

Greining á framförum nemenda innan vefkerfis sem býður upp á gagnvirk krossapróf

Ásta Jenný Sigurðardóttir¹ og Gunnar Stefánsson²

¹Verzlunarskóli Íslands og ²Raunvísindastofnun Háskólans, Háskóli Íslands

Vefútgáfa: 5. nóvember 2009

Ágrip — Svör nemenda við gagnvirkum krossaprófum voru rannsökuð til að meta lærdóm í gagnvirkum umhverfi. Gögnin eru úr efnisveitu (www.tutor-web.net) sem notuð var í námskeiði í Línulegri algebru og tölfraði við Háskóla Íslands. Tölfræðilegar aðferðir eru þróaðar til að prófa og leggja mat á lærdóm nemenda í kerfinu. Aðalmarkmið rannsóknarinnar er að athuga hvort og þá hversu mikil áhrif gagnvirk krossapróf hafa á lærdóm nemenda. Marktæk fylgni er á milli einkunna nemenda í krossaprófunum og einkunnar þeirra á lokaprófinu. Með því að skoða framvindu hvers nemenda innan ákveðins efnis komu ýmis mynstur í ljós. Sér í lagi greindum við talsverða framför í endurteknum spurningum. Með því að skoða sérstaklega spurningar sem innihéldu slembnar tölur er staðfest að þessi framför er ekki eingöngu vegna þess að nemendur leggja svörin á minnið.

Lykilorð: gagnvirk nám, nám með aðstoð tölvu, mat á lærdómi

1. Inngangur

Undanfarin ár hefur orðið mikil þróun í hugbúnaði til kennslu. Mörg þessara forritra eru gagnleg en þó er nauðsynlegt að hafa í huga að tæknin ein og sér tryggir ekki betri kennslu [6]. Í þessari grein er með námi átt við þá athöfn að læra en orðið lærdómur er notað um afrakstur námsins.

Rannsókn sem [11] gerði sýndi að netið hentar ekki síður til kennslu en kennslustofunnar og er mun skilvirkara. Enginn munur var á lærdómi þeirra tveggja hópa sem tóku þátt í rannsókninni en æfingarnar á netinu tóku einungis helming þess tíma sem fyrirlestrarnir tóku og kostuðu minna. Á hinn bóginn upplifði fyrirlestrahópurinn meiri hvatningu og var jákvæðari gagnvart kennslunni en nethópurinn. Þetta er í samræmi við rannsókn sem [9] gerðu en hún benti til þess að hópverkefni sem unnin eru í gegnum netið séu jafn gagnleg og hópverkefni þar sem hópurinn hittist, hvað varðar lærdóm, gæði og innihald lausnara og ánægju með gæði lausnarinnar. Þrátt fyrir þetta voru nemendur mun óánægðari með netreynsluna. Því þarf að hafa í huga þegar slík forrit eru metin að þó ekki sjáist munur á lærdómi, geta nemendur haft mjög ólíkar skoðanir á þeim.

Efnisveitan (tutor-web ¹) er opið vefkerfi sem býður meðal annars upp á gagnvirk krossapróf [12]. Þessi rannsókn kannar notkun þessara krossaprófa með því að skoða gögn sem var safnað í námskeiði í Háskóla Íslands. Án samanburðarhóps er erfitt að meta hrein áhrif krossaprófanna. Hins vegar er hægt að skoða hvernig nemendur nota kerfið og hvernig árangur þeirra á krossaprófunum tengist árangri þeirra á lokaprófinu. Hér verður þó að hafa í huga að sterkur námsmaður mun standa sig væl bæði á krossaprófunum og lokaprófunum. Þess vegna voru þróaðar aðferðir til að meta framför nemenda í krossaspurningunum yfir önnina. Svör við endurteknum spurningum og þá sérstaklega spurningum með slembitölum voru notuð til að leggja mat á lærdóm með tveimur mismundandi tölfræðilíkönum.

Sýnt hefur verið fram á að notkun æfinga og stuttara prófa í vefstuddri kennslu hefur jákvæð áhrif á niðurstöður lokaprófa [6]. Í sömu rannsókn komu ekki fram nein áhrif af fjölda dæma eða þeim tíma sem nemandinn varði í efniseiningu.

¹ <http://www.tutor-web.net>

Ein rannsókn [10] sýndi að þegar nemendur voru spurðir um þrjár bestu hliðar gagnvirkra prófa nefndu 17% tafarlausa svörun, 15% krossaspurningaformið og 13% að það tæki svo stuttan tíma að gera þau. Sama rannsókn sýndi einnig að 75% þeirra sem voru spurðir kusu tölvupróf fram yfir pappírspróf (þar sem svörun er seinni). Önnur rannsókn [5] sýndi að matsaðferðir með tafarlausri svörun efla nám og leiðrétta röng fyrstu svör. Þess vegna gætu krossaspurningar með tafarlausri svörun verið öflug leið til að auka lærdóm nemenda. Það er því áhugavert að skoða hvort hægt sé að meta lærdóm útfrá gögnum tengdum krossaspurningum efnisveitunnar. Rannsókn á fyrri gögnum úr sömu krossaprófum [12] sýndi hækkun í einkunnum nemenda innan hvers hluta krossaprófanna en engin frekari greining var gerð til að meta lærdóm.

Rannsókn sem [4] gerðu sýndi að krossaspurningar á netinu bæta frammistöðu á lokaprófi, ekki bara í krossaspurningunum heldur líka í fræðilegum spurningum. Rannsókn sem [2] gerðu sýndi að krossaspurningar eru ekki góðar til að spá fyrir um árangur á ritgerðarspurningum og öfugt. Í því tilfelli voru krossaspurningarnar einungis notaðar sem mælitæki í lok námskeiðsins en ekki sem tæki til að læra yfir önnina. Aðrar rannsóknir hafa á sama hátt sýnt að tafarlaus svörun er mjög gagnleg á hinum ýmsu stigum náms, þar með talið fyrri hluta háskólastigs [3].

Sérstök áhersla var lögð á krossaspurningar með slembitölum (slembispurningar) í eftirfarandi rannsókn þar sem hægt er að nota þær til að leggja mat á lærdóm án þess að möguleiki sé á að nemendur leggi svörin á minnið.

2. Gögnin

Gögnunum var safnað í námskeiði um línulega algebru og tölfræði við Háskóla Íslands. Í námskeiðinu var notuð efnisveita (<http://www.tutor-web.net>) sem innihélt m.a. gagnvirkar krossaspurningar sem gíltu í lokaekinn. Svör nemenda við þessum spurningum ásamt einkunnum þeirra á lokaprófi og nafnlausri spurningakönnun í lok annar eru gögnin sem voru notuð.

2.1. Efnisveitan (tutor-web)

Efnisveitan er kennsluvefur sem inniheldur glærur, ítarefni, dæmi, heimaverkefni og gagnvirk krossapróf.

Efniveitunni er skipt niður í kafla en hver þeirra inniheldur nokkra fyrirlestra um ákveðið efni. Hverjum fyrirlestri tilheyrir krossaspurningar sem ætlast er til að nemendur í námskeiðinu svari [12].

2.1.1. Krossaspurningarnar

Krossaprófshluti efnisveitunnar er byggður á gagnagrunni krossaspurninga með valmöguleika í slembinni röð. Flestar spurningarnar hafa 3 svarmöguleika en nokkrar þeirra hafa 4 þar sem síðasti möguleikinn er stundum „Ekkert ofangreindra”.

Nemendur geta svarað krossaspurningunum innan hvaða fyrirlesturs sem er, hvenær sem er fram að lokaprófinu. Hugtakið „krossapróf” verður hér notað fyrir þær spurningar sem nemendur fá í tengslum við hvern fyrirlestur. Nemandi getur opnað hvert próf eins oft og hann vill og séð einkunn sína. Þegar nemandi hefur svarað spurningu fær hann strax upplýsingar um hvort hann hafi svarað rétt, ásamt rétta svarinu ef hann valdi það ekki. Nemandinn fær eitt stig fyrir rétt svar en hálft stig er dregið frá ef hann svarar rangt. Einkunn innan hvers fyrirlesturs er reiknuð út frá 8 síðustu spurningum sem nemandinn svaraði innan fyrirlestursins. Tutor-web einkunn (tw-einkunn) hvers nemanda er meðaleinkunn allra fyrirlestranna. Einkunnir eru kvarðaðar í 100.

Til þess að halda utan um feril hvers nemanda verða þeir að skrá sig inn áður en þeir svara spurningunum. Nemendur fá spurningarnar í slembinni röð. Í byrjun hafa allar spurningar jafnar líkur innan hvers fyrirlesturs en spurningar úr fyrri fyrirlestrum hafa lægri líkur. Eina undantekning frá þessu er þegar nemanda vantar bara eitt rétt svar til að vera með fullt hús stiga, í því tilfelli fær nemandinn þá spurningu sem talin er vera erfiðust í fyrirlestrinum (metið út frá heildarsvörun við spurningunni).

2.2. Námskeiðið

Línuleg algebra og tölfræði í Háskóla Íslands eru 8 ECT² einingar og byggist á 45 fyrirlestrum. 62 nemendur luku námskeiðinu, 31 karlmaður og 31 kona. Þetta námskeið er skyldunámskeið fyrir meirihluta nemendanna en þeir koma frá mismunandi deildum háskólans (t.d. lyfjafræði, lífefnafræði, matvælafræði

² European Credit Transfer System

og tölvunarfræði). Tw-einkunnin gildi 10% í lokaeinkunn námskeiðsins, 3 skilaverkefni giltu samtals 30% og lokaprófið gildi 60%. Ætlast var til að hver nemandi svaraði að minnsta kosti 8 spurningum í hverjum fyrirlestri rétt. Til að fá próftökurétt þurftu nemendur að uppfylla að meðaltali a.m.k. 75% af þessu skilyrði. Þessar kröfur eru einungis til þess að fá nemendur til að nota kerfið þar sem þeir gætu fræðilega séð uppfyllt skilyrðin með því einu að giska nógu oft.

2.3. Spurningalistinn

47 nemendur svöruðu nafnlausri könnun sem var lögð fyrir í einum fyrirlestri í lok annar. Í könnuninni var spurt um kyn, námsvenjur, viðhorf til efnisveitunnar, hvort efnisveitan hefði aukið lærdóm o.s.frv.

2.4. Athugasemdir varðandi gögnin

Gögnin sem notuð eru í þessari rannsókn byggja á svörum nemenda sem voru skráðir í námskeiðið haustið 2004 og luku lokaprófi í desember sama ár.

Áður en nemendur svara spurningum í efnisveitunni skrá þeir sig inn í kerfið með því að nota netfangið sitt. Sumar færslur innihéldu minniháttar frávik (t.d. munaði einum staf í aðgangsorði eða veffang vantaði). Þessar færslur voru ekki teknar með þegar tw-einkunnin var reiknuð sjálfvirk meðan námskeiðið var í gangi en þær eru með í greiningunni sem var gerð eftir á.

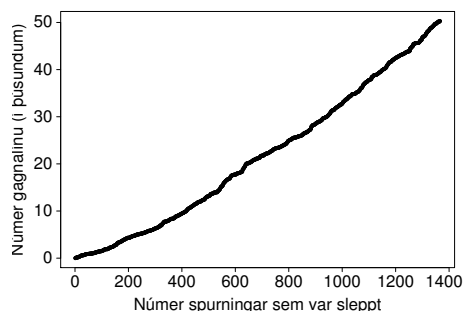
Ef nemandi tvísmellir á spurningahnappinn eða biður um spurningu og hættir án þess að svara þá skráir efnisveitan að spurningu hafi verið sleppt. Það er því ekki hægt að segja hvort nemandi hafi sleppt spurningu viljandi eða óvart.

Villur fundust í sumum spurningum meðan námskeiðið var í gangi en þær voru leiðréttar jafnóðum og öllum gögnum sem þá hafði verið safnað varðandi þá spurningu eytt.

3. Niðurstöður

Heildarfjöldi spurninga sem nemendur fengu var 50.434, þar af var 37.718 svarað rétt, 11.350 var svarað rangt og 1.366 svörum var sleppt. Það kom nokkuð á óvart að svo mörgum svörum væri sleppt þar sem nemendur fengu jafn mikinn frádrátt hvort sem þeir slepptu spurningu eða svöruðu henni rangt.

Þetta gæti stafað af því að nemendur kunna ekki vel á kerfið í byrjun en þá ætti tíðni slepptra svara



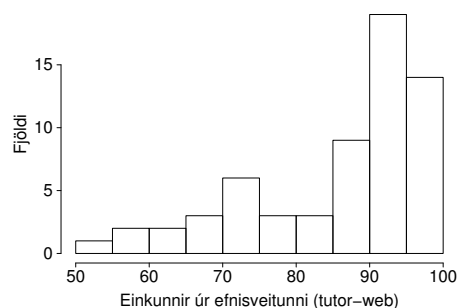
Mynd 1. Svörin sem voru notuð í þessari rannsókn voru númeruð eftir tímaröð. Myndin sýnir hvar í röðinni nemandur slepptu spurningum.

að vera mun hærri í byrjun. Til að kanna þetta nánar voru svör nemenda númeruð eftir tímaröð, $i = 1, \dots, n = 50.434$. Hægt er að skoða staðsetningu slepptar spurningar j, i_j með því að teikna pörin (j, i_j) , $j = 1, \dots, m = 1366$ eins og gert er í mynd 1. Ef spurningum væri aðallega sleppt í byrjun annar þá ætti ferillinn að vera mjög brattur í byrjun. Eins og sést á mynd 1, virðist fátt benda til að sú sé raunin. Raunar virðast slepptu spurningarnar dreifast nokkuð jafnt yfir önnina.

Meðal tw-einkunn allra nemenda var 86,40 (miðað við 8 síðustu svörin í hverjum fyrirlestri) en meðaltal allra svara innan vefkerfisins var 62,18. Dreifing tw-einkunnanna sést á mynd 2.

Meirihluti nemenda fékk mjög góða tw-einkunn en það ætti ekki að koma á óvart þar sem nemendur geta haldið áfram að svara þar til þeir ná ásætlanlegri einkunn.

Eitt lokapróf var haldið í lok annar. Á prófinu voru engar krossaspurningar heldur einungis spurningar og dæmi þar sem nemendur áttu að sýna útreikninga og svara með eigin orðum. Heildarfjöldi stiga í prófinu



Mynd 2. Dreifing einkunnna úr efnisveitunni (tw-einkunnir)

Tafla 1. Niðurstöður aðhvarfsgreiningar þar sem reynt er að spá fyrir um einkunn á lokaprófi út frá tw-einkunn og heildarfjölda tilrauna.

Stuðlar:	Mat	staðalskekkja	t gildi	Pr(> t)
(Skurðpunktur)	22,6515	15,5730	1,455	0,15110
tw-einkunn	0,8154	0,1952	4,177	9,88e-05***
Heildarfjöldi tilrauna	-1,2973	0,4334	-2,993	0,00403**

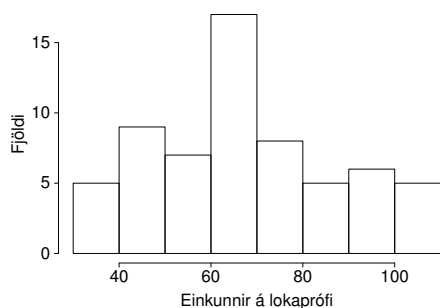
Marktækni: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

var 120 og þurftu nemendur að ná 100 stigum til að fá 10 í einkunn. Fimm nemendur fengu meira en 100 stig (mesti stigafjöldinn var 105) og í þessari rannsókn var stigafjöldi notaður sem einkunn á lokaprófi. Nemendur þurftu að fá a.m.k. 50 stig til að ná prófinu (þó einkunn sé venjulega námunduð frá 45 til 50). Meðaleinkunn úr prófinu var 68,42 og dreifing einkunna sést á mynd 3. Það er meiri dreifing í lokaprófseinkunnum en tw-einkunnum. Graf af lokaeinkunn og tw-einkunn sýnir að nemendur sem fengu háa tw-einkunn fengu einkunnir á bilinu 35 til 104 (sjá mynd 4), en fallið í þeim hópi er þó greinilega mun minna en annarra nemenda.

Línuleg aðhvarfsgreining var notuð til að spá fyrir um lokaeinkunn út frá tw-einkunn ásamt fjölda spurninga sem hver nemandi svaraði í efnisveitunni. Niðurstöðurnar eru í töflu 1.

Báðar breytur eru marktækar, jákvæð fylgni er á milli lokaeinkunnar og tw-einkunnar en neikvæð fylgni er milli lokaeinkunnar og fjölda svara.

Það að nemendur svari mjög mörgum spurningum í efnisveitunni samsvarar því ekki betri lokaeinkunn. Það geta að sjálfsögðu verið nokkrar skýringar á þessu en í grein [12] var einnig sýnt fram á þetta neikvæða samband og þar talið að ástæðan væri að slakir nemendur giskuðu oft til að reyna að fá háa einkunn.

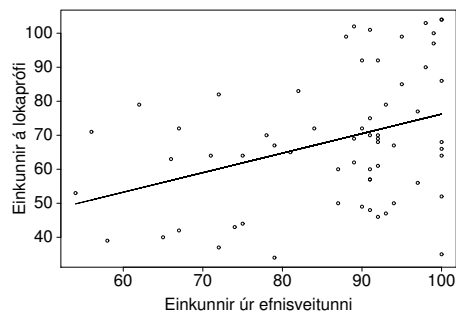
**Mynd 3.** Dreifing einkunna úr lokaprófinu

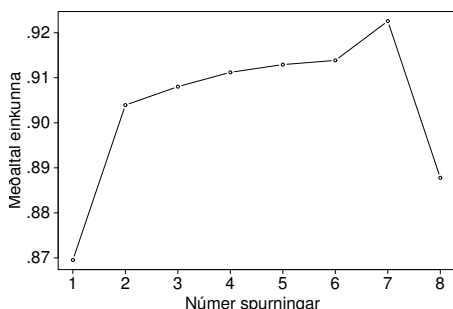
4. Spurningar með slembitölum

Þar sem nemendur geta fengið sömu spurninguna oft- ar en einu sinni er mögulegt að einhverjir þeirra svari rétt í annað eða þriðja sinn vegna þess að þeir muni svarið en ekki vegna þess að þeir kunni efnið (þar sem svarmöguleikum er raðað í handahófskennda röð í hvert skipti verða nemendur að muna svarið sjálft en ekki staðsetningu þess). Til þess að komast hjá þessu vandamáli voru settar inn spurningar með slembitölum. Þar með varð til nánast óendanlegt safn spurninga sem allar prófa það sama en krefjast þess að nemandinn reikni út svarið í hvert skipti.

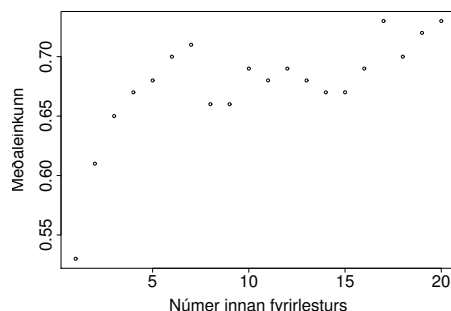
Af 772 spurningum innihalda 105 slembitölur. Meðaleinkunn þeirra spurninga var 66,39 en meðaleinkunn allra spurninga var einungis 62,18. Nú skyldi ætla að meðaleinkunn slembisurninganna væri lægri þar sem þar ekki hægt að leggja svörin á minnið en líklega kemur hér til að erfiðara er að semja spurningar með slembitölum og því ákveðin tilhneiging að nota þann möguleika frekar í auðveldari dæmum.

Fylgnin milli einkunna fyrir slembisurningar og lokaeinkunnar er 0,204 sem er ekki marktækt ($p=0,11$).

**Mynd 4.** Tengsl milli tw-einkunna og einkunna á lokaprófi.



Mynd 5. Einkunnir síðustu átta spurninganna. Meðaltal er tekið yfir alla nemendur og alla fyrirlestra (ef nemandi svaraði færri en 8 spurningum í einhverjum fyrirlestri var þeim svörum sleppt). Síðasta spurningin er lengst til hægri.



Mynd 6. Þróun einkunnar innan fyrirlestra, meðaleinkunnin er reiknuð yfir alla nemendur og alla fyrirlestra.

5. Lærdómur og forspárgildi

Mynd 5 sýnir meðaltalseinkunnir fyrir hverja af síðustu átta spurningunum. Meðaltalið er tekið yfir alla nemendur í öllum fyrirlestrum fyrir utan þau tilfelli er nemendur svörðum færri en 8 spurningum í fyrirlestri. Meðaltal síðustu spurningarinnar er lengst til hægri. Myndin sýnir að meðaltalið hækkar uns nemendur koma að síðustu spurningunni, þá lækkar einkunnin aftur. Þetta stafar sennilega af hönnun kerfisins því ef nemendur svara 7 spurningum í röð rétt þá fá þeir næst þá spurningu sem flestir hafa svarað rangt fram að því.

Ein af áhugaverðustu spurningum þessarar rannsóknar er, hversu mikið nemendur læra á því að nota efnisveituna. Í spurningalistanum í lok námskeiðisins sögðu 36 nemendur (77%) að þeir skildu spurningarnar betur eftir því sem tíminn liði og 42 sögðu að þeir lærðu á því að nota efnisveituna. Þetta lofar góðu en hér á eftir verða gögn frá efnisveitunni skoðuð til að athuga hvort þau styðji þessa tilfinningu nemenda. Þar sem svona könnun er að hluta til mat nemenda á námskeiðinu sem slíku þá var hún nafnlaus. Því er ekki hægt að tengja mat hvers nemanda á eigin frammistöðu við raunverulega frammistöðu hans í efnisveitunni. Þó þetta sé að einhverju leyti takmarkandi þá er hægt að gera nokkur próf og athuganir til að leggja mat á lærdóm.

Spurningunni hvort nemendum gengi betur með síðustu spurninguna í hverjum fyrirlestri en þá fyrstu svörðu einungis 27 játandi en 20 neitandi. Mynd 6 sýnir meðaleinkunnir spurninga eftir röð (meðaltalið nær yfir fyrirlestra og nemendur). Þar sést að einkunn-

ir hækka talsvert eftir fyrstu spurningunna þó þetta sé aðeins tilfinning 57% nemenda. Hér ber að hafa í huga að hafi nemandinn svarað 7 spurningum í röð rétt þá fær hann spurningu sem flestir hafa fram að því svarað rangt, því hefði verið betra að orða spurninguna öðruvísi til að fá svör við því hvort nemendum gengi almennt betur eftir því sem þeir svörðu fleiri spurningum.

5.1. Þróun svara við sömu spurningu

Önnur leið til að meta hvort nemendur læri á því að nota gagnvirk krossapróf er að skoða framför hvers nemanda innan hversrar spurningar. Til að skoða þetta bárum við saman fyrsta og annað svar nemenda við sömu spurningunni. Þar sem tilviljun ræður því hvaða spurningu nemendur fá er misjafnt hvað þeir svara hverri spurningu oft. Þess vegna skoðuðum við bara nemenda-spurninga pör þar sem nemandi hafði svarað viðkomandi spurningu a.m.k. tvisvar. Þetta gaf okkur 12.295 pör. Tafla 2 sýnir skiptinu þessara para.

59% fyrstu svaranna voru rétt en 41% röng. Ef allir nemendur hefðu alltaf giskað hefði hlutfall réttra svara átt að vera um 30% (flestar spurningar höfðu 3 möguleika en nokkrar þó 4). Það er því ljóst að hluti nemenda veit eða getur fundið svarið strax í byrjun,

Tafla 2. Fjöldi nemenda-spurninga para flokkað eftir fyrsta og öðru svari við spurningunum.

Fyrsta svar	Annað svar		Samtals
	Rétt	Rangt	
Rétt	6.486 (53%)	783 (6%)	7.269 (59%)
Rangt	3.580 (29%)	1.446 (12%)	5.026 (41%)
Samtals	10.066 (82%)	2.229 (18%)	12.295 (100%)

Tafla 3. Flokkun á líkum fyrir fyrsta og annað svar ef gert er ráð fyrir að enginn lærdómur eigi sér stað á milli svara.

Fyrsta svar	Annað svar		Samtals
	Rétt	Rangt	
Rétt	p^2	$p(1-p)$	p
Rangt	$(1-p)p$	$(1-p)^2$	$1-p$
Samtals	p	$1-p$	1

eða þeir geta að minnsta kosti fækkað möguleikunum þannig að það verði líklegra að þeir velji rétta svarið.

Í svari númer tvö fer hlutfall réttra svara upp í 80% en það gefur til kynna að einhvers konar lærdómur hafi átt sér stað. 1446 pör (11,8%) eru bæði með fyrsta og annað svarið rangt. Nemandi sem giskar, giskar stundum á rétt svar í fyrsta eða annað skiptið þannig að í raun eru enn fleiri pör þar sem enginn lærdómur átti sér stað.

Aðalspurningin er hvort við getum hafnað núlltilgátunni um að lærdómur sé enginn. Ef ekki var neinn lærdómur þá vissu nemendur annað hvort svarið fyrir eða þeir giskaðu á það og þá ættu að vera sömu líkur á að þeir svöruðu rétt í fyrsta og annað skiptið. Ef líkurnar á réttu svari eru p og lærdómur er enginn þá eru líkindin eins og í töflu 3.

Ef enginn lærdómur á sér stað eru jafnar líkur á því að giska fyrst rétt og svo rangt og öfugt, $p(1-p)$. Í okkar gagnasafni er fjöldi para í fyrri hópnum 783 en 3.580 í þeim seinni. Núlltilgátan er $H_0: p_{12} = p_{21}$ [7], þ.e. enginn lærdómur. Þar sem öllum pörum sem hafa sama svar var sleppt segir höfnun þessarar tilgátu einungis að það var einhver lærdómur en ekki hversu mikill.

Einnig er hægt að skoða spurninguna um lærdóm með því að finna sennileikamat fyrir p miðað við fjölkostadreifingu í töflu 3 og nota svo kí-kvaðrat próf [1] til að meta hvort líkanið sé rétt. Í þessu tilfelli er senni-

Tafla 4. Fjöldi nemenda-spurninga para flokkað eftir fyrsta og öðru svari við spurningunum. Þessi tafla inniheldur einungis síðasta parið sem tilheyrir hverjum nemanda.

Fyrsta svar	Annað svar		Samtals
	Rétt	Rangt	
Rétt	30	2	32
Rangt	22	8	30
Samtals	52	10	62

Tafla 5. Mat á η í útvíkkaða línulega líkaninu.

Stuðlar	Mat	Std. Error	z gildi	Pr(> z)
η	1,36	0,031	41,18	<2e-16 ***

leikamat fyrir p (miðað við fjölkostadreifingu)

$$\hat{p} = \frac{2n_{11} + n_{12} + n_{21}}{2(n_{11} + n_{12} + n_{21} + n_{22})} = 0.705$$

og þetta getum við notað til að skilgreina töflu með væntigildum fyrir kí-kvaðrat próf sömu tilgátu.

Einfalt kí-kvaðrat próf hentar ekki vel á þessi gögn því hver nemandi tilheyrir nokkrum pörum svo forsenda um að pörin séu óháð er ekki rétt. Til að komast framhjá þessum vanda var aðeins tekið síðasta parið sem tilheyrði hverju nemanda (tafla 4) og prófið gert aftur. Í báðum tilfellum eru niðurstöður marktækar ($p < 0,001$) en það gefur til kynna að gögnin sýni fram á lærdóm.

5.2. Læra nemendur?

Til að leggja mat á hversu mikill lærdómur á sér stað var notað útvíkkað línulegt líkan með logit falli (GLM) [8], þ.e. líkur á réttu svari voru settar upp sem

$$\frac{1}{1 + e^{-(\beta + \gamma_s + \delta_q + \eta x)}}$$

en hér er x raðnúmer svarsins.

Þar sem þegar hefur verið sýnt fram á að einhver lærdómur eigi sér stað þá ætti stikinn η að vera marktækur en meira máli skiptir að matið á η er tölusett mat á lærdómi. Niðurstöðurnar eru í töflum 5 og 6 en þar sést að heildarframmistaða nemanda, þyngd spurninga og framför nemenda eru öll mjög marktæk.

Mat á stuðlum líkansins er hægt að túlka á eftirfarandi hátt: Meðaleinkunn fyrstu svara ($x = 1$) í öllum fyrirlestrum fyrir alla nemendur er 42,36. Miðað við þrjú mismunandi svör og einkunnirnir 1 og

Tafla 6. Áhrif af að sleppa ákveðnum breytum úr líkaninu.

	Df	Deviance	AIC	LRT	Pr(Chi)
<none>	23943,2	25599,2			
γ_s	61	24559,6	26093,6	616,3	<2,2e-16 ***
δ_q	765	27783,7	27909,7	3840,4	<2,2e-16 ***
η	1	25794,8	27448,8	1851,6	<2,2e-16 ***
Marktækni: 0 **** 0,001 *** 0,01 ** 0,05 * 0,1 ' ' 1					

Tafla 7. Meðaleinkunn, staðalfrávik og heildarfjöldi þriðju svara flokkað eftir fyrsta og öðru svari.

	1. rétt, 2. rétt (RR)	1. rangt, 2. rétt (VR læra)	1. rétt, 2. rangt (RV giska)	1. rangt 2. rangt (VV giska)
Meðaleinkunn	87	82	59	49
Staðalfrávik	42	49	67	71
Heildarfjöldi	52	45	20	30

-0,5 fyrir hvert rétt eða rangt svar bendir þetta til að 61,6% af öllum fyrstu svörum séu rétt. Með því að nota þá einkunn sem einkunn „dæmigerða” nemandans samsvarar það því að gildi línulega þáttarins sé $\ln(\frac{0,616}{1-0,616}) = 0,47$, þ.e. við fáum $\beta + \gamma_s + \delta_q + \eta \cdot 1 = 0,47$ fyrir slíkan nemanda. Spáð hlutfall réttra svara í annað skiptið fæst svo með því að fara frá $x = 1$ til $x = 2$ í jöfnunni svo matið á einkunninni fyrir svar númer tvö er $\frac{1}{1+e^{-(0,47+1,36)}} = 0,86$ þ.e. 86%. Mat á framför í spurningu sem byrjar með meðaleinkunn er því frá 62% upp í 86% en þessi breyting er mjög marktæk eins og sést í GLM útkomunni.

5.3. Vitneskja, lærdómur, ágiskun

Auk þess að greina spurningapörin hér að ofan er hægt að skipta pörunum í flokka eftir svörum og skoða hvernig þriðja svarið í hverjum flokki er. Flokkarnir eru RR, RV, VR og VV þannig að í RR fóru þau pör sem höfðu bæði svörin rétt, í RV fóru þau pör þar sem fyrsta svarið var rétt en það seinna rangt og svo framvegis.

Þessir flokkar innihalda ákveðnar upplýsingar um hvort nemendur læra eða ekki. Þannig er t.d. nemandi sem lendir oft í VR hópnum nemandi sem lærir og nemandi í RR hópnum kann sennilega efnið eða lærir það um leið og hann svarar spurningu í fyrsta skiptið. Nemandi sem sýnir enga framför (er þá oft í RV eða VV hópnum) er greinilega ekki að læra neitt og verður

Tafla 8. Fjöldi nemenda-spurninga-para flokkaður eftir fyrsta og öðru svari við slembisprungunum.

Fyrsta svar	Annað svar		Samtals
	Rétt	Rangt	
Rétt	974 (59%)	129 (8%)	1,103 (67%)
Rangt	368 (22%)	172 (10%)	540 (33%)
Samtals	1.342 (82%)	301 (18%)	1.643 (100%)

Tafla 9. Mat á η í útvíkkaða línulega líkaninu fyrir slembisprungar.

Stuðlar	Mat	Std. Error	z gildi	Pr(> z)
η	0,94	0,091	10,28	< 2e-16 ***

því flokkaður sem nemandi sem giskar. Við munum notast lauslega við þessar hugleiðingar hér að neðan.

Tafla 7 sýnir meðaleinkunn þriðja svars miðað við ofangreinda flokkaskiptinu. Þar er vert að benda á að það er mjög lítt munur á þeim sem þekkja efnið (RR) og þeim sem læra (VR) en það ætti ekki að koma á óvart að þeir sem giska fá mun lægri einkunn. Með því að setja „giskarana” í einn hóp (RV+VV) fæst með t-prófi marktækt lægri meðaleinkunn fyrir þriðju spurningu heldur en þeir sem læra fá (VR).

5.4. Greining á svörum við slembisprungum

Augljós galli á greiningunni í síðasta kafla er að ekki er hægt að segja til um hvort framfarir nemandans séu vegna þess að hann læri eða hvort hann muni einfaldlega svörin við þeim spurningum sem hann hefur fengið áður. Þar sem ekki dugir að leggja svör við slembisprungum á minnið þá var samskonar greining gerð fyrir þær spurningar (sjá töflu 8).

Í slembisprungunum voru 67% fyrstu svaranna rétt en það hlutfall var einungis 59% fyrir allan spurningahópinn. Í svari númer tvö er enginn munur þar sem í báðum tilfellum er hlutfall réttra svara 82%.

Kí-kvaðrat próf á hornalínuna er mjög marktækt ($p < 0,001$), þannig að líkt og áður höfnum við núlltilgátunni um að lærdómur sé enginn ($H_0: p_{12} = p_{21}$).

Útvíkkað línulegt líkan var gert fyrir þessi gögn eins og í fyrri kafla. Niðurstöðurnar eru í töflum 9 og 10. Af þessu sést að ef 61,6% af fyrstu svörum eru rétt þá ættu $\frac{1}{1+e^{-(0,47+0,94)}} = 0,80$ eða 80% af svörum númer tvö að vera rétt.

Tafla 10. Áhrif þess að sleppa ákveðnum breytum úr líkaninu fyrir slembisprungarnar.

	Df	Deviance	AIC	LRT	Pr(Chi)
<none>		3057,7	3391,7		
γ_s	61	3233,4	3445,4	175,7	4,777e-13 ***
δ_q	104	3482,5	3608,5	424,8	< 2,2e-16 ***
η	1	3168,2	3500,2	110,5	< 2,2e-16 ***

Marktækni: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tafla 11. Meðaleinkunn, staðalfrávik og fjöldi þriðju svara slembisurninganna flokkað eftir fyrsta og öðru svari.

	1. rétt, 2. rétt	1. rangt, 2. rétt	1. rétt, 2. rangt	1. rangt 2. rangt
Meðaleinkunn	83	77	43	54
Staðalfrávik	47	54	73	70
Heildarfjöldi	19	14	8	11

Þetta er næstum því 20% aukning og þó þetta sé lægra en fyrri aukningin sem var uppí 86% þá innihélt fyrri greiningin þann möguleika að nemendur væru að bæta einkunnirnar með því að leggja svörin á minnið. Í þessari greiningu er hins vegar búið að útiloka þann möguleika og því erfitt að túlka þessa 20% hækkun sem eitthvað annað en lærdóm.

Við skoðun á meðaleinkunn þriðja svars fást svipaðar niðurstöður og í fyrri greiningu. Einkunnirnar eru aðeins lægri í þremur af fjórum flokkum en fjöldi svara er líka lítill (sjá töflu 11)

5.5. Samanburður á spurningum með og án slembitalna

Þegar bornar eru saman spurningar með og án slembitalna eins og gert er í töflu 12 þá gefur einfalt kí-kvaðrat próf marktækan mun (þetta er samt einungis hægt að líta á sem vísbendingu þar sem sami nemendinn kemur fyrir í mörgum nemenda-spurninga pörum)

Þessi munur bendir til þess að það verði meiri framför í spurningunum án slembitalna en það er í samræmi við að þar er hægt að leggja svörin á minnið.

6. Umræður

Í þessari grein hefur verið sýnt að einkunnir hækka ekki einungis innan hvers fyrirlesturs heldur einnig innan svipaðra spurninga þannig að að meðaltali þá gengur nemendum betur með spurningar sem þeir hafa fengið áður. Til að komast hjá því vandamáli að nemendur leggi svörin á minnið þá var samskonar samanburður gerður á spurningum sem innihéldu slembitölur. Með slembitölum var aukningin einnig há og marktæk en það bendir til þess að nemendur læra á meðan þeir svara spurningum í efnisveitunni. Þó það sé varasamt að draga ályktanir um orsök og afleiðingu af einföldum tölfræðilegum aðferðum þá er erfitt að ímynda sér annað en að greiningin hér á und-

Tafla 12. Samanburður á spurningapörum með og án slembitalna.

	Fyrsta rétt, annað rangt,	Fyrsta rangt, annað rétt,	Samtals
Slembisurningar	129 (26%)	368 (74%)	497 (100%)
Aðrar spurningar	654 (17%)	3.212 (83%)	3.866 (100%)

an sýni að notkun gagnvirkra krossaprófa auki þekkingu nemenda á efninu.

Slíkt kerfi hefur einnig ýmsa kosti fyrir leiðbein- andann þar sem það krefst lítillar vinnu eftir að búið er að gera spurningarnar. Gögnin sem voru notuð eru úr námskeiði þar sem einkunnir úr gagnvirkum krossa- prófum gilda í lokaeinkunn og þar sem kerfið sér um að reikna út þær einkunnir þá gefst meiri tíma til að þróa frekar efni námskeiðisins.

Einnig var sýnt að tw-einkunnin og heildarfjöldi tilrauna tengjast lokaprófseinkunn. Jákvæð fylgni er á milli lokaeinkunnar og tw-einkunnar en neikvæð fylgni er milli lokaeinkunnar og fjölda svara, þ.a. það að svara mörgum spurningum í efnisveitunni sam- svarar ekki góðri einkunn. Mikill fjöldi tilrauna gæti einfaldlega þýtt að nemendur giska eða læra „rétt nóg“ til að fá ásætlanlega einkunn frekar en að öðl- ast þann djúpa skilning sem þarf til að ná prófi.

Þar sem tilgangur krossaprófsins er að hjálpa nemendum að læra frekar en að mæla þekkingu þeirra og þar sem tafarlaus svörun eykur lærdóm [5], þá væri áhugavert að skoða hvort enn meiri lærdómur fengist með því að svara með leiðbeiningum í stað þess að gefa bara upp rétta svarið. Slíkar leiðbeiningar gætu t.d. innihaldið tengil í efnið sem spurningin er úr eða lýsingu á útreikningum.

Eitt af því sem ekki var skoðað var tími, þ.e. hvort það skipti máli hversu oft eða á hvaða tíma annarinnar nemendur svöruðu spurningunum. Þetta væri áhuga- vert að skoða þar sem nemendur hafa engin tímamörk fyrir krossaprófið önnur en enda annarinnar. Þó þetta veiti nemendum frelsi til að skipuleggja nám sitt þá er ekki víst að þeir ráði allir við þetta frelsi þar sem margir þeirra nota síðustu viku annarinnar til að vinna upp fjölda svara.

Í könnuninni í lok annar þegar nemendur voru spurðir hvort þeir kysu frekar að svara spurningum í efnisveitunni eða hvort þeir vildu hefðbundin heima- dæmi þá sögðu 31 nemandi að þeir vildu frekar efn- isveituna en 14 heimadæmin. Þar sem könnunin var

Tafla 13. Samanburður á svörum nemenda eftir því hvort spurningarnar innihéldu slembitölur eða ekki.

	Rétt	Rangt	Sleppt	Samtals
Spurningar með slembitölum	5.277 (77,6%)	1.283 (18,9%)	241 (3,5%)	6.801 (100%)
Aðrar spurningar	32.441 (74,3%)	10.067 (23,1%)	1.125 (2,6%)	43.633 (100%)
Samtals	37.718	11.350	1.366	50.434

nafnlaus er ekki ljóst hvort þeir sem kusu efnisveituna voru þeir sem græddu mest á því að nota hana. Það væri því áhugavert að endurtaka rannsóknina með upplýsingum um val nemenda. Annar möguleiki væri að bjóða tvo eða fleiri möguleika á æfingum og skoða svo muninn á þeim sem nota efnisveituna og þeim sem nota annað form á æfingum. Einnig væri hægt að hafa próf á undan og eftir til að skoða áhrif mismunandi æfinga.

Augljósasta umbótin á efnisveitunni væri að velja spurningarnar á skynsamari máta. Til dæmis væri hægt að velja spurningar með jöfnum líkum í byrjun og færa sig svo smá saman í þyngri spurningar þegar einkunninn hækkar og jafnvel velja spurningu sem þessi tiltekni nemandi hefur svarað rangt áður. Slíkar aðferðir við að vega spurningar væri hægt að byggja í upphafi á gögnum sem hefur verið safnað við notkun efnisveitunnar þó svo slík greining myndi ekki taka tillit til breyttrar hegðunar nemenda vegna breytinga í kerfinu. Þar sem nemendur virðast hafa tilhneigingu til að halda áfram að nota gagnvirka kerfið þar til þeir ná ásættanlegri einkunn þá væri líklegt til árangurs að þróa vigtunarkerfi sem væri þannig að að tw-einkunnir (a) hefðu forspárgildi varðandi loka-prófseinkunn og (b) væru lágar þar til nemandinn væri líklegur til að ná prófinu.

Þakkir

Þessi grein byggir á lokaverkefni Ástu Jennýar Sigurðardóttur í M. Paed námi hennar við Háskóla Íslands. Dr Lorna Taylor fær þakkir fyrir gagnlegar athugasemdir.

Summary: Web-based student responses are analyzed to evaluate learning, based on multiple choice quiz questions in a dynamic and interactive environment. Inference is based on data gathered while using a quiz system in a college-level undergraduate course on mathematics and statistics. Statistical techniques are developed to test and quantify learning during use of the system. The primary aim of the analysis is to verify whether and by how much a mandatory quiz system improves student performance.

Students' grades on the quiz were found to be significantly correlated with their performance on the final exam. By looking at individual progress within a subject several patterns were detected. In particular considerable improvement was measured on questions which students got repeatedly. By analysing questions containing randomly generated numbers it was confirmed that this does not merely involve rote learning of appropriate responses.

Heimildir

- [1] Agresti, A., (1990) *Categorical Data Analysis*, John Wiley & Sons, Inc.
- [2] Becker, W. og Johnston, C., (1999) *The relationship between multiple choice and essay response questions in assessing economics understanding*, *Economic Record* **75**(231), 34–357.
- [3] Dibattista, D., Mitterer, J.O. og Gosse, L. (2004), Acceptance by undergraduates of the immediate feedback assessment technique for multiple-choice testing *Teaching in Higher Education* **9**(1), 17–28.
- [4] Douglas, D. E. og van der Vyver, G., (2004) Effectiveness of e-learning course materials for learning database management systems: An experimental investigation, *The Journal of Computer Information Systems* **44**(4), 41–48.
- [5] Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B. og Brosvic, G. M., (2002) Immediate feedback assessment technique promotes learning and corrects inaccurate first responses., *The Psychological Record* **52**(2).
- [6] Janicki, T. og Steinberg, G., (2003) Evaluation of a computer-supported learning system, *Decision Sciences Journal of Innovative Education* **1**(2).
- [7] Kuebler, R. R. og Smith, H., (1976) *Statistics, A beginning*, John Wiley & Sons, Inc.
- [8] McCullagh, P. og Nelder, J., (1989) *Generalized Linear Models*, 2nd edn, Chapman & Hall.
- [9] Ocker, R. J. og Yaverbaum, G. J., (1999) Asynchronous computer-mediated communication versus face-to-face collaboration: Results on student learning, quality and satisfaction *Group Decision and Negotiation* **8**, 427–440.
- [10] Peat, M. og Franklin, S., (2002) Supporting student learning: the use of computer-based formative

- assessment modules, *British Journal of Educational Technology* **33**(5).
- [11] Schmeeckle, J.M. (2003) Online training: An evaluation of the effectiveness and efficiency of training law enforcement personell over the internet *Journal of Science and Technology* **12**(3).
- [12] Stefánsson, G., (2004) The tutor-web: An educational system for classroom presentation, evaluation and self-study, *Computers & Education* **43**, 315–343.

Um höfundana: Ásta Jenný Sigurðardóttir lauk BS og MPæd prófi í stærðfræði frá Háskóla Íslands. Hún er nú stærðfræðikennari í dagskóla og fjarnámi við Verzlunarskóla Íslands.

Gunnar Stefánsson er prófessor í stærðfræði.

Ásta Jenný Sigurðardóttir
Verzlunarskóla Íslands
Ofanleiti 1
103 Reykjavík
jenny@verslo.is

Gunnar Stefánsson
Raunvísindastofnun Háskólans
Dunhaga 3
107 Reykjavík
gunnar@hi.is
Móttekin: janúar 2009