

Objective Structured Clinical Examination - OSCE

En handbok för läkarprogrammets lärare



Karolinska
Institutet



Objective Structured Clinical Examination – OSCE

En handbok för läkarprogrammets lärare

Karolinska Institutet

Datum: 2022-05-19

Huvudförfattare: Wilhelmina Ekström, wilhelmina.ekstrom@ki.se, läkarprogrammets
Examinationskommitté

Bild framsidan: Attribution 2.0 Generic (CC by 2.0). EMT/Nursing Pediatric Emergency Simulation –
April 2013. Flickr.  



Innehåll

Objective Structured Clinical Examination - OSCE	1
Formativ bedömning enligt OSCE-modell.....	1
Att tänka på inför OSCE.....	2
Autenticitet	2
Poängsättning	2
Beslut om godkäntgräns	2
Samträning av bedömare	3
Psykometrisk analys av tentamensresultat	3
Återkoppling av tentamensresultat	3
Referenser	4
Att skapa OSCE	5
Grundplanering	5
Rekrytering	5
Bestäm stationer och skapa kriteriespecifika bedömningsmallar	5
OSCE processen.....	5
Mall: OSCE-station	8
Mall: Studentinstruktion	9
Mall: Information till bedömare	10
Mall: Information till standard patient (SP).....	11
Mall: Checklista till författargrupp.....	12
Mall: Bedömning.....	13
Mall: Checklista för kursadministratör.....	15
Referenser och rekommenderad litteratur	17

Objective Structured Clinical Examination - OSCE

OSCE är en summativ examination där studentens prestation bedöms i relation till lärandemålen. Resultatet är ett underlag för betygssättning: godkänd (G) eller underkänd (U).

OSCE är en stationsbaserad, strukturerad klinisk examination som fokuserar på tillämpning av framför allt färdigheter som bedöms i en simulerad miljö.^{1,2} Planering av OSCE bör ske enligt OSCE processen.² Denna startar med en blueprint och innehåller sedan val av stationsförfattare, val av stationstyp, skapa OSCE mall, skriva stationen, skapa bedömningsmall, skapa peer-bedömningsmall, praktisk testning av stationer med eventuell revision och psykometrisk analys.² Bedömning av studenten bör ske med en stationsbaserad poängsatt kriteriemall innehållande godkänt-gräns, global bedömning och gärna skriftlig återkoppling till studenten om användning av Qpercom. OSCE genomförs vid ett för kursen gemensamt tillfälle. Stationerna utgår från en klinisk kontext och bör examinera färdighet. De kan examinera en specifik färdighet, en riktad eller komplett patienthandläggning alternativt en mer komplex och sammansatt procedur. Stationens tidslängd brukar variera mellan 5–10 min/station och en välkonstruerad OSCE med adekvat reliabilitet omfattar 14–18 stationer.² På de kliniska terminerna på KI har framför allt 9–10 stationer använts där studenterna roterat i två olika varv. Stålltid mellan stationer brukar variera upp till två minuter.

Karantän kan användas i de fall då inte alla studenter kan examineras vid ett och samma tillfälle. Studentens uppgift vid stationen klargörs med en kort studentinstruktion som visas på dörren till stationen. OSCE genomförs individuellt och helst i en klinikliknande miljö till exempel på KTC. Vid underkänd prestation erbjuds restexamination.

Formativ bedömning enligt OSCE-modell

Formativ bedömning enligt OSCE-modell (fOSCE) innebär en strukturerad stationsbaserad träning av färdigheter och förhållningssätt med fokus på lärande, inte examination. Den kan ske löpande under kursen alternativt vid ett specifikt tillfälle. fOSCE eftersträvar longitudinell progressiv färdighetsträning vilket har visat sig ha betydelse för framtida summativa resultat.^{3,5,8,9} Studenten får vid fOSCE möjlighet att bekanta sig med formen för summativ OSCE som kommer att genomföras på efterföljande terminer vilket reducerar upplevelse av stress.⁷ Formativ OSCE innebär att:

- studenten tränar färdigheter och erhåller konstruktiv återkoppling på sin förmåga utan att summativ betygssättning sker.

fOSCE möjliggör för studenten att kunna bli bedömd för den egna prestationen i relation till lärandemålen för att avgöra vad denne behöver utveckla vidare. Vid fOSCE ska stationerna bygga på kursens färdighetsmål och vara framtagen på

samma sätt som för summativ OSCE. Stationsinnehåll kan vara grundläggande kommunikations- och statusfärdigheter, enklare laboratorieprocedur, anatomimonstration etc. Bedömning bör ske med kriteriebaserad bedömningsmall och kan utföras av peer-bedömare eller lärare.⁴ Stationernas innehåll och bedömningskriterier bör vara kända för studenterna i förväg så de kan förbereda sig.^{6,7} Stationen bör var tidsbestämd med avslutande återkoppling från lärare och/eller peer-bedömare. Studenten kan genomföra en fOSCE vid flera tillfällen tills adekvat färdighet uppnåts. Detta för att möjliggöra att examination av studenternas progression avseende praktisk färdighet ska kunna bedömas genom hela programmet.

fOSCE kan ske individuellt eller i mindre grupp och kan genomföras i lämplig lokal för färdighetsträning alternativt KTC.

Att tänka på inför OSCE

Autenticitet

Använd autentiska avidentifierade patientfall då stationsinnehåll skapas. OSCE-stationens innehåll ska vara direkt förankrat i lärandemål och bör ha ett scenario som är realistiskt förankrat i det sätt stationen presenteras. Stationens längd, vanligen 5-20 minuter ska matcha den svårighetsnivå som föreligger i lärandemålen och vara anpassad till det uppdrag studenten har att genomföra på stationen. Den standardiserade/riktiga patientens gestaltning av fallet ska vara konstant mellan stationerna och kräver samträning. Studenten ska ha exponerats för det övergripande innehållet på stationen under den gångna kursen. Innehåll på stationen kan förstärkas med till exempel bandage, sminkning av sår, hematom, blodsminkade kläder, verbal förstärkning, laborativärden, EKG, röntgenbilder, sjukvårdsutrustning etc.

Poängsättning

Poäng kan ges på detaljerad eller övergripande nivå inom respektive domän på stationen. Detaljerade poänglistor premierar noggrannhet, ofta lämpligt på lägre nivå där procedurer plockats ner till detaljerat innehåll för att säkerställa teknisk färdighet. Värdera vad som poängsätts (ett poäng för att tvätta händer, ett för att hälsa på patienten jämfört med kliniskt resonemang, process/beteende/tänkt utfall). Såväl vid tillämpning av checklista som vid övergripande poängsättning behöver noggrann kalibrering ske mellan bedömare avseende vad som erfordras på respektive nivå i poäng. Bedömning innebär en kognitiv utmaning där värdering och bedömning av prestation kan influeras av ett flertal faktorer. Samskattningsövningar och bedömandiskussioner inför examinationen minskar risken för bedömarbias.

Beslut om godkänthgräns

Detta avser den process som omfattar de beslut som ligger till grund för ett godkänt/underkänt examinationsresultat, dvs. definition av absolut lägsta nivå för kunskaps-/färdighets- och förhållningssätt som krävs för att ha nått erforderlig nivå i relation till kursens lärandemål.

På läkarprogrammet används absolut godkänthgräns. Vid absolut gräns, till exempel kriteriebaserad godkänd nivå, definieras denna som måluppfyllelse relaterat till formulerade kriterier. Samtliga studenter som uppfyller kriterierna erhåller godkänt resultat vid examinationen. Beslut om godkänthgräns kan ske före eller efter examinationens genomförande. Poänggräns kombineras ofta med en global skattning. Denna skattning överridder den kriteriespecifika bedömningen. Vid godkänt nivå satt innan examinationen sker en validering av stationens innehåll mot en tänkt nivå av en precis godkänd prestation till exempel med hjälp av Angoff.¹⁰ Detta är ett vanligt förfaringsätt vid summativa slutliga examinationer till exempel finala programexaminationer. I denna metod bedömer kliniker/individer i lärarlag hur stor andel av borderline studenter som skulle klara stationen. På detta sätt fastställs den absolut lägsta godkänt nivån. Metoden är relativt tidskrävande.

Andra vanliga metoder är skattningar för godkänthgräns definierad efter examinationen. Den vanligaste av dessa är borderline regression där studentens poäng plottas på y-axeln och global skattning på x-axeln och en linje med bästa passform (regressionslinje) anläggs. Grafens lutning ger god information om hur poängskattningen på examinationen har fungerat samtidigt som global godkänd/underkänd skattning kompenserar för svårigheter med kvantifiering i poängmallen. Den punkt där linjen från "borderline" korsar grafen indikerar "cut-off" för stationen.

Samträning av bedömare

Samträning av bedömare, och därmed en överenskommelse om krav på färdighetsnivå relaterat till måluppfyllelse, utgör en viktig del i all praktisk och teoretisk bedömning.

Psykometrisk analys av tentamensresultat

En övergripande och detaljerad analys av tentamensresultatet av måluppfyllelse avseende till exempel delkomponenter för poäng, samt övergripande måluppfyllelse ger en god grund för bedömning av examinationsinnehåll och resultat som kan relateras till lärandemål, relevanta lärandeaktiviteter och prestationsnivå.

Återkoppling av tentamensresultat

Återkoppling av resultat i fOSCE ges under pågående examination där genomförande/återkoppling kan ske individuellt eller i dialog i grupp. Vid summativ OSCE ges återkoppling skriftligt i sammanställning av återkoppling på samtliga stationer i examinationen. fOSCE som summativ OSCE bör strävan mot att utgöra en del i ett självreglerat progressivt lärande, med möjlighet till såväl övergripande som detaljerad återkoppling. Offentlighetsprincipen ses ibland som problematisk mot bakgrund av krav på framtagande av ständigt nytt examinationsinnehåll. Värt att beakta är dock även att en kriteriespecifik bedömningsmall även utgör en detaljerad återkoppling till studenten avseende tänkt måluppfyllelse/genomförande av delmoment/procedur.

Referenser

1. 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 Sep;35(9): e1437-46. doi: 10.3109/0142159X.2013.818634.
2. Kamran Z Khan I, Kathryn Gaunt, Sankaranarayanan Ramachandran, Piyush Pushkar. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. *Med Teach* 2013 Sep;35(9):e1447-63. doi: 10.3109/0142159X.2013.818635.
3. Townsend AH, McLivenny S, Miller CJ, Dunn EV. The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. *Med Educ*. 2001;35(9):841–6. <https://doi.org/10.1046/j.1365-2923.2001.00957.x>.
4. Brazeau C, Boyd L, Crosson J. Changing an existing OSCE to a teaching tool: the making of a teaching OSCE. *Acad Med*. 2002;77(9):932.
5. Lien HH, Hsu SF, Chen SC, Yeh JH. Can teaching hospitals use serial formative OSCEs to improve student performance? *BMC Res Notes*. 2016;9(1):464. <https://doi.org/10.1186/s13104-016-2266-1>.
6. Möltner A, Lehmann M, Wachter C, Kurczyk S, Schwill S, Loukanova S. Formative assessment of practical skills with peer-assessors: quality features of an OSCE in general medicine at the Heidelberg Medical Faculty. *GMS J Med Educ*. 2020 Jun 15;37(4):Doc42. doi: 10.3205/zma001335. eCollection 2020.
7. Schwill S, Fahrbach-Veeser J, Moeltner A, Eicher C, Kurczyk S, Pfisterer D, Szecsenyi J, Loukanova S. Peers as OSCE assessors for junior medical students - a review of routine use: a mixed methods study. *BMC Med Educ*. 2020 Jan 16;20(1):17. doi: 10.1186/s12909-019-1898-y. PMID: 31948425
8. Martin IG, Jolly B. Predictive validity and estimated cut score of an objective structured clinical examination (OSCE) used as an assessment of clinical skills at the end of the first clinical year. *Med Educ*. 2002; 36: 418– 25. Wiley Online Library PubMed Web of Science @ Google Scholar
9. Magdalene Hui Min Lee 1, Dong Haur Phua 2, Kenneth Wei Jian Heng. The use of a formative OSCE to prepare emergency medicine residents for summative OSCEs: a mixed-methods cohort study. *Int J Emerg Med*. 2021 Oct 1;14(1):62. doi: 10.1186/s12245-021-00383-4.
10. Mohammad Jalili, Sara M Hejri, John J Norcini. Comparison of two methods of standard setting: the performance of the three-level Angoff method. *Med Educ*. 2011 Dec;45(12):1199-208. doi: 10.1111/j.1365-2923.2011.0407

Att skapa OSCE

Grundplanering

- Bestäm datum, klockslag och lokal. Gör en tidslinje bakåt från detta datum.
- Skapa en organisatorisk blueprint för hela OSCE:n.
- Lista alla ansvarsområden, sista datum för respektive område och ansvarig.
- Bestäm mötestider.
- Besluta vem som har huvudansvar.

Rekrytering

- Rekrytera Standard Patienter (SP) och sätt datum och tid för träning av SP.
- Rekrytera lärare och sätt tid och datum för genomgång av fall/träning av lärare, korridorvakt/administratör.
- Säkra medverkan från SP och lärare.
- Besluta vem som har huvudansvar.

Bestäm stationer och skapa kriteriespecifika bedömningsmallar

- Hur fastställs innehållet på respektive station och övergripande?
- Vilken kompetens examineras relaterat till kursplan?
- Tänk på att en kompetens behöver bedömas på flera stationer.

OSCE processen

1. Bestäm ämnesområden och vilka kompetenser inom dessa ämnesområden som ska testas. Gör blueprint och mappa mot lärandemål. Relatera antal ämnesområden till hur stor omfattning respektive moment har på kursen. Samla berörda momentansvariga för planering.
2. Skapa fall/stationer.
3. Använd autentiska patientfall som är representativa för det som undervisats. Täcker stationerna de kompetenser som behöver bedömas? (jmf akut, allvarligt, vanligt). Finns såväl kunskap, förmåga som färdighetsintervention med bland uppgifterna? Är svårighetsnivån varierad? Är fallen representativa för det stora urvalet; jmf kön, ålder, socioekonomi etc.? Tänk på att i instruktionen till SP täcka upp med svar för de potentiella frågor studenten kan komma att ställa vad gäller symptompresentation/scenario. Bestäm ämnet för stationerna. Skapa fallen/stationerna gemensamt eller via remissrunda.

4. Skapa studentinstruktion, lärarinstruktion/stationsöversikt. Bestäm vem som ansvarar för studentinstruktionen.
5. Stäm alltid av relevans i stationens innehåll och kriterier för godkänt med EBM, rekommenderad litteratur och expert inom ämnet.
6. Skapa kvalitativa kriterier för bedömning i linje med lärandemål. Kriterier ska vara kända av bedömaren och studenter inom ramen för undervisningen. Bedömningsmallen ska hjälpa bedömaren att diskriminera vilken nivå studenten presterar på vilket studenten bör vara informerad om. Inkludera även global bedömning.
7. Global bedömning. Denna syftar till att ge upplysning om bedömarens övergripande upplevelse av studentens prestation och kan ibland fånga aspekter som inte riktigt täcks av beteendespecifika item. Använd till exempel UK, GK, Väl GK.
8. Inkludera gärna standard patientens bedömning av studentens prestation till exempel ”Rekommenderar varmt, Rekommenderar; Rekommenderar med reservation; Rekommenderar ej.
9. Gör en pilot med bedömningsformuläret (Be till exempel en kollega agera bedömaren, en student och agera själv patient.) Säkerställ genomförbarhet, tidsaspekt, samt att stationen verkligen examinerar det den syftar till att göra. Bedömningsmallen ska underlätta bedömarens förmåga att diskriminera studentens kunskap och färdighet. Korrigera vid behov i bedömningsmallen.
10. Används någon modell för skattning av GK gräns. Gå igenom mallen och skatta hur en student som precis klarar GK nivå skulle förväntas klara sig.
11. Gå igenom stationerna och bedömningskriterier med bedömaren några dagar innan OSCE för att nå samskattning.
12. Träna SP och gå igenom bedömningsformulär med bedömaren. Genomgång av station med SP kan ske digitalt ca 1-2 veckor innan OSCE.
13. Skapa rotationssnurror med studenter. Stäm av med administratör/vakt.
14. Skriv ut bedömningsformulär och foton på studenterna för att säkra identitet.
15. Tillsä att utrustning på respektive stationen finns/fungerar och kan hanteras av examinerande lärare. Stationerna kan med fördel iordningsställas och testas em/kväll innan OSCE.
16. Ringklocka/visselpipa/gonggong el. dyl. används för annonsering av start, 2 min kvar och avslut på tid på respektive station.
17. Använd gärna video. Studentens prestation kan bedömas av flera bedömaren. Återkoppling blir konkret och faciliterar reflektion.
18. Publicera/delge resultat till studenterna efter avslutad OSCE. Sammanfatta gärna en generisk övergripande feedback; vad förväntades på stationen,



vad gjordes i allmänhet bra, vad kunde gjorts annorlunda. Återkoppling kan ges muntligt direkt efter genomförd OSCE eller senare efter konferens via e-post.

19. Analysera stationen utefter resultat och erfarenhet. Återför resultat till lärandemål och undervisning.
20. Skapa en OSCE-bank med dina stationer och planera för nästa OSCE.

Mall: OSCE-station

Författare	
Ämne	
Termin	
Kompetens	
Anamnes	
Status	
Kommunikation	
Kliniskt resonemang	
Bedömning	
Åtgärd	
Procedur	
Duration	
Information till siteansvariga	
Statist	
Standard Patient (SP)	
Ålder	
Kön	
Utrustning till station	

Mall: Studentinstruktion

Du är läkarstudent/AT-läkare på en vårdcentral /akutmottagning. Du ska bedöma en XX-årig kvinna/man som söker för XX symtom/skada. Du förväntas ta en relevant anamnes/genomföra ett riktat status av XX/utifrån anamnes och statusfynd genomföra ett klinisk resonemang/utifrån anamnes och status föreslå rimlig diagnos/ utifrån anamnes och status föreslå rimlig/ genomföra angiven procedur.

Mall: Information till bedömare

Kort bakgrund till fallet/stationen	
Bedömarens roll/uppgift	
Beskrivning av uppgiften och förväntad prestation av student	
Information som får ges till student	<p><i>Exempel 1.</i> När studenten lyssnar med stetoskopet på buken berättar du att hen inte hör några tarmljud. Buken är tyst.</p> <p><i>Exempel 2.</i> Om studenten inte spontant berättar för patienten ... så fråga: definierad fråga om t.ex. differentialdiagnoser/handläggning efter vårdbesöket/uppföljning etc.</p>
Information som inte får ges till student	<i>Exempel:</i> Resultat av prestation.
Klinisk information som studenten bör få fram i anamnes/struktur i status etc.	

Mall: Information till standard patient (SP)

Vem ska hen vara?	
Ålder	
Kön	
Anamnes - Aktuell	
Debut (När?)	
Lokalisation (Var?)	
Hur upplever patienten sina besvär	
Dygnsvariation	
Analgetika	
Oro	
Anamnes - Generisk	
Socialt	
Pensionär/Arbete	
Gift/barn	
Boende	
Rökning	
Alkohol	
Motion	
Hereditet	
Tidigare/nuvarande sjukdomar	
Allergi	
Läkemedel	
Vad får SP fråga?	
Rekommendation av student	

Mall: Checklista till författargrupp

	Ja	Nej
Har ämnet undervisats under kursen?		
Var har det undervisats på kursen?		
Klarar studenten av stationer på x minuter?		
Är det som ska testas relevant för student termin x?		
Diskriminerar stationen mellan bra /dålig student?		
Är stationen autentisk?		
Motsvarar G på stationen kompetens i VFU?		
Testar stationen vad den är tänkt att testa?		
Går det att hitta en SP för denna station?		

Mall: Bedömning

Station:			
PATIENTKOMMUNIKATION 1			
X	Efterhör ej 0	Efterhör delvis: kriterier 0,25	Efterhör adekvat och fullständigt: Kriterier 0,5
Y	Efterhör ej 0		Sammanfattar rele- vant och avvägt 0,5
Anamnes			
X	Efterhör ej 0	Efterhör delvis: kriterier 0,25	Efterhör adekvat och fullständigt: Kriterier 0,5
Y	Efterhör ej 0	Efterhör delvis: kriterier 0,25	Efterhör adekvat och fullständigt: Kriterier 0,5
Z	Efterhör ej 0	Efterhör delvis: kriterier 0,25	Efterhör adekvat och fullständigt: Kriterier 0,5
Socialt:	Efterhör ej 0	Efterhör delvis: kriterier 0,25	Efterhör adekvat och fullständigt: Kriterier 0,5
Socialt; rökning, alkohol	Efterhör ej 0	Efterhör delvis: kriterier 0,25	Efterhör adekvat och fullständigt: Kriterier 0,5
Tidigare/nuvarande sjukdom	Efterhör ej 0	Efterhör delvis: kriterier 0,25	Efterhör adekvat och fullständigt: Kriterier 0,5
Läkemedel/allergi	Efterhör ej 0	Efterhör delvis: kriterier 0,25	Efterhör adekvat och fullständigt: Kriterier 0,5
Status			
X	Utför ej 0	Utför delvis: kriterier 0,25	Utför adekvat och fullständigt: Kriterier 0,5
Y	Utför ej 0	Utför delvis: kriterier 0,25	Utför adekvat och fullständigt: Kriterier 0,5
Z	Utför ej 0	Utför delvis: kriterier 0,25	Utför adekvat och fullständigt: Kriterier 0,5

Kliniskt resonemang			
	Utför ej 0	Utför delvis: kriterier 0,25	Utför adekvat och fullständigt: Kriterier 0,5
Patientkommunikation 2			
	Utför ej 0	Utför delvis: kriterier 0,25	Utför adekvat och fullständigt: Kriterier 0,5
Bedömning och åtgärd			
	Utför ej 0	Utför delvis: kriterier 0,25	Utför adekvat och fullständigt: Kriterier 0,5
	Utför ej 0	Utför delvis: kriterier 0,25	Utför adekvat och fullständigt: Kriterier 0,5
	Utför ej 0	Utför delvis: kriterier 0,25	Utför adekvat och fullständigt: Kriterier 0,5
Reflektion			
	Utför ej 0	Utför delvis: kriterier 0,25	Utför adekvat och fullständigt: Kriterier 0,5

Godkänt: x poäng

Global bedömning:

Global bedömning av studentens prestation	Underkänd	Med tvekan godkänd	Godkänd	Väl godkänd
--	-----------	--------------------	---------	-------------

Om underkänd motivera:

Återkoppling till studenten:

Vilka aspekter av uppgiften utfördes SÄRSKILT bra?

Vilka aspekter av uppgiften utfördes MINDRE bra?

Mall: Checklista för kursadministratör

	Ja	Nej	Kommentar
Verifiera vilka/hur många studenter som ska delta i OSCE- fastställ deadline för anmälan			
Verifiera lärarbehov/SP/statister			
Skapa rotationssnurror			
Färgsätt respektive rotationssnurror om behov av flera snurror			
Skapa schema för respektive färgsatt snurra med minifoto av student			
Skapa OSCE ID för respektive student			
Trycka ut papper backup av bedömningsdokument			
Sortera in dokument: Information till student/Information till statist alt SP/Information till bedömare och bedömningsmall i kuvert			
Skapa PowerPoint för demo av OSCE snurror till momentansvariga/övriga kursgivare			
Skicka ut mail/information till momentansvarig en vecka innan OSCE			
Skicka ut "Dummy" exempel till lärare			

Station	Moment (hp)	Station/uppgift	Anamnes	Status	Kliniskt resonemang	Bedömning och åtgärd	Kommunikation	Praktisk färdighet
1	Kirurgi (7,5 hp)	Sutur (sår i ansikte)				x		x
2	Kirurgi	Bröstanamnes och palpation Sp	x	x	x		x	x
3	Kirurgi	kärldata				x		x
4	Ortopedi (3 hp)	Hand (de Quervain)	x	x				
5	Ortopedi	(Fotleds-distorsion)		x		x		
6	Uro (1,5 hp)	KAD, spoldropp			x	x		x
7	Anestesi (1,5 hp)	Urosepsis akuten		x	x	x		x
8	Bild & funk (1 hp)	Appendicit			x	x		
9	KUA/IPE (3 hp)	Nytt patientfall			x	x	x	

Referenser och rekommenderad litteratur

- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.). *Educational Measurement, 2nd ed.*, pp. 508-600. Washington DC: American Council on Education.
- Ben-David, M. F. (2000). AMEE Guide No. 18: Standard setting in student assessment. *Medical Teacher 22*(2): 120-130.
- Billings, M. S., DeRuchie, K., Hussie, K., Kulesher, A., Merrell, J., Morales, A., Paniagua, M. A., Sherlock, J., Swygert, K. A., & Tyson, J. (2020). *NBME item-writing guide: Constructing written test questions for the health sciences*. Philadelphia: National Board of Medical Examiners.
- Bandaranayake, R. C. (2008). Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. *Medical Teacher 30*(9-10): 836-845.
- Case, S., & Swanson D. (2002). *Constructing Written Test Questions For the Basic and Clinical Sciences*. Philadelphia: National Board of Medical Examiners. <http://www.nbme.org/publications/item-writing-manual.html>
- Cohen-Schotanus, J., & van der Vleuten, C. P. (2010). A standard setting method with the best performing students as point of reference: practical and affordable. *Medical Teacher 32*(2), 154-160.
- Cook, S. (2008). Writing outstanding MCQs that match your objectives: why keep assessing your student's performance a secret? *Singapore General Hospital Proceedings*, 17(3).
- Ebel, R. L. (1972). *Essentials of Educational Measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Homer, M., & Darling, J. C. (2016). Setting standards in knowledge assessments: Comparing Ebel and Cohen via Rasch. *Medical Teacher 38*(12): 1267-1277.
- Kilgour, J. M., & Tayyaba, S. (2015). An investigation into the optimal number of distractors in single-best answer exams. *Advances in Health Sciences Education: Theory and Practice*, 21(3), 571–585.
- Medical Council of Canada. (2010). Guidelines for the development of multiple-choice questions.
- Millman, J., & Greene, J. (1989). The specifications and development of tests of achievement and ability. In: R. L. Linn (Ed.), *Educational measurement* (3rd Ed.). Phoenix, AZ: American Council on Education.
- Raymond, M. R. (2016). Job analysis, practice analysis, and the content of credentialing examinations. In S. Lane, M. R. Raymond, T. M. Haladyna (Eds.) *Handbook of test development. 2nd ed.* New York (NY): Routledge.
- Raymond, M. R., & Grande, J. P. (2019). A practical guide to test blueprinting. *Medical Teacher*, 41(8), 854-861.

- Sam, A. H., Westacott, R., Gurnell, M., Wilson, R., Meeran, K., & Brown, C. (2019). Comparing single-best-answer and very-short-answer questions for the assessment of applied medical knowledge in 20 UK medical schools: Cross-sectional study. *BMJ Open*, 9(9), e032550–e032550.
- Swanson, D. B., Case, S. M. (1997). Assessment in basic science instruction: directions for practice and research. *Advances in Health Sciences Education: Theory and Practice*, 2, 71-84.
- Tavakol, M.; & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55.
- Williams, C. (2020). Evaluating multiple choice questions for readiness assurance tests and application activities. University of Western States Portland Oregon.

