Health Professions Education

Volume 3 | Issue 2 Article 4

2017-09-30

Validating OSCE Performance: The Impact of General Intelligence

Paul F. Wimmers
David Geffen School of Medicine, University of California, 60-051 Center for Health Sciences, 10833 Le Conte Avenue, Los Angeles 90095, CA, United States, pwimmers@ucla.edu

Guido F. Schauer
The University of Texas at Dallas, United States

Recommended Citation

Wimmers, Paul F. and Schauer, Guido F. (2017) "Validating OSCE Performance: The Impact of General Intelligence," *Health Professions Education*: Vol. 3: Iss. 2, Article 4.
DOI: 10.1016/j.hpe.2016.12.002
Available at: https://hpe.researchcommons.org/journal/vol3/iss2/4

Available online at www.sciencedirect.com
ScienceDirect
Health Professions Education 3 (2017) 79–84
www.elsevier.com/locate/hpe

Validating OSCE Performance: The Impact of General Intelligence

Paul F. Wimmers[a,*], Guido F. Schauer[b]

[a]David Geffen School of Medicine, University of California, 60-051 Center for Health Sciences, 10833 Le Conte Avenue, Los Angeles 90095, CA, United States
[b]The University of Texas at Dallas, United States

Received 28 June 2016; accepted 17 December 2016
Available online 22 December 2016

Abstract

Purpose: To investigate the relationship between medical students' eductive ability as measured by the advanced version of the Raven's Progressive Matrices (RPM) test, reproductive ability as measured by performance on the United States Medical Licensing Examination (USMLE) Step I, and Objective Structured Clinical Examination (OSCE) performance.
Method: Thirty-two third-year medical students took the Advanced Progressive Matrices (APM) online, which consists of two parts: (1) a practice set of 12 items, and (2) 36 items which become progressively more difficult as the test proceeds. Several models representing different causal structures are tested and compared.
Results: Comparison of the different structural models revealed that eductive reasoning ability better predicted OSCE performance than reproductive ability.
Discussion: The relationship between APM and OSCE performance indicates that more in-depth research in domain-general abilities is important.

Keywords: General intelligence; Objective Structured Clinical Examination (OSCE); Raven's Progressive Matrices (RPM); Structural Equations Modeling (SEM); United States Medical Licensing Examination (USMLE) Step I

1. Introduction

The popularity of the Objective Structured Clinical Examination (OSCE) in medicine has grown substantially since its inception in the late 70s.[1] A survey conducted by Hauer, Hudgson, Kerr, Teherani, and Irby[2] showed that approximately 84% of the medical schools in the US use standardized patients in the assessment of students' clinical skills. And although the utilization and practicality of OSCEs are well-discussed in the literature, systematic investigations of validity and reliability are limited and show mixed results.[3,4] OSCEs reliability coefficients commonly range between .4 and .8.[5] Threats to reliability, in part caused by inconsistent performance of students across cases, could be avoided by increasing test time and the number of stations.[6,7] For good reliability, minimum test lengths of 3–4 h and a minimum of 10 stations are suggested.[8,9] The fact that clinical competence is measured over multiple (and often very short) stations and spread over various clinical tasks makes it hard to

*Corresponding author.
E-mail address: pwimmers@ucla.edu (P.F. Wimmers).
Peer review under responsibility of AMEEMR: the Association for Medical Education in the Eastern Mediterranean Region

http://dx.doi.org/10.1016/j.hpe.2016.12.002

establish the validity of an OSCE. Comparing performance of a complex tool like an OSCE with a single construct will yield low-results.[3]

In the current study we use skill components of an OSCE and hypothesize that components of intelligence predict OSCE performance to some extent. Spearman[10] identified a common or general underlying factor "g" in any mental ability. He further identified two components of g named "eductive ability" and "reproductive ability". Eductive mental activity involves making meaning out of confusion; developing new insights; going beyond the given to perceive that which is not immediately obvious; forming largely non-verbal constructs that facilitate the handling of complex problems involving many mutually dependent variables. Reproductive ability involves mastering, recalling and reproducing the largely verbal material which forms a cultural store of explicit knowledge.

The Raven's Progressive Matrices Tests has been widely used for decades as a measure of *eductive ability*—"the ability to evolve high-level constructs which make it easier to think about complex situations and events".[11,12] Carpenter, Just, and Shell[13] described the Raven's Progressive Matrices as "a classic test of *analytic intelligence* … the ability to reason and solve problems involving new information, without relying extensively on an explicit base of declarative knowledge derived from either schooling or previous experience".

USMLE Step 1, on the other hand, is a knowledge-based test that assesses whether students understand and can apply important concepts of the medical sciences and in this study is used to identify students' reproductive ability. Some studies have shown a weak till moderate relationship between OSCE performance and USMLE Step I.[14,15]

Three hypothesized models that reflect different causal relationships between the OSCE components, USMLE Step I and the Raven Progressive Matrices were analyzed and compared via Structural Equation Modeling (SEM).

2. Material and methods

The data used in this study were collected from third-year medical students (class of 2011). Students were recruited by email. All third-year students received an email request to participate in study that looks at the effect of general reasoning skills on USMLE Step I and Clinical Performance Exam (CPX, or more commonly known as OSCE). Thirty-three took the Raven's Advanced Progressive Matrices test. One student was excluded for exceeding the allotted test time without completing the test. Students received a $25 gift card for their participation.

2.1. Measurement instruments

2.1.1. Raven's Progressive Matrices

Raven's Progressive Matrices (RPM) tests were designed to measure abstract reasoning ability. The tests are non-verbal, picture-based, and require minimal instructions to administer. RPM tests correlate well with many other measures of intelligence and in factor analyses have been found to be the most central measure of g_f, or general fluid intelligence.[13]

Of the several versions of the RPM, the Advanced Progressive Matrices (APM) test is most appropriate for above average adults. Each APM test problem is comprised of black-and-white line drawings of figures in three-by-three arrays, with each part containing one or more figural elements—except for the lower right part, which is left blank. The task for the problem solver is to determine which part from a set of eight figure choices (which include perceptual and relational lures) would best complete the figure pattern.

Each part within an array is related to its neighbors by discernible rules, such that how the array should be completed (from the available options) can be determined by grasping and applying these rules. Simpler arrays can sometimes also be completed by perceptual pattern completion.

APM test items are divided into two sets, which differ only in difficulty. Both sets begin with easy problems, with each successive problem generally more difficult than the last. Most Set I problems are relatively easy (for above average adults), whereas Set II problems range from relatively easy to very difficult. For this study, Set I, with 12 problems, was used to familiarize students with the format of the test. Set II, which has 36 items, were used for testing, proper. Experimental subjects had 45 min to complete the test. The score received on the APM and used in our analysis was the number of correctly solved problems on Set II.

2.1.2. USMLE Step I

United States Medical Licensing Examination (USMLE) Step 1 is a knowledge-based test that assesses whether students understand and can apply important concepts of the sciences basic to the practice of medicine, with special emphasis on principles and mechanisms underlying health, disease, and modes of therapy. The scores issued by USMLE are used in our analysis.

2.1.3. Objective Structured Clinical Examination

The Objective Structured Clinical Examination (OSCE) involves a series of short clinical encounters with standardized patients (SP) as part of the California Consortium of the Assessment of Clinical Competence (Known at UCLA as "Clinical Performance Exam"). The four major skills of clinical performance measured by the examination are history taking (HX), physical examination (PE), information sharing (IS), and patient-provider interaction (PPI). There are a total of eight stations (cases) on the OSCE. The OSCE is scored by SPs using a checklist. Standardized patients are trained to consistently portray the role of the patient and to use the checklists reliably. Each skill component

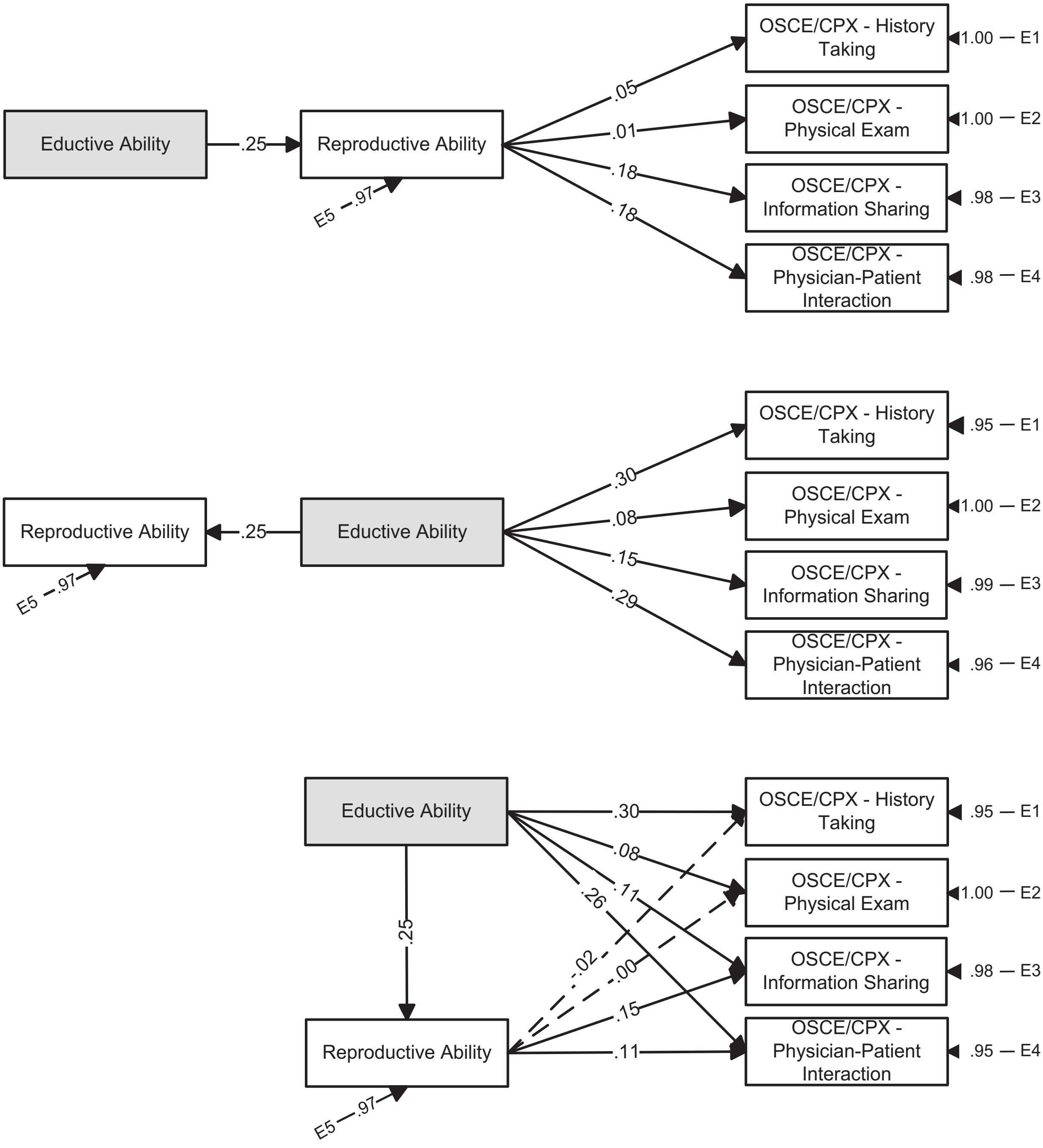

Fig. 1. Three hypothetical structural clinical performance models with cause indicators.

is composed of checklist items that are both unique to the case as well as common across cases. The number of items under each component varies from case to case. Skill component scores are calculated by averaging the number of points received on each item across all cases. The scores on the four individual components are used in our analysis.

UCLA IRB has approved research protocol by expedited review.

2.2. Statistical analysis

Two hypothesized models were analyzed and compared via Structural Equation Modeling (SEM) using the software package EQS 6.1.[16] SEM models reflect different hypothesized causal relationships between the OSCE components, USMLE Step I and the Raven Progressive Matrices. These causal processes are represented by the regression equations. The hypothesized models can be tested statistically in a simultaneous analysis of the entire system of variables to determine the extent to which it is consistent with the data (for a practical guide, see Byrne[17]). The fit indices provided by the SEM package were used to determine whether the model adequately fit the data. The chi-square (χ^2) indicates that the existing model's covariance structure is different from the observed covariance matrix. Bentler's Comparative Fit Index (CFI) compares the fit of the particular model under test with a model in which none of the variables are related; a CFI of .90 or higher has typically been taken as indicating good fit between the model tested and the data. The Standardized Root Mean-Square Residual (SRMR) represents the average standardized discrepancy between observed and model-implied relations; a value below .08 indicates good fit. Finally, Steiger's Root Mean Square Error of Approximation (RMSEA) adjusts for a model's complexity. A value below .05 indicates proper fit.

We identified two possible causal relationships between the variables. Since eductive ability is at the center of performance it represents the independent variable in both models. Fig. 1 shows these two hypothesized models. The first model hypothesizes that eductive ability (as measured by the Raven Progressive Matrices test) predicts to some degree reproductive ability (as measured by USMLE Step I), which predicts OSCE performance. The second model places eductive ability in the center which predicts OSCE and reproductive ability. However, reproductive ability has no direct influence on OSCE performance. The third model hypothesizes that eductive ability and reproductive ability both have a direct impact on OSCE performance whereby eductive ability also predicts reproductive ability.

3. Results

Descriptives and correlations among all variables used are shown in Table 1. The 8-station OSCE's reliability is consistent over the years (Cronbach's alpha ~ .7).

Model 1 displays a poor fit. $\chi^2(10) = 11.744$, $p = .3063$; SRMR = .121 RMSEA = .075 (90% confidence interval from .000 to .213); AIC = −8.256; and CFI > .000.

Model 2 $\chi^2(10) = 7.204$, $p = .706$; SRMR = .102; RMSEA = .000 (90% confidence interval from .000 to .147); AIC = −12.796; and CFI < 1.000. This model has an adequate overall fit.

Model 3 is in principle a more restricted version of model 2. Fit indices are: $\chi^2(6) = 6.113$, $p = .847$; SRMR = .092; RMSEA = .025 (90% confidence interval from .000 to .232); AIC = −5.887; and CFI = .857. This model shows a near acceptable fit, and the Wald Test for dropping parameters estimates that the overall fit of the model would increase if regression weights between reproductive ability and history taking and reproductive ability and physical examination would be fixed to zero. The overall improved fit when both parameters are removed (dashed lines in Model 3 of Fig. 1) is: $\chi^2(8) = 6.129$, $p = .633$; SRMR = .093; RMSEA < .000 (90% confidence interval from .000 to .173); AIC = −9.871; and CFI < 1.000. The adjusted model 3 has an almost identical fit as model 2 and is not of any help in differentiating the models. However, conceptually it seems justified to include the parameters between reproductive

Table 1
Correlation matrix of all used variables (calculated with IBM SPSS 18).

Variable	Mean N=32	SD	1	2	3	4	5	6	7
1. Raven's APM	28.0	4.5	–	.25	.31*	.08	.15	.29	.47**
2. USMLE Step I	229.8	19.1		–	.05	.01	.18	.18	.16
3. OSCE-HX	.73	.050			–	.08	−.01	−.17	.70**
4. OSCE-PE	.67	.077				–	−.11	−.21	.48**
5. OSCE-IS	.67	.088					–	.26	.06
6. OSCE-PPI	.73	.065						–	.33*
7. OSCE-Total	.71	.031							–

*Correlation is significant at the .05 level (1-tailed).
**Correlation is significant at the .01 level (1-tailed).

ability and history taking and reproductive ability and physical examination. We accept this more restricted version of model 2 as our model of choice.

4. Discussion

The modified Model 3 indicates that mostly eductive ability, and to a lesser extent reproductive ability, predict performance on the individual components of the OSCE. As said, eductive mental activity involves making meaning out of confusion; developing new insights; going beyond the given to perceive that which is not immediately obvious; in mostly non-verbal constructs that facilitate solving of complex problems and reproductive ability involves mastering, recalling and reproducing mostly verbal material. It is interesting to note that eductive ability primarily loads on history taking and physician–patient interaction while reproductive ability loads (to a lesser extent) on information sharing and physician–patient interaction. The correlational structure (Table 1) confirms these associations.

An important consideration is whether basic reasoning abilities such as those measured by the APM are amenable to training. As it is a measure of intelligence, the general assumption has been that such basic abilities are largely unchangeable. Yet in a program of study, Schauer[18] found that controlling allocation of attentional resources significantly improves performance on APM-like problems, and that such control can be learned.

As the current study shows a significant relationship between APM and OSCE performance, the implications for medical reasoning performance of training in basic, domain-general abilities, such as those discussed in Schauer[18] would be a substantial and an important line of study to pursue.

A limitation of this study is the relative small sample size. The generally agreed-on value is 8–10 participants for every free parameter estimated when a reasonable estimate about the magnitude of the effect in the population is unknown. However the models used in this study are all simple and lack complex latent relationships. The values of the estimated parameters are comparable to the correlation matrix of Table 1, confirming that our models' estimates were stable and not affected by the low sample size.

5. Conclusions

The direct relationships between the components of the OSCE and other constructs need further mapping. The study of Lee and Wimmers,[15] for example, showed that history taking as measured with an OSCE had associations with clinical knowledge as measured on a NBME medicine subjects exam, the patient-provider interaction component of the OSCE had a moderate significant correlation with empathy, and communication component of an OSCE with aspects of emotional intelligence (i.e., Feelings, Empathic Concern, and Perspective Taking).[19,20] This study adds to the complexity of the OSCE by showing direct relationships with components of intelligence.

References

1. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ* 1975;1(5955):447–451.
2. Hauer KE, Hodgson CS, Kerr KM, Teherani A, Irby DM. A national study of medical student clinical skills assessment. *Acad Med* 2005;80:S25–S29.
3. Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med* 1990;2:58–76.
4. Patricio M, Juliao M, Fareleira F, Young M, Norman G, Vaz Carneiro A. A comprehensive checklist for reporting the use of OSCEs. *Med Teach* 2009;31(2):112–124.
5. Swanson DB, Norcini JJ. Factors influencing reproducibility of tests using standardized patients. *Teach Learn Med* 1989;1: 158–166.
6. Schuwirth LW, van der Vleuten CP. The use of clinical simulations in assessment. *Med Educ* 2003;37(1):S65–S71.
7. Wimmers PF, Fung C-C. The impact of case specificity and generalisable skills on clinical performance: a correlated traits-correlated methods approach. *Med Educ* 2008;42(6):580–588.
8. Wass V, Van der Vleuten C, Shartzer J, Jones R. Assessment of clinical competence. *Lancet* 2001;357:945–949.
9. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: a review of metrics â€“ AMEE guide no. 49. *Med Teach* 2010;32(10):802–811.
10. Spearman C. Theory of general factor. *Br J Psychol* 1946;36: 117–131.
11. Raven J, Raven JC, Court JH. *Raven Manual: Section 1, General Overview,1998 Edition*. Oxford, UK: OPP Limited; 1998.
12. Raven JC. *Progressive Matrices: a Perceptual Test of Intelligence*. London: H.K. Lewis; 1938.
13. Carpenter PA, Just MA, Shell P. What one intelligence test measures: a theoretical account of the processing in the Raven progressive matrices test. *Psychol Rev* 1990;97(3):404–431.
14. Simon SR, Volkan K, Hamann C, Duffey C, Fletcher SW. The relationship between second-year medical students' OSCE scores and USMLE Step 1 scores. *Med Teach* 2002;24(5):535–539.
15. Lee M, Wimmers PF. Clinical competence understood through the construct validity of three clerkship assessments. *Med Educ* 2011;45(8):849–857.
16. *EQS 6 Structural Equations Program Manual* [Computer Program]. Encino, CA: Multivariate Software, Inc.; 2005.
17. Byrne BM. *Structural Equation Modeling with EQS: Basic Concepts, Applications, and Programming*, 2nd ed., MahWah, NJ: Lawrence Erlbaum Associates; 2006.
18. Schauer GF. *Attentional Scope and Performance in Solving relational figure matrix puzzles*. Los Angeles: Dissertation Abstracts International; 2009.

19. Wimmers PF, Stuber M. Measuring medical students' empathy and attitudes towards patient-centered care with an existing clinical performance exam (OSCE). In: *Proceedings of Annual Meeting of the American Educational Research Association (AERA)*, April 30–May 4. Denver, Colorado; 2010.
20. Stratton TD, Elam CL, Murphy-Spencer AE, Quinlivan SL. Emotional intelligence and clinical skills: preliminary results from a comprehensive clinical performance examination. *Acad Med* 2005;80(10):S34–S37.