Accurate documentation of the patient encounter is essential for patient care and is a core competency for medical students. Current guidelines require students to master using a standard format to “record pertinent clinical information accurately, chronologically, comprehensively, and free of extraneous information.” Nevertheless, students make errors in the post-encounter note (PEN), even in high-stakes clinical skills assessments (CSAs). Errors of omission (under-reporting) are common and attributed to students forgetting to record important data and/or not recognizing the significance of a key item. Studies consistently report that students obtain but fail to report 20%–30% of key information. The prevalence of errors of commission, ie, reporting an item that was not obtained in the patient encounter (over-reporting), is much less clear. In pilot tests for national examinations, over-reporting was uncommon, varied significantly by case, and was attributed mainly to problems in the design of scoring instruments. Conversely, some authors have noted high rates of over-reporting, especially for physical examination items, and have attributed these to deliberate fabrication by students. This raises serious issues of cheating on an examination and has been identified as a marker for problems in future professional integrity. Both unprofessional behavior as a student and serious deficiencies in communication on national licensing examinations have been associated with future complaints to professional licensing authorities.

The National Board of Osteopathic Medical Examiners recently failed eight students in the Comprehensive Osteopathic Medical Licensing Examination (COMPLEX-USA Level 2-PE) on the basis of fabrication of clinical notes.

In view of the divergent reports about the prevalence and reasons for over-reporting plus the potentially devastating consequences for students, we conducted an in-depth analysis of over-reporting in our end of third-year CSA. We limited the study to the medical history because of concerns about precision in validating student performance on physical examination. Camera angles and other technical considerations can limit observations; further, many physical examination components require subjective reports from a standardized patient (SP) or an observer as to how well a student performance meets an established standard.

Methods

Data were collected on four of the 12 cases in the end of third-year CSA. In developing each case, faculty identified 15 key items in the medical history. The student’s ability to obtain each item was measured by SP report (obtained/not obtained) on a

From the Department of Family and Community Medicine (Drs Walling, Moser, and Dickson), Office of Research (Ms Zackula), and Standardized Patient Program (Ms Fussell), University of Kansas School of Medicine-Wichita.
A standardized checklist entered into the electronic database immediately after each encounter. For each case, students completed a post-encounter note (PEN) in a format based on that used by USMLE-SE of history, physical examination, differential diagnosis, and diagnostic workup. Faculty scoring of PENs included assessing each of the 15 key history items as reported/not reported using a checklist almost identical to that used by the SPs. Minor differences between the SP and faculty checklists were necessary to accommodate differences in how an item might be discussed with a patient and recorded in a clinical record. The retrospective analysis of these data was determined to be exempt by the Institutional Review Board’s Human Subjects Committee.

A descriptive analysis compared the SP and faculty-scored PEN checklists for each of the key items in the medical history. To minimize interobserver errors, one SP and one faculty scorer were assigned to each case. A randomly selected sample of 28% of SP checklists was independently scored by trained staff members watching performances or video recordings of the encounter. A randomly selected sample of 25% of PENs for each case was independently scored by one of the other faculty scorers. All SP checklists and PEN scores were directly entered into an electronic data management system. Those for the four study cases were abstracted and compared for the electronic data management system. Those for the four study cases were abstracted and compared for the study. This comparison generated four possible categories for each key item in the history, ie, obtained/reported, obtained/not reported, not obtained/reported, and not obtained/not reported (Table 1).

For each instance of over-reporting (not obtained/reported), one of the authors reviewed the video of the encounter and the original SP checklist to validate the scoring as “not obtained” by the SP. The same faculty member also reviewed the original PEN written by the student and the PEN scoring form to validate the scoring as “reported.” A narrative explanation of each over-reported event was prepared. These explanations were analyzed by case and by type of error.

**Results**

The performance of the class of 50 students on the 15 history items over four cases generated 3,000 data entries. Initially, 73 events were identified where a student reported an item in the PEN that was scored as not obtained on the SP checklist (over-reported). This implied that about 4% of reported items were “false” or had not been obtained during the SP encounter (Table 1). This result is similar to or slightly lower than reported in previous studies.

The analysis attributed 69 of the 73 over-reporting events (94.5%) to non-student factors (Table 2).

The largest source of over-reporting was SP error. Thirty instances were identified in which the video documented that the item was obtained, but the SP recorded it as “not obtained” on the checklist. Twenty of these errors were in a single case. Over the 5-day exam (50 student encounters), we identified seven episodes of this SP misreporting an event as “not obtained” when video review documented it being obtained and 13 episodes when he reported a volunteered item as “not obtained.”

Problems in case design accounted for 24 over-reporting errors. In 15 of these, lack of specificity in items related to symptoms (mainly in the neurological examination) explained different scores by SPs and physician scorers. In another five errors, the key item could be discovered by the student without directly asking the checklist question. The remaining four errors attributed to case design resulted from differences in the wording of the checklist and PEN scoring template.

Data entry errors in the process of abstracting the study data from the large dataset on SP programs to a database for the analysis contributed 14 items of over-reporting. One item was identified where the faculty scorer recorded “documented” for an item not present in the PEN.

Three of the four over-documentation errors attributed to students occurred in the memory loss case. One student documented “no alcohol use” in the PEN, but the video did not contain any discussion of the topic. Another student documented “independent in all IADLs” but the video contained no references to the patient’s ability to do any of the specific components of the Activities of Daily Living assessment. A third student documented “no significant previous medical history” without any exploration of prior health problems. The fourth over-documentation error involved an SP who was portraying upper respiratory symptoms. One student documented “sore throat and congestion” but did not ask the patient about these specific symptoms.

**Discussion**

The rates of under- and over-reporting in this study are very comparable to those reported in a recent study of 219 students completing
a single CSA case but lower than rates reported by those studies that have included physical examination items.4,11

As anticipated, the biggest source of error was in SP performance, particularly on the headache case. Previous studies have noted clustering of over-reporting in specific cases in studies of both real and standardized patients.5,11 Because the headache case was complex and portrayed by an inexperienced SP, it had been identified for more frequent monitoring during the CSA, and performance fell within criteria for quality control during direct observation. During the CSA monitoring we did not appreciate that the SP had misunderstood instructions for reporting items that were volunteered or obtained by students without directly asking. He consistently reported these items as “not obtained.” Since the student knew the information, it was included in the PEN and flagged as over-reporting. The importance of training SPs in recoding data on checklists has been noted but may receive less attention than training in performance of the clinical scenario.12,13 Despite training, SPs make mistakes during complex CSAs. The most comprehensive study of SP error involved a headache case and reported an overall rate of errors of omission by SPs of 1.6%.14 Some authors report that later time of day and being the first or second case of the session slightly increase SP error rate,15 but others found no effect of these variables or SP gender on students’ scores.16

Case design issues were predominantly due to case complexity or poorly worded checklist items. Each case is re-edited every year by a team of at least four faculty members. Faculty members tend to add details either to enhance clinical complexity or to ensure the SP is prepared for any potential question. Increasing detail can complicate accurately defining an item as “obtained.” In addition, all changes must be carefully mapped to ensure they fit with the complete scenario and scoring instruments. We found one script change that inadvertently provided the student with a key item and another case in which a change had been made to the SP checklist but not to the corresponding PEN scoring document. These disappointing findings emphasize the need for scrupulous attention to detail in case design, especially of checklist and scoring instruments. In particular, well-established cases should be monitored to ensure they do not become unnecessarily complex over time by the accretion of details. Our headache case exemplifies the “perfect storm” of SP potential errors, ie, an inappropriately complex case containing script inconsistencies, portrayed by an inexperienced SP whose first language was not English, volunteered too much, and had misunderstood how to score volunteered items.

We were surprised to encounter data entry errors. The SP database was not designed to facilitate research but did appear to have the capability to compare two matching templates when the study was designed. Nevertheless, problems were encountered that necessitated exporting data to another system for analysis. In this process, data for two students were miscoded.

The remaining events where students clearly over-reported could all be attributed to inappropriate generalization rather than deliberate falsification of information. The checklist item called for specific clinical information that students had either assumed from the type of patient (eg, that elderly ladies don’t use alcohol) or a cluster of related symptoms (upper respiratory symptoms) or abilities (“doing well” equating to independent in I/ADLs). The tendency of students to over-interpret general information is well-recognized.17 While understandable, the clinical note is a significant legal and clinical document that must be appropriately specific. “Even inadvertent over-reporting erodes the integrity of the record.”11

We undertook this intensive review because of the serious implications for students accused of falsely reporting information in the medical record. The findings confirm our concern that over-reporting should not be attributed to student falsification of information unless all other possibilities have been excluded. Despite an apparently adequate system for quality control in a well-established CSA, almost all over-reporting in the medical history could be attributed to non-student factors. The study provided us with important new

Table 2: Source of Over-reporting Error by Case

<table>
<thead>
<tr>
<th>Case</th>
<th>Data Entry</th>
<th>Case Design</th>
<th>Standardized Patient</th>
<th>PEN Scorer</th>
<th>Student</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal pain</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Depression</td>
<td>3</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Headache</td>
<td>8</td>
<td>3</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>33</td>
</tr>
<tr>
<td>Memory loss</td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>14 (19.2%)</td>
<td>24 (32.9%)</td>
<td>30 (41.1%)</td>
<td>1 (1.4%)</td>
<td>4 (5.5%)</td>
<td>73</td>
</tr>
</tbody>
</table>

PEN—post-encounter note
insights and motivation to improve exam quality.

ACKNOWLEDGMENTS: This study was presented at the 2011 Society of Teachers of Family Medicine Conference on Medical Student Education.

CORRESPONDING AUTHOR: Address correspondence to Dr Walling, University of Kansas School of Medicine-Wichita, Department of Family and Community Medicine, 1010 N. Kansas, Wichita, KS 67214. 316-293-2607. Fax: 316-293-2696. awalling@kumc.edu.

References