

# Examinationsformer under läkarutbildningens kliniska del.

Lars Börjesson

*Institutionen för kliniska vetenskaper, avdelningen för Kirurgi  
Sahlgrenska Akademin, GU*

## Abstract

### Bakgrund:

För att bedöma studentens kompetens utefter läkarutbildningen behövs olika examinationsformer. Uppsatsen problematiserar salsskrivning och muntlig examination (analyseras tillsammans såsom Traditionell examination); formerna praktiskt prov, examination med standarpatient och examination med patient analyseras tillsammans (Performance-examination). De båda examinationstyperna analyseras avseende validitet – reliabilitet, relevans – verklighetsanknytning, ytligt-djupt lärande, samt resursåtgång.

### Resultat / kommentar:

#### *Traditionell examination:*

Salsskrivning har högst reliabilitet. Det är svårt att konstruera frågor med hög validitet sent i en läkarutbildning. Traditionell muntlig examination har läge reliabilitet, men kan hålla hög validitet. En svaghet med Traditionell examination är att den inte prövar förmågan att omsätta teori i praktik. Salsskrivning har hög effektivitet, medan muntlig examination är resurskrävande.

#### *Performance-examination:*

Denna examinationsform hög validitet; ett verkligt patientmöte allra högst. Standarpatienter ger möjlighet att fokusera frågeställningar under ett patientmöte. Att producera säkra bedömningsmatriser vid denna examinationsform är krävande. Bedömning av professionell kompetens kräver performance-examination. För att pröva djup kunskap krävs vanligen någon form av samtal som följer upp det praktiska momentet. Examinationsformen är resurskrävande.

### **Konklusion:**

Flera examinationsformer behövs under utbildning av läkare. Salstentamen är effektiv och har hög reliabilitet; den används för att pröva basal teoretisk kunskap. För att pröva den framtida yrkesrollen behövs performance-examinationer. Användning av standardpatienter har fördelar framför riktiga patientmöten.

## Bakgrund

Läkarutbildning måste ständigt anpassats till medicinsk kunskapsutveckling (Eklöf, 2013), men utöver detta även till samhällets förväntningar (Andersson et al., 2012; Wahlqvist & Rödger, 2013). Läkarutbildningen har sakta glidit från att ha varit en ganska renodlad akademisk utbildning mot en yrkesutbildning. I denna yrkesutbildning är djup ämneskunskap kärnan, men allt större vikt har med tiden lagts vid sk ”professionell utveckling”. Mål för denna är t ex helhetssyn utifrån ett vetenskapligt och humanistiskt synsätt, självkänedom och empati, etisk reflektion och förmåga till lagarbete och samverkan.

Utbildningen leder fram till läkarlegitimation och med detta ett betydande ansvar. Det är därför en självklarhet att denna innefattar en rad examinationer. Yrkesrollen är helt uppenbart så mångfacetterad att enbart en examinationsform inte kan användas. En färsk statlig utredning om den svenska läkarutbildningen konstaterar *”En central utgångspunkt för de lärosäten som har tillstånd att utfärda läkarexamen är därför att säkra den kunskapsnivå, kliniska erfarenhet och professionella kompetens som legitimationen är en garanti för. Det innebär att användningen av olika bedömnings- och examinationsmetoder behöver utvecklas.....detta är de enskilda lärosätenas ansvar”* (Lindgren, 2013).

Tanken med detta arbete är därför att, utifrån egen erfarenhet och litteratur, reflektera över styrkor och svagheter av några av de examinationsformer vi använder inom läkarutbildningen vid SA. Mitt eget kärnämne är kirurgi, som i grunden är ett praktiskt ämne med stark förankring i teori. Det är därför ett typiskt moment där olika examinationsformer har en självklar plats. Inom denna kurs har vi genom åren använt oss av:

1. Salsskrivning. Denna har skett som sk MEQ-skrivning (Multiple Essay Questionens; Miller 1990). Ett patientfall presenteras och studenten besvarar relativt öppna frågor av typen, ”Vad gör Du nu?” eller ”Vad svarar Du patienten?”. När en fråga är besvarad lämnar studenten ifrån sig svaret och nästa fråga, som utgår från samma fall, tar vid.
2. Muntlig examination. Två eller tre studenter möter två lärare som ställer frågor, eller kanske mer korrekt, för ett kollegialt och akademiskt samtal.
3. Praktiskt prov (OSCE; objektiv strukturerad klinisk examination; Harden, 1988). Avgränsade, vanligen praktiska uppgifter, som studenten förväntas lösa på ca 10 minuter. Ett sammanlagt prov består ofta av ca 10 sådana uppgifter (stationer).
4. Examination med ”Standardpatient (SP; van der Vleuten et al, 1990). Denna form är fortfarande under utveckling, men innebär i korthet att studenten skall lösa en simulerad

situation, vanligen ett patientmöte. Patienten spelas av skådespelare, som själv kan vara sjukvårdsutbildad, eller inte.

5. Examination med patient (Murhpy et al., 2009). Studenten möter en verklig patient med ett verkligt problem. Uppgiften vanligen uppenbar, nämligen att utifrån kontext lösa problemet, eller i alla fall föra handläggningen framåt på ett ändamålsenligt sätt.

## **Resultat & Diskussion**

För att analysera och problematisera de olika tentamensformerna har jag valt att sammanfoga salsskrivning och muntlig tentamen (traditionell examination) och OSCE, SP och examination med patient (performance-examination).

### Traditionell examination

#### *Validitet-reliabilitet*

De yttre förutsättningarna för en traditionell salsskrivning optimerar reliabiliteten; alla studenter gör samma prov under samma tid i samma lokal. Däremot är det svårt att konstruera skriftliga frågor med hög validitet för examination av kliniska mål i slutdelen av en läkarutbildning. Vi har använt MEQ-konstruktion, en modell som även används i den sista skriftliga examinationen av läkare i Sverige (AT-examinationen; Miller, 1990; AT-nämndens hemsida). Formen ger en möjlighet att konstruera frågor med hög validitet då studenten tvingas fatta beslut utifrån givna förutsättningar på ett likartat sätt som i verkligheten. Största problemet är att konstruera frågorna så att olika typer av missförstånd undviks (och reliabiliteten därför sjunker). Inte sällan är fallen så komplexa att en grupp studenter kommer att vikta information i fallet på annat sätt än det som konstruktören hade tänkt. Utifrån denna tolkning kan handläggningen sedan bli rätt, men använder man en strikt bedömningsmatris blir kanske resultatet på frågan underkänt.

Vid traditionell muntlig examination är förutsättningarna annorlunda. Här är största problemet reliabiliteten. För kursen i Kirurgi var det just svag reliabilitet som gjorde att vi fasade ut denna tentamensform. Vår modell var traditionell och ”gammaldags”, där vi lät examinerarna välja ämne och egentligen helt fritt ställa frågor. Det enda ramverket var kursens mål. Med självklarhet blev detta inte rättvist mellan studenter. Man kan naturligtvis anföra att en enskild student i en legitimationsutbildning skall klara frågor utifrån målen.

Problemet är kravet på examinatorn om momentan ”rättning” av svaren. Det är alltid enklare att godkänna en student och möjligheten att ställa förenklade följdfrågor är upp till examinatorn. Den utveckling vi såg i Kursen i Kirurgi var att vissa examinatorer i stort sett aldrig underkände en muntlig tentamen. Detta är naturligtvis inte rimligt utifrån samhällets krav (Target domain; Crooks et al., 1996; Kane et al., 1999). Validiteten vid muntlig examination har däremot stora möjligheter att bli hög. Kanske är det på sätt och vis den bästa formen av examination för att skapa sig en uppfattning om studenten verkligen har fått förståelse för ämnet. De missförstånd som inte sällan ses vid skriftliga frågor kan också undvikas genom att examinatorn förtydligar sammanhangen; utifrån just denna aspekt ökar reliabiliteten.

#### *Relevans och verklighetsanknytning.*

Utgår man från samhällets krav på en läkare (Target domain; Crooks et al., 1996; Kane et al., 1999) är det uppenbart att traditionell skriftlig/muntlig examination inte är tillräcklig. De teoretiska målen är emellertid omfattande och lämpar sig väl för denna examinationsform. Svårigheten är att pröva om studenten klarar av att omsätta teori till praktik för att lösa ett problem. En traditionell examination kan lätt ifrågasättas som irrelevant, något som examination med en verklig patient i princip är helt ”immun” mot.

#### *Ytligt och djupt lärande*

Ett sätt att se på möjligheten att bedöma och stimulera till djupt lärande är utgå från taxonominivån på både frågor och svar (Bloom et al, 1956). Det är lätt att säga att detta enbart är upp till examinatorn. Men att upprepat konstruera skriftliga frågor som på ett säkert sätt (dvs med hög reliabilitet) prövar kunskap på hög taxonominivå är krävande. Muntlig examination ger i detta avseende mycket bättre förutsättningar.

#### *Resursåtgång*

Salstentamen är rimligen den mest effektiva formen av alla. Resurskonsumtionen ligger på konstruktion och rättning. Ju större kurser ju högre effektivitet. Få examinatorer torde konstruera en helt ny skrivning för en enstaka omtentamensstudent. Muntlig examination är i hög grad resurskrävande; vi brukade använda ett format med två studenter och två examinatorer och beräkna ca 90-120 minuter för detta. Reliabiliteten ökar om man använder få och samtränade examinatorer. Detta innebär i sin tur hög belastning på få individer och det

är då lätt att kretsen vidgas, men sjunkande kvalitet (och en högre andel godkända studenter) som följd.

### Performance-examination

#### *Validitet - reliabilitet*

Simulerade patientmöten (SP; van der Vleuten et al, 1990) är vanliga i utländska läkarutbildningar, inte minst i USA och Canada. De används även frekvent i slutexaminationer för legitimation (Smee et al., 2001). Utgångspunkten för fallkonstruktionen blir med självklarhet ett verkligt patientmöte, där korrekt handläggning av den reella situationen rimligen borde kunna utgöra ”Target domain” för examinationen (Kane et al., 1999). En konsekvens av detta synsätt skulle vara att i stället sträva efter att examinera med riktiga patienter (punkten 5 ovan). Utgår man från validiteten i examinationen har dock SP klara fördelar. Viktigast är att komplexiteten av situationen kan manipuleras och därmed kan examinationen fokuseras; självklara delar kan tas bort (Kane et al.1999). Det blir då enklare för examinatorerna att observera (mäta) ”rätt” detaljer – validiteten torde därför öka.

Validiteten styrs i hög grad av vad man på förhand beslutat sig för att mäta, dvs någon form av bedömningsmatris (Miller, 1990). Det är just vid konstruktionen av denna matris som det är förvånansvärt lätt att hamna i att man ”gör det mätbara viktigt”. Det är vanligen enkelt att observera detaljer, t ex under en kroppsundersökning; moment som utförs bockas av på en checklista och ger ”poäng”. Att i stället ”göra det viktiga mätbart” är att försöka skapa förutsättningar för att mäta om studenten förstått varför momentet utförs och vad konsekvensen blir vid olika utfall av det standardmässiga undersökningsmomentet. Ett sätt att närma sig detta är att efter det att studenten träffat SP får vederbörande diskutera situationen med ”en äldre kollega”; man simulerar en rapportfunktion. Under rapporten kan ”den äldre kollegan” närma sig studenten med frågor som syftar till att öka taxonominivån på examinationen, dvs man söker t ex efter förståelse och generaliserbarhet i studentens resonemang.

Slutligen måste ändå matrisen på ett eller annat sätt summeras och utfallet placeras i förhållande till ett betyg. Detta har jag personligen tyckt vara det svåraste. Om olika delar av en matris skall kunna summeras kräver detta en mycket noggrann viktning av de ingående momenten. Detta är svårt (för att inte säga omöjligt). Är en korrekt upptagen sjukhistoria viktigare än en korrekt kroppsundersökning? Som regel är det nog farligt att arbeta med en fix

viktning om inte den enskilda matrisen använts många gånger och viktningen justerats efter hand. Ett sätt att arbeta med problemet är att använda checklistan som ett strukturerat protokoll, men när det gäller att bedöma om insatsen är godkänd eller inte förhålla sig till en global skala (Pell et al., 2010). Det är helheten som skall bedömas och denna är inte enbart en enkel summering av detaljerna. Avgörande är helt enkelt om utfallet blev ändamålsenligt; synsättet sätter validitet över reliabilitet. Att summera detaljerna hade rimligen ökat reliabiliteten. Reliabiliteten torde dock i alla fall vara betydligt högre än vid riktiga patientmöten, som är ”omöjliga” att få rättvisa mellan olika studenter, i alla fall om man väger in antalet studenter i läkarutbildning vid SA.

Konflikten mellan validitet och reliabilitet kan lätt problematiseras med just ett riktigt patientfall som utgångspunkt för examination. Ett sätt att se den examinationsformen är att den i hög utsträckning är valid. En riktig patient med ett riktigt problem ger förutsättningar för att ”mäta rätt saker”; det är sedan upp till examinatorn att göra mätningen. Patienten (i förlängningen samhällets krav) är knappast intresserad av reliabilitet för just ett legitimationsyrke; ur det perspektivet är det utfallet av det enskilda patientmötet som skall bedömas. Relationen till andra examinationer med andra patienter, har låg eller ingen relevans. Det viktiga för examinatorn är att förhålla sig till komplexiteten av fallet i relation till studentens utbildningsnivå. Samma patientfall kan mycket väl användas under examination av allt i från studenter till specialistläkare; det är ”bara” att anpassa bedömningsmatrisen.

### *Relevans och verklighetsanknytning.*

Det torde i hög grad vara relevans och verklighetsanknytning som drivit utvecklingen av performance-tester (Lindgren, 2013). Jämförs de olika examinationsformerna utifrån dessa aspekter kommer examination med patient alltid att stå som segrare; examinationen är då rimligen redan förlagd i en ”Target domain”. Fördelen med SP (och i viss mån även med OSCE) är att det ger möjlighet att pröva och träna svåra, verklighetsnära situationer upprepat. Ett sådant moment kan vara att hantera ”det oväntade”, kanske dessutom under stress. Detta kan vara mycket lärorikt, men har sin främsta plats inom ramen för formativ examination, där det är viktigt att feed-back momentet görs ändamålsenligt; det finns annars en risk att det totala utfallet blir negativt för studenten.

Som nämnts i inledningen har kraven på lärosäten som ansvarar för läkarutbildning ökat när det gäller att försäkra sig om studenternas ”professionella kompetens” (Lindgren, 2013). I

begreppet ingår t ex självkänedom, förmåga att bemöta människor i svåra situationer, empati, etisk reflektion och förmåga till lagarbete och samverkan. Det enda sättet att försäkra sig om kompetensen är någon typ av examination. På andra lärosäten har man låtit de lekmän (skådespelare) som agerat som SP skatta t ex bemötande och empati och på så vis just försökt ”göra det viktiga mätbart”. Vid muntlig AT-examination i Sverige utgår man också från ett verkligt patientmöte. I den aktuella mallen ingår att patienten skattar AT-läkarens prestation (AT-nämnden; hemsida). Den ”professionella kompetensen” sett som ett sammantaget begrepp är inte lätt att examinera oavsett examinationsform, men någon typ av performance-examination torde vara en förutsättning för att efterleva universitetskanslersämbetes krav.

### *Ytligt och djupt lärande*

I allt lärande torde det vara en målsättning att sträva efter djupt lärande (Marton & Säljö, 2000). Examinationen kan användas för att bedöma djupet av studentens kunskap, men även för att styra studenten mot djupt lärande. Olika examinationsformer ger här olika möjligheter. Vid en OSCE-examination är det svårt att pröva annat än ytlig kunskap; taxonominivån på såväl frågor som svar blir vanligen låg. Examinationsformen lämpar sig bäst för att säkerställa att alla studenter verkligen klarar ett visst (vanligen praktiskt) moment; man vill försäkra sig om en lägstanivå för gruppen. Vid examination med patient och SP krävs vanligen att examinationen efterföljs av någon typ av strukturerat samtal, under vilket riktade frågor kan pröva djupare kunskap; detta kräver en ökad resurs.

### *Resursåtgång*

Performance-tester är resurskrävande. Om t ex 120 studenter skall genomgå en OSCE med 10 stationer räcker inte tiden till för att göra det under en dag om enbart kontorstid används (om inte parallella stationer byggs upp). När vi genomför SP-examination i Kirurgi (ca 120 studenter) räknar vi med att sysselsätta minst 20 lärare på heltid varje dag under tre dagar. Examination med patient är rimligen mest krävande av alla former; inte sällan avsätts två lärare och dessutom med självklarhet en aktiv patientmedverkan. För att utnyttja den formens alla fördelar är vanligen tidsåtgången minst en timma/student. Även om man tycker att sjukvården ständigt skulle tillhandahålla lämpliga patienter kräver en examination med patient ordentlig förberedelse, där observationer och bedömningsmatris anpassas utifrån den aktuella patientens problem och studentens nivå.



## Konklusion

Precis som den statliga utredningen slår fast är det uppenbart att olika typer av examination behövs under en läkarutbildning. Under kursen i Kirurgi har vi beslutat att behålla skriftlig salstentamen. Denna har hög reliabilitet och med kurser på över 100 studenter är den effektiv. Vi är medvetna om att det är svårt att pröva djup kunskap, utan ser detta som ett test för att säkra en lägsta-nivå på teoretisk kunskap – basen för att sedan lösa ett kliniskt problem.

För att pröva studenten i den mer komplexa läkarrollen utvecklar vi examination med standardpatient. Formen har hög relevans och formativt värde. Konstruktion av fallen kräver betydande engagemang för att skapa förutsättningar för att göra det viktiga mätbart och att göra detta med hög reliabilitet.

## Referenser

Andersson S-O et al. (2012). I. *Professionell utveckling inom läkaryrket*. Liber förlag.

AT-nämndens hemsida: <http://ki.se/utbildning/at-provet>.

Bloom BS et al. (1956) Taxonomy of educational objectives. Cognitive domain. New York: Mc Kay.

Crooks et al. 1996. Threats to the valid use of assessments. *Assessment in Education*, Vol. 3, No. 3.

Eklöf M (2013). I: *Lindgren, För framtidens hälsa – en ny läkarutbildning, Bilaga 3. Statens offentliga utredningar*; SOU 2013:15.

Harden (1988) What is an OSCE? *Medical Teacher*: 10:1, 19-22.

Kane et al., 1999. Validating Measures of Performances. *Educational Measurement: Issues and Practice*.

Lindgren S (2013), *För framtidens hälsa – en ny läkarutbildning*. Statens offentliga utredningar; SOU 2013:15.

Marton F & Säljö R (2000) Kognitiv inriktning vid inläring. I: *Hur vi lär*. Prisma.

Miller GE (1990) The assessment of clinical skills/competence/performance. *Academic Medicine*: 65:9, S63-S67.

Murphy DJ et al. (2009) The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom. *Advances in Health Sciences Education*: 14:219-232.

Pell et al. (2010) How to measure the quality of the OSCE: A review of metrics – AMEE guide no. 49. *Medical Teacher*. 32: 802-11.

Smee SM & Blackmore DE (2001) Setting standards for an objective structured clinical examination: the borderline group method gains ground on Angoff. *Med Educ* 35:1009 –1010.

van der Vleuten C & Swanson D (1990) Assessment of clinical skills with standardized patients: State of the art. *Teaching and Learning in Medicine: An International Journal*: 2:2, 58-76.

Wahlqvist M & Rödger S (2013). I: *Lindgren, För framtidens hälsa – en ny läkarutbildning*, Bilaga 4. Statens offentliga utredningar; SOU 2013:15.