

# Manglende sammenhæng mellem præstationer i et virtuelt og i et virkeligt miljø

Lars Konge<sup>1,2</sup> & Mikael Bitsch<sup>1,2</sup>

## RESUME

**INTRODUKTION:** Simulationsbaseret uddannelse giver indlysende fordele for patienterne og de uddannelsessøgende læger. Virtual reality-simulatorer er ofte kostbare, og evidensen for deres effekt er tit mangelfuld, bl.a. som følge af studier med dårlig metodik og få forsøgspersoner. I alle medicinske, simulerede trænings- og evalueringsprogrammer er det et spørgsmål om, hvorvidt programmerne kan overføres på den kliniske virkelige verden. For at belyse denne problemstilling gennemførte vi et studie, hvor vi sammenlignede testpersoners præstation i en bowling simulator med deres præstation på en virkelig bowlingbane.

**MATERIALE OG METODER:** Femogtyve testpersoner spillede to runder bowling på en Nintendo Wii og 25 dage senere på en rigtig bowlingbane. Sammenhængen mellem score i første og anden runde (*test-retest-reliability*) og mellem scoren på simulatoren og i virkeligheden (*criterion validation*) blev undersøgt. Desuden testede vi for evt. forskelle på mænd og kvinders præstationer.

**RESULTATER:** Med simulatoren kunne det ret præcist måles, hvor godt man bowlede (på simulatoren) – *intraclass correlation coefficient* = 0,76. Derimod var der overhovedet ingen sammenhæng mellem testpersonernes virkelige bowlingevner og deres score i simulatoren (Pearsons  $r = 0,06$ ). Vi fandt ingen signifikante forskelle på mænds og kvinders evner.

**KONKLUSION:** Simulationsbaseret test og uddannelse skal baseres på evidens. Det er nødvendigt med flere gode studier, hvor et passende antal forsøgspersoner indgår. Bowlingkompetence bør ikke baseres på måling med Nintendo Wii. Simulerede trænings- og evalueringsprogrammer bør valideres inden brug for at sikre sammenhæng med den virkelige verden.

Formålet med dette studium var at undersøge om testpersoners præstationer på en virtual reality-simulator kunne forudsige, hvordan de klarede sig i den virkelige verden.

## BAGGRUND

I Sundhedsstyrelsens målbeskrivelser for alle de kirurgiske specialer og for den kliniske basisuddannelse beskrives træning i færdighedslaboratorium inklusive simulation som en mulig læringsmetode [1]. Det første danske færdighedslaboratorium åbnede i København i 1996, og de findes nu i betydeligt omfang i alle danske regioner [2]. Her kan såvel medicinstudere-

rende som uddannelsessøgende læger øve sig på noget andet end patienterne – f.eks. avancerede virtual reality-simulatorer (**Figur 1**). Der er mange potentielle fordele ved simulationsbaseret uddannelse: forbedret patientsikkerhed, mindre tidsforbrug på operationsstuerne, ingen ventetid på »den rigtige patient« og mindre stressfyldt læringsmiljø for den uddannelsessøgende [3]. Simulerede, standardiserede øvelser bruges også til at vurdere den uddannelsessøgende med mhp. konstruktiv feedback og certificering [4]. Den store anvendelighed og de mange potentielle fordele har gjort, at der er investeret mange penge i anskaffelse, vedligeholdelse og drift af simulatorer – ofte på beskeden videnskabelig baggrund. Den viste bronkoskopisimulator (Figur 1) (pris: ca. 600.000 kr.) er f.eks. kun valideret i to mindre, randomiserede studier med hhv. seks og ti deltagere [5, 6]. I oversigtsartikler om effekten af simulationsbaseret uddannelse har man konkluderet, at studier af en højere kvalitet er nødvendige, for at man skal kunne belyse den [7, 8]. Specielt efterlyses større studier, hvor man har brugt objektive, validerede scoringsystemer til at sammenligne præstationer på simulatorer med præstationer i virkeligheden (såkaldte *skills transfer*-studier). Dette gav os ideen til et bowlingstudie, hvor man relativt let kunne skaffe mange, egnede, frivillige forsøgspersoner, og hvor der findes en billig, transportabel simulator (Nintendo Wii), der bruger samme internationalt anerkendte scoringsystem som i rigtig bowling.

## MATERIALE OG METODER

Forsøgspersonerne var 25 læger, psykologer, medicinstuderende og administrativt personale ansat på Center for Klinisk Uddannelse, Københavns Universitet og Region Hovedstaden, Den Lægelige Videre- og Efteruddannelse. Alle deltagere udfyldte et skema vedr. køn, alder og tidligere bowlingerfaring på henholdsvis Nintendo Wii og i virkeligheden. Erfaring blev undersøgt vha. en fempunkts Likert-skala (helt uenig → helt enig) og følgende to udsagn: »Jeg er en meget rutineret bowlingsspiller« og »Jeg har spillet rigtig meget Wii-bowling«.

Herefter gennemførte deltagerne to runder bowling på simulatoren (en bowlingrunde består af 12-21

## ORIGINALARTIKEL

1) Center for Klinisk Uddannelse, Det Sundhedsvidenskabelige Fakultet, Københavns Universitet, og  
2) Den Lægelige Videre- og Efteruddannelse, Region Hovedstaden



FIGUR 1

Eksempel på en virtual reality-simulator (bronkoskopi).



FIGUR 2

Skærbillede fra Nintendo Wii-bowlingsimulator.



skud), og scoren blev noteret (Figur 2). Der blev ikke givet mulighed for opvarmning, men der var sørget for forplejning i form af øl eller sodavand efter eget valg. Præcis 25 dage senere foregik den rigtige bowling. Femten deltagere nåede at gennemføre to fulde runder, mens ti deltagere kun bowlede en runde. Igen var der ingen mulighed for opvarmning, og forplejningen var den samme.

#### Statistik

For at undersøge i hvor høj grad deltagere scorede ens i hhv. første og anden runde (*test-retest reliability*) udregnede vi en *intraclass correlation coefficient* (ICC) for både simulatoren ( $n = 25$ ) og den virkelige verden ( $n = 15$ ). For at undersøge, om kvinderne kla-

TABEL 1

Demografiske data.

| Køn, n (n = 25) |        | Alder, år, gns.<br>(min.-maks.) | Bowlingfering, score, gns.<br>(min.-maks.) |               |
|-----------------|--------|---------------------------------|--|---------------|
| Mand            | Kvinde |                                 | Nintendo Wii                               | virkeligheden |
| 11              | 14     | 31,0 (23-55)                    | 1,5 (1-4)                                  | 1,3 (1-3)     |

rede sig forholdsvis bedre i simulatoren, hvor den fysiske komponent er mindre krævende, end de klarede sig på den virkelige bowlingbane, sammenlignede vi resultater for hhv. mænd og kvinder med *independent samples* t-test. En evt. *learning by testing*-effekt blev undersøgt ved at sammenligne resultater for første og anden runde med *paired samples* t-test. Sammenhængen mellem præstationen på simulatoren og præstationen i virkeligheden blev undersøgt med Pearsons koefficient. Denne blev valgt frem for en ICC, da vi havde en formodning om, at forsøgspersonerne generelt ville score højere på simulatoren, og vi ikke var interesserede i at sammenligne absolutte scorere, men kun i at finde korrelation mellem scorere. Alle udregninger blev lavet i SPSS version 18.0 (SPSS Inc; Chicago IL), og en p-værdi under 0,05 blev anset for at være signifikant.

#### RESULTATER

Erfaringsniveauet var generelt meget lavt, og der var ingen meget erfarne. De demografiske data fremgår af Tabel 1.

Der var en rimelig god overensstemmelse mellem scoren i første runde og scoren i anden runde på simulatoren – ICC = 0,76 (Figur 3A). Overensstemmelsen var lidt dårligere i virkelighedens verden – ICC = 0,62 (Figur 3B).

Mændene scorede i gennemsnit 136 point, og kvinderne scorede 131 point på simulatoren. I virkeligheden scorede mænd i gennemsnit 96 point, og kvinder scorede 81 point. Forskellene var ikke signifikante ( $p = 0,68$  i simulatoren og  $p = 0,084$  i virkeligheden).

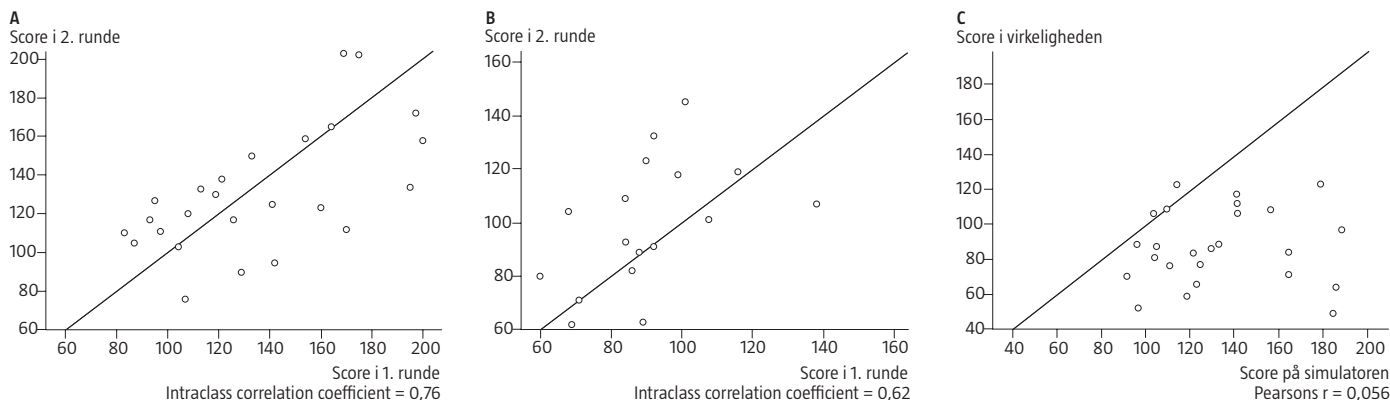
I simulatoren var gennemsnitscoren pr. runde næsten den samme, 135 point i første runde og 131 point i anden runde. I virkeligheden var scoren i anden runde noget højere end scoren i første runde (hhv. 100 point og 89 point), hvilket kunne tyde på en *learning by testing*-effekt. Forskellen var dog ikke signifikant,  $p = 0,09$ .

Scoren var signifikant lavere i virkeligheden end i simulatoren, hhv. 87 point og 133 point,  $p < 0,001$ . Selvom den absolutte score var forskellig, kunne der



FIGUR 3

A. Sammenhæng mellem score i første runde og score i anden runde på simulatoren. B. Sammenhæng mellem score i første runde og score i anden runde i virkeligheden. C. Sammenhæng mellem score på simulatoren og score i virkeligheden.



godt være en sammenhæng mellem scoren på simulatoren og scoren i virkeligheden, hvis de samme personer scorede relativt lavt/højt begge steder. Dette var dog ikke tilfældet, Pearsons korrelation,  $r = 0,056$ . Dvs. at der ikke var skyggen af sammenhæng ( $p = 0,79$ ) (Figur 3C).

## DISKUSSION

Simulatorer giver mulighed for fokuseret træning af tekniske procedurer i et stressfrit miljø uden risiko for patienterne [9]. De kan også anvendes til standardiseret test af uddannelsessøgende med henblik på feedback og certificering [10, 11]. Inden man bruger tid og penge på simulationsbaserede trænings- og evalueringssystemer, må man sikre sig, at disse er pålidelige og valide, og at der er overensstemmelse mellem det virtuelle og det virkelige miljø.

Vores studie viste, at Nintendo Wii-bowlingsimulatoren temmelig præcist målte, hvor gode folk var til at bowle i simulatoren (*test-retest reliability*-ICC = 0,76). Ved mere komplekse procedurer vil scoren ofte afhænge af subjektive vurderinger – f.eks. hvis man på en skala fra 1 til 5 skal vurdere, hvor skånsomt en kirurg behandler vævet. I disse tilfælde bør man lade to eller flere bedømmere foretage scoringen uafhængigt af hinanden og undersøge graden af overensstemmelse i vurderingerne (*inter-rater reliability*) [12]. I bowling er det heldigvis ret nemt at vurdere, om en kegle står op eller ligger ned, så vi har ikke undersøgt dette nøjere. Muligheden for at levere fuldstændig objektive og reproducerbare tal for, hvor god en procedure har været, er en potentiel stor fordel ved virtual reality-simulatorer. Dette gælder imidlertid kun, hvis dataene fra simulatoren har klinisk relevans. Et studie om en bronkoskopisimulator viste, at

komplette nybegyndere efter kort tids træning kunne gennemse alle segmenterne og derved opnå scoren 100% [13]. Dette skete dog ved, at de bevægede sig omkring i bronkietræet som en mus i en labyrint og kiggede ind i alle huller, uden at vide hvor de var. At give læger ret til at bronkoskopere udelukkende ud fra simulatorscoren ville svare til at give kørekort til dem, der var gode til at spille bilspil på computeren.

Det er altså ikke nok, at en simulator måler præcist – målingen skal også være valid. Dette kan undersøges på flere forskellige måder, f.eks. ved at sammenligne scoren på simulatoren med præstationen i den virkelige verden (*criterion validation*) [14]. Her kommer Nintendo Wii fuldstændig til kort – flere testpersoner, der klarede sig godt i simulatoren, viste sig at være elendige til at bowle i virkeligheden (Figur 3C). Dette viser, at bowlingsimulatoren ikke kan bruges til at måle, hvor god man er til at bowle i virkeligheden. Hvorvidt simulatoren kan bruges til at træne fremtidens bowlingspillere er uafklaret, og dette kræver en randomiseret undersøgelse, hvor interventionen er simulatortræning, og effekten måles i den virkelige verden. Denne type studier er specielt blevet gennemført inden for laparoskopi. *Larsen et al* fandt, at simulatortrænede nybegyndere opererede lige så godt som intermedieært erfarne kirurger [15], hvorimod *Hogle et al* i tre konsekutive studier ikke fandt nogen effekt af simulatortræning [16]. I en Cochrane-rapport konkluderes det, at virtual reality-træning i laparoskopi kan anvendes som et supplement til traditionel mesterlære, men at vi behøver flere randomiserede studier af høj kvalitet [17].

## KONKLUSION

Simulationsbaseret uddannelse kan generelt anbefa-

les som et godt supplement til traditionel mesterlære. Beslutninger om, hvordan vi uddanner fremtidens speciallæger, bør baseres på samme grad af evidens som beslutninger om, hvordan vi behandler vore patienter. Dette kræver flere gode simulationsstudier. På baggrund af resultaterne fra nærværende studie kan Nintendo Wii ikke anbefales som redskab til at forudsige bowlingspilleres præstation i virkeligheden.

**KORRESPONDANCE:** Lars Konge, Kongestien 72, 2830 Virum.  
E-mail: lkonge@yahoo.dk

**INTERESSEKONFLIKTER:** Ingen

#### LITTERATUR

1. www.sst.dk (31. maj 2010).
2. Thomsen L, Nielsen HH, Pedersen AM et al. Færdighedslaboratorier. Ugeskr Læger 2004;166:1986-9.
3. Reznick RK, MacRae H. Teaching surgical skills – changes in the wind. N Engl J Med 2006;355:2664-9.
4. Tavakol M, Mohagheghi MA, Dennick R. Assessing the skills of surgical residents using simulation. J Surg Educ 2008;65:77-83.
5. Ost D, DeRosiers A, Britt EJ et al. Assessment of a bronchoscopy simulator. Am J Respir Crit Care Med 2001;164:2248-55.
6. Blum MG, Powers TW, Sundaresan S. Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. Ann Thorac Surg 2004;78:287-91.
7. Sutherland LM, Middleton PF, Anthony A et al. Surgical simulation: a systematic review. Ann Surg 2006;243:291-300.
8. Sturm LP, Windsor JA, Cosman PH et al. A systematic review of skills transfer after surgical simulation training. Ann Surg 2008;248:166-79.
9. Grantcharov TP, Reznick RK. Teaching procedural skills. BMJ 2008;336:1129-31.
10. Martin JA, Regehr G, Reznick R et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997;84:273-8.
11. Issenberg SB, McGaghie WC, Hart IR et al. Simulation technology for health care professional skills training and assessment. JAMA 1999;282:861-6.
12. Axelsson RD, Kreiter CD. Reliability. I: Downing SM, Yudkowsky R, red. Assessment in Health Professions Education. New York: Routledge, 2009:57-73.
13. Colt HG, Crawford SW, Galbraith O, III. Virtual reality bronchoscopy simulation: a revolution in procedural training. Chest 2001;120:1333-9.
14. Streiner DL, Norman GR. Health Measurement Scales – a practical guide to their development and use. Oxford: Oxford University Press, 2008:274.
15. Larsen CR, Soerensen JL, Grantcharov TP et al. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ 2009;338:b1802.
16. Hogle NJ, Chang L, Strong VE et al. Validation of laparoscopic surgical skills training outside the operating room: a long road. Surg Endosc 2009;23:1476-82.
17. Gurusamy KS, Aggarwal R, Palanivelu L et al. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 2009;(1): CD006575.



## AKADEMISKE AFHANDLINGER



Jasper Neergaard Jacobsen:

### Molecular aspects of impaired wound healing

Ph.d.-afhandling

**E-MAIL:** jpij@ssi.dk

**UDGÅR FRA:** Statens Serum Institut.

**FORSVARET FANDET STED:** den 3. december 2010.

**BEDØMMERE:** Tonny Karlsmark, Magnus Ågren og Madeleine Durbeej-Hjalt, Lund Universitet, Sverige.

**VEJLEDER:** Karen A. Krogfelt.

Karen Marie Lyng:

### Computerizing clinical practice guidelines

Guidance in clinical work practice and implications for computerization

Ph.d.-afhandling

**E-MAIL:** kamaly@dadlnet.dk

**UDGÅR FRA:** Datalogisk Institut, Københavns Universitet.

**FORSVARET FINDER STED:** den 20. december, kl. 10.00, lille Auditorium, Universitetsparken, Nørre Alle 1, København.

**BEDØMMERE:** Silvana Quaglini, Anders Grimsmo og Jørgen Bansler.

**VEJLEDER:** Finn Kensing og Thomas Hildebrandt.