



HÁSKÓLI ÍSLANDS

B.S. ritgerð
í sálfræði

Greining prófatriðabanka úr samræmdum prófum í stærðfræði

Greining á dýpt atriðabanka

Bergsveinn Andri Halldórsson og Davíð Kolka

Leiðbeinandi: Sigurgrímur Skúlason, aðjunkt
Júní 2022

SÁLFRÆÐIDEILD

**Greining prófatriðabanka úr samræmdum prófum
í stærðfræði**
Greining á dýpt atriðabanka

Bergsveinn Andri Halldórsson og Davíð Kolka

Lokaverkefni til B.S.-prófs í sálfræði
Leiðbeinandi: Sigurgrímur Skúlason, aðjunkt

Sálfræðideild
Heilbrigðisvísindasvið Háskóla Íslands
Júní 2022

Greining prófatriðabanka úr samræmdum prófum í stærðfræði: Er nægjanlegt magn af prófatriðum í prófatriðabankanum til að standa undir aðlögunarhæfu langtímaprófi?

Ritgerð þessi er 15 eininga lokaverkefni til B.S.-prófs við sálfræðideild á Heilbrigðisvísindasviði Háskóla Íslands

© 2022 Bergsveinn Andri Halldórsson og Davíð Kolka
Ritgerðina má ekki afrita nema með leyfi höfunda.

Ágrip

Menntun barna er mikilvæg fyrir þróun samfélags í heild sinni og þarf að sjá til þess að skólagangan sé að veita góða menntun. Menntamálastofnun (MMS) sér um skipulagningu á námsmáti barna og sér til þess að gæði menntunar sé sem mest. MMS hefur séð um fyrirlögn á samræmdu könnunarprófum á landsvísu til að kanna stöðu nemenda og hvort þeir nái ákveðnum hæfnisviðmiðum. Stöðugt er verið að reyna bæta námsmátið á Íslandi og nú stendur til að þróa aðlögunarhæft langtímapróf sem kæmi í stað samræmdra könnunarprófa. Markmiðið er að þróa lóðréttan kvarða þar sem allir nemendur væru metnir á sama kvarða. Til þess að þróa nýtt aðlögunarhæft próf þarf að vera til staðar góður prófatriðabanki með nægilegt magn af prófatriðum á mismunandi þyngdarbilum. Mikilvægt er að athuga hvort prófatriðin hafa nægilega mikla sundurgreiningu fyrir hvert þyngdarbil. Í þessari rannsókn er til skoðunar prófatriðabanki MMS með gögnum úr gömlum samræmdum prófum í stærðfræði sem á að nota í þróun á nýju prófi.

Niðurstöðurnar í greiningu á prófatriðabanka þar sem flokkaskipting varð á þyngd og sundurgreiningu urðu að það þyrftu að vera fleiri prófatriði á báðum mörkum þyngdarbilsins auk prófatriða með meiri sundurgreiningu. Það á sérstaklega við í 7. bekk þar sem hlutfallsmunur við 10. bekk í efstu tveim sundurgreiningaflokkum var mikill. Fyrir framtíðarrannsókn væri hægt að skipta prófatriðum í bankanum eftir hlutum innan samræmda prófs í stærðfræði og skoða flokkaskiptingu þeirra. Möguleiki er á notkun fjölþrepaþrófs sem virðist leysa vandamál sem geta komið með aðlögunarhæfu sniði.

Abstract

Education for children is important for the development of society and the school system must ensure that it is providing a good education. Menntamálastofnun (MMS) organizes the assessment of students in Iceland and ensures that the quality of education is as high as possible. MMS has overseen the execution of the Icelandic National Assessment Tests (i. Samræmdu könnunarpróf) which is an external assessment of how well the students score relative to their peers and whether they meet certain aptitude criteria. MMS is constantly trying to improve the assessment of students and is now developing a new adaptive test to replace the Icelandic National Assessment Tests in the future. There are plans to develop a vertical scale from which assessment of all students would be issued on the same scale. In order to develop a new adaptable test, it is necessary to have a good test bank with a sufficient amount of test items in various difficulty ranges. It is important to check whether the test items have sufficient discrimination for each difficulty range. In this study we examine a MMS test item bank that includes data from older versions of the Icelandic National Assessment Test of mathematics, which will be used in the development of a new test.

The results of the analysis, after a range classification of item difficulty and discrimination, showed that more test items were needed at both ends of the difficulty range as well as test items with a greater discrimination. This is specifically relevant in the 7th grade, where the proportional difference with the 10th grade in the top two categories of discrimination was large. For future research, it would be possible to divide test items in the item test bank according to parts within the Icelandic National Assessment Test of mathematics and examine their range classification as well. There is the possibility of using a multi-stage test format that seems to solve problems that may arise from an adaptable test format.

Efnisyfirlit

Ágrip	4
Abstract	5
Efnisyfirlit	6
Töfluskrá	7
Inngangur	9
1.1 Námsmat	10
1.1.1 Samræmd próf	11
1.2 Þróun námsmats og samræmdra prófa	12
1.3 Stígandamælingar	13
1.3.1 Lóðrétt skölun	13
1.4 Stöðluð próf	15
1.4.1 Áhættupróf (e. High-Stakes Tests) og próf í fylkjum BNA (e. State Testing Programs)	15
1.4.2 Common Core Assessment	16
1.4.3 Róttækar breytingar á námsmati	16
1.5 Sundurgreining og þyngd atriða (e. item)	17
1.6 Prófatríðabanki	18
1.6.1 Hönnun prófa	19
1.6.2 Snið prófa	20
1.7 Markmið rannsóknarinnar	21
Aðferð	22
Þátttakendur	22
Mælitæki	22
Frankvæmd	23
Úrvinnsla gagna	23
Niðurstöður	25
Umræða	37

Heimildskrá.....	40
------------------	----

Töfluskra

Tafla 1 Flokkaskipting fyrir þyngd.....	24
Tafla 2 Flokkaskipting fyrir sundurgreiningu.....	24
Tafla 3 Lýsandi tölfræði þyngd prófatriða á samræmdu prófum í stærðfræði hjá 7. og 10. Bekk	25
Tafla 4 Lýsandi tölfræði þyngd prófatriða í samræmdum prófum í stærðfræði, 7. og 10. bekkur aðskildir	25
Tafla 5 Lýsandi tölfræði sundurgreiningar prófatriða í samræmdum prófum í stærðfræði, 7. og 10. bekkur sameinaðir	26
Tafla 6 Lýsandi tölfræði sundurgreiningar prófatriða í samræmdum prófum í stærðfræði, 7. og 10. bekkur aðskildir	26
Tafla 7 Lýsandi tölfræði á þyngd prófatriða í samræmdum prófum í stærðfræði hjá 7. bekk.....	27
Tafla 8 Lýsandi tölfræði á sundurgreiningu prófatriða í samræmdum prófum í stærðfræði hjá 7. bekk.....	28
Tafla 9 Lýsandi tölfræði á þyngd prófatriða í samræmdum prófum í stærðfræði hjá 10. bekk	29
Tafla 10 Lýsandi tölfræði á sundurgreiningu prófatriða í samræmdum prófum í stærðfræði hjá 10. bekk.....	30
Tafla 11 Krosstafla með dreifingu prófatriða hjá 7. og 10. bekk.....	31
Tafla 12 Krosstafla með dreifingu prófatriða í prósentum hjá 7. og 10. bekk.....	32
Tafla 13 Krosstafla með dreifingu prófatriða hjá 7. bekk.....	33
Tafla 14 Krosstafla með dreifingu prófatriða í prósentum hjá 7. bekk.....	34
Tafla 15 Krosstafla með dreifingu prófatriða hjá 10. bekk.....	35
Tafla 16 Krosstafla með dreifingu prófatriða í prósentum hjá 10. bekk.....	36

Inngangur

Menntun er samfélögum um allan heim gríðarlega mikilvæg. Þau samfélög sem hafa hátt stig menntunar sýna fylgni og empírísk gögn um betri lífsgæði auk þess sem því fylgir meiri tæknivæðing, skilningur á samfélagsmálum sem og betri fjárhagstaða og heilsa. Menntun skiptir líka máli við að auka þekkingu og getu einstaklinga á ýmsum sviðum. Góð menntun eykur getu viðkomandi til að takast á við ýmis verkefni í lífinu en einnig eykur það atvinnumöguleika. Þekking í stærðfræði er eitt af mikilvægustu sviðum menntunar (Eurostat, 2021; Javed, Javed og Khan, 2016; Powdthavee, Lekfuangfu og Wooden, 2015; Winters, 2011).

Stærðfræði nýtist öllum, hvort sem það er tengt algebru, reiknifræði, rúmfræði eða tölfræði. Einstaklingar nota stærðfræði við hversdagslegar athafnir eins og að elda sér mat, í viðskiptum, hreyfingu og á margvíslegan beinan eða óbeinan hátt (Otten, 2011). Mögulega væri samfélagið týnt og vanþróað án stærðfræðinnar. Stærðfræðiþekking hefur þróast í gegnum tíðina þar sem aðferðir, formúlur og jöfnur verða flóknari og bæta við þekkingargrunn. Mikilvægt er að þessi þekking tynist ekki og því er miðlun á milli kynslóða nauðsynleg fyrir áframhaldandi bætingu á þekkingargrunn stærðfræðinnar og til að tryggja auðveldri aðlögun að lífinu fyrir þá sem eru yngri (Otten, 2011).

Á Íslandi er það Menntamálastofnunin (MMS) sem sér um skipulagningu á stærðfræðinámi barna frá leikskólaaldri fram að fullorðinsaldri. Á leikskólaaldri er börnum kynnt stærðfræðileg viðfangsefni eins og t.d. tölur, tákni og mynstur. Á grunnskólaaldri er fræðsluefni skipt eftir þremur árgangstigum sem eru yngsta stig (1-4. bekkur), miðstig (5-7. bekkur) og elsta stig (8-10. bekk). Að lokum er framhaldsskólastigið þar sem kennsla getur verið mismunandi eftir námsbraut hvers nemanda en kjarnafögin eru íslenska, enska og stærðfræði en á öllum námsbrautum þarf að huga að þessum þrem fögum sem einnig er prófað úr í samræmdum prófum í grunnskóla (Mennta- og menningarmálaráðuneytið 2012; Mennta- og menningarmálaráðuneytið, 2015; Mennta- og menningarmálaráðuneytið, 2021).

Á öllum þessum mismunandi stigum er markmiðið það að stuðla að mótun sex grunnþátta menntunar en þau eru læsi, sjálfbærni, heilbrigði og velferð, lýðræði og mannréttindi, jafnrétti og sköpun. Síðan eru búin til hæfniviðmið sem eiga við hvert stig árgangs sem nemandi ætti að uppfylla. Í stuttu máli má draga þá ályktun að þyngdarstig námsins verður erfiðara eftir því sem nemandinn verður eldri en til að meta getustig nemenda í náminu þarf að mæla getustig þeirra,

þar er notast við námsmatið sem er þó ekki notast við í leikskólum (Mennta- og menningarmálaráðuneytið 2011; Mennta- og menningarmálaráðuneytið, 2015; Mennta- og menningarmálaráðuneytið, 2021).

1.1 Námsmat

Ein af meginstoðum skólakerfisins er námsmatið. Námsmat gegnir margvíslegu hlutverki þ.á.m. að meta árangur og framfarir barna og ungmenna, veita upplýsingar fyrir leiðbeinendur um námið og hvernig markmiðum þess verði náð. Einnig að meta hvernig nemendum tekst að ná almennum hæfniviðmiðum aðalnámskrár, stuðla að námshvatningu, örva nemendur til framfara og meta hvaða aðstoð þeir þurfa. Að lokum eru skilyrði sem gæta skal að námsmatið uppfylli eins og að það á að vera áreiðanlegt, óhlutdrægt, heiðarlegt og sanngjarnt gagnvart nemendum (Mennta- og menningarmálaráðuneytið, 2021).

Það þýðir að meta þarf alla þætti námsins, þekkingu, leikni og hæfni, með hliðsjón af hæfniviðmiðum í aðalnámskrá. Þetta á að gera með fjölbreyttum aðferðum sem endurspeglar markmið skólastarfs. Einnig er mikilvægt að nemandi og allir aðilar í kring sem koma að, geti beitt námsmati til að geta nálgast markmið og upplýsingar um námsgengi sem geta leitt til bætinga við stundun náms til frambúðar (Mennta- og menningarmálaráðuneytið, 2021).

Námsmatið í sínu nýjasta formi á grunnskólastigi er á matskvarða sem er skipt upp eftir bókstöfum frá A, B+, B, C+, C og D til að leggja mat á getustig nemenda. Námsmatið metur lykilhæfni nemenda og hversu vel nemendum gengur að ná hæfniviðmiðum á námssviðinu. Í bókstafsmatskvarðanum lýsir (A) framúrskarandi hæfni, (B) góðri hæfni, (C) sæmilegri hæfni og (D) hæfni sem nær ekki viðmiðum sem lýst er í (C), einnig eru fjórir undirflokkar í bókstöfum (B) og (C) en í hæsta undirflokknum, sem er þá (B4) og (C4) í þeim tilvikum þá er nemandi ofarlega í sínu bókstafsfernisviði og fær plús (+) (Mennta- og menningarmálaráðuneytið, 2021; Menntamálastofnun, 2016)).

Matskvarðinn er notaður við útskrift nemenda úr grunnskóla og einnig í lok 4. og 7. bekkjar og í almennu námsmati. Annað dæmi um hvernig notkun þessa kvarða getur verið er þegar nemandi í 9. bekk sýnir A eða B getu við hæfniskröfu 10. bekkjar þá er samkvæmt aðalnámskrá grunnskóla möguleiki að biðja um útskrift árinu fyrir en 10 ára skyldunámi lýkur (Mennta- og menningarmálaráðuneytið, 2021).

Námsmat skóla er tvenns konar, innra og ytra mat. Þeir sem sjá um innra mat skóla eru starfsmenn skólans, nemendur og foreldrar. Þar gerir skólinn sjálfur faglega ígrundun og metur hversu vel tekst að ná þeim viðmiðum sem sett eru og gæði þjónustunnar sem er veitt. Aðrir sjá um ytra mat skólans þar sem lagt er mat á starfsemi skólans og hversu vel gengur að fylgja aðalnámskrá og hvort skólinn veiti þá þjónustu sem hann á að gera. Samræmd próf er ein aðferð ytra mats til að meta stöðu nemenda í skólum á landsvísu. MMS sér um ytra mat allra leikgrunn- og framhaldsskóla á Íslandi fyrir mennta- og menningarmálaráðuneytið. (Mennta- og menningarmálaráðuneytið 2011; Mennta- og menningarmálaráðuneytið, 2015; Mennta- og menningarmálaráðuneytið, 2021).

1.1.1 Samræmd próf

Í 4., 7. og 9. bekk grunnskóla eru samræmd könnunarpróf lögð fyrir alla nemendur og könnuð kunnátta í íslensku og stærðfræði. Nemendur í 9. bekk taka einnig próf í ensku í samræmdu prófunum en prófin eru lögð fyrir á sama tíma í öllum skólum landsins. Tilgangur samræmdu prófana samkvæmt 3. gr. í reglugerð um fyrirkomulag þeirra í grunnskólum er að

- „ a) Athuga, eftir því sem kostur er, að hvaða marki hæfniviðmiðum aðalnámskrár í viðkomandi námsgrein eða námsþáttum hafi verið náð,*
- b) Vera leiðbeinandi um áherslur í námi einstakra nemenda,*
- c) Veita nemendum, foreldrum, skólum og menntayfirvöldum upplýsingar um námsárangur og námsstöðu nemenda,*
- d) Veita upplýsingar um hvernig einstakir skólar og skólakerfið í heild stendur í þeim námsgreinum og námsþáttum sem prófað er úr. “ (Reglugerð um fyrirkomulag og framkvæmd samræmdra könnunarprófa í grunnskóla nr. 173/2017).*

Allt þetta nýtist við að betrubæta skólastarf eða skólakerfið sjálft (Menntamálastofnun, 2016; Mennta- og menningarmálaráðuneytið, 2020; Sigurgrímur Skúlason 2020).

Þar sem rannsóknin mun byggja á gögnum úr samræmdum prófi í stærðfræði ber að nefna hvernig uppbygging þessa prófs er. Það eru þrír til fjórir prófhlutar í samræmda prófinu í stærðfræði en eru mismunandi eftir aldurstigum. Prófhlutarnir geta verið fimm alls og nefnast

þeir Rúmfræði og mælingar og Reikningur og aðgerðir á öllum stigum. Tölur og talnaskilningur er á yngsta og miðstigi. Hlutföll og prósentur og Algebra eru á unglíngastigi (Sigurgrímur Skúlason, 2020).

1.2 Þróun námsmats og samræmdra prófa

Þegar aðalnámskrá grunnskóla kom út árið 2011 þá var ein af stærstu breytingum, einkunnagjöfin. Breyta átti frá því að vera með einkunnagjöf í tölustöfum yfir í bókstafi. Ástæðan fyrir því voru nýju hæfniviðmiðin á bakvið hvern einasta bókstaf og að nýi kvarðinn myndi draga úr mismun milli einkunnagjafa hjá skólum. Það tók nokkur ár að innleiða ferlið en í dag er í öllum grunnskólum landsins sá kvarði notaður í einkunnagjöf prófa og í lok skólaársins (Mennta- og menningarmálaráðuneytið, 2021; Menntamálastofnun, 2016).

Samræmt námsmat hófst árið 1907 með prófi á kunnáttu barna, en þau skiptust í fullnaðarpróf sem var eins konar lokapróf og landspróf í lestri og reikning. Samræmdu prófin eins og við þekkjum þau í dag þar sem prófin skiptast milli árgangsstiga hófust nú ekki fyrr en 1946 þá var 3., 6. og 8. bekkir látnir þreyta próf þar sem 8. bekkur þreytti fullnaðarpróf til að meta innritun í landsprófsdeild (Mennta- og menningarmálaráðuneytið, 2020).

Þessi próf voru svo lögð niður árið 1976 en þá voru samræmd próf tekin upp í staðin sem urðu inntökupróf. Frammistaða nemenda í samræmdu prófunum hefðu þá áhrif í hvaða skóla og á hvaða braut nemendur gætu farið í. Framhaldsskólinn var nú opinn öllum en prófin höfðu áhrif á hvaða skóla nemendur kæmust inn í (Mennta- og menningarmálaráðuneytið, 2020).

Fjöldi prófa var breytilegur milli ára en þegar þau voru einungis tvö var prófað í íslensku og stærðfræði. Prófað var einnig í ensku og dönsku þegar prófin voru fjögur og einnig í náttúrugreinum og samfélagsgreinum þegar prófin voru orðin sex. Árið 2009 var samræmdu prófunum breytt í þrjú próf og hlutverk prófana nú stöðupróf en ekki staðsetningarpróf. Nú höfðu prófin ekki lengur áhrif á námsferil nemenda og var einungis til upplýsingar um stöðu nemenda í fræðigreinunum. (Mennta- og menningarmálaráðuneytið, 2020; Menntamálastofnun, 2016).

Að lokum má nefna að nýlegasta tilbreytingin var vorið 2017. Fram að þeim tíma höfðu nemendur í 10. bekk þreytt samræmd próf en þá varð sú ákvörðun tekin að láta bæði 9. og

10.bekk taka samræmd próf það ár en eftir næsta haust yrði það einungis 9.bekkur látinn taka samræmt próf (Mennta- og menningarmálaráðuneytið, 2021).

1.3 Stígandamælingar

Hægt er að mæla stíganda milli bekkja með því að leggja fyrir mismunandi útgáfur af sama prófinu á hverju ári og bera svo saman niðurstöðurnar úr prófunum milli ára í nokkur ár. Námsefnið á milli bekkja er mjög breytilegt og ekki er hægt að hafa eitt próf sem inniheldur allt efnið og láta alla nemendur taka sama prófið. Það væri of mikið af prófatriðum sem væri of erfið eða of létt fyrir suma nemendur og væri ekki að mæla raunverulega getu nemenda. Einnig væri þannig próf of tímafrekt og ekki hentug aðferð (Kolen og Brennan, 2014).

Lausnin við þessum vanda er að búa til próf sem notar mismunandi þyngdarstig til að mæla raunverulegu getu nemenda. Hvert þyngdarstig er hannað þannig að það innihaldi það námsefni sem er viðeigandi fyrir þann aldursflokk. Til að mæla loks stígandann er fengin niðurstaðan úr hverju prófi fyrir alla aldursflokka og sett í einfaldan staðlaðan kvarða (e. Single score scale) sem er notuð í lóðréttri skölun. Það veitir upplýsingar um stígandann milli ára. (Kolen og Brennan, 2014).

1.3.1 Lóðréttt skölun

Í lóðréttri skölun eru próf sem mæla sama fyrirbærið (e. construct) en eru mismunandi í þyngd sett á sama skalann. Notast er við staðlaðan kvarða á landsvísu til samanburðar á einstaklingum sem taka mismunandi gerðir af prófum og athugað hvort nemandi er undir eða yfir meðaltali. Hægt er að mæla bætingarnar í prófunum með því að leggja fyrir mismunandi útgáfur af sama prófinu í hverjum bekk. Ekki er hægt að hafa sama prófið fyrir alla bekki þar sem það gæti verið of erfitt fyrir nemendur á yngri stigum og of auðvelt fyrir nemendur á eldri stigum. Nemendur á yngri stigum myndu þá gefast upp og eldri nemendur myndu missa áhugann (Kolen og Brennan, 2014).

Til þess að prófin kanni raunverulegt getustig nemenda á hverjum aldri þurfa þau að vera hæfilega erfið og hafa mismunandi þyngd atriða. Síðan er hægt að sjá bætingar hjá nemendum með því að bera saman niðurstöðurnar úr prófunum í mismunandi þyngd. Ekki er hægt að nota

jöfnun til að framkvæma lóðréttan mælikvarða eða að bera saman það sem 4.bekkur og 9.bekkur fær því innihald þeirra prófa er mismunandi (Kolen og Brennan, 2014; Penfield, 2014).

Nemendur á hverju stigi þreyta próf sem eru hönnuð fyrir þeirra getustig og hversu vel þau kunna það efni sem er viðeigandi fyrir þann aldur. Með hverju ári verða prófin erfiðari og efnið verður flóknara. Spurningar á prófunum skarast oft milli ára þar sem oft er verið að kenna sama efnið í bekkjum sem eru nálægt hvort öðru í aldurstigi (Kolen og Brennan, 2014; Mennta- og menningarmálaráðuneytið, 2021).

Ef skoða á hvernig lóðrétt skölun er notuð þá er hægt að skoða ITBS (Iowa Test of Basic Skills) prófið þar sem eru sjö flokkar af mismunandi prófapáttum, merkt *a* til *g*. Í 3. bekk þreyta hóparnir *a* og *b* flokkinn, *b* og *c* eru teknir fyrir í 4. bekk svo *c* og *d* í 5. bekk og svo koll af kolli. Hver bekkur tekur þá tvo flokka, einn flokk sem inniheldur auðvelt efni og svo einn flokk sem væri flókið efni fyrir nemendur (Kolen og Brennan, 2014).

Nemendur ná meiri framförum í því efni sem er kennt yfir önnina. Ef kennt er samlagning í 3. bekk er búist við framförum í því efni í lok árs. Þegar nemendur fara í 4. bekk og læra deilingu er meiri framför í því efni heldur en framför í samlagningu. Hægt væri að láta alla nemendur í öllum bekkjum taka sama prófið með alla flokka og öll erfiðleikastig til að kanna getu nemenda en það er ekki hagkvæmt. Prófið tæki of langan tíma og nemendur sem eru á yngri stigi myndu fljótlega gefast upp og nemendur á eldra stigi myndu missa áhugann (Kolen og Brennan, 2014).

Hægt er að nota þrjár mismunandi aðferðir við lóðrétt skölun. Almennt atriðasnið (e. Common item design) er aðferð þar sem bekkir þreyta próf á því erfiðleikastig sem er viðeigandi. Þættirnir í prófunum skarast svo á milli bekkja þar sem prófin tengjast frá ári til árs. Þetta snið hefur þann kost að vera auðveldast í notkun. Snið samanburðarhópa (e. Equivalent groups design) er önnur aðferð en þá ræður tilviljun því hvort nemendur taka próf fyrir þeirra stig eða í stigum fyrir neðan eða ofan. Nemendur í 4. bekk taka annað hvort 4. bekkjar próf eða 5. bekkjar próf (Kolen og Brennan, 2014).

Að lokum er skölunarprófssnið (e. Scaling test design) en það er próf sem nær yfir efni hjá öllum aldurshópum en er sérhannað til að vera það stutt að hægt sé að taka það í einni lotu. Ef það er verið að kanna stíganda frá 3. til 8. bekk nær prófið yfir allt það efni í öllum aldurstigum og fá allir bekkirnir sama prófið (Kolen og Brennan, 2014).

1.4 Stöðluð próf

Til þess að próf teljist vera staðlað próf þurfa allir nemendur að svara sömu spurningunum á sama tíma og endurgjöfin þarf að vera gerð á sama hátt. Uppbygging staðlaðra prófa getur verið mismunandi, þau geta verið uppbyggð með t.d. sönnun-ósönnunum fullyrðingum, opnum spurningum, fjölvalsspurningum eða af öðrum mismunandi hönnunum spurninga. Fjölvalsspurningar og satt/ósatt fullyrðingar eru algengustu form staðlaðra prófa þar sem það er auðvelt að leggja það fyrir stóran hóp og yfirferð svara því auðveldari. Stöðluð próf eins og t.d. samræmd próf eru helsta aðferð til að framkvæma ytra mat á nemendum (Furr, 2018; Mennta- og menningarmálaráðuneytið, 2021)

1.4.1 Áhættupróf (e. High-Stakes Tests) og próf í fylkjum BNA (e. State Testing Programs)

Áhættupróf eru próf sem hafa þann eiginleika að það sé tekin mikilvæg ákvörðun út frá niðurstöðu prófa og geta haft í kjölfarið afleiðingar t.d. að komast inn í eftirsóttarverðan skóla ef nemandi fær háar einkunnir. Þetta samræmist því hvernig samræmdu prófin virkuðu langt fram á síðustu öld á Íslandi þegar þau voru notuð sem inntökupróf í framhaldsskóla (Moon o.fl., 2007; Mennta- og menningarmálaráðuneytið, 2020).

Það var talsvert seinna sem þetta skipulag kom fram í Bandaríkjunum þar sem fyrir 1980 voru stöðluð próf einungis notuð til sýningar fyrir foreldra og kennara sem leiddi til vandamáls varðandi áhugaleysi fyrir prófinu sjálfu. Áhættupróf var svar við því vandamáli. Innleiðing af því voru próf í fylkjum Bandaríkjanna. Þau fara fram á landsvísu sem bera saman frammistöðu barna innan fylkja, svipað eins og landspróf á Íslandi áður fyrr. Athyglisvert er að skoða hvernig það virðist sem fyrirkomulagið á Íslandi og í Bandaríkjunum hafi þróast í andstæðar áttir á þessu sviði (Moon o.fl., 2007; Mennta- og menningarmálaráðuneytið, 2020).

Þetta var þó ekki fullkomin nálgun en það kom seinna í ljós sá galli að þegar byrjað var að bera saman nemendur á landsvísu við þá á heimsvísu að nemendur voru ekki að standa sig vel á heimsvísu mælikvarða. Mögulegar orsakir fyrir því voru ósamræmi milli yfirvalda og menntastofnuna og starfsmanna þeirra, erfiðleikar við að finna jafnvægi milli nýrra staðla og eldri námskrá sem leiddi til mismunandi innleiðingu staðla hjá hverju fylki. Að lokum var einnig sú mögulega útskýring að kennarar þurftu að túlka of mikið í óljósa staðla, eitthvað þurfti til þess að bæta námsmatið (Kendall, 2011).

1.4.2 Common Core Assessment

Árið 2009 var sett fram ný áætlun í Bandaríkjunum sem nefnist Common Core State Standards Initiative (CCSSI) sem átti að vera endurbæting frá námsmati byggt á stöðlum. Markmið þess var að brúa bilið á milli nemenda á mismunandi landsvæðum og stéttum og auka námsgetu þeirra með því að miða við viðmið á heimsvísu frekar en viðmið á landsvísu. Með þessari afmörkun átti að upplýsa alla í kringum nemandann, sérstaklega kennurum á hvað þyrfti til að tryggja hvernig standa ætti að og getu til að eiga bjarta framtíð (Kendall, 2011).

Áður fyrr var námsmat í Bandaríkjunum einungis kannað á fylkisvísu þar sem stöðluð próf voru notuð til mælingar á nemendum í sérhverju fylki, enda var verið að fylgja áætlun fylkis prófa. Stöðluð próf í fylkjum gáfu þá niðurstöðu að nemendur væru að standa sig vel. Þegar hins vegar geta nemenda var skoðuð miðað við aðra nemendur í öðrum þjóðum þá stóðu nemendurnir í Bandaríkjunum sig mun verr og því var ákveðið að bregðast við því með Common Core Assessment. (Moon o. fl. 2007; Kendall, 2011; Conley, 2014).

1.4.3 Róttækar breytingar á námsmati

Sumstaðar eru miklar breytingar í gangi og framundan hvað varðar kennslu um allan heim en orsakirnar eru taldar vera heimsvæðing og aukin rafræn tækni sem hefur aukið upplýsingaflæði. Það hefur einnig komið í ljós að núverandi uppsetning skólakerfisins er ekki að virka eins vel lengur og hefur stöðug aukin frammistaða nemenda staðnað (Hill og Barber, 2014).

Breytingar eru nú þegar hafnar í skólum en í mismiklu magni og eru sex þættir sem þessar umbreytingar takast á við innan skólakerfa. Þau eru námsgeta, kennsluáætlun, menntunarstefna, námstækifæri, kennsluhættir og kennslugæði. Breytingarnar hafa í huga námsframvindu einstaklingsins frekar en að alhæfa um nemendur, hafa meira aðgengi, sveigjanlegra nám og sterkara sambands milli skólaaðila (Hill og Barber, 2014).

Á Íslandi var nýlega komið á starfshóp frá mennta- og menningarmálaráðherra sem hafði það hlutverk að vinna að tillögu að framtíðarstefnu um markmið, hlutverk, framkvæmd og fyrirkomulag samræmdra könnunarprófa. Sá hópur skoðaði og lagði til róttækar breytingar sem virðast fá innblástur úr anda Renaissance Assessment en vilji var til staðar að leggja niður samræmd próf og taka upp matsferil í staðinn. Í matsferlinum þá væri lögð meiri áherslu á stutt og hnitmiðuð rafræn próf þar sem valkvætt væri fyrir nemendur að taka þau á sveigjanlegum

tíma og jafnvel oftast en einu sinni ásamt því að aðgengi væri aukið fyrir kennara svo lengi sem prófið væri ekki áhættuþróf (Hill og Barber, 2014; Mennta- og menningarmálaráðuneyti, 2020).

1.5 Sundurgreining og þyngd atriða (e. item)

Tvær kenningar sem eru vinsælustu viðmið próffræðingar eru klassíska raungildislíkani (e. Classical Test Theory) og svarferlallíkani (e. Item Response Theory). Í klassísku raungildislíkani er áreiðanleiki skoðaður með tengslum milli mældra stiga (e. observed scores), raungilda (e. true cores) og mælivillu (e. measurement error). Ef mæld stig væru ekki með villu þá myndi fást raungildi. Ekki er búist við að endurteknar mælingar með sama prófi veiti sömu niðurstöðu en frekar að tiltekin líkindi séu fyrir að niðurstaða lendi á ákveðnu bili sem nefnist staðalvilla mælinga (e. standard error of measurement). Líkanið er með fimm forsendur sem ekki verður fjallað um því snið samræmda prófsins sem rannsóknin afmarkast við tengist meira svarferlallíkani (Furr, 2018).

Svarferlallíkani er annar samtíðarkostur til móts við klassíska raungildislíkanið en í því líkani er einnig reynt að finna raungildi út frá mældum gildum en aðalmunurinn liggur í því að það er ekki úrtaksháð. Svarferlallíkani er mikið notað við gerð rafræna, samræmdra og staðlaðra prófa og forsenda hennar er sú að hægt sé að halda prófun áfram út frá svári við spurningu hjá einstaklingi. Sveigjanleikinn í líkaninu býður upp á að í atriðagreiningu er hægt að nota líkanið til að mæla eiginleika og gæði atriða, þróa mót og meta eiginleika skora sem koma frá matinu með mótun á svarferli (e. Item Response Function) sem nýtist í hönnun og yfirferð á atriðabanka. Í svarferli skiptast atriðin í a og b -breytu, a metur sundurgreiningu og b metur þyngd atriðis, metið er frá 0 sem táknar ákveðið stig af eiginleika sem er verið að mæla, niðurstöður fyrir a -breytu liggja oftast frá 0 til 2 en í b -breytu er algengt að mæla frá -3 til 3, en það eru undantekningar. Mælingar eru einfaldari fyrir atriði með tvískipt (e. dichotomous) frekar en fjölskipt (e. polytomous) skor. (Furr, 2018; Kolen, Brennan, 2014; Penfield, 2013)

Sundurgreining gefur í skyn hversu mikið atriðið segir til um eiginleika sem matstækið er að mæla, vanalega er leitast eftir a.m.k. 0,8 í þeim stuðli sem segir að atriðið hefur nægilega sundurgreiningu, < 1,5 væri mjög mikil sundurgreining. Þyngd atriðis gefur upplýsingar um hvaða stig á eiginleikanum er atriðið gagnlegt til að mæla fyrir. Ef t.d. væri verið að mæla hóp með lága getu í eiginleika þá væri ekki mikið gagn af því að hafa einungis atriði sem eru mjög

þung. Meginkosturinn við svarferlálíkanið í samanburði við klassíska raungildislíkanið er að það fást ríkari upplýsingar um hvað atriðið mælir og greinir nákvæmlega um, því það mælir hvern punkt á kvarðanum en varast skal forsendu um staðbundið sjálfstæði (e. Assumption of local independence). Það þarf stór úrtök í svarferlálíkani en 500 er almennt talið lágmark en því fleiri í úrtakinu því betra (Kingston, Scheuring og Kramer, 2013; Penfield, 2013).

Fleiri kostir sem svarferlálíkan hefur umfram klassíska raungildislíkanið er að það nær yfir margs konar tölfræðileg líkön sem sýna fram á tengsl milli svörunar við atriði, eiginleika próftakanda og fjölda einkenna atriða. Svarferlálíkan gagnast mikið við gerð, túlkun og endurbætingar á prófum. Sérstaklega er þróun tækninnar grundvöllur að auknum gæði prófa úr svarferlálíkani. Þó virðist sem að munur milli klassíska raungildislíkans og svarferlálíkans sé óljós ef litið er á prófum með færri atriðum en svarferlálíkanið nýtist betur í prófum með a.m.k. 20 atriði. (Furr, 2018; Jabrayilov o.fl., 2016)

Samræmd könnunarpróf í stærðfræði samanstendur af fjölvalsspurningum (e. Selected Response Questions) með oftast fjóra svarliði. Þessi tegund spurningar er mest að finna á prófinu, eða spurningum með opnum svarlið (e. Constructed Response Questions). Spurningar með opnum svarlið eru erfiðari í úrvinnslu en uppsetning þeirra í samræmda prófinu er fremur einföld og dregur þar með úr skekkju sem gæti verið vegna mismunandi matsgjafar yfirferðaraðila. Sjálf hönnun samræmda prófa byggir á inntakstöflu (e. table of specifications) sem segir til um skiptingu prófhluta. Öll atriði eru uppsett til að orðalag sé hvorki truflandi né að mismuna á hópum og að öll atriði séu réttmæt. Í samræmdum prófum eru þeir hlutar prófsins sem hafa fá atriði ákveðin vandamál með áreiðanleika sem leiðir til þess að ekki sé hægt að túlka út frá þeim um stöðu nemenda í náminu (Rodriguez og Haladyna, 2013; Sigurgrímur Skúlason, 2020;).

1.6 Prófatriðabanki

Prófatriðabankar eru notaðir til að auðvelda vinnu og til þess að það þurfi ekki að hanna ný prófatriði í hvert skipti sem verið er að leggja fram próf. Ef passað er upp á gæði og að auðvelt sé að ritstýra og uppfæra atriði þá verður mikið notagildi af bankanum. Notast er mikið við tölvur eða algrími til að koma á sjálfvirkni í ferlinu í þeim hlutum sem krefjast miklar endurtekningar. Að auki eru þau nýtt í tölfræðilegar athuganir á prófatriðum en það þarf samt mannleg vinnubrögð einnig í það ásamt í hönnun prófatriða. Allir sem undirbúa mat reglulega

hafa not af prófatriðabanka. Prófatriðabanki hefur þrjú mikilvæg hlutverk; innritun atriða og geymsla, endurheimt fyrir skoðun atriða, sniðmáta próf og ritstýringu eða uppfærslu atriða, og að lokum viðhald atriða. Það er hins vegar mismunandi nothæfi fyrir alla. Kennari í bekk þarf ekki aragrúa atriða í samanburð við einhvern sem er að halda stöðluð próf á landsvísu (Parshall o.fl., 2002; Ward og Murray-Ward, 1994).

Nokkrir af kostum við notkun prófatriðabanka í stöðluðu prófi á landsvísu er að ekki þarf að skrifa upp ný atriði í hvert skipti, það tryggir betri gæði en próf sem eru hönnuð út af nauðsyn og að lokum auðveld regluleg uppfæring atriða. Í prófatriðabanka þarf að varast að próftakar geti ekki lagt á minnið prófatriði vegna tíðrar birtingar, að hafa mismunandi útgáfur prófs við fyrirlagningu, síðari próf séu jafngilt og viðhalda gæði mats (Ward og Murray-Ward, 1994).

Ef prófatriðabanki er hannaður fyrir aðlögunarhæft snið prófa þá er reynt að forðast þess að sum atriði verði ekki nýtt, hægt er að búa til algrím sem tryggir að atriði koma nægilega oft upp en hægt er að hanna eða uppfæra prófatriðabanka til að þess sé ekki nauðsyn. Þrjú skref eru í því ferli. Fyrst þá eru inntök efnis fyrir aðlögunarhæfa prófið greind og allir eiginleikar inntaksins sem koma fram í þeim, auðkenndir. Niðurstaðan leiðir til skilgreiningar á hvaða efni er verið að vinna með fyrir hönnun prófa. Í öðru skrefi er Monte Carlo aðferð notuð til að líkja eftir fyrirlagningu prófsins. Í lokaskrefinu er ákjósanlegasta sniðið fengið af fjölda skipta sem atriði var tekin úr hverri skiptingu í hermunum, gott er að endurtaka þetta ferli til að hanna hágæða prófatriðabanka þar sem þekkingin á efninu verður meiri í hvert skipti sem ferlið er framkvæmt (van der Linden og Glas, 2010)

1.6.1. Hönnun prófa

Rafræn próf eru sífellt að verða algengari þar sem þau bjóða fram á auðveldari úrvinnslu og meiri möguleika í hönnun spurninga. Mikilvægt er að fara ekki of langt með notkun nýrra möguleika sem rafræn próf bjóða upp einungis vegna þess að það er mögulegt, það verður að mæla hvort tölfræðin styðst við notkun. Mat atriða í rafrænu prófi fer fram í þriggja þrepa ferli þar sem atriði eru metin af sérfræðingum um efnið, skoðað er tölfræðilega eiginleika og að lokum eru framkvæmdar hermanir til að athuga hvort atriði henta í ákveðin snið prófs. Eftir það er viðhald líkt og í prófatriðabanka mikilvægt. Þegar er verið er að endurskoða atriði þá er athugað hvort það tengist því sem snið prófs leitast eftir, hvort það sé úrelt, hvort þau séu rétt

kóðuð við núverandi snið og hvort að atriðið sé ekki að veita samhengisáhrif við annað atriði (Parshall o.fl., 2002; van der Linden og Glas, 2010).

Ef val á svarferlalíkan styðst í gegnum gögn og sögulegt samhengi þá er hægt að nýta það í tölfraðilega vinnslu á atriðum. Úr því er gert stutt viðmiðunarpróf sem notast er við hermanir. Hermanir geta gefið upplýsingar sem segja hvort ákveðið snið prófs sé raunhæft því hönnuðir prófa vilja alltaf hafa próf styttri, nákvæmari og öruggari (Parshall o.fl., 2002).

1.6.2. Snið prófa

Hið hefðbundna línulega próf hefur við högg að sækja frá öðrum valkostum í hönnun prófs á borð við aðlögunarhæft og fjölþrepa próf. Línulegt próf mælir einstaklinga á meðalstigi eiginleika vel en er hins vegar ekki nákvæmt fyrir einstaklinga á sitt hvorum mörkunum. Munurinn á línulegu og aðlögunarhæfu sniði er sú að það kemur aðlögun eftir hvert atriði í aðlögunarhæfu sniði, ólíkt við línulega sniðið þar sem allir svara sömu spurningum og enginn aðlögun verður. Aðlögunarhæft próf einblínir betur á mælingar um hæfni einstaklinga á öllum stigum eiginleikans, sérstaklega þá sem eru í efri mörkum auk þess að nota færri prófatriði en línulegt próf (Hendrickson, 2007).

Aðlögunarhæf próf geta haft eftirfarandi vandamál. Hætta er á að prófið sé að breyta forsendu staðbundins sjálfstæðis með því mæla annað en það sem prófinu er einungis ætlað að mæla sem getur leitt til uppblásins áreiðanleika. Það geta komið upp atriði sem svarar öðru óvart, til að forðast þess verður að setja upp viðeigandi algrím. Þau geta einnig verið fyrir ójafnri skiptingu á efni. Ef ekki er horft upp á hvenær atriði koma upp í prófinu eða hvaða tímasetningu próftakar sitja undir það þá þarf mikinn fjölda atriða og aðferðir til að meðhöndla uppljóstrun atriða. Þau bjóða ekki upp á yfirferð svara fyrir skil að hálfu próftaka. Ef próf leyfir yfirferð þá geta próftakar auðveldað sér prófið með því að svara atriðum ranglega á falskan hátt og í kjölfar þess gert sér próf með minni þyngd, svarað auðveldari spurningum rétt í yfirferðinni sem leiðir til uppblásnar mælingar á eiginleikanum. Að lokum þá krefjast þau einnig miklar gagnastýringar og tölvuafls (Hendrickson, 2007; van der Linden og Glas, 2010).

Vegna þessa vandamála hefur verið mótað nýtt snið prófs sem nefnist fjölþrepa próf. Í fjölþrepa prófi er innleitt kosti bæði línulegs og aðlögunarhæfu sniði prófa með því að nota hóp atriða sem eru lögð fram eftir aðlögun og meta þau atriði saman. Fjölþrepa próf byrjar á hóp af atriðum með víðu bili þyngdar til að koma með upphaflegt mat á getu einstaklings. Síðan kemur

seinni hópur atriða eftir aðlögun sem á að aðgreina einstaklinga á svipuðu getustigi og síðan fer eftir fjölda þrepa sem óskað er eftir hvort að prófi er lokið (Hendrickson, 2007).

Í samanburði við línuleg próf þá eru þau betri að mæla hóp með stóru bili á eiginleika, er hraðara í úrvinnslu og hefur verið sýnt fram á að veita jafna eða hærri réttmæti fyrir ályktanir um spá og samleitni miðað við línuleg próf. Í samanburði við aðlögunarhæft próf þá hafa prófhönnuðir meiri stjórn yfir gæðum uppbyggingar og framkvæmd prófs. Það gefur kost á að leysa þau vandamál sem nefnt voru við aðlögunarhæft próf með því að t.d. gefa áreiðanlegri skor, lækka líkur á samhengisáhrifum, gefa kost upp á að próftakar fái svipað efni og leyfa skoðun á öðrum ótölfræðilegum eiginleikum, koma í veg fyrir uppljóstrun atriða með því að fækka fjölda aðlögunarforma prófsins, býður upp á takmarkaða endurskoðun og að lokum minni þörf fyrir gagnastýringu og tölvuafli sem leiðir til nákvæmari framkvæmd prófs (Hendrickson, 2007).

1.7 Markmið rannsóknarinnar

Þegar verið er að hanna stöðupróf í stærðfræði þá er byrjað á því að setja upp prófatriðabanka með atriðum sem notuð hafa verið í samræmdum könnunarprófum. MMS er að vinna að gerð prófatriðabanka fyrir nýtt matstæki í stærðfræði með þeirri nálgun. Með þeim prófatriðabanka er ætlað að koma langtíma- eða lóðréttum kvarða sem á að nýtast í að geta birt einkunnir einum sameiginlegum kvarða fyrir alla bekki. Notast var við svarferlallíkön til að meta próffræðilega eiginleika atriðanna en fylgnin fundin þar var mjög sterk. Ásamt því var beitt samtímalausn (e. concurrent calibration) til að koma öllum atriðum á sama kvarða. Til frambúðar er ætlað að skoða stíganda milli hvers bekks frá 4. til 10. bekks. Áður en að því verður komið er tilvalið að skoða prófatriðin sem liggja að baki undirbúnings á nýju matstæki (Sigurgrímur Skúlason, 2022; Unnar Geirdal o.fl., 2021).

Prófatriðabanki frá MMS með gögnum frá samræmdum prófum í stærðfræði á árunum 2000-2013, auk 2019 (einungis með prófatriði 4. og 7.bekk) og 2020 verður til skoðunar. Markmiðið er að greina nægjanlegt magn af prófatriðum í prófatriðabankanum til að standa undir aðlögunarhæfu prófi til langtíma. Mikilvægt er að huga sérstaklega að því hvort vanti atriði með nægjanlega sundurgreiningu fyrir einhver þyngdarstig. Þetta verður skoðað með flokkaskiptingu. Rannsóknarspurning okkar einblínir á því hvort nægilegt magn prófatriða séu innan hvers stigs þyngdar og sundurgreiningar til að nýtast í aðlögunarhæft próf til frambúðar?

Aðferð

Þátttakendur

Þátttakendur voru nemendur í 7., 9. og 10. bekk sem tóku samræmd könnunarpróf í stærðfræði á árunum 2000 til 2013, ásamt 2019 og 2020. Gagnasafnið veitti engar upplýsingar um fjölda þátttakenda eða persónuauðkenni þeirra.

Mælitæki

Mælitæki rannsóknarinnar voru samræmd próf í stærðfræði frá árunum 2000 til 2013, 2019 og 2020 sem eru lögð fyrir alla nemendur í 4., 7. og 10. bekk. og öll gögnin frá prófunum fengin frá MMS. Um 3700 til 4500 nemendur taka prófin í hverjum árgangi í 140 til 190 grunnskólum (Inga Úlfisdóttir o.fl., 2015).

Það eru þrjú til fjórir prófhlutar í samræmdu könnunarprófi í stærðfræði. Prófhlutarnir eru Reikningur og aðgerðir, Rúmfræði og mælingar, Tölur og talnaskilningur, Hlutföll og prósentur og að lokum Algebra. Reikningur og aðgerðir og Rúmfræði og mælingar eru á öllum aldursstigum. Tölur og talnaskilningur eru á yngsta og miðstigi en Hlutföll og prósentur og Algebra eru einungis á unglingsstigi (Sigurgrímur Skúlason, 2020).

Öll prófatriði í prófhluta Reikningur og aðgerðir notast við grunnreikniáðferðir og unnið er með tölulegar upplýsingar og er innihald prófatriða mismunandi eftir aldursstigi. Rúmfræði og mælingar athugar hæfnistig nemenda í öllu sem tengist umfangi og stærð fyrirbæra í umhverfinu, og notkun mælieininga á þessi fyrirbæri. Tölur og talnaskilningur reynir á hæfni nemenda í að nota eiginleika talna og talnakerfis. Í prófhlutanum Hlutföll og prósentur er unnið með prósentureikning og skilning á líkindum og hlutföllum. Viðfangsefnið í prófhlutanum Algebra eru mynstur og reglur sem finna má í umhverfinu og að vinna með óþekktar stærðir og jöfnur (Sigurgrímur Skúlason, 2020).

Það er notast við þrjár mismunandi gerðir af einkunnagjöf í samræmdu könnunarprófunum. Raðeinkunn er notuð til að túlka frammistöðu einstakra nemenda og sýnir hlutfall nemenda sem eru með sambærilega eða slakari einkunn úr prófinu. Í samræmdum prófum fyrir 2016 er notast við einkunnir sem eru á kvarðanum 1-10, eftir það var byrjað að notast við hæfniseinkunn. Samræmdar einkunnir líkjast hefðbundnum einkunnum sem notað er í skólum en ekki er hægt að túlka þær með sama hætti. Að lokum eru normaldreifðar einkunnir, þær nota meðaltalið 30

með staðalfrávikíð 10 og eru notaðar til að skoða meðaltöl landshluta og skóla (Inga Úlfsdóttir o.fl., 2015).

Það er reynt að forprófa öll próf með nemendum sem eiga að vera líkir þeim nemendum sem taka könnunarprófin. Þar er athugað þyngd prófatriða sem fæst með því að mæla hlutföll réttra svara. Það er reynt að hafa prófatriðin í prófunum í breytilegri þyngd. Prófatriði með 50% til 75% hlutfall réttra svara eru gagnlegastar þar sem þau greina milli þeirra sem hafa ákveðna þekkingu og þeirra sem gera það ekki (Inga Úlfsdóttir o.fl., 2015).

Framkvæmd

Athugað var hvort væri nægjanlegt magn af prófatriðum í prófbankanum og hvernig sundurgreining í hverju prófatriði sé. Öll gögn voru veitt af MMS. Við úrvinnslu voru notaðar upplýsingar úr prófabanka MMS fyrir stærðfræði, atriðastikar (e. item parameters) fyrir prófatriði sem lögð voru fyrir 2000 til 2013, ásamt 2019 og 2020. Þessir atriðastikar voru metnir út frá svörum nemenda sem þreyttu samræmd könnunarpróf árin 2000 til 2013, ásamt 2019 og 2020.

Úrvinnsla gagna

Gögnin voru endurkóðuð í SPSS (Statistical Package for the Social Sciences) þar sem þyngd og sundurgreiningu atriða var skipt í flokka, 24 fyrir þyngd og 7 fyrir sundurgreiningu. Upphaflega var reynt að hafa 11 sundurgreiningarflokka en við frumskoðun gagna þá kom í ljós að fjóru flokkarnir á sitthvorum enda markanna voru með of takmarkaða skiptingu, þar af leiðandi voru þeir sameinaðir niður í tvo flokka fyrir sitthvorn enda markanna.

Töflur voru unnar í Microsoft Excel og Microsoft Word. Til að byrja með var skoðuð lýsandi tölfræði þar sem borin voru saman meðaltöl, lágsta og hæsta gildi, staðalfrávik og miðgildi á milli prófa á árabílinu 2000-2013. Ákveðið var að notast ekki við 2019 og 2020 gögnin vegna þess að 2019 innihélt einungis gögn úr 7. bekk. Árið 2020 var með of mikið bil milli ára og nýtist því ekki að skoða það í samanburði við hin árin ásamt því að það innihélt breytingu í unglíngastigi um hvaða bekk tæki prófið. Eftir flokkaskiptinguna voru gerðar krosstöflur til að bera saman flokkana í 7, 9. og 10. bekk saman og í sitthvorum lagi, þar voru prófatriði notuð úr öllum árunum, 2019 og 2020 meðtöldum.

Tafla 1*Flokkaskipting fyrir þyngd*

Þyngd	neðri mörk	efri mörk
1	neikvæð	-3
2	-3	-2.5
3	-2.5	-2
4	-2	-1.75
5	-1.75	-1.5
6	-1.5	-1.25
7	-1.25	-1
8	-1	-0.75
9	-0.75	-0.5
10	-0.5	-0.25
11	-0.25	0
12	0	0.25
13	0.25	0.5
14	0.5	0.75
15	0.75	1
16	1	1.25
17	1.25	1.5
18	1.5	1.75
19	1.75	2
20	2	2.25
21	2.25	2.5
22	2.5	2.75
23	2.75	3
24	3	

Tafla 2*Flokkaskipting fyrir sundurgreiningu*

Sundurgreining	Neðri mörk	Efri mörk
1	Frá 0.5 til neikvæðar	0.5
2	0.5	1
3	1	1.25
4	1.25	1.5
5	1.5	1.75
6	1.75	2.25
7	2.25	2.25+

Niðurstöður

Tafla 3

Lýsandi tölfraedi á þyngd prófatriða á samræmdu prófum í stærðfræði, 7. og 10. bekk sameinaðir

Ár	Meðaltal	Lægsta gildi	Hæsta gildi	Staðalfrávik	Miðgildi	Fjöldi
2000	-0.59	-6.68	2.42	1.30	-0.41	130
2001	-0.51	-3.62	2.11	1.06	-0.47	132
2002	-0.51	-4.91	1.67	1.29	-0.29	95
2003	-0.67	-4.16	3.20	1.22	-0.64	96
2004	-0.57	-5.83	1.66	1.24	-0.44	94
2005	-0.66	-3.69	2.36	0.92	-0.62	110
2006	-0.56	-4.02	1.68	0.91	-0.55	101
2007	-0.68	-3.95	1.11	0.85	-0.68	96
2008	-0.13	-3.78	24.89	2.68	-0.33	100
2009	-0.37	-4.60	2.47	1.22	-0.26	99
2010	-0.27	-5.62	2.32	1.36	-0.03	100
2011	-0.26	-24.73	2.67	2.75	0.15	99
2012	-0.36	-4.16	1.77	1.14	-0.24	104
2013	-0.26	-4.45	1.96	0.90	-0.15	106

Tafla 4

Lýsandi tölfraedi á þyngd prófatriða í samræmdum prófum í stærðfræði, 7. og 10. bekkur aðskildir

Bekkur	Meðaltal	Lægsta gildi	Hæsta gildi	Staðalfrávik	Miðgildi	Fjöldi
7	-0.72	-6.68	2.47	1.25	-0.62	680
10	-0.23	-24.73	24.89	1.58	-0.23	782
Samtals	-0.46	-24.73	24.89	1.46	-0.37	1462

Á töflu 3 má skoða lýsandi tölfraedi prófatriða á samræmdum prófum á hverju ári frá 2000-2013 þar sem 7. og 10. bekkur eru sameinaðir. Á töflu 4 er hægt að skoða meðaltöl á gildum ásamt lágsta og hæsta gildi auk skiptingu atriða þar sem 7. og 10. bekkur eru aðskildir.

Þegar litið er yfir þyngd prófa milli ára þá sést í fljótu bragði að þyngdarstuðlar prófa voru lægri í byrjun aldarinnar, á bilinu 0.50-0.70 en tóku stökk í þyngd árið 2008 og hafa síðan flöktað í kringum 0.25 til 0.40 bilið. Miðgildið hefur einnig átt við svipaða þróun. Prófin í 7. bekk voru léttari fyrir nemendur í 7. bekk en prófin í 10. bekk voru fyrir nemendur í 10. bekk. Samræmda prófið 2008 var með hæsta gildið og 2011 var með lágsta. Sömu prófin voru með

afbrigðileg staðalfrávik. Í samanburði bekkja þá voru atriðin í 10.bekk þyngri um 0.5 á stuðulskvarða, ásamt með hærri staðalfrávik og miðgildi.

Tafla 5

Lýsandi tölfraedi á sundurgreiningu prófatriða í samræmdum prófum í stærðfræði, 7. og 10. bekkur sameinaðir

Ár	Meðaltal	Lægsta gildi	Hæsta gildi	Staðalfrávik	Miðgildi	Fjöldi
2000	1.55	0.34	3.13	0.59	1.42	130
2001	1.64	0.47	3.82	0.64	1.56	132
2002	1.59	0.57	4.26	0.58	1.53	95
2003	1.63	0.61	3.09	0.50	1.57	96
2004	1.75	0.74	3.26	0.56	1.66	94
2005	1.80	0.64	3.14	0.54	1.78	110
2006	1.81	0.47	3.28	0.55	1.76	101
2007	1.82	0.77	3.20	0.58	1.77	96
2008	1.87	0.06	3.39	0.59	1.86	100
2009	1.58	-0.43	3.03	0.59	1.59	99
2010	1.58	-1.24	3.24	0.63	1.60	100
2011	1.72	-0.70	3.28	0.61	1.73	99
2012	1.61	0.10	3.87	0.61	1.51	104
2013	1.82	0.69	3.22	0.58	1.81	106

Tafla 6

Lýsandi tölfraedi á sundurgreiningu prófatriða í samræmdum prófum í stærðfræði, 7. og 10. bekkur aðskildir

Bekkur	Meðaltal	Lægsta gildi	Hæsta gildi	Staðalfrávik	Miðgildi	Fjöldi
7	1.51	-1.24	4.26	0.50	1.49	680
10	1.85	-0.43	3.87	0.62	1.84	782
Samtals	1.69	-1.24	4.26	0.59	1.66	1462

Á töflu 5 má skoða lýsandi tölfraedi sundurgreiningar prófatriða á samræmdum prófum frá 2000-2013 þar sem 7 og 10. bekkur eru sameinaðir. Á töflu 6 er hægt að skoða meðaltöl á gildum ásamt lágsta og hæsta gildi auk skiptingu atriða þar sem bekkirnir eru aðskildir.

Sundurgreining hefur haldist stöðugari milli ára heldur en þyngdin, þó að einhver stökk verða milli ára. Almennilega eru prófin á bilinu 1.55-1.90, með engum afbrigðilegum

staðalfrávikum eða miðgildum. Minnst sundurgreinandi prófatriðið fannst í 7.bekkjarsamræmda prófinu 2010, mest sundurgreinandi prófatriði fannst einnig í 7.bekk árið 2002. Í samanburði bekkja þá var hærri sundurgreining í 10.bekk ásamt hærri staðalfráviks og miðgildis.

Tafla 7

Lýsandi tölfraedi á þyngd prófatriða í samræmdum prófum í stærðfærði hjá 7. bekk

Ár	Meðaltal	Lægsta gildi	Hæsta gildi	Staðalfrávik	Miðgildi	Fjöldi
2000	-0.94	-6.68	1.38	1.49	-0.68	68
2001	-0.63	-3.62	2.11	1.14	-0.52	58
2002	-1.12	-4.91	1.67	1.39	-1.02	42
2003	-1.25	-4.16	1.11	1.17	-1.29	43
2004	-0.70	-5.83	1.18	1.53	-0.41	44
2005	-0.81	-3.69	1.97	1.01	-0.91	54
2006	-0.91	-4.02	1.68	0.95	-0.80	51
2007	-0.81	-3.95	0.67	0.83	-0.78	46
2008	-0.36	-3.78	2.27	1.02	-0.33	50
2009	-0.77	-4.60	2.47	1.33	-0.62	50
2010	-0.51	-5.62	2.11	1.67	-0.01	41
2011	-0.30	-3.63	1.98	1.15	-0.20	42
2012	-0.58	-4.16	1.69	1.22	-0.37	44
2013	-0.36	-4.45	1.96	1.11	-0.20	47

Í töflu 7 má skoða lýsandi tölfraedi á þyngd prófatriða í samræmdum prófum 7. bekk. Meðalþyngd prófa er óstöðug milli ára en hún fer frá -1.25 á einum tímapunkti alveg upp í -0.30. Lægsta og hæsta prófatriðið í þyngd hefur svipaða þróun. Staðalfrávikkið hefur verið verið frá 0.83 til 1.67 á milli ára þar sem lægsta staðalfrávikkið var þegar hæsta gildið var sem lægst. Miðgildið hefur bilið frá -1.29 til -0.01.

Tafla 8

Lýsandi tölfraedi á sundurgreiningu prófatriða í samræmdum prófum í stærðfærði hjá 7. bekk

Ár	Meðaltal	Lægsta gildi	Hæsta gildi	Staðalfrávik	Miðgildi	Fjöldi
2000	1.27	0.34	2.24	0.38	1.25	68
2001	1.33	0.47	2.53	0.41	1.26	58
2002	1.44	0.58	4.26	0.61	1.34	42
2003	1.48	0.71	2.73	0.46	1.41	43
2004	1.52	0.74	2.58	0.43	1.47	44
2005	1.55	0.64	2.52	0.45	1.55	54
2006	1.63	0.55	2.63	0.44	1.56	51
2007	1.67	0.77	2.54	0.39	1.71	46
2008	1.75	0.66	2.57	0.40	1.82	50
2009	1.44	0.25	2.69	0.49	1.42	50
2010	1.30	-1.24	2.51	0.60	1.33	41
2011	1.63	-0.70	2.40	0.52	1.67	42
2012	1.67	0.44	3.40	0.64	1.55	44
2013	1.63	0.69	3.00	0.56	1.61	47

Í töflu 8 má skoða lýsandi tölfraedi á sundurgreiningu prófatriða í samræmdum prófum 7. bekkis. Sundurgreining hækkar stöðugt frá lágpunktinum 1.27 árið 2000 alveg þangað til 2008 þar sem hápunktinum 1.75 er náð. Því fylgdi falli í sundurgreiningu alveg fram að mesta mun milli ára á prófunum 2010 og 2011 þar sem hækkun varð. Lægsta gildið er oft í jákvæðri tölu nema árin 2010 og 2011 þar sem koma inn neikvæðar tölur annars er bilið vanalega milli 0.25 til 0.77. Hæsta gildið fer þrisvar yfir 3.00 en árið 2002 á hæsta og eina sundurgreinandi prófatriðið yfir 3.50. Staðalfrávikkið er stöðugt milli ára á bilinu 0.38 til 0.64, sama má segja um miðgildið sem er á bilinu 1.25 til 1.82.

Tafla 9

Lýsandi tölfræði á þyngd prófatriða í samræmdum prófum í stærðfærði hjá 10. bekk

Ár	Meðaltal	Lægsta gildi	Hæsta gildi	Staðalfrávik	Miðgildi	Fjöldi
2000	-0.21	-2.13	2.42	0.92	-0.24	62
2001	-0.41	-2.85	1.57	0.98	-0.44	74
2002	-0.03	-3.85	1.43	0.97	0.10	53
2003	-0.21	-3.14	3.20	1.06	-0.17	53
2004	-0.46	-2.99	1.66	0.91	-0.44	50
2005	-0.52	-2.33	2.36	0.82	-0.52	56
2006	-0.21	-1.67	1.62	0.73	-0.24	50
2007	-0.56	-2.34	1.11	0.86	-0.53	50
2008	0.10	-2.57	24.89	3.66	-0.29	50
2009	0.04	-3.21	1.87	0.95	-0.01	49
2010	-0.10	-3.01	2.32	1.09	-0.04	59
2011	-0.23	-24.73	2.67	3.50	0.37	57
2012	-0.20	-3.87	1.77	1.06	-0.13	60
2013	-0.17	-1.74	1.55	0.70	-0.13	59

Tafla 9 sýnir lýsandi tölfræði á þyngd prófatriða í samræmdum prófum 10. bekk. Meðaltalið fer tvisvar í jákvæða tölu, ef litið er framhjá því þá er þyngd vanalega einhvers staðar á bilinu -0.52 til -0.03. Gildin sem standa hvað mest út eru lægsta gildið -24.73 á prófinu 2011 og hæsta gildið 24.89 á prófinu 2008. Ekkert gildi fer annars undir -4.00 eða yfir 4.00. Staðalfrávikin er fremur stöðugt á bilinu 0.70 til 1.09 ef litið er framhjá árunum 2008 og 2011 með staðalfrávikin 3.66 og 3.50. Miðgildið er oftast í neikvæðri tölu nema árið 2011 þar sem miðgildið var 0.37.

Tafla 10

Lýsandi tölfraði á sundurgreiningu prófatriða í samræmdum prófum í stærðferði hjá 10. bekk

Ár	Meðaltal	Lægsta gildi	Hæsta gildi	Staðalfrávik	Miðgildi	Fjöldi
2000	1.86	0.72	3.13	0.63	1.81	62
2001	1.89	0.62	3.82	0.67	1.76	74
2002	1.70	0.57	3.06	0.54	1.71	53
2003	1.75	0.61	3.09	0.51	1.85	53
2004	1.95	1.03	3.26	0.59	2.08	50
2005	2.04	0.82	3.14	0.52	1.95	56
2006	2.00	0.47	3.28	0.59	2.04	50
2007	1.96	0.77	3.20	0.69	1.81	50
2008	1.99	0.06	3.39	0.72	2.00	50
2009	1.72	-0.43	3.03	0.66	1.71	49
2010	1.78	0.18	3.24	0.58	1.82	59
2011	1.78	-0.38	3.28	0.67	1.81	57
2012	1.56	0.10	3.87	0.59	1.48	60
2013	1.98	0.70	3.22	0.55	1.94	59

Í töflu 10 má skoða lýsandi tölfraði á sundurgreiningu prófatriða í samræmdum prófum 10. bekk. Sundurgreining prófatriða nær tvisvar yfir 2.00 í meðaltali en er með lágpunkt 1.56 árið 2012 sem er einnig eina árið þar sem sundurgreining var meiri í prófatriðum á samræmdu prófi 7. bekk en í 10. bekk. Lægsta gildið fer einnig tvisvar í neikvæða tölu og einu sinni yfir 1.00 í jákvæðum tölum. Hæsta gildið er alltaf yfir 3.00 en aldrei yfir 4.00. Staðalfrávik er stöðugt á bilinu 0.52 til 0.72. Miðgildið fer þrisvar yfir 2.00. Í eitt skipti fer það undir 1.71 með 1.48 á prófinu 2012 en vanalega er það á bilinu 1.71 til 2.08 ef ekki er tekið það ár með.

Tafla 11*Krosstafla með dreifingu prófatriða hjá 7. og 10. bekk*

	Sundurgreining							Samtals	
	1	2	3	4	5	6	7		
Þyngd	1	10	20	10	1	1	0	1	43
	2	4	4	4	5	3	0	0	20
	3	0	11	13	20	9	4	0	57
	4	1	3	10	12	8	4	1	39
	5	0	7	17	10	13	6	0	53
	6	2	9	12	21	16	25	6	91
	7	0	9	29	19	24	37	13	131
	8	0	5	12	28	22	25	29	121
	9	0	5	18	24	29	55	34	165
	10	0	9	19	26	24	55	33	166
	11	0	7	20	21	27	62	29	166
	12	0	14	23	32	27	45	42	183
	13	0	7	14	20	16	53	29	139
	14	0	2	12	14	25	37	25	115
	15	2	5	7	8	12	20	11	65
	16	0	6	7	8	10	16	19	66
	17	1	6	7	4	6	13	10	47
	18	2	2	3	3	7	9	6	32
	19	0	2	1	4	2	3	1	13
	20	0	1	0	0	0	1	0	2
	21	1	3	1	0	0	0	4	9
	22	0	0	0	0	0	0	1	1
	23	0	0	0	0	0	0	0	0
	24	1	0	0	0	0	1	0	2
Samtals		24	137	239	280	281	471	294	1726

Tafla 12*Krosstafla með dreifingu prófatriða í prósentum hjá 7. og 10. bekk*

		Sundurgreining							
		1	2	3	4	5	6	7	Samtals
Þyngd	1	23.3	46.5	23.3	2.3	2.3	0.0	2.3	100.0
	2	20.0	20.0	20.0	25.0	15.0	0.0	0.0	100.0
	3	0.0	19.3	22.8	35.1	15.8	7.0	0.0	100.0
	4	2.6	7.7	25.6	30.8	20.5	10.3	2.6	100.0
	5	0.0	13.2	32.1	18.9	24.5	11.3	0.0	100.0
	6	2.2	9.9	13.2	23.1	17.6	27.5	6.6	100.0
	7	0.0	6.9	22.1	14.5	18.3	28.2	9.9	100.0
	8	0.0	4.1	9.9	23.1	18.2	20.7	24.0	100.0
	9	0.0	3.0	10.9	14.5	17.6	33.3	20.6	100.0
	10	0.0	5.4	11.4	15.7	14.5	33.1	19.9	100.0
	11	0.0	4.2	12.0	12.7	16.3	37.3	17.5	100.0
	12	0.0	7.7	12.6	17.5	14.8	24.6	23.0	100.0
	13	0.0	5.0	10.1	14.4	11.5	38.1	20.9	100.0
	14	0.0	1.7	10.4	12.2	21.7	32.2	21.7	100.0
	15	3.1	7.7	10.8	12.3	18.5	30.8	16.9	100.0
	16	0.0	9.1	10.6	12.1	15.2	24.2	28.8	100.0
	17	2.1	12.8	14.9	8.5	12.8	27.7	21.3	100.0
	18	6.3	6.3	9.4	9.4	21.9	28.1	18.8	100.0
	19	0.0	15.4	7.7	30.8	15.4	23.1	7.7	100.0
	20	0.0	50.0	0.0	0.0	0.0	50.0	0.0	100.0
	21	11.1	33.3	11.1	0.0	0.0	0.0	44.4	100.0
	22	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0
	23	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	24	50.0	0.0	0.0	0.0	0.0	50.0	0.0	100.0
Samtals		1.4	7.9	13.8	16.2	16.3	27.3	17.0	100.0

Tafla 11 sýnir dreifingu prófatriða á 7. og 10. bekk með fjölda eftir skiptingu flokka nefnda í úrvinnslu gagna. Dreifingin á hvern flokk er yfir 20 prófatriði á þyngd alveg fram að flokki 19 þar sem þurrkinn kemur, flokkar 20, 22 og 24 ná ekki yfir fleiri en 2 prófatriði og einn flokkur er ekki með eitt einasta prófatriði. Sundurgreiningin er lægri í fyrri flokkunum og er fremur jöfn í flokkum 4, 5 og 7 en er töluvert hærri í flokki 6. Á töflu 12 má sjá prósentudreifingu flokkaskiptingarnar á prófatriðum í prófatriðabankanum fyrir 7. og 10. bekk. Þessi tafla nýtist til skoðunar vægisskiptingar prófatriða í hverjum flokk þar sem bekkir eru sameinaðir.

Tafla 13*Krosstafla með dreifingu prófatriða hjá 7. bekk*

		Sundurgreining							Samtals
		1	2	3	4	5	6	7	
Þyngd	1	5	19	10	1	0	0	0	35
	2	1	2	3	4	1	0	0	11
	3	0	8	10	15	6	2	0	41
	4	1	2	5	11	7	1	0	27
	5	0	4	11	4	9	4	0	32
	6	2	7	7	10	7	18	3	54
	7	0	4	22	11	12	19	4	72
	8	0	2	4	14	15	14	11	60
	9	0	5	13	17	16	23	4	78
	10	0	6	4	14	14	24	6	68
	11	0	2	15	11	16	27	2	73
	12	0	7	13	15	15	13	9	72
	13	0	2	6	12	5	25	2	52
	14	0	2	6	3	15	10	5	41
	15	1	2	3	5	4	9	2	26
	16	0	4	4	4	3	5	1	21
	17	1	2	4	1	0	4	3	15
	18	1	1	1	2	5	2	2	14
	19	0	0	1	2	1	0	0	4
	20	0	1	0	0	0	1	0	2
	21	0	0	0	0	0	0	2	2
	22	0	0	0	0	0	0	0	0
	23	0	0	0	0	0	0	0	0
	24	0	0	0	0	0	0	0	0
Samtals		12	82	142	156	151	201	56	800

Tafla 14*Krosstafla með dreifingu prófatriða í prósentum hjá 7. bekk*

		Sundurgreining							
		1	2	3	4	5	6	7	Samtals
Þyngd	1	14.3	54.3	28.6	2.9	0.0	0.0	0.0	100.0
	2	9.1	18.2	27.3	36.4	9.1	0.0	0.0	100.0
	3	0.0	19.5	24.4	36.6	14.6	4.9	0.0	100.0
	4	3.7	7.4	18.5	40.7	25.9	3.7	0.0	100.0
	5	0.0	12.5	34.4	12.5	28.1	12.5	0.0	100.0
	6	3.7	13.0	13.0	18.5	13.0	33.3	5.6	100.0
	7	0.0	5.6	30.6	15.3	16.7	26.4	5.6	100.0
	8	0.0	3.3	6.7	23.3	25.0	23.3	18.3	100.0
	9	0.0	6.4	16.7	21.8	20.5	29.5	5.1	100.0
	10	0.0	8.8	5.9	20.6	20.6	35.3	8.8	100.0
	11	0.0	2.7	20.5	15.1	21.9	37.0	2.7	100.0
	12	0.0	9.7	18.1	20.8	20.8	18.1	12.5	100.0
	13	0.0	3.8	11.5	23.1	9.6	48.1	3.8	100.0
	14	0.0	4.9	14.6	7.3	36.6	24.4	12.2	100.0
	15	3.8	7.7	11.5	19.2	15.4	34.6	7.7	100.0
	16	0.0	19.0	19.0	19.0	14.3	23.8	4.8	100.0
	17	6.7	13.3	26.7	6.7	0.0	26.7	20.0	100.0
	18	7.1	7.1	7.1	14.3	35.7	14.3	14.3	100.0
	19	0.0	0.0	25.0	50.0	25.0	0.0	0.0	100.0
	20	0.0	50.0	0.0	0.0	0.0	50.0	0.0	100.0
	21	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0
	22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	23	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Samtals		1.5	10.3	17.8	19.5	18.9	25.1	7.0	100.0

Tafla 13 og 14 sýna dreifingu fjölda og í prósentum á prófatriðum í flokkaskiptingu úr 7. bekk. Flokkar 2, 17 og 18 eru með 15 atriði eða undir. Flokkar 19-21 eru með tvö eða fjögur atriði á flokk og frá flokk 22 og ofar er ekkert atriði sem finnst úr prófatriðum 7. bekk. Dreifingin er lægri í neðsta og efsta sundurgreiningar flokkunum. Eini sundurgreiningarflokkurinn sem er með meiri en 4% mun í dreifingu í samanburði við krosstöflu 7 og 10. bekk er flokkur 7 með 10% mun.

Tafla 15*Krosstafla með dreifingu prófatriða hjá 10. bekk*

		Sundurgreining							Samtals
		1	2	3	4	5	6	7	
Þyngd	1	5	1	0	0	1	0	1	8
	2	3	2	1	1	2	0	0	9
	3	0	3	3	5	3	2	0	16
	4	0	1	5	1	1	3	1	12
	5	0	3	6	6	4	2	0	21
	6	0	2	5	11	9	7	3	37
	7	0	5	7	8	12	18	9	59
	8	0	3	8	14	7	11	18	61
	9	0	0	5	7	13	32	30	87
	10	0	3	15	12	10	31	27	98
	11	0	5	5	10	11	35	27	93
	12	0	7	10	17	12	32	33	111
	13	0	5	8	8	11	28	27	87
	14	0	0	6	11	10	27	20	74
	15	1	3	4	3	8	11	9	39
	16	0	2	3	4	7	11	18	45
	17	0	4	3	3	6	9	7	32
	18	1	1	2	1	2	7	4	18
	19	0	2	0	2	1	3	1	9
	21	1	3	1	0	0	0	2	7
	22	0	0	0	0	0	0	1	1
	23	0	0	0	0	0	0	0	0
	24	1	0	0	0	0	1	0	2
	Samtals		12	55	97	124	130	270	238

Tafla 16*Krosstafla með dreifingu prófatriða í prósentum hjá 10. bekk*

		Sundgreining							Samtals
		1	2	3	4	5	6	7	
Þyngd	1	62.5	12.5	0.0	0.0	12.5	0.0	12.5	100.0
	2	33.3	22.2	11.1	11.1	22.2	0.0	0.0	100.0
	3	0.0	18.8	18.8	31.3	18.8	12.5	0.0	100.0
	4	0.0	8.3	41.7	8.3	8.3	25.0	8.3	100.0
	5	0.0	14.3	28.6	28.6	19.0	9.5	0.0	100.0
	6	0.0	5.4	13.5	29.7	24.3	18.9	8.1	100.0
	7	0.0	8.5	11.9	13.6	20.3	30.5	15.3	100.0
	8	0.0	4.9	13.1	23.0	11.5	18.0	29.5	100.0
	9	0.0	0.0	5.7	8.0	14.9	36.8	34.5	100.0
	10	0.0	3.1	15.3	12.2	10.2	31.6	27.6	100.0
	11	0.0	5.4	5.4	10.8	11.8	37.6	29.0	100.0
	12	0.0	6.3	9.0	15.3	10.8	28.8	29.7	100.0
	13	0.0	5.7	9.2	9.2	12.6	32.2	31.0	100.0
	14	0.0	0.0	8.1	14.9	13.5	36.5	27.0	100.0
	15	2.6	7.7	10.3	7.7	20.5	28.2	23.1	100.0
	16	0.0	4.4	6.7	8.9	15.6	24.4	40.0	100.0
	17	0.0	12.5	9.4	9.4	18.8	28.1	21.9	100.0
	18	5.6	5.6	11.1	5.6	11.1	38.9	22.2	100.0
	19	0.0	22.2	0.0	22.2	11.1	33.3	11.1	100.0
	21	14.3	42.9	14.3	0.0	0.0	0.0	28.6	100.0
	22	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0
	23	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	24	50.0	0.0	0.0	0.0	0.0	50.0	0.0	100.0
	Samtals		1.3	5.9	10.5	13.4	14.0	29.2	25.7

Tafla 15 og 16 sýna dreifingu fjölda og í prósentum á fjölda prófatriða í flokkaskiptingu úr 10. bekk. Flokkar 1, 2, 4, 19-24 eru með 15 atriði eða undir og í tilviki flokks 23 er ekkert prófatriði. Dreifingin fer stigvaxandi úr neðsta fram að næstefsta sundurgreiningarflokk en efsti sundurgreiningarflokkurinn er samt með næstmesta hlutfallið. Eini sundurgreiningarflokkurinn sem er með meiri en 4% mun í dreifingu í samanburði við töflu 12 er flokkur 7 með 8.7% mun. Í hlutfallssamanburð milli prófatriða 7. og 10.bekks er talsvert meira hlutfall atriða í sundurgreiningarflokki 6 og 7 hjá 10. bekk en hinir 5 flokkarnir eru hlutfallslega hærri hjá 7. bekk þó að flokkur 1 er einungis með 0.2% mun.

Umraeða

Eftir að hafa greint prófatriðabankann þá hafa niðurstöðurnar leitt eftirfarandi í ljós. Lýsandi tölfræðin um samræmdu prófin segir okkur að próf hafa orðið erfiðari með tímanum en það gæti verið vegna breytinga á stefnu varðandi að hafa ekki eins auðveld atriði sem flestir próftakar gætu svarað. Í samræmi við það þá getum við séð að miðgildi prófa hafði fylgt sömu stefnu sem segir okkur eitthvað svipað. Það að staðalfrávik var afbrigðilegt á tveim árum hefur líklegast orsakast vegna innihalds þeirra á lægsta og hæsta gildi prófatriðabankans. Áhugavert er samt að miðgildið hafi ekki verið afbrigðilegt á einu af þessum árum en eina jákvæða miðgildið í allri töflunni kom árið 2011. Bilið á milli lægsta og hæsta gildis á prófinu 2008 var lengra um 1,27 í samanburði við árið 2011. Mögulegt er að það séu orsök fyrir afbrigðilega miðgildinu.

Sundurgreiningin í sameinaðri lýsandi tölfræði í 7. og 10. bekk lítur stöðug út og samkvæmt því sem leitast er eftir þá höfðu öll próf mikla sundurgreininga að meðaltali (Penfield, 2013). Með einstaka undantekningu á lægstu og hæstu gildum.

Ef að skoðuð er lýsandi tölfræði úr meðaltölum hvers gildis á töflunum fyrir sitthvorn bekkinn með hliðsjón af þeirri framtíðarstefnu að sameina þau tvö próf, þá ætti að reyna semja próf í kringum -0,46 í þyngd og sundurgreiningu í a.m.k. upp á 1,69. Þetta er ekki heilagt en gott fyrir hönnuði þess prófs að hafa í huga. Einnig var athyglisvert að fjögur samræmd próf í 7. bekk myndu vera erfiðari í þyngd en nokkur þeirra sem komu fram í 10. bekk. Ekki hlutlægislega séð en í samanburði hafi samræmda prófið verið erfiðara fyrir nemandan í 7. bekk heldur en samræmda prófið var fyrir nemandan í 10. bekk þau ár. Það voru nokkur ár þar sem prófið í 7. bekk var svipað þungt og enprófið í 10. bekk, árið 2011 væri augljósasta dæmið þar sem munurinn var einungis 0.06, þó að það geti hafa orsakast enn og aftur vegna lægsta þyngdargildis í öllum bankanum sem kom fram hjá 10. bekk. Það var bara árið 2010 sem miðgildið var hærra hjá prófi 7. bekk. Einhver ár kom líka fyrir að sundurgreining samræmda prófsins hjá 7. bekk var ekki yfir 1.5 en það var í kringum það bil, þannig þó að það sé ekki ákjósanlegast þá er það samt ásættanlegt og ekki mikið til að sakast um.

Í sameinaðri krosstöflu 7. og 10. bekkjar kom fram að fjöldi prófatriða varð færri en 15 eftir þyngdarflokk 18 og eru of fá atriði til þess að sundurgreina þyngdarflokkana eftir þyngdarflokk 18. Ef stefna MMS tekst með því að sameina bekkir sem taka samræmd próf á einn kvarða þá væri ákjósanlegt að bæta við fleiri prófatriðum í þyngdarflokkana 19-24 til að hægt væri að

framkvæma atriðagreiningu á þeim og úr því meta hvort þau sé með nægilega sundurgreiningu (Furr, 2018; Kolen, Brennan, 2014; Penfield, 2013). 7. bekkur hefur úr sérstaklega litlu að velja með prófatriði í hárrí þyngd en bæði atriði í hæsta þyngdarflokki sem 7. bekkur fær prófatriði úr lendir hátt á sundurgreiningarkvarða. Það leiðir til þess að mögulega ríkir óvissa með hvar þyngri prófatriði í flokkum sem 7. bekkur náði ekki í myndu lenda á í sundurgreiningu eða hvenær of þung atriði veita enga hjálp hvað varðar sundurgreiningu.

Með 10. bekk þá eru færri prófatriði en 15 í nokkrum af lægri þyngdarfloknum. Velta má fyrir sér hvort það sé vegna þess að prófatriðin eru ekki nægilega vel samín eða hvort það sé ekkert gagn af því að nota atriði í of lágri þyngd fyrir 10. bekk. Það vantar einnig upp á atriði frá flokki 19 hjá 10. bekk auk þess að það virðist nokkuð vandamál með sundurgreiningu hjá sumum prófatriðum í þeim flokkum. Svipað á við hjá 7. bekk en ekki sama hlutfalli. Prófatriði hafa betri sundurgreiningu í 10. bekk heldur en 7. bekk en 54,9% prófatriða hjá 10. bekk lenda í flokki 6 og 7 í sundurgreiningu samanborið við einungis 32,1% hjá 7. bekk í sömu flokkum. Það er hlutfallslega meiri prósentu hjá 7. bekk í öllum flokkum frá 1-5 þó að ekki sé það mikill munur í sundurgreiningarflokki 1.

Ef þessi sameinaði kvarði verður að veruleika og notast væri við prófatriðabankann sem verið er að skoða þá ætti í stuttu máli að reyna bæta við fjölda prófatriða í flokki 19 og ofar en einnig ætti að endurskoða sum prófatriði sem falla í sundurgreiningaflokk 1 þar sem það fellur töluvert undir ásætlanlega 0.8 bilinu og þó að flokkur 2 fari upp í 1.0 þá eru þar einnig atriði sem ná ekki ásætlanlegum mörkum (Penfield, 2013). Mögulega eftir endurskoðun prófatriða gæti þurft að bæta við nokkur prófatriði sem skora í efri bilum flokks 2, eða fara í flokk 3 og 4 en frábært væri ef sem flest atriði komast í flokk 5 eða ofar í sundurgreiningu. Sérstaklega ef tekin yrði ákvörðun eftir endurskoðun að taka út öll prófatriði sem falla undir a.m.k. 0.8 bil í sundurgreiningu.

Þar sem MMS er að fara yfir hvort að nota eigi prófatriðabankann í aðlögunarhæft próf til langtíma þá ættu þeir sem hanna prófið að hafa í huga að endurskoða prófaatriðabankann með þeim sex vandamálum sem geta komið fram í aðlögunarhæfu prófi. Hanna ætti viðeigandi algrím við lausn þessa vandamála. Við viljum einnig stinga upp á að notast frekar við fjölþrepaðnið í stað aðlögunarhæfs ef ekki er nú þegar búið að skoða þann valmöguleika. Það virðist leysa öll vandamál varðandi aðlögunarhæfa sniðið ef prófhönnuðir eru nægilega hæfir

og veita mannlega yfirsýn á eiginleika prófa sem tölva eða algrími geta allavega ekki enn gert (Hendrickson, 2007).

Hvað varðar framtíðarstefnu MMS þá virðist sem að spennandi tímamót séu framundan hvað varðar námsmatið á Íslandi. Til dæmis varðandi rótækar breytingar sem virðist næsta þróunarskref í uppbyggingu námsmats. Hugmyndir um matsferil og notkun kvarða sem inniheldur alla bekk hljóma spennandi. Þó verður að gæta að því að varast að taka ekki upp of stutt rafræn próf til að forðast hugmyndir um hvort klassískt raungildislíkan gagnist betur en svarferlalíkan (Furr, 2018; Jabrayilov o.fl., 2016; Mennta- og menningarmálaráðuneyti, 2020; Sigurgrímur Skúlason, 2022).

Varðandi rannsóknina sem var hér framkvæmd þá þá væri hægt að skoða þetta nánar og íhuga að bæta við prófatriðum frá 2014-2019 og 2021. Þetta veitir meiri möguleika á að skoða skiptingu í prófatriðabankanum með mun fleiri gögn frá árunum þar sem 9. bekkur þreytti samræmd próf. Þar sem þetta er einnig prófatriðabanki úrsamræmdum prófum í stærðfræði þá væri áhugavert að kóða inn í prófatriðabankann, skiptingu prófatriða út frá því hvaða hluti prófsins eru úr. Þar af leiðandi væri hægt að skoða hvar helsta vandamálið er varðandi þyngd og sundurgreiningu.

Áhugavert verður að fylgjast með því hvað framtíðarrannsóknir munu leiða af sér, og sambærileg rannsókn á 4. og 7. bekk. Líklegast er að næsta skref verði að skoða stígandann milli hvers bekk en fróðlegt verður að sjá hvað er í raun að tapast á því að hafa svo vítt bil milli bekkja í samræmdum prófum (Sigurgrímur Skúlason, 2022). Hvort það leiði til þess að hanna verði enn fleiri prófatriði fyrir bankann, sérstaklega ef hann á að vera nýttur í framkvæmd prófa, á öllum bekkjum grunnskólastigs meðfram 4. bekk hvort sem að þau verði í sniði aðlögunarhæfs eða fjölþrepa prófs.

Heimildaskrá

- Conley, D.T. (2014). *The Common Core State Standards: Insight into Their Development and Purpose*. Council of Chief State School Officers.
- Eurostat. (2021). *Quality of life indicators - education*. Eurostat: Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Quality_of_life_indicators_-_education#Education_in_the_context_of_quality_of_life
- Furr, R., M. (2018). *Psychometrics* (3.útg). SAGE Publications.
- Hendrickson, A. (2007). *An NCME Instructional Module on Multistage Testing*. *Educational Measurement: Issues and Practice*, 26 (2), (bls. 44–52). <https://doi.org/10.1111/j.1745-3992.2007.00093.x>
- Hill, P. og Barber, M. (2014). *Preparing for a Renaissance in Assessment*. Pearson
- Inga Úlfisdóttir, Finnbogi Gunnarsson, Sigurgrímur Skúlason og Júlíus K. Björnsson (2015). *Gerð og framkvæmd samræmdra prófa*. Námsmatsstofnun.
- Jabrayilov, R., Emons, W. og Sijtsma, K. (2016). Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. *Applied Psychological Measurement*, 40(8), 559–572. <https://doi.org/10.1177/0146621616664046>
- Javed, S., Javed, S. og Khan, A. (2016). Effect of Education on Quality of Life and Well Being. *International Journal of Indian Psychology*. 3(3). (bls. 2349-3429). <https://doi.org/10.25215/0304.053>.
- Kendall, J. S. (2011). *Understanding Common Core State Standards*. ASCD.
- Kingston, N. M. og Kramer, L. B. (2013). High-Stakes Test Construction and Test Use. Í Little, T. D. (ritstj.), *The Oxford Handbook of Quantitative Methods, Volume 1: Foundations*. 189205. Oxford Library of Psychology. <https://doi.org/10.1093/oxfordhb/9780199934874.013.0010>
- Kingston, N. M., Scheuring, S. T. og Kramer, L. B. (2013). Test development strategies. Í K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (ritstj.), *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology*. (bls. 165–184). American Psychological Association. <https://doi.org/10.1037/14047-009>
- Kolen, M. J. og Brennan, R. L. (2014). *Test Equating, Scaling, and Linking: Methods and Practices* (3. útg.).

- Mennta- og menningarmálaráðuneytið (2020). *Framtíðarstefna um samræmt námsmat Tillögur starfshóps um markmið, hlutverk, framkvæmd og fyrirkomulag samræmdra könnunarprófa.*
- Mennta- og menningarmálaráðuneytið (2021). *Aðalnámskrá grunnskóla* (3. útg.).
- Mennta- og menningarmálaráðuneytið. (2012). *Aðalnámskrá leikskóla 2011.*
- Mennta- og menningarmálaráðuneytið. (2015). *Aðalnámskrá framhaldsskóla 2011* (2. útg.).
- Menntamálastofnun. (2016). *Að gefa lokaekinn í grunnskóla eftir nýjum matskvarða.*
- Moon, T., Brighton, C. M., Jarvis, J. og Hall, C. J. (2007). *State Standardized Testing Programs: Their Effects on Teachers and Students.* National Research Center on the Gifted and Talented.
- Otten, Samuel. (2011). Cornered by the Real World: A Defense of Mathematics. *The Mathematics Teacher*. 105, (bls. 20-25).
<https://doi.org/10.5951/mathteacher.105.1.0020>
- Parshall, C. G., Spray, J. A., Kalohn, J. C. og Davey, T. (2002). *Practical considerations in Computer-Based Testing.* Springer
- Penfield, R. D. (2013). Item analysis. Í K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (ritstj.), *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology.* (bls. 121–138). American Psychological Association. <https://doi.org/10.1037/14047-007>
- Powdthavee, N., Lekfuangfu, W. N., & Wooden, M. (2015). What's the good of education on our overall quality of life? A simultaneous equation model of education and life satisfaction for Australia. *Journal of behavioral and experimental economics*, 54, (bls. 10–21). <https://doi.org/10.1016/j.socec.2014.11.002>
- Reglugerð um fyrirkomulag og framkvæmd samræmdra könnunarprófa í grunnskóla nr. 173. (2017). <https://www.reglugerd.is/reglugerdir/eftir-raduneytum/menntamalaraduneyti/nr/20450>
- Rodriguez, M. C. og Haladyna, T. M. (2013). Objective testing of educational achievement. Í K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (ritstj.), *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology.* (bls. 305–314). American Psychological Association. <https://doi.org/10.1037/14047-018>

- Sigurgrímur Skúlason (2020). *Uppbygging samræmdra könnunarprófa, framkvæmd þeirra og úrvinnsla*. Menntamálastofnun
- Sigurgrímur Skúlason (2022). *Áætlun um val á prófatriðum fyrir rannsóknarpróf í stærðfræði*. Menntamálastofnun.
- Sigurgrímur Skúlason, Guðrún Birna Einarsdóttir, Auðun Valborgarson, Nói Kristinsson og Sverrir Óskarsson (2021). *Samræmd könnunarpróf í 9. bekk skólaárið 2020–2021*. Menntamálastofnun
- Unnar Geirdal, Guðmundur Arnkelsson, Auðun Valborgarson og Sigurgrímur Skúlason. (2021, 15. október). *Forsendur fyrir samtímalausn á svarferlalíkani fyrir prófatriðabanka í stærðfræði sem byggir á prófatriðum úr samræmdum könnunarprófum í 4. og 7. bekk og á unglingsstigi* [fyrirlestur]. Menntakvika, Menntavísindasvið Háskóla Íslands.
- Van der Linden, W. J. og Glas, C. A. W. (2010). *Elements of Adaptive Testing*. Springer
- Ward, A. W. og Murray-Ward, M. (1994). An NCME Instructional Module: Guidelines for the Development of Item Banks. *Educational Measurement: Issues and Practice*, 13, (bls. 34-39). <https://doi.org/10.1111/j.1745-3992.1994.tb00785.x>
- Winters, J.V. (2011) Human capital, higher education institutions, and quality of life, *Regional Science and Urban Economics*. 41(5), (bls. 446-454). <https://doi.org/10.1016/j.regsciurbeco.2011.03.001>.