

薬学 OSCE におけるコミュニケーション能力評価の信頼性に関する検討

窪田愛恵,^{a,d} 矢野義孝,^{*,a,d} 森本 剛,^b 関 進,^b 高田香織,^b
蔵本伸生,^b 前田祐子,^c 赤池昭紀,^a 平出 敦^b

Reliability of Evaluation for Communication Competency in Pharmaceutical Objective Structured Clinical Examination (OSCE)

Yoshie KUBOTA,^{a,d} Yoshitaka YANO,^{*,a,d} Takeshi MORIMOTO,^b
Susumu SEKI,^b Kaori TAKADA,^b Nobuo KURAMOTO,^b
Yuko MAEDA,^c Akinori AKAIKE,^a and Atsushi HIRAIDE^b

^aGraduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida Shimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan, ^bCenter for Medical Education, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan, ^cSchool of Health Sciences, Faculty of Medicine, Kyoto University, 53 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan, and ^dCenter for Integrative Education of Pharmacy Frontier, Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida Shimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan

(Received November 21, 2008; Accepted February 19, 2009)

In the objective structured clinical examination (OSCE), which for pharmaceutical students training, adequate methods should be used for evaluating a student's skill and aptitude for good communication in a medical interview. However, the reliability of the evaluation methods used in the pharmaceutical OSCE has not been investigated sufficiently. In this study, we reviewed the evaluation scores and video recordings obtained in a pharmaceutical OSCE trial, and examined the reasons for disagreement in the scores between two raters. We had two experienced raters in medical communication re-evaluate the students using the video images, and compared their scores with those on the examination day. The ratio of disagreement was 14.5% (87/600 items in 30 students), and the reason for disagreement could not be identified for 63 items that evaluated communication skills such as 'actively listen' and 'empathy'. A comparison of the scores on examination day and those on re-evaluation revealed a possible reason for the disagreement; the use of a checklist, *i.e.* binary scores, with criteria that differed between the raters. We suggest that the items used for a detailed performance evaluation be selected carefully and that rating scales be used in order to perform an adequate evaluation, especially regarding communication skill and aptitude.

Key words—objective structured clinical examination (OSCE); communication; medical interview; pharmacist training; checklist; rating scale

緒 言

薬学教育においては、医学教育と同様に平成 21 年度から共用試験が実施されることになっている。そのうち特に学生の技能・態度を評価する試験である客観的臨床能力試験 (Objective Structured Clinical Examination; OSCE) に関してはその実施体制

の確認あるいは課題や評価内容の検討を目的としたトライアルが全国の大学で実施されている。¹⁻³⁾ 薬学 OSCE で実施される課題は 5 領域から 6 課題と予定されており、⁴⁾ 患者との良好なコミュニケーションを行うための技能・態度についても取り上げられ医療面接課題として患者・来局者対応及び情報提供 (服薬指導) といった内容の課題が検討されている。

われわれは既に、医学部におけるコミュニケーション教育に取り組む一方で医学英語教育に関連した異文化コミュニケーションに関する研究や医療面接における非言語的コミュニケーション情報の解析に

^a京都大学薬学研究科, ^b京都大学医学研究科医学教育推進センター, ^c京都大学医学研究科人間健康科学系専攻, ^d京都大学薬学研究科統合薬学フロンティア教育センター

*e-mail: yano@pharm.kyoto-u.ac.jp

関する研究⁵⁻⁸⁾を行ってきた。これを基盤に学部横断的な取り組みとして、京都大学薬学部の OSCE において特に医療面接に関して医学部及び薬学部の教員が連携を取り、事前教育や模擬患者養成など実施体制の確立を目指してきた。これらの経験を通じて、医師養成のための医学教育と薬学部における薬剤師教育とは職種の違いによる技能面の教育内容には違いがあるものの、患者や医療スタッフ等とのコミュニケーションに関する教育においては医療人教育として共通であることを認識している。医学部 OSCE においてもコミュニケーション能力の評価に関する報告があるが、⁹⁻¹⁵⁾ 薬学 OSCE においても医学部と同様に 1 人の学生に対して複数の評価者が目の前の学生の実技を実地試験として評価する形態を取る。評価シートは個別的な評価項目である「細目評価 (detailed performance)」及び全体の印象で判断する「概略評価 (global rating)」に大きく分かれている。評価の信頼性を高めるためにはそれぞれの項目における評価結果が評価者間でよく一致していることが望ましく一致度¹⁶⁾がその指標として用いられる。われわれは既に薬学 OSCE において医療面接の技能・態度を評価する際、「身だしなみ」といった総合的な判断を要する項目や「姿勢はよいか」「傾聴の態度が見られるか」といった項目に関して一致度が低いことを報告し、また経験の少ない評価者ではたとえ評価マニュアルを理解したとしても 5 分間という短い課題遂行時間の中で適切な評価が行えない可能性も示唆した。¹⁷⁾ このように薬学 OSCE においてはその評価方法の確立や評価者の養成が重要な問題となっている。そこで本研究では、実際の薬学 OSCE トライアルの成績を基に評価の現状を分析し、薬学 OSCE において適切なコミュニケーションの技能・態度の評価を行うために必要な条件について評価項目や評価手法の面から検討を行った。

方 法

京都大学薬学部において 2008 年 3 月 1 日に実施した OSCE トライアルでの結果を用いた。このトライアルでは初回面接 (薬局での患者対応)、散剤調剤、水剤調剤、調剤鑑査、注射薬混合 (無菌操作)、薬局での薬剤交付 (情報提供) の 6 課題からなる試験を実施し、京都大学薬学部学生 (平成 17 年度入学、実施当時 3 年生) 29 名と大学院生 2 名

の合計 31 名が受験生として参加した。このうち患者対応に関する技能・態度を問う「初回面接」と題した課題実施状況を録画した。課題の内容及び評価項目は、薬学共用試験センターで作成されている標準的な課題を基に本学で独自に設定したシナリオに従って実施した。

OSCE トライアル当日の評価 (evaluation on examination day) において 2 名の評価者間の評価が一致しなかった評価項目について録画画像・音声で確認することにより不一致の理由を調査した。そこで、1) 見落としなど明らかに評価者の判断ミスと考えられるもの、2) 模擬患者が回答を誘導したために評価が分かれたと考えられるもの、3) 受験生の発言内容があいまいであったため評価が分かれたと考えられるもの、4) そのいずれにも当てはまらないものについて分類した。

次に OSCE トライアルの約 2 ヶ月後に、録画画像・音声を利用できなかった 1 名を除く 30 名の学生の課題実施状況について、コミュニケーションの技能・態度に限定した再評価を 2 名の評価者により録画画像・音声を再生して行い当日評価の検証を行った。これを検証評価 (re-evaluation) と称することとする。このときの 2 名の評価者は、(社)医療系大学間共用試験実施評価機構が開催する評価者認定講習会を受講した認定評価者であり、医学部、薬学部双方で OSCE の評価者としての経験を有し医療コミュニケーション教育に携わる京都大学医学部教員とした。

当日の評価では細目評価の中でコミュニケーション能力に特に関連する 6 項目「声の大きさ」、「アイコンタクト」、「傾聴」、「共感」、「丁寧な言葉遣い」を取り上げ、検証評価ではさらにコミュニケーション能力に関連する項目を補足して 12 項目とした (Table 1)。また当日評価では細目評価項目にはチェックリスト (「1:できた」、「0:できなかった」で評価) を用い、概略評価のみ 6 段階の評定尺度を用いたのに対し、検証評価では細目評価項目、概略評価ともに 5 段階の評定尺度を用いた。評定尺度を用いることによりコミュニケーションの技能・態度を総合的に評価し、評価結果の微妙な差を捉えることをねらった。当日評価は評価者として医学部教員 5 名と薬学部教員 5 名の、延べ 10 名の評価者がそれぞれペアとなって各 6-7 名の受験生を評価したの

Table 1. List of Evaluation Items in the OSCE Trial

| Evaluation on examination day | Re-evaluation using video recordings |
|---|---|
| Detailed performance (checklists) | Detailed performance (5-point rating scales) |
| Opening: 4 items | |
| Information gathering: 8 items | |
| Closing: 1 item | |
| Suitable appearance for medical interview: 1 item | |
| Communication competencies: 6 items | Communication competencies: 12 items |
| Posture, Gestures, Movement | (1) Appropriate distance to patient |
| Voice volume | (2) Posture, Gestures, Movement |
| Eye contact | (3) Voice volume |
| Actively listen | (4) Speed |
| Empathy | (5) Eye contact |
| | (6) Nodding, Non-verbal facilitation |
| | (7) Actively listen |
| | (8) Empathy |
| | (9) Balance in interview between speaking and listening |
| | (10) Coherence |
| | (11) Fluent and smooth conversation |
| | (12) Linguistic politeness |
| Linguistic politeness | |
| Global rating (6-point rating scale) | Global rating (5-point rating scale) |

に対し検証評価では2名の評価者が30名のすべての受験生を評価した。

当日評価と検証評価の間の比較検討の方法については、最初に細目評価と概略評価との関連性を検討した。なお細目評価と概略評価の相関の検討については2名の評価者のスコアの合計を用いた。次に、個々の細目評価項目について当日評価スコアと検証評価スコアとの関連性を調べた。具体的には上記の6項目について当日評価と検証評価のスコアに関して Spearman の相関係数を算出した。

データの集計及び統計解析には統計解析ソフトウェア jmp® (SAS Institute Japan), あるいは Microsoft Excel を用いた。

結 果

当日評価において2名の評価者間での評価が一致しなかった項目は87件(全20項目×30名=600件中14.5%)であり、録画画像・音声で検討したところ評価者の明らかな判断間違いによるものが15件、模擬患者による回答誘導によるものが3件、受験生のあいまいな発言によるものが5件であった。しかし残りの63件ではこうした原因が特定できな

かった。特定できなかった63件のうち不一致が多かった項目は、「みだしなみ」11件、「傾聴」10件、「共感」8件、「声の大きさ」5件、「姿勢」4件、「アイコンタクト」3件であり、コミュニケーション能力に関連する評価項目において評価者間の不一致が数多くみられた。特に「傾聴」、「共感」では30名の受験生のうちの17名においていずれかの項目が不一致であった。

以上のように当日評価では評価者間で一致率の低い評価項目がみられた。その原因を明らかにするために実施した録画画像・音声による検証結果を以下に示す。

まず当日評価と検証評価で細目評価と概略評価との相関性について比較検討を行った。当日評価スコアについて、すべての細目評価項目と概略評価との相関 [Fig. 1(a)] は $\rho=0.706$ ($p<0.01$) であり、検証スコアについて同様の検討をしたところ、すべての細目評価項目と概略評価との相関 [Fig. 1(b)] は $\rho=0.831$ ($p<0.01$) であった。これらの相関係数には有意な差はみられないものの、検証評価においては評定尺度を用いているため細目評価のスコアが広く分布しているのに対して、当日評価ではチェ

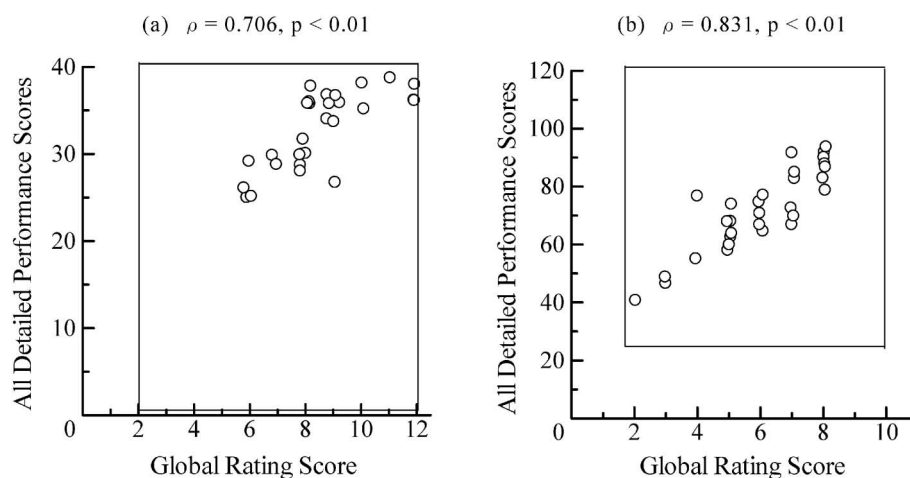


Fig. 1. Relationships between the Global Rating Scores and the Detailed Performance Scores

Sum of the scores by two raters are plotted. For the detailed performance scores, sum of the scores for all items are used. (a): scores on examination day, (b): scores on re-evaluation. The lines indicate the possible ranges of the plots.

ックリストを用いているため細目評価の取りえる幅が狭くなっている。さらにコミュニケーション能力に関するそれぞれの細目評価について当日評価と検証評価とを比較検討した (Fig. 2)。いずれの細目評価でも検証スコアはおおよそ 3-9 の範囲で広く分布した。これに対し当日評価スコアの合計は 3 段階のみであり、例えば「丁寧な言葉遣い」では、当日評価ではほとんどが 2 であったのに対して、検証評価では広くばらついて両者は関連しなかった。また「傾聴」の項目では当日スコア合計が 0, 1 あるいは 2 いずれの場合でも検証スコアは 4-8 程度にばらついており相関は全くみられず客観性の面からみて評価が難しいことが示唆された。当日評価スコアと検証評価スコアの関連を示す Spearman の相関係数 (ρ) は、「2:姿勢: $\rho=0.192$ 」,「3:声の大きさ: $\rho=0.566$ 」,「5:アイコンタクト: $\rho=0.409$ 」,「7:傾聴: $\rho=0.003$ 」,「8:共感: $\rho=0.559$ 」,「12:丁寧な言葉遣い: $\rho=0.020$ 」であり、「3:声の大きさ ($p<0.01$)」,「5:アイコンタクト ($p<0.05$)」,「8:共感 ($p<0.01$)」のみで有意な相関がみられた。

その内容をさらに詳細に検討するため「傾聴」に関して同じ受験生 6 名を評価したときの評価者 2 名の当日評価スコアと検証評価スコアとの関係を一部の評価において検討した [Fig. 3(a)]。1 名の評価者 (左) は検証評価スコアが 6 付近に「1:できた」と「0:できなかった」の境界があるが、他方の評価者 (右) では全受験生で「1:できた」と判定しており境界レベルがより低いと考えられる。また

「共感」について同じ受験生で検討してみると Fig. 3(b) のようになり、評価者間で比較的類似したパターンを示すものの両評価者での境界点と思われる検証スコアが 4 付近で当日評価が一致していない。つまり、たとえそれぞれの評価者内で評価基準に一貫性があったとしても評価者間での基準に一貫性がない場合には全体としての客観性が低下し、このように当日評価と検証評価のスコアに相関がみられなくなる。また当日評価スコアはチェックリスト (2 段階評価) によるスコアであり、このことが特に境界点付近において評価結果が一致しない原因の 1 つになっていた。

考 察

評価の客観性とは誰が実施しても評価結果が同じかどうか (結果の一致性) をいうが,¹⁸⁾ OSCE において評価の客観性を高めることは、評価結果の説明責任を果たす意味で極めて重要である。われわれは薬学 OSCE におけるコミュニケーション能力に関する評価の客観性について包括的な検討を進めている。前回の報告では、特に現在用いられている細目評価においては極めて繁忙な OSCE 評価の状況できめ細かな評価が難しいのではないかという点を指摘した。¹⁷⁾ 今回の報告はわれわれの一連の研究の中で評価の客観性に焦点を当て評価者間の一致度の視点から検討したものである。

OSCE 当日の結果に関して、2 人の評価者の不一致は細目評価全体の 14.5% という結果であった

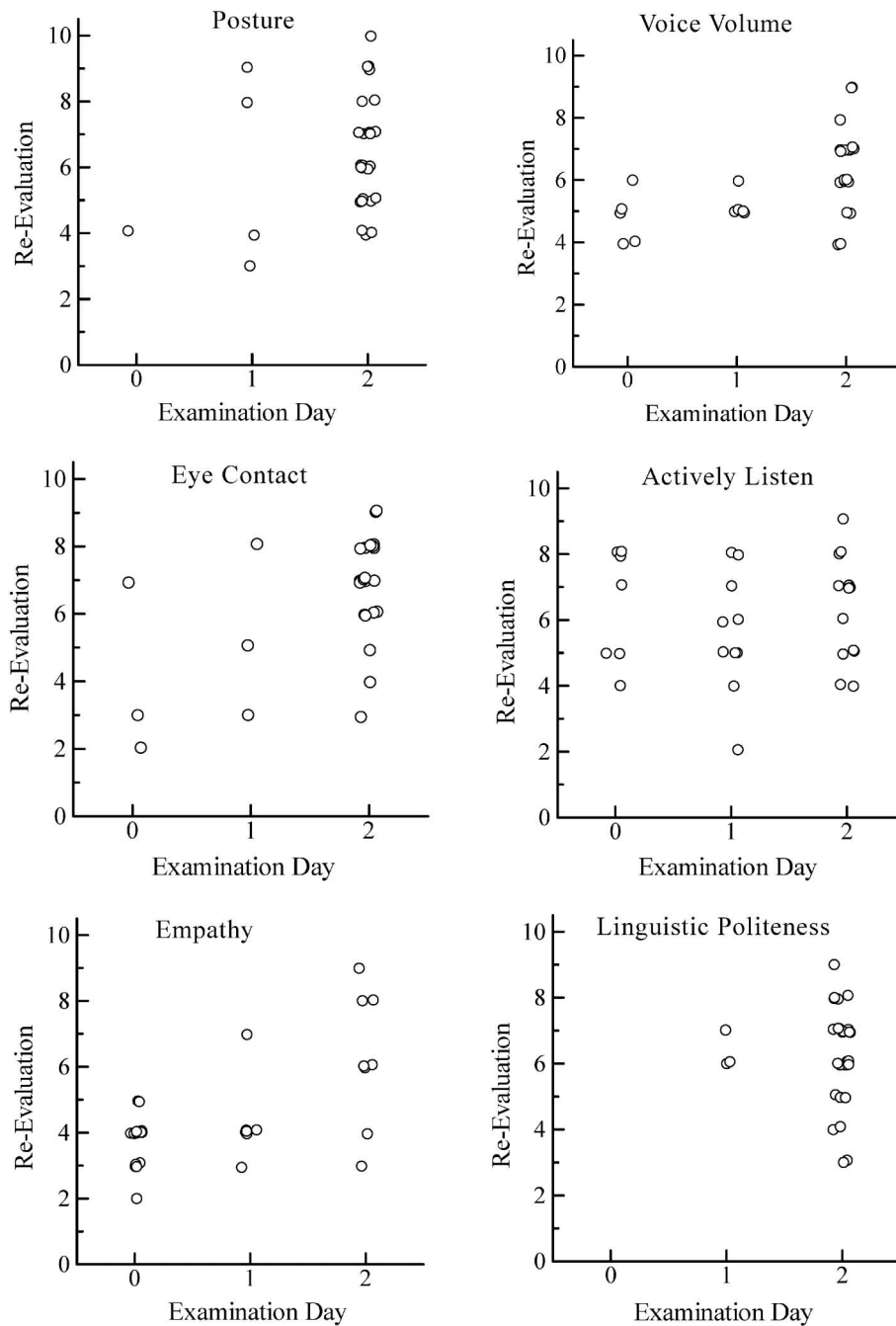


Fig. 2. Relationships between the Scores on Examination Day and the Scores on Re-evaluation for Each Item of Detailed Performance

Sum of the scores by two raters are used.

が、このことは今回の細目評価がチェックリストを用いていることを考慮すると過小評価できない。すなわちこの不一致とは単に相対的な判断のずれというのではなく可か不可かの両極に割れているからである。その不一致の原因に関して、録画画像・音声で原因を特定できなかったケースが87件中63件にも及んだことはさらに注視すべきことである。録画画像・音声で第3者が検討して、評価者の判断間違

いや模擬患者の誘導など明らかに特定できる原因を指摘できないということは、原因を評価者同士の内的な判断基準の相違に求めることが自然である。それを裏付けるように、この原因が特定できなかった不一致はコミュニケーションの能力を評価する細目項目で著明であり、特に傾聴、共感に関しては不一致が目立った。すなわち、傾聴、共感といった部分の評価は評価者の内的な判断基準に依存している部

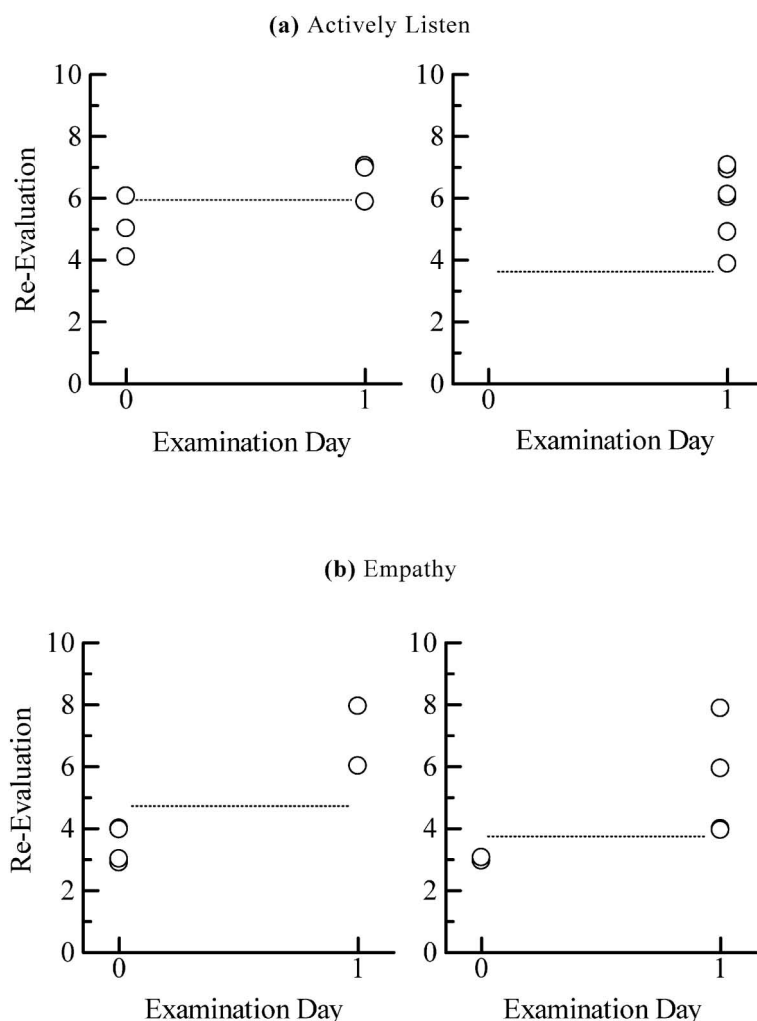


Fig. 3. Relationship between the Scores on Examination Day and the Scores on Re-evaluation for Items of (a) Actively Listen, and (b) Empathy, in Each Rater

The scores on the examination day by two raters who evaluated the same 6 students are compared with the scores on re-evaluation. Plots show individual scores.

分があるからである。このことはコミュニケーション能力の評価の困難さを物語っているとみられる。例えば傾聴の項目を整理してみると、評価が一致せずかつその原因が特定できなかったケースは30名の受験生のうち1/3に上っており、さらに傾聴と共感のいずれかがこれに当てはまる受験生は過半数に上った。これらの高次な能力を評価することが容易でないことを浮き彫りにしている。

細目評価は、評価者に対して一定の確認項目を具体的に示すことにより質の高い評価を実現する手掛かりになるものといえるが、われわれは現時点では細目評価に重点を置くよりも概略評価に重点を置いた評価の方が妥当であることを示唆した。¹⁷⁾ 今回、細目評価と概略評価の相関を検討した結果では、当日評価と検証評価ともに両者はよく相関していた。

しかし当日評価では細目評価の幅は狭く、広くスコアが分布した概略評価と対照的であった。最終的に、受験者が実際に患者と接することを想定して総合評価した場合に、特にコミュニケーション能力に関する細目評価の幅をある程度広く取り、きめ細かい評価によりその質を高めることは評価の水準を向上させるために不可欠と考えられる。医療面接の課題では、コミュニケーション能力の評価に加えて聞き取り項目の確認に関する評価も含まれているが、短い課題遂行時間の中ですべての評価を適切に行うことは難しい。聞き取り項目に関しては単にその項目について触れたかどうかではなくむしろ、重要な事項を聞くことができたか、確認できたかといった点を評価することが重要である。このように考えると、例えば面接終了後に受験生に聞き取った項目を

書き取る時間を設け、この受験者の記載したシート等を評価するなどして別の方法で評価し、実際の面接では評価者は検証評価のときのようにコミュニケーション能力の評価に集中するといった考え方もあってよいのではないと思われる。

このように考えると、やはり傾聴や共感といったより深い意味でのコミュニケーション能力の評価をどのように行うかということが問題となる。これらの項目に関してそれぞれ当日評価スコアと検証評価スコアとを比較検討した結果 (Fig. 3) は、すべてのケースに当てはまる訳ではないが極めて示唆的である。すなわち、先に推測したように細目評価での不一致は評価者の内的な判断基準に依存しているために、チェックリストを用いた評価では評価が可と不可に割れてしまう。しかし、評定尺度を用いた評価によれば比較的妥当な評価に近づくことが予想される。コミュニケーション能力に関する細目評価にきめ細かな評価を行った場合、これらの細目評価の内容が積み重なって概略評価への確かな手掛かりになっている可能性がある。なお、チェックリストと評定尺度との比較を行う際には同じ評価者で両方の評価を行うことが理想的であるが、実際には評定尺度とチェックリストで独立性が保たれないことも懸念される。このような問題を解決する方法としては、それぞれの評価を行うためにかなりの時間をおくことや多人数の評価者を割り付けたりすることも考えられるが、この点については今後の課題の1つである。

以上、薬学 OSCE における患者・来局者対応、情報提供のステーションの評価に関しては、共感、傾聴などの高次なコミュニケーション能力を問う細目評価において評価者間の不一致が問題であった。模擬面接をみて評価する細目評価に関してはコミュニケーション能力に関する評価項目に集中し、これらの細目評価に評定尺度も導入して実施することにより的確な概略評価の手掛かりとなるきめ細かな評価体系を構築できるのではないかと考える。

謝辞 京都大学薬学部における OSCE トライアルの実施に当たりご支援頂いた、京都府薬剤師会、京都府病院薬剤師会、京都大学医学部附属病院薬剤部、京都大学模擬患者の会、京都薬科大学、摂南大学薬学部、同志社女子大学薬学部、京都大学薬

学研究科、の各位に深謝致します。

この論文内容の一部は、文部科学省平成 18 年度採択「地域医療等社会的ニーズに対応した質の高い医療人養成推進プログラム (医療 GP) —先端医療の育・創薬を先導する薬剤師養成—」の支援によって行われました。

REFERENCES

- 1) Kogo M., Koyama N., Negoro T., Aoki K., Saito I., Kobayashi Y., Mashimo J., Sasaki K., Tobe T., Yamamoto T., Kiuchi Y., Sato H., *Yakugaku Zasshi*, **127**, 905–917 (2007).
- 2) Saito M., Nakajima Y., Kosano H., Kurihara J., Watanabe M., Fujii Y., Yanagawa Y., Yamaoka K., Inoue K., *Jpn. J. Pharm. Health Care Sci.*, **33**, 741–747 (2007).
- 3) Saito I., Mashimo J., Sasaki K., Kogo M., Kurata N., Yamamoto T., Kiuchi Y., Sato H., Tobe T., *Jpn. J. Pharm. Health Care Sci.*, **34**, 805–810 (2008).
- 4) Irie T., *J. Jpn. Soc. Hosp. Pharm.*, **43**, 1361–1364 (2007).
- 5) Maeda Y., Yamaoka A., Tabuchi M., Kubota Y., Hiraide A., *Med. Educ. (Japan)*, **37**, 83 (2006).
- 6) Seki S., Morimoto T., Maeda Y., Kuramoto N., Kubota Y., Takada K., Hiraide A., *Med. Educ. (Japan)*, **38**(S1), 39 (2007).
- 7) Maeda Y., Morimoto T., Kuramoto N., Kubota A., Seki S., Katada K., Hiraide A., *Med. Educ. (Japan)*, **38**, 23 (2007).
- 8) Kobori E., Maeda Y., Kubota Y., Seki S., Takada K., Kuramoto N., Hiraide A., Morimoto T., *Gen. Med.*, **9**, 5–12 (2008).
- 9) Suzuki E., Ito M., Aoyagi Y., Fuse I., Tanaka K., Naito M., Yamamoto M., *Med. Educ. (Japan)*, **34**, 37–44 (2003).
- 10) Fukumoto Y., Murakami F., Kobayakawa S., Ono S., Murakami Y., Tamura I., Kawasaki M., *Med. Educ. (Japan)*, **35**, 229–234 (2004).
- 11) Regehr G., MacRae H., Reznick R. K., Szalay D., *Acad. Med.*, **73**, 993–997 (1998).
- 12) Hodges B., McIlroy J. H., *Med. Educ.*, **37**, 1012–1016 (2003).
- 13) Yedidia M. J., Gillespie C. C., Kachur E., Schwartz M.D., Ockene J., Chepaitis A. E.,

- Snyder C. W., Lazare A., Lipkin Jr. M., *JAMA*, **290**, 1157–1165 (2003).
- 14) Mukohara K., Kitamura K., Wakabayashi H., Abe K., Sato J., Ban N., *BCM Med. Educ.*, **4**, 1–6 (2004).
- 15) Jefferies A., Simmons B., Tabak D., Mcilroy J. H., Lee K.-S., Roukema H., Skidmore M., *Med. Teach.*, **29**, 183–191 (2007).
- 16) Landis J. R., Koch G. G., *Biometrics*, **33**, 159–174 (1977).
- 17) Kubota Y., Yano Y., Morimoto T., Takada K., Kuramoto N., Seki S., Maeda Y., Akaike A., Hiraide A., *Jpn. J. Pharm. Health Care Sci.*, **34**, 1004–1010 (2008).
- 18) “Health Professional Workshop Guide” ed. by Japan Society for Medical Education, FD Committee, 2008.