SD)	226 ± 21.6	226 ± 19.3	224 ± 19.9	226 ± 19.6
Academic year (% in 2009-2010)	54 (58%)	52 (55%)	38 (64%)	39 (60%)

Note: "Assigned" students are all students who consented to participate. "Assessed" students are all assigned students who appeared for the standardized patient assessment at the end of the subinternship. There were no statistically significant differences between students assigned to intervention vs control, students assessed in intervention vs control, intervention students assigned vs assessed, or control students assigned vs assessed.

Table 2: Rates of probing and appropriate treatment plans in control and intervention groups.

	Unadjusted proportion (95% CI)			Adjusted proportion (95% CI)*		
	Control (236 encounters)	Intervention (258 encounters)	p	Control (236 encounters)	Intervention (258 encounters)	p
Probing						
Biomedical red flag	0.77 (0.72-0.82)	0.77 (0.72-0.82)	.851	0.81 (0.76-0.86)	0.80 (0.75-0.85)	0.729
Contextual red flag	0.61 (0.55-0.67)	0.86 (0.82-0.90)	<.001	0.62 (0.54-0.69)	0.90 (0.87-0.94)	<.001
Appropriate treatment plan						
Baseline variant	0.97 (0.93-1.00)	0.97 (0.93-1.00)	0.922	0.96 (0.92-1.00)	0.98 (0.96-1.00)	0.175
Biomedical variant	0.58 (0.45-0.71)	0.53 (0.41-0.65)	0.609	0.66 (0.53-0.79)	0.54 (0.42-0.67)	0.181
Contextual variant	0.24 (0.13-0.35)	0.67 (0.55-0.79)	<.001	0.22 (0.12-0.32)	0.69 (0.57-0.81)	<.001
Biomedical/contextual variant	0.21 (0.10-0.32)	0.27 (0.16-0.38)	0.448	0.16 (0.06-0.25)	0.27 (0.16-0.38)	0.107

Table note: Rates are per encounter, across 124 participants, with 95% confidence intervals in parentheses. p-values for comparisons of unadjusted rates are based on χ^2 tests with 1 degree of freedom. p-values for comparisons of adjusted rates are 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.