Necessity of introducing postencounter note describing history and physical examination at clinical performance examination in Korea

Jonghoon Kim

Office of Medical Education, Inha University School of Medicine, Incheon, Korea

Purpose: Information gathering ability had been evaluated mainly via checklists in clinical performance examinations (CPX). But, it is not proved yet if students write the information correctly in postencounter note (PN), although they asked questions or performed physical examinations (PE) about the information when they interacted with standardized patients in CPX. This study addressed the necessity of introducing PN to evaluate the ability in CPX.

Methods: After patient encounters, students were instructed to write the findings of history taking and physical examination that they considered as important information in approaching the patient's problems in PN. PNs were scored using answer keys selected from checklist items, which were considered to be recorded in PN by CPX experts.

Results: PNs of six CPX cases from 54 students were analyzed. Correlation coefficients between the key-checklist scores and PN scores of six cases were moderate to high (0.52 to 0.79). However, students frequently neglected some cardinal features of chief complains, pertinent findings of past/social history and PE, and pertinent negative findings of associated symptoms in PNs, which were checked as 'done' in the keys of checklists.

Conclusion: It is necessary to introduce PN in CPX to evaluate the students' ability of synthesis and integration of patient information.

Key Words: Patient simulation, Medical records, Clinical competence

INTRODUCTION

Gathering information from patients is an essential step for good medical practice. However, selecting valuable information among gathered information and recording it appropriately are also important because the information from history taking (Hx) and physical examination (PE) is often used not only by the person who gets it directly from a patient, but also accessed by many personnel who may participate in caring the patient through the record [1]. Therefore, the information should be recorded completely in the postencounter note (PN) for delivering information correctly. In addition, someone who read PN, including Hx, PE, and provisional diagnosis, may be able to recheck the appropriateness of clinical reasoning to generate provisional diagnosis by analyzing the relationship between them.

Evaluating the ability of medical students to gather

important information from patients through Hx and/or PE is one of the essential contents of clinical performance examination (CPX). There are several studies to evaluate the ability of students or doctors in describing the information that they gathered during the encounter with standardized patients (SPs) [2,3,4,5,6,7]. However, until now, the ability has been evaluated only via checklists, completed by SPs and/or faculties, not by PN, in many medical schools in Korea. Students are requested to describe only the provisional diagnosis and future plan in PN. I think that there are, at least, two assumptions for not requesting Hx and PE information in PN.

At first, the examiners may assume that 'what students listen and find through Hx and PE, which are checked in the checklist' can be optimal representatives of 'what students consider as important information, which should be described in PN.' However, all checklist items are not necessarily described in PN. There are items that should be elicited through Hx or PE, but does not need to be recorded in PN. Therefore, it is yet to be proved if the students' ability to select and describe the information that should be recorded in PN can be assessed through the checklist. Second, the examiners may assume that it is possible to evaluate the clinical reasoning ability through a described provisional diagnosis and future plan in PN. However, shared information that the students received mostly before CPX from other students, who already had finished the same CPX, was revealed as a provisional diagnosis in Korea [8,9,10]. It means that many students already knew the provisional diagnosis of the case before they took the CPX. Therefore, it is very difficult to figure out if the students describe the diagnosis and plan in PN, through clinical reasoning during their own medical encounter with SP, or just through shared information from other students. So, these two assumptions for not requesting Hx and PE information in PN are not proved until now.

The purpose of this study is to verify the first assumption. If the assumption is proved to be true, the necessity of introducing PN of Hx and PE findings in CPX may be not high. But, if the first assumption is proved to be not true or suspicious, PN of Hx and PE findings should be introduced in CPX to evaluate the students' ability to select and describe the information that should be recorded, regardless of the results of the study to verify the second assumption.

SUBJECTS AND METHODS

A total of 54 final-year medical students of participated in a six-station CPX. The presenting problems of the stations used in CPX are described in Table 1. Each problem was portrayed by five SPs. SPs were trained until they could portray their role properly, and complete the checklist accurately. SPs completed their presenting-problem-specific checklist just after a 10-minute encounter with a student. The SPs had 5 minutes to complete the checklist, in which the student wrote PN outside of the station at the same time. The students were instructed to write the findings of Hx and PE in PN that they considered as important information in approaching the patient's problems. For scoring PN, key items were selected from the checklists of Hx and PE of each presenting problem by a consensus of five CPX experts (Table 1). For example, in the case of insomnia, nine items (such as onset of insomnia, current medication, and seven more items for example) were selected as key items among 17 Hx checklist items. Experts were all medical doctors who were faculty members at medical schools, and had in-depth experiences in authoring the scenarios and checklists of CPX, training SPs and managing the process of CPX. For analysis, key checklist items were classified into two criteria; the six presenting

problems (Table 1) and categories (Table 2). Two major categories were Hx and PE. Hx items were subdivided as cardinal features of presenting the problem (such as onset, duration, location, and so forth), associated symptoms and other findings (such as past, family and social histories, and so forth). Associated symptoms and other findings were further subdivided as positive and negative findings. PE items were also subdivided as positive and negative findings.

After completion of CPX, all PNs were examined whether the contents of the key checklist items were properly described in PN. The PN scores of an individual student were generated by counting the key checklist items properly described in PN. The checklist scores of an individual student were generated by counting the items checked as 'done' among the key checklist items. The scores of an individual student were rearranged and summed up according to the presenting problems and subcategories. To measure the level of concordance bet-

ween the PN scores and the checklist scores, according to the presenting problems and subcategories, correlation coefficients were calculated. To verify if there is any difference among the correlation coefficients, multiple comparisons between the correlation coefficients were performed at an internet site called, 'Tests for equality of several correlation coefficients' [11].

To measure the agreement between the key checklist items written in PNs and the same items checked as done in checklists, according to the subcategories of Hx and PE, the items were counted from all completed checklists and written PNs, and rearranged by four domains for 2×2 tables to calculate κ (kappa). The four domains were as follows: checked as 'done' in the checklist and written in PN, checked as 'done' in the checklist but not written in PN, checked as 'not done' in the checklist but written in PN, and checked as 'not done' in the checklist and not written in PN (Table 3). High kappa means that the incidence, that items which were checked as done in

Table 1. Presenting Problems and Number of Key Checklist Items of Each Problem

Presenting problem	No. of key checklist items	
	History taking	Physical examination
Fever	6	2
Heartburn	9	1
Polyuria	7	2
Abdominal pain	12	1
Insomnia	9	0
Dyspnea	8	2
Total	51	8

Table 3. 2x2 Table of Numbers of Items to Calculate Kappa

		No. of items written in patient note	
		Written	Not written
No. of items checked in checklist	Done	a[a]	b[b]
	Not done	c[c]	d[d]

[a]Checked as 'done' in the checklist and written in postencounter note (PN), [b]Checked as 'done' in the checklist but not written in PN, [c]Checked as 'not done' in the checklist but written in PN, [d]Checked as 'not done' in the checklist and not written in PN.

Table 2. Categories, Subcategories, and Sub-Subcategories of Key Checklist Items

Categories	Subcategories	Sub-subcategories	No. of key checklist items
History taking	Cardinal features of presenting problem		24
	Associated symptoms	Positive findings	12
		Negative findings	7
	Other findings	Positive findings	6
		Negative findings	2
Physical examination		Positive findings	5
		Negative findings	3

checklists were also written in PN, is high. To verify if there is any difference in the kappa according to the subcategories of Hx and PE, multiple comparisons were performed by the method described in an article [12].

To minimize the error of SP who completed checklist after a 10-minute encounter with a student, if I found the items checked as 'done' in the checklist and not written in PN, or checked as 'not done' in the checklist and written in PN, videos of students of the stations were reviewed to confirm if the students elicited the checklist items, and the items were corrected as a result of confirmation. In addition, the percentage of students who showed complete concordance in the key checklist items between PN and checklists were calculated.

RESULTS

The results, according to the presenting problems, were

summarized in Table 4. The correlation coefficients were moderate to high, and showed no statistically significant difference between the presenting problems. The kappas were fair to good, but showed statistically significant difference between some presenting problems. The percentages of students who showed complete concordance between PN and checklists were less than 40%.

The correlation coefficients according to the subcategories of Hx and PE are summarized in Tables 5 and 6. The correlation coefficients were moderate to high, and showed no statistically significant difference between the subcategories. In associated symptoms, the correlation coefficient of positive findings was significantly higher than that of negative findings.

The kappas (calculated from 2×2 tables of Tables 7-10) according to the subcategories of Hx and PE are summarized in Tables 5 and 6. The matches between completed checklists and PNs in key checklist items of subcategories of history taking were plotted at 2×2

Table 4. The Results according to the Presenting Problems

Presenting problem	Correlation coefficient	Kappa	Complete concordance (%)
Fever	0.70	0.60	19
Heartburn	0.55	0.45[a]	9
Polyuria	0.78	0.77[b,c]	31
Abdominal pain	0.55	0.65[b,d]	19
Insomnia	0.58	0.59[c,d]	24
Dyspnea	0.57	0.77	35

[a]p<0.05 fever vs. heartburn, [b]p<0.05 polyuria vs. abdominal pain, [c]p<0.05 polyuria vs. insomnia, [d]p<0.05 abdominal pain vs. insomnia.

Table 5. The Results according to the Subcategories of History Taking

		History taking					
		Associated symptoms			Other findings		
	Cardinal features	Positive findings	Negative findings	Total	Positive findings	Negative findings	Total
Correlation coefficient	0.48	0.84[a]	0.52	0.71	0.58	0.64	0.65
Kappa	0.62	0.80[a]	0.43	0.66[b]	0.64	0.63	0.64
Complete concordance (%)	2			9			22

[a]p<0.05 vs. pertinent (-) findings of same categories, [b]p<0.05 vs. cardinal features of chief complain, other findings and physical examination (correction for multiple comparison).

Table 6. The Results according to the Subcategories of Physical Examination

	Physical examination		
	Positive findings	Negative findings	Total
Correlation coefficient	0.62	0.34	0.58
Kappa	0.65[a]	0.56	0.62
Complete concordance (%)			22

[a]$p<0.05$ vs. pertinent (-) findings of same categories.

Table 7. Match between Completed Checklists and Postencounter Notes in Key Checklist Items of Cardinal Features of Presenting Problems

		Patient note		Total
		Written	Not written	
Checklist	Done	844	197	1,041
	Not done	4	251	255
Total		848	448	1,296

Table 8. Match between Completed Checklists and Postencounter Notes in Key Checklist Items of Associated Symptoms

		Patient note		Total
		Written	Not written	
Checklist	Done	273 [117]	62 [113]	335 [230]
	Not done	2 [4]	311 [144]	313 [148]
Total		275 [121]	373 [257]	648 [378]

Data are presented as positive findings [negative findings].

Table 9. Match between Completed Checklists and Postencounter Notes in Key Checklist Items of Other Findings

		Patient note		Total
		Written	Not written	
Checklist	Done	238	74	312
	Not done	1	119	120
Total		239	193	432

tables in Table 7-9. The match between completed checklists and PNs in key checklist items of physical examinations was plotted in Table 10. The kappas were fair to good, but the kappa of associated symptoms was significantly higher than other subcategories and PE. The kappas of positive findings in associated symptoms

Table 10. Match between Completed Checklists and Postencounter Notes in Key Checklist Items of Physical Examinations

		Patient note		Total
		Written	Not written	
Checklist	Done	121 [83]	49 [32]	170 [115]
	Not done	0 [3]	100 [44]	100 [47]
Total		121 [86]	149 [76]	270 [162]

Data are presented as positive findings [negative findings].

and PE were significantly higher than those of negative findings. The kappas according to the presenting problems were also calculated from 2×2 table, but not presented in tables.

The percentages of the students who showed complete concordance between PN and checklist were less than 25% (Tables 5, 6).

DISCUSSION

Although the correlation and agreement between PN and checklists of the key checklist items, according to the presenting problems or categories, were not low, there were significant discrepancies between them. There were significant differences at the level of agreement (kappa) between some presenting problems (Table 4). According to the results from categories of information, students neglected many key checklist items in their PNs, which were checked as 'done' in the checklists. They frequently neglected the cardinal features (Tables 5, 7), negative findings of associated symptoms (Tables 5, 8) and PE (Tables 6, 10), compared to positive findings. Only small percentages of students showed complete concordance of their individual data between the checklist and PN, and the percentages were different by the presenting problems and categories. It means that many students did not write important findings in PN (corresponded to medical record in real practice), even

though they asked questions or performed physical examinations about the findings when they interacted with SPs in CPX. They neglected many important findings. So, this result indicated that, without PN of Hx and PE, it is difficult to evaluate if the students reached the provisional diagnosis with reasonable clinical reasoning process.

Researches regarding the agreement between the checklist and PN had been conducted. Physical examinations were focused on one study [13]. They checked the PE that students took on SPs and analyzed PNs to identify if the students described the contents of PE on PNs. A total of 207 PNs from three presenting problems were analyzed, and the contents of PNs and items of checklists were identical to only 4% of the checklists and PNs, which were considered to be a very low rate compared with our study (22% in PE). The difference may be due to the selection of the checklist items. In their study, all PE checklist items were included in research. However, in our study, only a part of the checklist items, which were determined as the key information by CPX experts, were included, and other items that would be considered as not so important, were excluded. I believe that students could not estimate if they would get the key information before they interpreted the findings of Hx or PE. Thus, some of the checklist items should be performed, but do not have to be described in PN. If the checklists consist of not only items that contain key information, but also items that should be elicited, but might be turned out to be not so important information, the items that contain such key information should be selected for this kind of research. Although the percentages of our study were higher than the study of Szauter et al. [13], the percentage itself is still very low and conclusion of the two studies are not different in that assessing a written documentation of findings was needed for evaluating the clinical abilities of students.

Worzala et al. [14] also reported the results of comparing 219 students' written notes and checklists to examine the congruence between them, using a single case of 21 Hx and PE checklist items. In their study, students documented the findings for 71% of items checked as 'done' in the checklists and 6% of their documented findings were checked as 'not done' in the checklists (over-documentation). Comparing their result with our study, the rate of documented findings checked as 'done' in the checklist is similar (76% in our study), but less than 1% of students' documented findings were checked as 'not done' in the checklists of our study (Tables 7–10). It may also be due to a selection of the checklist items that used as keys to score the PN. In their study, 21 items, which were almost all the checklist items (out of 24 items), were selected in a single case. In our study, although it was various according to the presenting problems, only about half of the checklist items were selected, and some items that would be considered as keys in their study might not be selected as the key items in our study. Therefore, many documented findings that were decided as inconsistent with the checklists in their study would be disregarded from consideration in our study. In other words, the items selected in our study might be considered as more essential items that are directly related with the patients' problems, and the possibility that the students would write them in PN might be higher than their study. However, the possibility cannot be confirmed until all the checklist items of two studies are compared.

There are researches regarding the correlation between the checklist scores and PN scores of Hx and PE. Boulet et al. [15] reported the correlation between the checklist scores and PN scores of information gathering in the ECFMG® Clinical Skills Assessment (CSA®). They analyzed 61,497 completed checklists and PN, and the

correlation coefficient between the two scores was 0.51. Berg et al. [16] also conducted a similar study in the USMLE® step 2 Clinical Skill, and the correlation coefficient was 0.51. I presume that the differences between the correlation coefficients of these studies and our study are due to the methods of scoring PN. In USMLE step 2 CS or ECFMG CSA, a nine-point rating rubric was used to evaluate PN as a whole; whereas keys were selected from the checklists to evaluate the contents of PN in our study. Therefore, the scores from the completed checklists are more directly related to the PN score in our study, and the correlation coefficients are likely to be higher.

The methods for evaluating PN are divided into two main methods; scoring PN as a whole, which were used in USMLE® step 2 Clinical Skill until recently, and analyzing the contents of PN by using the keys, which were described in studies of Szauter et al. and Worzala et al. [13,14]. The latter method was used in our study. However, there are some unique features in our study compared with the other studies. I selected key items to evaluate PN from the checklists, and therefore, many checklist items of impertinent findings were not selected as the keys. I think that these items are not suitable to be used as the keys because, although the students should try to elicit the items in the process of information gathering from patients, the information, that turns out to be not related to the problems after Hx or PE, is not necessarily described in PN. Therefore, if it is planned that the checklist items are used as the keys to evaluate PN, the items that contain pertinent findings should be selected from the whole checklist items. How to select the key items from checklist may be an important issue. The method used in our study lays in extension of authoring scenario and checklist. The scenarios and checklists used in our study were finalized through a consensus of CPX experts. I think that the same process would be needed in a selection of key items from the checklists.

I think that the contents of PN should be analyzed not only according to the presenting problems, but also to the attributes of PN. It is necessary to estimate what are the weak or strong aspects of the students in describing the PNs. According to the results of our study, correlations and agreements between the checklists and PNs were quite different according to the attributes. The students neglected the pertinent negative findings more frequently than the positive findings in associated symptoms and PE, and neglected the associated symptoms less frequently than other attributes of PN. These results reflect that the students neglect the negative findings more often in their process of clinical reasoning to approach the patient's problems, and they concentrate more on the associated symptoms, especially positive symptoms. Walling et al. [17] also reported similar results. They compared reporting rates for pertinent positive and negative items on two cases in a clinical skills assessment taken by all 55 third-year students. Scoring of PNs found significant differences in the reporting rates for positive (75%) and negative (52%) items.

There are some discrepancies between completed checklists and corresponding PN. Therefore, an evaluation through the checklist cannot be representative of an evaluation by PN. I think that it is necessary to introduce PN, including gathered information in CPX, to evaluate the students' ability of synthesis and integration of patient information. But this study was performed in only one medical school with small number of examinees and it is not possible to generalize the results of this study without further studies. So, further studies using more examinees of other medical schools, more cases of different clinical presentations and different time plans (for example, giving more time for completing PN)

should be performed to generalize the results of this study and, especially, to use the PN of this study in national medical license examination.

In addition, by analyzing PNs, it is possible to explore how students prioritize important information they have obtained from patients and to provide feedback to students on how to record the information. But, additional effort of the faculties will be needed to get these advantages. So it is necessary to develop convenient scoring systems to evaluate PN of gathered information, and method that faculties can use to provide feedback to their students.

REFERENCES

1. Weber DO. Charting a course toward legible medical records: perfect paperwork can mean financial savings, better patient care. Physician Exec 2002; 28: 8-13.

2. MacRae HM, Vu NV, Graham B, Word-Sims M, Colliver JA, Robbs RS. Comparing checklists and databases with physicians' ratings as measures of students' history and physical-examination skills. Acad Med 1995; 70: 313-317.

3. Dresselhaus TR, Peabody JW, Lee M, Wang MM, Luck J. Measuring compliance with preventive care guidelines: standardized patients, clinical vignettes, and the medical record. J Gen Intern Med 2000; 15: 782-788.

4. Macmillan MK, Fletcher EA, Champlain AF, Klass DJ. Assessing post-encounter note documentation by examinees in a field test of a nationally administered standardized patient test. Acad Med 2000; 75(10 Suppl): S112-S114.

5. Boulet JR, McKinley DW, Norcini JJ, Whelan GP. Assessing the comparability of standardized patient and physician evaluations of clinical skills. Adv Health Sci Educ Theory Pract 2002; 7: 85-97.

6. Norcini J, Boulet J. Methodological issues in the use of standardized patients for assessment. Teach Learn Med 2003; 15: 293-297.

7. Durning SJ, Artino A, Boulet J, Van Der Vleuten CP, La Rochelle J, Arze B, Schuwirth L. Making use of contrasting participant views of the same encounter. Med Educ 2010; 44: 953-961.

8. Park HK, Kwon OJ. Sharing of information among students and its effect on the scores of clinical performance examination (CPX). Korean J Med Educ 2005; 17: 185-196.

9. Kim J, Lee K, Yoo D, Yang E. The effects of information sharing between students on results of clinical performance examination. Korean J Med Educ 2006; 18: 239-247.

10. Kim JH. Does sharing information before a clinical skills examination impact student performance? Med Teach 2010; 32: 747-753.

11. Tests for equality of several correlation coefficients [Internet]. Hossein Arsham; c2013 [updated 2013 July 1; cited 2013 October 10]. Available from: http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/MultiCorr.htm.

12. Donner A, Eliasziw M, Klar N. Testing the homogeneity of kappa statistics. Biometrics 1996; 52: 176-183.

13. Szauter KM, Ainsworth MA, Holden MD, Mercado AC. Do students do what they write and write what they do? The match between the patient encounter and patient

note. Acad Med 2006; 81(10 Suppl): S44-S47.

14. Worzala K, Rattner SL, Boulet JR, Majdan JF, Berg DD, Robeson M, Veloski JJ. Evaluation of the congruence between students' postencounter notes and standardized patients' checklists in a clinical skills examination. Teach Learn Med 2008; 20: 31-36.

15. Boulet JR, Rebbecchi TA, Denton EC, McKinley DW, Whelan GP. Assessing the written communication skills of medical school graduates. Adv Health Sci Educ Theory Pract 2004; 9: 47-60.

16. Berg K, Winward M, Clauser BE, Veloski JA, Berg D, Dillon GF, Veloski JJ. The relationship between performance on a medical school's clinical skills assessment and USMLE Step 2 CS. Acad Med 2008; 83(10 Suppl): S37-S40.

17. Walling A, Moser SE, Dickson G, Zackula RE. Are students less likely to report pertinent negatives in post-encounter notes? Fam Med 2012; 44: 22-25.