

## Athugun á samræmdu einkunnunum 5,0 og 6,0

Sigurgrímur Skúlason  
Námsmatsstofnun

Próffræðilegir eiginleikar samræmdra einkunna voru kannaðir með aðferðum sem eru til þess hannaðar að gera einkunnir á ólíkum útgáfum sama prófs jafngildar. Var rannsóknin afmörkuð við einkunnirnar 5,0 og 6,0. Skilgreind var grunnútgáfa samræmdra prófa í íslensku og stærðfræði og einkunnir á sex öðrum útgáfum prófanna tengdar einkunnum á þeim. Niðurstöður sýna að neðri mörk einkunnanna 5,0 og 6,0 liggja mjög nærri sömu mörkum og í grunnútgáfunni í sumum útgáfum en víkja talsvert frá þeim í öðrum útgáfum. Einnig kemur fram að hlutfall nemenda sem fær tiltekna einkunn eða hærri einkunn víkur frá grunnútgáfu um 0,7% til 25,8%. Niðurstöður eru ræddar í ljósi réttmætis samræmdra einkunna og út frá jafnrétti nemenda með tilliti til ákvarðana sem byggjast á samræmdum einkunnum. Ljóst er að samræmdar einkunnir sem notaðar hafa verið síðan árið 1985 standa ekki undir þeim kröfum sem gerðar eru til einkunna sem gegna jafn veigamiklu hlutverki og þær.

Þegar aðgangur nemenda að æðra skólastigi byggir á frammistöðu þeirra á því skólastigi sem þeir eru að ljúka er sjálfsgöð krafa að lágmarksviðmið taki mið af sömu kunnáttu eða færni burtséð frá því í hvaða námsgrein nemandi tekur próf eða á hvaða ári hann lýkur skólastiginu. Í þessu felst ein hlið á réttmætri notkun á niðurstöðum prófs (Messick, 1989). Með þeim hætti er tryggður jafn réttur til frekara náms, t. d. að loknum grunn- eða framhaldsskóla, í þeim skilningi að frammistaða sem gefur rétt til framhaldsnáms eitt árið gæfi það jafnframt árið á undan og árið á eftir. Á sama hátt væri tryggður jafn réttur til framhalds á öllum skilgreindum námsbrautum. Í reglugerð um samræmd próf (414/2000) eru viðmiðunarmörk eða lágmarkskröfur hinsvegar skilgreind á einkunnakvarða sem byggir á hlutfalli prófatriða sem nemandi leysir. Við núverandi fyrirkomulag er nemendum í reynd mismunað eftir því á hvaða ári þeir ljúka grunnskóla.

Fljótt á litið virðist sjálfgefið að nemandi

sem fær einkunnina 7,5 á samræmdu prófi eitt vorið standi jafnfætis nemandi sem fékk 7,5 á samræmdu prófi tveimur árum fyrr. Þetta þarf ekki að vera rétt. Þegar litið eru á einkunnir þessara nemenda virðist sem þeir ráði báðir yfir svipaðri kunnáttu eða færni og eigi því báðir að standa jafnfætis þegar að útskrift af einu skólastigi eða inntöku á næsta skólastig kemur. Forsendur af þessu tagi, þó rangar séu, liggja til grundvallar ákvæðum um lágmarksviðmið (falleinkunn) í reglugerð um samræmd próf í 10. bekk grunnskóla (414/2000).

Það er algengara en ekki að fleiri en ein útgáfa af sama prófi sé notuð. Þannig eru til að mynda samin og notuð tvö samræmd próf á ári í hverri námsgrein sem prófað er í 10. bekk (414/2000). Eru þá aðalpróf og sjúkrapróf tvær útgáfur sama samræmda prófsins. Einnig eru próf sem lögð eru fyrir ár frá ári ólíkar útgáfur sama prófs. Að jafnaði samanstanda kunnáttupróf sem notuð eru í Bandaríkjunum (CAT, ITBS, MAT) af tveim-

ur eða þremur sambærilegum útgáfum prófsins (Anastasi, 1989). Inntökupróf fyrir háskólastig þar í landi eru að jafnaði notuð í fimm til sjö sambærilegum útgáfum á ári, t.d. ACT, SAT, TOEFL og GRE. Þeir sem nota prófin við inntöku nemenda á námsbrautir verða því að geta treyst á að einkunnir á öllum útgáfum hafi sömu eiginleika.

Niðurstöður prófa eru notaðar bæði til upplýsingar og sem grundvöllur fyrir ákvarðanir (Messick, 1989). Kennarar taka ákvarðanir tengdar því sem gerist í skólastofu á grunni athugana, prófa og annars námsmats, skólar taka ákvörðun um hvort árangur nemenda telst fullnægja kröfum og enn aðrir taka ákvarðanir um hverjir komast inn á námsbrautir þar sem takmarka verður fjölda nemenda. Þegar afdrifaríku ákvarðanir, eins og þær síðasttöldu, eru teknar skiptir miklu að sérkenni mælitækja hafi sem minnst áhrif (Messick, 1989). Til að tvær útgáfur sama prófs geti gefið sambærilegar niðurstöður verða þau að veita sömu upplýsingar um nemendur. Próffræðileg skilgreining á sambærilegum útgáfum sama prófs (*equivalent test forms*) felur í sér að útgáfurnar hafi sama meðaltal, dreifitölu (og um leið staðalfrávik), skekkju (*skewness*) og ris (*kurtosis*). Þetta felur í sér að einkunnir á báðum prófum dreifast eins (Feldt og Brennan, 1989). Í reynd er ávallt einhver stigatölumunur á ólíkum útgáfum sama prófs. Prófatríði sem reyna eiga á sömu færni eru að jafnaði misþung, blæbrigðamunur er í samsvörun við námsmarkmið og ónákvæmni í orðalagi eða gallar í spurningum gera sumar spurningar þungri eða léttari en til var ætlast. Þessi vandamál eru vel þekkt í kunnáttu- og sálfræðilegu mati og hafa verið þróaðar próffræðilegar aðferðir til að gera sambærilegar útgáfur prófs jafngildar í þeim skilningi að ekki skipti máli fyrir nemanda hvaða prófútgáfu hann tekur vegna þess að hann fengi sömu einkunn samkvæmt þeim báðum (Angoff, 1972; Peterson, Kolen & Hoover, 1989). Frá próffræðilegum sjónarhóli eru jafngildar einkunnir skilgreindar með

þeim hætti að einkunnir séu jafngildar ef nemandi hefur ávallt jafnar líkur á að fá sömu einkunn á hvorri útgáfu prófs sem er (Kolen og Brennan, 1995; Lord, 1980). Til dæmis að nemandi sem tæki samræmt próf í stærðfræði frá 1997 ( $M = 5,0$ ) og 1998 ( $M = 6,2$ ) hefði jafnar líkur á að ná lágmarksviðmiði, einkunn 5,0, á báðum prófum þrátt fyrir mun á þyngd prófanna.

Samræmdar einkunnir sem notaðar hafa verið frá 1985 til dagsins í dag eru skilgreindar í reglugerð (65/1985) og eru þær línuleg föll – svona um það bil – af stigatölum. Þær endurspeglar því breytileika í mismun í áherslum á námsmarkmið og svo þyngd prófatríða og breytileika í dreifingu stigatalna sem þessu fylgir. Viðmiðunarmörk eða lágmarks-einkunnir eru því í raun ólík eftir útgáfum prófanna. Einkunnastigar eins og samræmdar einkunnir skortir því alla þá eiginleika sem æskilegir eru í einkunnum sem gegna svo veigamiklu hlutverki og megna ekki að endurspeglar frammistöðu með þeim hætti að hún sé sambærileg milli prófa (Brennan & Kolen, 1989; Dorans, 2002; Peterson, Kolen og Hoover, 1989). Hugsanlegt er, til dæmis, að röng ákvörðun um inntöku nemenda á námsbraut væri vegna þess að þungt próf gæfi til kynna slakari kunnáttu en raunin er eða að létt próf bendi til meiri kunnáttu en í raun er. Það skiptir því miklu máli fyrir suma nemendur hvora útgáfu prófs þeir taka – nokkur fjöldi nær viðmiðunarmörkum á léttari prófútgáfu en nær þeim ekki á þungri prófútgáfu. Alvarleg vandamál geta skapast út af þessum veikleika samræmdra einkunna vegna þess að inntökuskilyrði framhaldsskólans eru bundin þeim. Enn hættara er við vandamálum á framhaldsskólastiginu ef slíkar einkunnir verða einnig notaðar á samræmdum prófum þess skólastigs.

Vegna vandamála af því tagi sem fjallað hefur verið um hér að ofan hafa verið þróaðar próffræðilegar aðferðir til að gera einkunnir sem byggja á ólíkum útgáfum sama prófs jafngildar. Aðferðirnar byggja á að aðlaga dreifingu stigatalna á einni útgáfu prófs að

dreifingu stigatalna á annarri útgáfu prófsins áður en einkunnir eru reiknaðar eða að aðlaga stigatölur allra útgáfa að fræðilegri dreifingu (Angoff, 1971; Kolen og Brennan, 1995). Dæmi um próf þar sem dreifing stigatalna á einni útgáfu er notuð til að skilgreina grunnkvarða er ACT prófið. Einkunnir þess voru endurskilgreindar árið 1989 með þeim hætti að skilgreina tengsl stigatalna við einkunnakvarða á grunni einnar prófútgáfu, dreifing stigatalna á síðari útgáfum prófsins hefur síðan verið löguð að dreifingu þessa grunnkvarða áður en einkunnir eru reiknaðar (Brennan og Kolen, 1989; Kolen og Hanson, 1989). Dæmi um einkunnir sem styðjast við fræðilega dreifingu eru samræmdar einkunnir sem notaðar voru á árunum 1976 til 1983 (Mentamálaráðuneytið, 1979; 398/1977; 141/1980), normaldreifðar einkunnir sem í dag eru notaðar til að reikna meðaltöl á samræmdum prófum og raðeinkunnir sem notaðar hafa verið í tengslum við samræmd próf frá árinu 2000 (Sigurgímur Skúlason, 2000).

Tilgangur þessarar athugunar var að kanna að hvaða marki samræmdar einkunnir eru jafngildar frá einu ári til annars. Verkefnið var afmarkað við tvær einkunnir, 5,0 og 6,0. Þær voru valdar þar sem þær liggja nærri viðmiðunarmörkum sem máli skipta við lok grunnskóla og inntöku nemenda í framhaldsskóla. Þær tengjast því mikilvægum ákvörðunum sem geta haft áhrif á framtíð nokkurs fjölda nemenda. Jafnframt fær mikill

fjöldi nemenda árlega einkunnir sem þessar. Eiginleikar þessara einkunna eru því áhugaverðari en eiginleikar annarra samræmdra einkunna.

## Aðferð

### Gagnasöfn.

Gagnavinnsla byggir á stigatölum samræmdra prófa í íslensku og stærðfræði frá árunum 1995 til 2001. Í hverju gagnasafni eru niðurstöður 3668 til 4275 nemenda. Fjöldi nemenda sem þreyttu hvert próf, meðaltal og staðalfrávik stigatalna, og áreiðanleika má finna í 1. töflu.

### Mælitæki.

Frá því samræmd próf voru fyrst lögð fyrir árið 1977 hafa nokkrum sinnum orðið umtalsverðar breytingar á uppbyggingu þeirra, fjölda og tegundum prófverkefna, hlutverki þeirra í skólakerfinu og umgjörð við framkvæmd þeirra. Prófin sem notuð voru þau ár sem þessi rannsókn nær til eru að flestu leyti efnislega svipuð. Stærðfræðiprófin samanstanda öll af 60 til 70 tveggja og þriggja punkta prófatriðum. Öll árin meta prófin námsþættina reikning og aðgerðir, algebru, rúmfræði og tölfræði. Innbygðis hlutföll í vægi eða fjölda prófverkefna í hverjum námsþætti eru tiltölulega svipuð frá ári til árs. Hlutfall fjölvalsspurninga og opinna

**1. tafla.** Meðaltöl, staðalfrávik og fjöldi nemenda sem þreyttu samræmd próf í stærðfræði og íslensku árin 1995 til 2001.

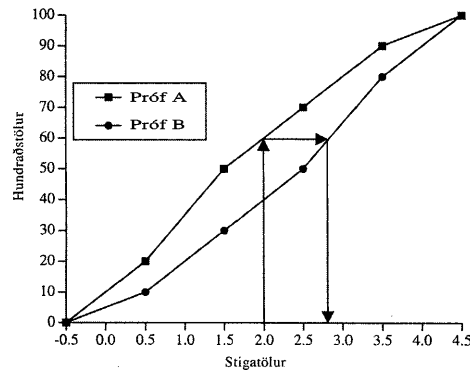
	Íslenska			Stærðfræði		
	M	sf	N	M	sf	N
2001	6,4	2,3	3687	5,8	2,4	3668
2000	6,6	1,2	3879	5,1	2,4	3870
1999	6,5	1,5	4071	5,7	2,3	4059
1998	6,6	1,6	4176	6,2	2,3	4188
1997	6,5	1,6	4176	5,0	2,2	4188
1996	5,8	1,8	4239	5,5	2,4	4275
1995	5,3	1,7	4213	5,8	2,1	4216

prófatriða (dæma) breytast mjög milli prófa á tímabilinu (Námsmatsstofnun, 2001b; RUM, 1996b, 1997b, 1998b, 1999b, 2000b).

Samræmd próf í íslensku á þessu tímabili samanstanda af stafsetningu, ritun, málfræði, og bókmenntum eða lestri. Vægi námsþáttanna stafsetning og ritun var stöðug á tímabilinu en vægi námsþáttanna málfræði og bókmennta og lesskilningur breytast talsvert innbyrðis eftir árum. Efnistöð haldast óbreytt í stafsetningu og málfræði á þessu tímabili. Ritun breytist að því leyti að matsreglur taka miklum breytingum á einum tímapunkti. Bókmenntir og lesskilningur breytast mjög milli árunna 2000 og 2001, fallið er frá því að prófa úr fyrirfram ákveðnum skáldverkum sem nemendur hafa lesið og farið að prófa lesskilning og bókmenntaskilning út frá ólesnum textum. Fjöldi prófatriða er breytilegur og hlutföll fjölvalsspurninga eykst almennt á kostnað eyðufyllinga og opinna spurninga frá eldri prófum til þeirra yngstu (Námsmatsstofnun, 2001a; RUM, 1996a, 1997a, 1998a, 1999a, 2000a).

### Aðferð

Notuð var hundraðstalnateging (*equi-percentile equating*) til að gera einkunnir jafngildar. Þessi aðferð gerir allar stigatölur dreifingar jafngildar dreifingu á grunnkvarða. Aðferðin styðst við hundraðstölur (*percentile*), sem eru í tölfraðilegum skilningi ákveðinn eiginleiki stigatalna í hverri dreifingu einkunna. Stigatölur nýrrar útgáfu af prófinu eru tengdar þeirri stigatölu á grunnkvarða sem hefur sömu hundraðstölu (1. mynd). Aðferðin felst í ólínulegri vörpun stigatalna á grunnkvarða og byggir á klassíska próffræðilíkaninu (*classical test theory*). Til eru einfaldari aðferðir sem felast í línulegri aðlögun einnar dreifingar að annarri, til dæmis meðaltal eða meðaltal og staðalfrávik. Þessar aðferðir fela að jafnaði í sér síaukna skekkju þegar fjær dregur meðaltali, en mun minna úrtak þarf til að beyta þeim en hundraðstölu tengingu. Aðferðir sem byggja á svarferla-



1. mynd. Stigatölur á tveim útgáfum prófs tengdar gegnum hundraðstölur.

líkönum (*item response theory*) gera kleift að tengja ólíkar prófútgáfur með línulegri aðlögun á tölfraðilegum eiginleikum prófatriða. Slíkar aðferðir byggja á mjög stífum tölfraðilegum forsendum og er því ekki beitt hér. (Angoff, 1971; Kolen og Brennan, 1995; Peterson, Kolen og Hoover, 1989).

Aðferðir til að gera einkunnir jafngildar (*equating*) gera ráð fyrir ákveðnum forsendum um gagnasöfnun (Kolen og Brennan, 1995). Gengið var út frá forsendum um að tilviljun ráði hvaða útgáfu prófs hver nemandi þreytir. Það sem um eða yfir 96% nemenda taka samræmd próf í íslensku og stærðfræði á hverju ári og prófútgáfur byggja ávallt á sömu námsmarkmiðum má líta svo á að tilviljun ráði hvaða útgáfa prófanna er lögð fyrir á hverju ári. Þar af leiðandi er réttlætlanlegt að beita þessari aðferð.

## Niðurstöður

Skilgreindur var grunnkvarði með því að velja miðlungspungt samræmt próf í stærðfræði og íslensku. Einkunnirnar 5,0 og 6,0 (48-52 og 58-62 ef lítið er á stigatölur) á sex prófútgáfum í íslensku og stærðfræði voru gerðar jafngildar þessum útgáfum. Samræmd próf í

**2. tafla.** Jafngildar einkunnir samræmdra prófa í stærðfræði og íslensku árin 1995 til 2001.

	Stærðfræði		Íslenska	
	5,0 (48-52 stig)	6,0 (58-62 stig)	5,0 (48-52 stig)	6,0 (58-62 stig)
2001	48	59	49	58
2000	39	49	53	61
1999	48	58	48	58
1998	53	64	49	59
1997	40	49	47	59
1996	43	54	39	50
1995	50	59	34	45

Íslensku og stærðfræði voru skilgreind sem grunnútgáfur prófanna í þessari rannsókn. Voru þau valin út frá ákveðnum tölfræðilegum forsendum. Meðaltal landsmeðaltalna (samræmdra einkunna) í íslensku og stærðfræði á árunum 1993 til 2000 reyndist 6,21 í íslensku en 5,63 í stærðfræði. Meðaltöl beggja þessara samræmdu prófa árið 1999 liggja tiltölulega nærri þessu meðaltali allra meðaltala árána 1993 til 2000 (6,4 í íslensku og 5,7 í stærðfræði). Þau eru því „dæmigerð“ samræmd próf í vissum skilningi. 3. og 4. mynd sýna meðaltöl hvers árs fyrir sig og má þar sjá að prófin 1999 liggja um miðju dreifingar meðaltalna hvors prófs.

Fundin voru jafngildi neðri marka samræmdu einkunnanna 5,0 og 6,0 á sex útgáfum samræmdra prófa (1995 til 1998, 2000 og 2001) við útgáfu prófanna frá 1999 (sjá 2. töflu). Talsverður munur reyndist á þeim stigatölum á einstökum útgáfum sem teljast jafngildar, til dæmis er lægsta stigatala sem telst jafngild neðri mörkum 5,0 (4,8) í stærðfræði 39 stig en sú hæsta 53. Í íslensku eru sambærilegar tölur 34 stig og 53 stig. En einnig kemur fram að jafngildar einkunnir liggja oft nálægt grunnútgáfu sem stuðst er við hér, til að mynda munar einu til tveimur stigum í íslenskuprófunum frá 1997, 1998 og 2001 þegar 1999 útgáfan er notuð sem grunnútgáfa. Svipaðar niðurstöður má draga fram úr þeim niðurstöðum um einkunnina 6,0

sem sýndar eru í sömu töflu.

Þessar einkunnir má einnig skoða út frá öðru sjónarhorni. Líta má á einkunnirnar 5,0 og 6,0 sem ákveðin viðmiðunarmörk og finna hlutfall nemenda sem standast þau ef sérhver útgáfa prófsins er notuð óleiðrétt. Þetta hefur í reynd verið gert frá árinu 1985 (Sigurgrímur Skúlason, 2000). Þess má geta að sú aðferð sem beitt var við að finna jafngildar einkunnir í þessari rannsókn miðar að því að gera þessi hlutföll eins lík og kostur er. Slíkar niðurstöður eru sýndar í 3. töflu. Taflan sýnir hlutfall nemenda sem fær annarsvegar 5,0 og hærri einkunnir og hinsvegar 6,0 og hærri einkunnir. Sé einkunnin 5,0 í stærðfræði tekin sem dæmi, þá fengu á 1999 útgáfu prófsins 65,5% þessa einkunn eða hærri. Hæst hlutfall nemenda sem fékk 5,0 eða herra var árið 1998, 72,0%, en lægst árið 2000, 53,2%. Svipaða hluti má draga fram fyrir aðra dálka töflunnar. Víða liggja þessi hlutföll nokkuð nærri útgáfu prófsins frá 1999, til dæmis í íslensku árin 1997, 1998 og 2001. Þar munar frá 1% og upp í 4%. Þótt þessi munur virðist lítil í prósentum talið hefur hann þau áhrif í heilum árgangi að um 40 til 150 nemendur falla ofan eða neðan viðmiðunarmarkanna sem felast í þessum einkunnum ef ein útgáfa prófs er notuð fremur en önnur. Í öðrum tilvikum, þar sem skeikar tugum prósentu milli útgáfa af sama prófi falla fleiri hundruð nemenda ýmist ofan eða neðan viðmiðunarmarkanna eftir því

hvaða útgáfa prófs er notuð. Ef bornar eru saman niðurstöður íslenskuprófanna frá 1996 og 2000 má álykta að um 800 nemendur hefðu talist standast viðmiðunarkröfur samkvæmt öðru prófanna en ekki hinu.

## Umraeda

Markmið þessarar rannsóknar var að kanna tiltekinn próffræðilegan eiginleika samræmdra einkunna, eiginleika sem felst í því að tiltekin einkunn hafi sömu merkingu frá einni útgáfu prófs til annarrar. Niðurstöður sýna ótvírætt að óstöðugleiki hrjáir samræmdar einkunnir. Hvort tveggja er að niðurstöður fyrir einstakar útgáfur prófa skera sig mjög úr og einnig að oft víkja jafngildar einkunnir um fáar stigatölur frá grunnvarða. Hvorutveggja endurspeglast í nokkrum mun á hlutfalli nemenda sem standast - eða hefðu staðist - viðmiðunarkröfur á hverju ári. Það má ljóst vera að einkunnir á samræmdum prófum standast almennt ekki áður nefnt viðmið um hvort einkunnir á prófútgáfum eru jafngildar – að ekki skipti máli fyrir nemanda hvaða útgáfu prófs hann tekur.

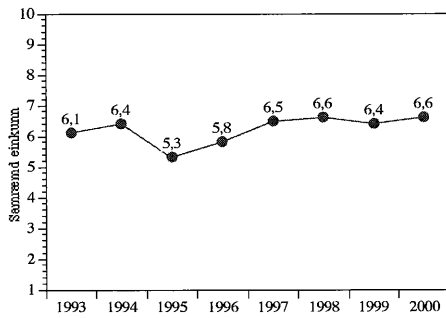
Til að átta sig á hve mikill munur er á ólíkum útgáum samræmdra prófa má benda á að um 40 einstaklingar standa að baki hverju einu prófútgáfu nemenda. Fyrir hver 2,5 prósent sem munar á tveimur prófútgáfum liggja því

um 100 einstaklingar ofan við viðmiðunarmörk í einni dreifingu en neðan við þau í annarri dreifingu. Það er réttlætismál frá sjónarhóli nemenda jafnt sem skólanna að einkunn sem nemandi fær fyrir sömu frammistöðu þýði það sama hvort sem hann tekur próf eitt árið eða það næsta. Það er óverjandi að einkunnastigi geti haft þau áhrif að ákveðin frammistaða leiði til þess að nemandi telst standast ákveðin lágmarksviðmið eitt árið en að sama frammistaða telst ófullnægjandi annað ár eins og þessar niðurstöður sýna að er reyndin með samræmdar einkunnir. Fyrir nemanda getur þetta þýtt að hann ætti kost á að hefja tiltekið nám ef hann þreytir prófin eitt árið en ekki ef hann þreytir próf annað ár. Fyrir framhaldsskóla þýðir þetta að sum ár eiga nemendur sem skortir nauðsynlegan grunnkost á að hefja nám á skilgreindum námsbrautum. Önnur ár gæti hið gagnstæða átt sér stað, að nemendur sem hefðu fullnægjandi undirbúning ættu ekki rétt á að hefja nám á þessum brautum. Þessi veikleiki takmarkar mjög réttmæti samræmdra einkunna og nauðsynlegt er að breyta fyrirkomulagi þessara einkunna.

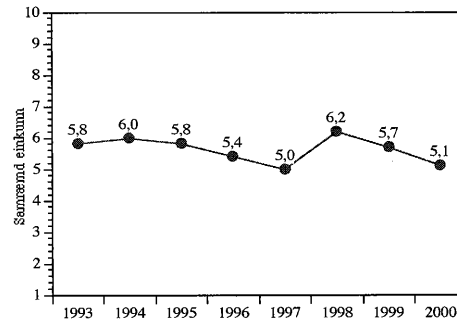
Margar aðferðir eru til sem gera kleift að skilgreina einkunnastiga sem lausir eru undan þeim vandkvæðum sem hér var fjallað um. Byggja má á fyrirfram skilgreindum dreifingum, til dæmis normaldreifingu, eða skilgreina tvo viðmiðunarpunkta sem ávallt

### 3. tafla. Hlutföll nemenda sem hefðu fengið tiltekna einkunn eða hærrí og frávik frá grunnútgáfu.

	Stærðfræði				Íslenska			
	%	5,0 Frávik frá 1999	%	6,0 Frávik frá 1999	%	5,0 Frávik frá 1999	%	6,0 Frávik frá 1999
2001	66,2	0,7	52,3	0,8	85,5	0,8	66,3	-0,5
2000	53,2	-12,3	39,5	-12,0	93,5	8,8	75,7	8,9
1999	65,5	-	51,5	-	84,7	-	66,8	-
1998	72,0	6,5	60,0	8,5	86,6	1,9	70,1	3,3
1997	68,1	2,6	52,8	1,3	85,2	0,5	69,6	2,8
1996	59,6	-5,9	45,8	-5,7	71,9	-12,8	52,1	-14,7
1995	68,1	2,6	52,8	1,3	60,1	-24,6	41,0	-25,8



2. mynd. Meðaltöl samræmdra einkunna í íslensku 1993-2000.



3. mynd. Meðaltöl samræmdra einkunna í stærðfræði 1993-2000.

fái sömu einkunn eins og vikið var að áður (Angoff, 1971). Hvoru tveggja felur í sér að horfst er í augu við það að ákveðinn hluti nemenda nær ekki þeim viðmiðunarmörkum sem sett eru. Eiginleikar viðmiðunarmarkna eru þá formlega ljósir og nemendur sitja við sama borð óháð því hvaða útgáfu prófanna þeir taka. Samræmdar einkunnir sem notaðar voru á árunum 1977 til 1984 hafa það fram yfir samræmdar einkunnir sem skilgreindar voru í reglugerð árið 1985 að einkunnir þýddu um það bil það sama frá ári til árs og að viðmiðunarmörk um fullnægjandi frammistöðu voru formlega ljós fyrirfram. Enn aðrar aðferðir gefa kost á að byggja upp réttmæti viðmiðunarmarkna og traust nemenda og skóla á þeim með því að tengja þau markmiðum aðalnámsskrár grunnskóla (Mentamálaráðuneytið, 1999). Með því að skilgreina færni viðmið (*achievement levels*) sem byggja á námsmarkmiðum má staðsetja lágmarksviðmið jafnt sem viðmið fyrir góða frammistöðu þannig að tengsl þeirra við færni og kunnáttu nemenda gefi kost á réttmætari notkun einkunna (Cizek, 2001; Jaeger, 1989).

Niðurstöður þess samanburðar sem fjallað hefur verið um hér eru afstæðar í þeim skilningi að tölugildi niðurstaðna – tölur í 1. og 2. töflu – verða breytilegar eftir því hvaða útgáfa samræmdra prófa er skilgreind sem grunnform. Niðurstöðurnar eru hinsvegar traustar að því leyti að þær sýna þann stöðug-

leika eða óstöðugleika sem er í samræmdum einkunnum. Þær sýna að nauðsynlegt verður að skilgreina aðra grunn-einkunnakvarða fyrir samræmd próf.

Sú aðferðafræði sem notuð var hér er nokkuð flókin og krefst mikils fjölda einstaklinga til að ná fram stöðugleika í efri og neðri hluta dreifingarinnar. Ýmsar niðurstöður mætti fá með einfaldari aðferðum. Styrkur þessarar aðferðar er að hún dregur fram hvaða tölfraðilegar forsendur liggja í raun til grundvallar samanburði af þessu tagi. Auk þess er hér á ferð sú aðferðafræði sem eðlilegt væri að nota við tengingu einkunna ef slíkt væri í raun gert á samræmdum prófum í 10. bekk eða á samræmdum prófum á framhaldsskólasigi. Endanlegt val á aðferð færi í reynd eftir því með hvaða hætti unnt yrði að safna gögnum til að tengja próf saman og í hvaða tilgangi einkunnir verða notaðar.

Vandamál af því tagi sem bent hefur verið á hér að ofan verða enn alvarlegri í umhverfi sem búast má við að skapist kringum væntanleg samræmd próf á framhaldsskólastigi en þau eru í grunnskólum. Margt er enn óljóst um fyrirkomulag þeirra prófa en líklegt að lögð verði fleiri en ein útgáfa samræmdra stúdentsprófa fyrir á hverju ári og að litið verði á einkunnir á báðum útgáfum prófsins sem jafngildar af aðilum sem taka þurfa ákvarðanir þar sem þær verða hluti af gögnum um einstaklinga. Til dæmis við val nemenda

inn á námsbrautir eða við ráðningu í störf. Einnig er nokkuð algengt að einstaklingar taki sér hvíld frá námi í eitt eða tvö ár, jafnvel fleiri. Þegar þessir einstaklingar koma aftur til náms verður gerð krafa um að litið verði á einkunnir þeirra sem jafngildar einkunnum nemenda er tóku aðrar útgáfur prófsins. Það er því réttlætismál fyrir nemendur jafnt sem þá sem taka þurfa ákvarðanir á grunni einkunna að þær verði jafngildar. Slíkt mun ekki gerast nema lagður verði grunnur að því strax á undirbúningstigi prófanna. Því er ljóst að mjög brýnt er að taka strax í upphafi upp traustar aðferðir til að tengja einkunnir á ólíkum útgáfum samræmdra prófa á framhaldsskóla stigi.

### Aftanmálgrein

Fyrri útgáfa flutt sem erindi á ráðstefnunni *Rannsóknir – Nýbreytni – Þróun: 5. málþing Rannsóknastofnunar Kennaraháskóla Íslands í Reykjavík, 12. október 2001.*

### Abstract

The psychometric properties of the Icelandic Nationwide Examination (INE) score scale in 10<sup>th</sup> grade were investigated using equating methods. The INE uses a new test form each year but the INE scores are not linked from one year to another. The investigation was limited to the scores of 5,0 and 6,0 because the scores are relevant for grade school graduation reporting and admission to college bound secondary school educational tracks. Data from seven test forms administered during the years 1995 to 2001 were used and the random groups equipercentile equating methods applied. Due to the fact that about 96% of each cohort take the tests each year and the tests are interchangeable, the random sampling assumption was judged to be tenable. The 1999 test forms were defined as basic score scales and the other test forms equated to this test forms. The results indicate

that raw scores currently scaled to the same INE score differ considerably for some of the test forms. When compared with the 1999 test form the difference in the proportions of students passing the critical grade points ranges from 0,7% to 25,8%. The results are discussed in relation to the validity of the INE scores.

### Heimildir

- Angoff, W. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ritstj.), *Educational measurement* (2. útg.). Washington, DC: American Council on Education.
- Anastasi, A. (1989). *Psychological testing* (5. útg.). New York: Mcmillan.
- Brennan, R. L. og Kolen, M. J. (1989). Scaling the ACT assessment and P-ACT+: Rational and goals. Í R. L. Brennan (Ritstj.), *Methodology used in scaling the ACT Assessment and P-ACT+*. Iowa City, IA: Americal College Testing Program.
- Cizek, C. J. (Ritstj.) (2001). *Setting performance standards: Concepts, methods and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Dorans, N. J. (2002). Recentering and rescaling the SAT score distributions: How and why. *Journal of Educational Measurement*, 39, 59-84.
- Feldt, L. S. & Brennan, R. (1989). Reliability. Í R. L. Linn (Ritstj.), *Educational measurement* (3. útg.). New York: American Council on Education /Mcmillan.
- Jaeger, R. (1989). Certification of student competence. In R. L. Linn (Ritstj.), *Educational measurement* (3. útg., pp 485-514). New York: American Council on Education /Mcmillan.
- Kolen, M., J. & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York: Springer.
- Kolen, M. J. og Hanson, B. A. (1989). Scaling the ACT Assessment. Í R. L. Brennan (Ritstj.), *Methodology used in scaling the ACT Assessment and P-ACT+*. Iowa City, IA: Americal College Testing Program.



- Menntamálaráðuneytið. (1979). *Skýrsla um niðurstöður samræmdra prófa í grunnskólum skólaárið 1978-1979*. Reykjavík: Höf.
- Menntamálaráðuneytið. (1999). *Aðalnámsskrá grunnskóla*. Reykjavík: Höf.
- Messick, S. J. (1989). Validity. Í R. L. Linn (Ritstj.), *Educational measurement* (3. útg.). New York: American Council on Education /McMillan.
- Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. Princeton, NJ: Educational Testing Service.
- Námsmatsstofnun (2000a). *Samræmt próf í Íslensku*. Reykjavík. Höf.
- Námsmatsstofnun (2000b). *Samræmt próf í stærðfræði*. Reykjavík. Höf.
- Peterson, N., Kolen, M., & Hoover, H. D. (1989). Norms, scales and equivalent scores. In R. L. Linn (Ritstj.), *Educational measurement* (3. útg.). New York: American Council on Education /McMillan.
- RUM (1995a). *Samræmt próf í Íslensku 1995*. Reykjavík.
- RUM (1995b). *Samræmt próf í stærðfræði 1995*. Reykjavík.
- RUM (1996a). *Samræmt próf í Íslensku 1996*. Reykjavík.
- RUM (1996b). *Samræmt próf í stærðfræði 1996*. Reykjavík.
- RUM (1997a). *Samræmt próf í Íslensku 1997*. Reykjavík.
- RUM (1997b). *Samræmt próf í stærðfræði 1997*. Reykjavík.
- RUM (1998a). *Samræmt próf í Íslensku 1998*. Reykjavík.
- RUM (1998b). *Samræmt próf í stærðfræði 1998*. Reykjavík.
- RUM (1999a). *Samræmt próf í Íslensku 1999*. Reykjavík.
- RUM (1999b). *Samræmt próf í stærðfræði 1999*. Reykjavík.
- RUM (2000a). *Samræmt próf í Íslensku 2000*. Reykjavík.
- RUM (2000b). *Samræmt próf í stærðfræði 2000*. Reykjavík.
- Reglugerð um námsmat í grunnskóla, 65/1985.*
- Reglugerð um samræmd próf í 10. bekk, 414/2000.*
- Reglugerð um námsmat í grunnskóla, 141/1980.*
- Reglugerð um námsmat í grunnskóla, 398/1977.*
- Sigurgrímur Skúlason (2000). *Einkunnir notaðar í tengslum við samræmd próf*. Reykjavík: Rannsóknastofnun uppeldis- og menntamála. (sjá einnig <http://www.namsmat.is/samrpr-umeinkunnir.html>)