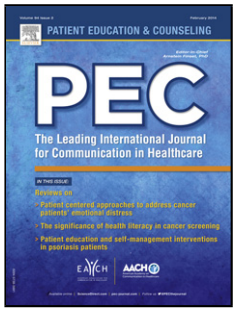

Title: True communication skills assessment in interdepartmental OSCE stations: Standard setting using the MAAS-global and EduG

Authors: Winny Setyonugroho, Thomas Kropmans, Ruth Murphy, Peter Hayes, Jan van Dalen, Kieran M. Kennedy

PII: S0738-3991(17)30416-0
DOI: http://dx.doi.org/doi:10.1016/j.pec.2017.07.003
Reference: PEC 5734

To appear in: *Patient Education and Counseling*

Received date: 26-5-2016
Revised date: 12-6-2017
Accepted date: 6-7-2017

Please cite this article as: Setyonugroho Winny, Kropmans Thomas, Murphy Ruth, Hayes Peter, van Dalen Jan, Kennedy Kieran M.True communication skills assessment in interdepartmental OSCE stations: Standard setting using the MAAS-global and EduG.*Patient Education and Counseling* http://dx.doi.org/10.1016/j.pec.2017.07.003

True Communication Skills Assessment in Interdepartmental OSCE Stations : Standard Setting Using The MAAS-Global and EduG.

Winny Setyonugroho[1]; Thomas Kropmans[2]; Ruth Murphy[2]; Peter Hayes[2]; Jan van Dalen[3]; Kieran M Kennedy[2]

[1]Faculty of Medicine and Health Sciences of the Universitas Muhammadiyah Yogyakarta, Indonesia, [2]School of Medicine, College or Medicine, Nursing & Health Sciences, National University of Ireland Galway, [3]Skills laboratory, Faculty of Health, Medicine and Life Sciences, Maastricht University, The Netherlands.

Corresponding author at**:**

Thomas JB Kropmans

thomas.kropmans@nuigalway.ie

College of Medicine, Nursing & Health Sciences

School of Medicine, Comerford Building Room 204, Clinical Science Institute, National University of Ireland Galway; Galway, Ireland

Phone: +353 91 494340

Fax: +353 91 495512

Contributors

Winny Setyonugroho, MT.

wsetyonugroho@umy.ac.id

wsetyonugroho@gmail.com

Faculty of Medicine & Health Sciences, Universitas Muhammadiyah Yogyakarta, Indonesia

Dr. Ruth Murphy M.B., MRC Psych

ruth.murphy@nuigalway.ie

School of Medicine, Clinical Science Institute, National University of Ireland Galway; Galway, Ireland

Dr Peter Hayes MB,BCh,BAO,B.Med.Sc.,MICGP,Cert.Med.Ed.

Peter.hayes@nuigalway.ie

School of Medicine, Clinical Science Institute, National University of Ireland Galway; Galway, Ireland

Thomas JB Kropmans

thomas.kropmans@nuigalway.ie

School of Medicine, Comerford Building Room 204, Clinical Science Institute, National University of Ireland Galway; Galway, Ireland

Dr. Jan van Dalen

j.vandalen@maastrichuniversity.nl

Discipline of Skills laboratory, Faculty of Health, Medicine and Life Sciences, Maastricht University, The Netherlands.

Kieran M Kennedy, MB BCh BAO., MMedSci MHSc, MSc MICGP.

kieran.kennedy@nuigalway.ie

School of Medicine, Comerford Building, Clinical Science Institute, National University of Ireland Galway; Galway, Ireland

Highlights

- Comparing assessment outcome of communication skills (CS) is challenging
- Mapping items to the MAAS-Global is crucial
- Comparing non-standardized items is impossible
- Automated clinical skills assessments in OMIS needs a curriculum mapping tool for learning outcomes

Abstract

Background

Comparing outcome of clinical skills assessment is challenging. This study proposes reliable and valid comparison of communication skills **(1)** assessment as practiced in **Objective Structured Clinical Examinations (2). The aim of the present study is to compare CS assessment, as standardized according to the MAAS Global, between stations in a single undergraduate medical year.**

Methods

An OSCE delivered in an Irish undergraduate curriculum was studied. We chose the MAAS-Global as an internationally recognized and validated instrument to calibrate the OSCE station items. The MAAS-Global proportion is the percentage of station checklist items that can be considered as 'true' CS. The reliability of the OSCE was calculated with G-Theory analysis and nested ANOVA was used to compare mean scores of all years.

Results

MAAS-Global scores in psychiatry stations were significantly higher than those in other disciplines ($p<0.03$) and above the initial pass mark of 50%. **The higher students' scores in psychiatry stations were related to higher MAAS-Global proportions when compared to the general practice stations.**

Conclusion

Comparison of outcome measurements, using the MAAS Global as a standardization instrument, between interdisciplinary station checklists was valid and reliable.

Practice implications

The MAAS-Global was used as a single validated instrument and is suggested as gold standard.

keywords : Communication skills; OSCE; Objective structured clinical examination; MAAS-Global; OMIS; OSCE Management Information System

Background

Comparison of the outcomes of communication skills (CS) assessments in undergraduate medical education is challenging. Assessments are intended to measure differences between individual students, however their outcomes are also influenced by difference between cohorts of students. A synthesis of the literature demonstrates that a large proportion of medical errors and adverse events in postgraduate professional practice can be explained by communication factors between clinicians and patients and between health care providers themselves (1, 3). Schoenthaler et al mentioned in her recent systematic review that only few papers targeted patient–practitioner communication and assessed the impact on cardiovascular-related clinical outcomes, limiting the ability to determine effectiveness in other areas such as well-being and compliance (4). Referring to previous research, we support Schoenthalers' conclusion that additional rigorous research supported by theoretical frameworks and validated measurement is required to understand the potential of patient–practitioner communication to improve cardiovascular-related clinical outcomes. **Effective communication is known to be extremely important to ensure safe and effective clinical practice (5).** Training and assessment of clinical communication starts at undergraduate level and the purpose of this training is to adequately prepare undergraduate students for professional practice. This appears to be a major challenge for educational institutions. **Nevertheless, with a careful prospective design, an Objective Structured Clinical Examination (OSCE) may be developed to** simultaneously assess multiple competencies including CS. Holistic ratings as well as checklists may help to evaluate physician competencies (CanMeds) in a reliable and valid manner in the OSCE.

In previously published research we explored 27 domains of CS being assessed and discussed in the international medical education literature (6). Comparing the existing evidence about reliable and valid assessment to our existing assessment practice, suggests to us that assessment of CS with either a checklist or global rating scale is like measuring these crucial skills using a rubber band because those measurement instruments are not standardized. We previously attempted to calibrate all our existing clinical skills assessment forms used in previous OSCEs to assess 'aspects of CS' and adopted the MAAS-Global as a

previously validated (2) and reliable assessment tool and the "gold standard" within our Medical School (7, 8). In previous research, we calibrated all clinical skills items of our OSCE forms in the year 4 OSCEs to arrive at adequate comparison of the CS component of each OSCE station (7). **CS are commonly measured in an OSCE where students' ability to communicate with the standardized patient is measured by examiners. Unlike language tests which already are measured with standardized tests (e.g. Teaching of English as a Foreign Language - TOEFL), CS do not have such a standardized test. Direct comparison of CS attained from station results is not possible, since every station uses unique and non-validated station checklists. Therefore calibration of CS items is required in order to make comparison possible (7).**

In the present study we introduce a conversion method for existing non-standardized OSCE station checklists, using the MAAS-Global as a standarization tool. In that way, we assess the true CS measured within our existing station checklists. This study proposes a method for reliable and valid comparison of CS by comparing 'raw scores of communication skills' measured with 'uncalibrated' assessment forms with calibrated 'true' measurement of CS (7). **Therefore, the aim of the present study is to compare CS assessment, as standardized according to the MAAS Global, between stations in a single undergraduate medical year. This may allow examination results to become comparable between students, disciplines and different groups.**

Method

Overview

This retrospective study analyzed the penultimate OSCE of the undergraduate medical programme at the National University of Ireland in Galway, Ireland. The data from the station forms developed by the **Department of Psychiatry** and those of the Department of General Practice were retrieved from our online OSCE Management Information System (OMIS). Data from three academic terms, 2010/2011, 2011/2012, and 2012/2013 was analyzed. The penultimate OSCE is administered in February/March and April, for two consecutive groups of students. **The settings of OSCE circuits in both General Practice and Psychiatry parts of the OSCE are different. For General Practice stations, the number of stations in all circuits in all academic terms is 10. The station duration is 5 minutes for each station. Meanwhile, for Psychiatry stations, each circuit in term 2010/2011 incorporates 4 stations while in terms 2011/2012 and 2012/2013 each circuit consists of 5 stations. The station duration of psychiatry stations is twice as long as those in General Practice (i.e. 10 minutes each).** Both disciplines use different OSCE settings, e.g. number of stations, sequence of stations, or scoring rubrics. **The examiners for both disciplines are recruited from the staff of academic clinicians at the School of Medicine and therefore comprise of practicing General Practitioners and Psychiatrists.**

The MAAS-Global as a Standard

The MAAS-Global is designed as a generic instrument to rate physicians' CS. The MAAS-Global consists of 17 items divided into 3 sections. Seven items in section 1 refer to appropriate skills in the specific phases of clinical consultations. Items are related to introduction, follow-up consultation, a request for help, physical examination, diagnosis, management, and evaluation of the consultation. These items are a reflection of the logical order of consultation phases. Section 2 focuses on general CS which occur throughout the consultation, consisting of 6 items. Those items are: exploration, emotions, information giving, summarizations, structuring, and empathy. Section 3 is intended to examine the mastery of the medical content during medical

consultation. This section consists of 4 items: history taking, physical examination, diagnosis, and management which represent phases of the consultation.

Calibration Method

The conversion method used in this study is a process where the unstandardized measurement instruments are calibrated with the MAAS-Global, hence allowing the examination results to become comparable between students, disciplines and different groups. The first step of this method is to calibrate each measurement instrument i.e. station checklist, with the MAAS-Global. The calibration method has been proven to be a reliable procedure according to our previous study (6). This calibration result is what we will call the MAAS-Global proportion, which is a percentage amount of checklist items that are considered to be CS items according to the MAAS-Global. The second step of this conversion method is converting the students' score. **The students' score is their total item score for an individual station.** Students' scores in each station multiplied by MAAS-Global proportion is the MAAS-Global score. The total MAAS-Global score of the OSCE is the average of the station's MAAS-Global score over all relevant OSCE stations assessing any type of CS.

In conditions where not all of the OSCE stations have CS items, the total MAAS-Global score is the average of station with MAAS-Global score only (i.e. in April 2012, from 5 stations in Department of Psychiatry, only of 4 stations a MAAS-Global score could be calculated. Hence the total MAAS-Global score is average of 4 stations' MAAS-Global score). **In this study, the result of a calibrated OSCE using the MAAS-Global is referred to as the MAAS-Global score. We will present the MAAS-Global score followed by MAAS-Global proportion and section percentage. For example, MAAS-Global score of 65 with a MAAS-Global proportion of 75 which consists of 14% section 1, 29% section 2 , and 57% section 3, will be then written as MAAS-Global score 65 [MG75-14-29-57] (7).**

Statistical Analysis

The results of each OSCE were combined and converted to a percentage scale. Generalizability Theory was calculated independently for each circuit of each department to determine reliability of the test. Nested

ANOVA was used for the comparison of the mean of six circuits and Tukey's post-hoc test was used to distinguish which of the circuits introduced difference. **Statistical significance was set at the 0.03 level of probability in accordance with the use of Bonferroni adjustments**. A software package EduG (Swiss Society for Research in Education Working Group. Edumetrics - Quality of measurement in education) has been used to perform G-Theory analysis and SPSS version 21 was used to analyze ANOVA.

Results

Overview of OSCEs

Total students for academic terms 2010/2011, 2011/2012, 2012/2013 were 116, 123, and 141 respectively (total n=380). **Calibration of all station checklists from both disciplines revealed** that 9 out of 10 General Practice stations contain CS, except for one held in February of the 2010 – 2011 terms. **During the period 2010 – 2011, all psychiatry stations contained CS items while one station did not contain CS during the 2011 - 2012 and 2012 – 2013 terms. Those psychiatry and General Practice stations which did not contain CS items involved other tasks, for example, students watching a video and completing a short "tick-box" report on their observations of a thought-disordered patient.**

Table 1 **presents a summary of mean and standard deviation values for the overall OSCE score, MAAS-Global score, MAAS-Global proportion, sections of the MAAS-Global in percentages and MAAS-Global items within each section.** For the General Practice stations, the MAAS-Global proportion ranged between 65 to 75 percent, with an average of 68 for 3 academic terms. **It is apparent that section 3 of the MAAS-Global (mastery of the medical content during the consultation) represents the largest portion of the CS assessed in these stations. For sections 1 (sequential skills of a consultation) and 2 (generic skills) between 11 to 19 percent and between 29 to 35 percent, respectively, were considered to be 'true' CS items according to the MAAS-Global standard. It is also shown that the** General Practice' stations incorporate almost all of the MAAS-Global items in all circuits. By contrast, the Department of Psychiatry incorporates only item 5 (diagnosis) and item 10 (information giving) of section 1 and 2. **The overall majority (77 to 100%) of the station checklist items were**

considered to be representative of section 3 of MAAS-Global (i.e. mastery of the medical content during the consultation).

Statistical analysis

Referring again to Table 1, it can be seen that average OSCE scores (scale 0-100) of the six circuits of the Department of General Practice ranged from 58 (sd=5.5) to 63.9 (sd= 6.3). The MAAS-Global scores average ranged from 43.8 (sd=4)[MG75-14-29-57] to 48.2 (sd=4.5)[MG66-15-30-55]. Whereas the average OSCE scores of the Department of Psychiatry ranged from 61.2 (sd=8.4) to 69.8 (sd=9.9). The MAAS-Global scores for the Psychiatry stations tended to be higher, ranging from 58.5 (sd=7.9) [MG73-1-4-95] to 66.2 (sd=8.2)[MG73-2-4-96].

A Nested ANOVA was conducted to compare the effect of disciplines and circuits which nested within disciplines. Table 2 shows the analysis of variance for both OSCE score and MAAS-Global score. For the OSCE score, there was a significant effect for department at the corrected $p < 0.03$ level [$F(1, 748) = 84.12$, $p < .001$] and there was a significant effect for circuit which was nested within the department [$F(10, 748) = 10.58$, $p < .001$]. While for MAAS-Global scores we found similar results, there was significant effects for department and circuit which was nested within the department, $F(1, 748) = 1080.83$, $p < .001$ and $F(10, 748) = 7.50$, $p < .001$ respectively. Thus, the CS measurements differed, with statistical significance, between departmental stations.

A generalizability analysis was performed separately for each circuit of both disciplines. From Table 2, it can be seen that the generalizability coefficient (G) varied from 0.59 to 0.75 for the general practice stations and 0.54 to 0.73 for the psychiatry stations.

In a Decision Study (D-study) analysis, a hypothetical design with 15 stations from the each of the Departments of general practice and psychiatry were calculated (Table 2). By increasing the number of OSCE

stations, the range of reliability measurements improved for general practice to 0.68-0.82 and for psychiatry to 0.79-0.85, depending upon which term or circuit was analyzed.

Discussion and conclusion

Discussion

The aim of the present study is to compare CS assessment, as standardized according to the MAAS-Global, between stations in a single undergraduate medical year. We sought to determine the 'true' differences in CS outcomes as assessed in a penultimate OSCE of undergraduate medical students by comparing 'raw' scores and the standardized 'MAAS-Global' scores. The reliability of the MAAS-Global scores, according to G-theory, was moderate to good and could be improved by increasing the number of stations to at least 10 for the Psychiatry OSCE. It is apparent that 10 stations, each of 5 minutes duration, generate a higher G-coefficient than 5 stations of 10 minutes duration. This supports the concept mooted in previous publications that a larger number of stations lead to higher reliability (9). Our D-study for 15 stations for both disciplines revealed that the reliability of Department of Psychiatry is relatively higher that Department of General Practice. The possible explanation for higher reliability in Department of Psychiatry may be that the station duration is longer than those for the Department of General Practice (10-12).

There is an apparent difference in MAAS-Global items being covered between stations developed by General Practice and those developed by the Department of Psychiatry. Whilst in General Practice learning outcomes of CS training are more generic and apparently cover different stages of the consultation (Section 1 items of the MAAS-Global) being a reflection of the logical order of consultation phases with items related to introduction, follow-up consultation, a request for help, physical examination, diagnosis, management, and evaluation of the consultation. Section 2 focuses on general CS which are used throughout the consultation, consisting of 6

items. Those items are: exploration, emotions, information giving, summarizations, structuring, and empathy. In the Department of Psychiatry stations, there was an emphasis upon the medical content of the consultation, in contrast with the Department of General Practice, where there was a broader emphasis upon a wider range of CS items as defined by the MAAS-Global. In the Department of Psychiatry, there were more CS items that were specifically addressing history taking, physical examination, diagnosis, and management of diseases. As it is difficult to map the learning outcomes of CS training for departmental 'CS assessment forms' it is not certain whether the average coverage of 15, 31 and 51% MAAS-Global sections respectively covers learning outcomes related to CS that specifically focus on different stages of the consultation (section 1), general CS (section 2) and mastering the medical content of the consultation (section 3). Although we have automated clinical skills assessments in our School of Medicine, and we are able to calibrate our assessment forms in an automated fashion, we do not yet have a curriculum mapping tool linking specific items or competencies with the curriculum learning outcomes. It is obvious that Psychiatry stations with respectively 6, 5 and 89% MAAS-Global sections are assessing something different to General Practice stations. Furthermore, General Practice was found to be assessing a combination of communication (69%) and technical skills (31%), whereas Psychiatry is predominantly assessing CS (79%). In previous research the Generalized Kappa for reviewers agreement about calibration of forms was high for the General Practice stations (0.83), but was even higher for the Psychiatry stations (0.99) (7). Thus, we can be quite sure that the Psychiatry stations are assessing section 3 CS items, whereas the nature of the General Practice CS items is more open for debate.

Regarding content validity, a typical General Practice consultation entails both technical skills and CS, however in a learning situation it is necessary to confirm whether our assessment strategy verifies that students have achieved their learning goals and whether standard setting has been successful (13). The pass mark in Schools of Medicine in Ireland is generally regarded as 50%. Whilst pass marks vary between universities worldwide, there is not much evidence available to confirm the validity or ratio behind the use of static pass marks. With the OSCE Management Information System, we use Borderline Regression Analysis to incorporate the difficulty

of stations and variability between examiners. Marks are presented in terms of a regression outcome and, also, as a static pass mark of 50% [13]. Nevertheless, the average (SD) MAAS-Global score for General Practice stations is 46 (4) versus 62 (8) for Psychiatry measured over three academic terms. These scores indicate that 68% of our cohort of students being assessed achieved a 'true' score for CS between 42 (average minus 1 standard deviation) and 50% (average plus 1 standard deviation) which would indicate a performance below formal standard setting. For the Psychiatry stations, in which 68% of students of different cohort(s) achieved a score between 54 and 70, this is less significant an issue than in General Practice. A MAAS-Global score below 50% does not indicate a fail score. The MAAS-Global score was attained from OSCE score multiplied with MAAS-Global proportion, hence the actual maximum MAAS-Global score is similar to MAAS-Global proportion i.e. not equal to 100%. Furthermore, as explained in Table 1, each OSCE had a different MAAS-Global proportion. International literature pertaining to OSCE assessments addresses tendencies towards global marking whereas the assessment forms analyzed in the present study are very much item based 'tick box' assessment forms. Further research should consider matching the various intersections of our assessment forms addressing respectively section 1, 2 and 3 items of the MAAS-Global in order to generate a global communication score for CS [7, 8]. With respect to this dataset, we suggest the use of an overall MAAS-Global score as outlined in Table 1, 2nd row, column 4 – 8 being 43.8 (4) MAAS Global proportion 75%; with 14 % section 1; 29% section 2 and 57% section 3 items i.e. MAAS-Global score 44 [MG75-14-29-57], indicates a score below the pass mark of 50% in an OSCE addressing 75% CS and 25% technical (other) skills, not being CS.

Calibration of CS assessment forms will assist with identifying incremental change in student's CS performance. Sensitivity to change in clinimetrics is a well-established area of research. Sensitivity to change is part of the classical psychometric analysis of diagnostic, prognostic and therapeutic discriminative and evaluative instruments. There is however a notable shortage of assessment tools and psychometric characteristics to be used for progress monitoring in clinical skills assessments. Considerable evidence is available to measure growth in medical knowledge [13]. Recently, Turan and Valcke c.s. (2013) developed a

Medical Achievement Self-efficacy Scale (MASS) for students of the Ghent Curriculum (14). The latter scale is related to the general competency frameworks of CanMEDs and the Five-star Doctor and predict progress test outcome, however clinical skills assessments are not included. Multiple mini-interviews predict clerkship and licensing examination performance, including OSCE performance, but again evidence regarding the measurement of change in clinical skills assessments is lacking (15). **We suggest that true measurement of a student's progress in CS development is only possible if a standardized tool is employed. We have suggested a potential standardization method that utilizes the MAAS-Global to standardize existing OSCE checklists (6, 7).**

The procedure of calculating MAAS-Global scores based on the MAAS-Global standard is labour-intensive. This process currently entails two steps, calibrating the OSCE station-checklists and re-calculating each students MG score for each of the stations. To date our OSCE Management Information System does not have an automated mechanism for this process. Future development of the software could incorporate the option to map OSCE rubrics and calculate the MG score directly. **However, it might be easier to develop specific MAAS-Global based CS stations which would avoid the need to standardize items in the first instance. We recognize that comparing communication OSCE scores between disciplines might only be of limited value since communication competencies, as other medical competencies, are context or task specific. However, we suggest that future research should consider possible use of the MG score as one of the criteria for standard setting of an OSCE.** In such a case, not only would students have to pass the overall OSCE cut-score, but they would also have to pass a minimum MG score.

Conclusion

Comparison of outcome measurements, using the MAAS Global as a standardization instrument, between interdisciplinary station checklists was valid and reliable.

Practice implications

This study has demonstrated the process for distilling an MG score from an overall OSCE score. Secondly, we demonstrated the true characteristics of CS based on a standardized instrument (i.e. the MAAS-Global). It is now possible to compare CS assessment outcomes from different settings (i.e. rubrics or different modules) of OSCEs. Moreover, this new approach should be considered as a possible standard procedure to assess CS in OSCEs and to improve quality of measurement. Future research should be undertaken to explore how to incorporate the 'true' CS score as one criterion for passing the conjunctive standard.

Acknowledgement

The author would like to thank Professor Jean Cardinet (GENEVA Centre for Psycho-Educational Research), developer of EduG for his continuous contribution to our email discussions about the Generalizability Theory and EduG.. We wish him and his family all the best during his final journey. His contribution to the journey of WS, as a PhD student, was invaluable.

Winny Setyonugroho, as the lead author, received a PhD scholarship from the Directorate General of Higher Education, Ministry of Research, Technology and Higher Education, Republic of Indonesia.

Authors' contributions

WS, KK and TK developed the concept for the study. WS, RM, and PH participated in data acquisition. WS and TK carried out analysis and interpretation of data. WS, RM, PH, KK, and TK, prepared the manuscript. KK, TK, and JD, carried out critical revisions.

Declaration of interest

The authors declare there is no conflict of interests.

References

1. Beck RS, Daughtridge R, Sloane PD. Physician-patient communication in the primary care office: a systematic review. J Am Board Fam Pract. 2002;15(1):25-38.
2. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. Med Educ. 2011;45(12):1181-9.
3. Phillips C. Communication: the first tool in risk management for long-term care. J Am Med Dir Assoc. 2004;5(2):123-6.
4. Schoenthaler A, Kalet A, Nicholson J, Lipkin M, Jr. Does improving patient-practitioner communication improve clinical outcomes in patients with cardiovascular diseases? A systematic review of the evidence. Patient Educ Couns. 2014;96(1):3-12.
5. Weldon SM, Korkiakangas T, Bezemer J, Kneebone R. Communication in the operating theatre. Br J Surg. 2013;100(13):1677-88.
6. Setyonugroho W, Kennedy KM, Kropmans TJ. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. Patient Educ Couns. 2015.
7. Setyonugroho W, Kropmans T, Kennedy KM, Stewart B, van Dalen J. Calibration of communication skills items in OSCE checklists according to the MAAS-Global. Patient Educ Couns. 2016;99(1):139-46.
8. van Es JM, Schrijver CJ, Oberink RH, Visser MR. Two-dimensional structure of the MAAS-Global rating list for consultation skills of doctors. Med Teach. 2012;34(12):e794-9.
9. Smith V, Muldoon K, Biesty L. The Objective Structured Clinical Examination (OSCE) as a strategy for assessing clinical competence in midwifery education in Ireland: a critical review. Nurse Educ Pract. 2012;12(5):242-7.
10. Friedman Ben David M, Davis MH, Harden RM, Howie PW, Ker J, Pippard MJ. AMEE Medical Education Guide No. 24: Portfolios as a method of student assessment. Med Teach. 2001;23(6):535-51.
11. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. Med Teach. 2013;35(9):e1447-63.
12. Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. Med Teach. 2013;35(9):e1437-46.
13. Hejri SM, Jalili M, Muijtjens AM, Van Der Vleuten CP. Assessing the reliability of the borderline regression method as a standard setting procedure for objective structured clinical examination. J Res Med Sci. 2013;18(10):887-91.
14. Eva KW, Reiter HI, Rosenfeld J, Norman GR. The ability of the multiple mini-interview to predict preclerkship performance in medical school. Acad Med. 2004;79(10 Suppl):S40-2.
15. Essers G, Dielissen P, van Weel C, van der Vleuten C, van Dulmen S, Kramer A. How do trained raters take context factors into account when assessing GP trainee communication performance? An exploratory, qualitative study. Adv Health Sci Educ Theory Pract. 2015;20(1):131-47.

Figure Caption

Table 1. Summary of OSCE score Mean, MAAS-Global score Mean, COMMUNICATION SKILLS, MAAS-Global proportion, section of MAAS-Global in percentage, and MAAS-Global items.

Discipline of General Practice										
Academic Year	OSCE Circuit	OSCE Mean (SD)*	MG Mean (SD)*	MG Proportion	MAAS-Global (%)			MAAS-Global Items		
					Section 1	Section 2	Section 3	Section 1	Section 2	Section 3
2010/2011	February	58 (5.5)	43.8 (4)	75	14	29	57	1,3,4,5,6	8,9,10,11,12, 13	14,15,17
	April	58.6 (4.9)	45.8 (3.6)	69	19	29	52	1,4,5,6	8,9,10,11,12, 13	14,15,17
2011/2012	February	63.2 (4.3)	43.9 (3.3)	65	16	35	49	1,4,5,6	8,9,10,11,12, 13	14,15,17
	April	61.3 (5.8)	46.6 (4.2)	67	11	35	54	1,4,5,6	8,9,10,11,12, 13	14,15,17
2012/2013	February	63.9 (6.3)	48.2 (4.5)	66	15	30	41	1,4,5,6	8,9,10,11,12, 13	14,15,17
	April	61.6 (5.8)	47.1 (4.4)	71	15	30	55	1,3,4,5,6	8,9,10,11,12, 13	14,15,17
Discipline of Psychiatry										
2010/2011	February	69.8 (9.9)	62.6 (8.9)	90	13	6	81	5	10	14.17
	April	64.6 (8.8)	60.4 (8.2)	92	0	0	100			14,15,17
2011/2012	February	68.9 (8.3)	66.2 (8.2)	73	2	4	94	5	10	14,15,16,17
	April	61.2 (8.4)	60.6 (7.5)	74	5	6	89	5	10	14,16,17
2012/2013	March	69.1 (9.9)	61.4 (8.6)	70	16	7	77	5	10	14.17
	April	63.1 (8.5)	58.5 (7.9)	73	1	4	95	5	10	14,16,17

* score scale 1-100
abbreviation :
MG : MAAS-Global
SD : standard deviation

Table 2. Analysis of variance for both OSCE score and MAAS-Global score (compare the effect of disciplines and circuits which nested within disciplines).

OSCE score			
Effect	Sum of Squares	df	Mean Square
Department	4694.63	1	4694.63
Circuit nested within Department	5909.67	10	590.97
Error	41745.01	748	55.81
MAAS-Global score			
Effect	Sum of Squares	df	Mean Square
Department	45342.89	1	45342.89
Circuit nested within Department	3148.7	10	314.87
Error	31380.04	748	41.95

Table 3. Summary of Generalizability Coefficient and Decision Study (with 10 and 15 stations)

		Discipline of General Practice		Discipline of Psychiatry		
Academic Terms	Circuit	G	D-study (15)*	G	D-study (10)*	D-study (15)*
2010/2011	February	0.75	0.82	0.54	0.74	0.85
	April	0.73	0.8	0.61	0.79	0.85
2011/2012	February	0.59	0.68	0.65	0.79	0.85
	April	0.73	0.8	0.56	0.72	0.79
2012/2013	February	0.74	0.81	0.73	0.84	0.89
	April	0.7	0.77	0.68	0.81	0.87

abbreviation list :

D-Study : Decision Study

G : G coefficient

Calibration results