

Kitsa ja laia matemaatika riigieksami linkimine (omavahel ja erinevate aastatega) kasutades ülesannete kolmikute eeltestimisel erinevaid gruppe

Einar Rull

SA Innove analüütik

Üheks testitavaks kolmikuks oli kokku pandud üks eelmise aasta (2016) laia matemaatikakursuse eksami ülesanne ning üks selle aasta (2017) kitsa matemaatikakursuse eksami ja üks laia matemaatikakursuse eksami ülesanne. Kõik kolmiku ülesanded olid seotud ühe teemaga, nt algebraga.

Arvutused põhinevad psühhomeetriast teada tõsiasjadele

Esiteks tõdeme, et eksami summaarset tulemust saab vaadelda kui sooritusprotsenti, mis kujuneb **üksikülesannete sooritusprotsentide (kaalutud) keskmistest**. Sellest järeldeb muuhulgas, et kui mingi ühe ülesande, mille kaal on nt 10% kogu eksamist, eeltestimise korraldamine ebaõnnestub, on selle mõju keskmisele 10 korda väiksem, kui selle ülesande enese parameetri erinevus eksami keskmisest.

Teine kasulik vaatepunkt põhineb teadmisel, mis on laiemalt tuntuks saanud ka seoses PISA tulemuste tõlgendamisega, mille järgi 39 punkti vahemik PISA skaalal vastab ühele aastale õppimisele. Teiste sõnadega vastab põhikooli lõpus 0,39 standardhälbele 1 aasta õppimist PISA skaalal. Enam-vähem nii on see kõikide ainetestide puhul selles vanuses. Uuringud näitavad, et gümnaasiumi lõpuks on keskmine edasiliikumine vähenenud veidi alla 0,3 standardhälbe. Oluline kasutatav teadmine siit on, et **punktivahemikke on võimalik tõlgendada skaalal keskmiste õppimiseks kulunud ajavahemikena**, v.a otspunktide ümbrus, kus intervallskaala on moonutatud (vt PISA 2012 uuringu kokkuvõtte; HTM 3. detsember, 2013. aasta pressikonverentsi materjalid lk 5 https://www.innove.ee/UserFiles/%C3%9CIdharidus/PISA%202012/PISA_2012_uuringu_tulemuste_kokkuvote.pdf).

Kolmandaks eelduseks on tõsiasi, et matemaatika lõpptulemuse skaala moodustub paljudest ülesannetega tekitatud alamskaaladest, kus igaühel võib edasiliikumine toimuda erineval kiirusel. Kuid me võime seda ignoreerida ja tuginedes lineaarsusele võtta arvutustes edasiliikumise kiiruseks igal skaalal summaarse keskmise kiiruse, sest meid huvitavad vaid tulemused summaarse tulemuse osas, mitte üksikülesannete lõikes. Hariduslikes uuringutes ongi seetõttu laialt levinud lähenemine, kus õppekavapõhiste testide tõlgendamisel vaadeldakse tihti õpilaste edenemist ainult kogu testi summaarsel kognitiivsel skaalal.

Hariduslikes mõõtmistes kasutatakse ka tõsiasja, et kui üks ja sama õpilaste grupp teeb kahte erinevat ülesannet, siis vahe näitab nende ülesannete raskuste erinevust, mida me oskame ümber arvutada nt ajaskaalale, kui palju kuluks keskmiselt aega harjutamiseks, et sama hästi nagu on lahendatud ühte ülesannet, lahendataks ka teist ülesannet.

Samuti kasutatakse tõsiasja, et kui kaks erinevat gruppi lahendavad sama ülesannet, siis näitab tulemuste vahe seda, kui palju peaks nõrgem grupp veel keskmiselt õppima, et sooritada ülesannet sama hästi nagu tugevam grupp.

Kahest eelnevast tõsiasjast tuleneb järeldus, et kui kaks erinevat gruppi teevad kahte erinevat ülesannet, siis mõlemal juhul tulevad ülesannete raskuste sooritusprotsentide erinevused samad, seda siis kui tegemist on intervallskaalaga.

See annab meile võimaluse kasutada erinevate ülesannete paaride raskuste omavaheliseks võrdlemiseks ülesannete kolmikutes erinevaid valimeid.

Eelnevat kokku võttes muutub meile arusaadavaks psühhomeetrias kasutatav tõsiasi, et õpilaste võimekust ja ülesannete raskust mõõdetakse samal skaalal. Ülesande raskus on võrdne õpilase võimekusega, kes keskmiselt lahendab selle ülesande ära 50% tõenäosusega. Õpilase võimekus on võrdne ülesande raskusega, mille õpilane lahendab ära 50% tõenäosusega.

Lisame veel tähelepaneku, et hajuvusskaalal, kus standardhälve on ühikuks, ei sõltu ülesannete raskus hajuvusest, sest 66% õpilastest jääb $\pm$ standardhälbe vahemikku Gaussi jaotuse definitsiooni põhjal.

Intervallskaala ääreefektide mõju vähendamine

Vahed jäävad erinevate gruppide puhul samaks, kui segama ei hakka ääreefektid, mis rikuvad intervallskaalat. Nt saavutavad lihtsa ülesande puhul nii nõrgem kui ka tugevam grupp mõlemad maksimaalse tulemuse, kuid sellest ei järeldu, et grupid oleksid võrdsed. Teine võimalus on, et tugevad õpilased sooritavad kõik ülesanded võrdsele, st maksimaalsele tulemusele. Kuid sellest ei järeldu, et need ülesanded oleksid võrdselt rasked. Seega tuleks intervallskaalat rikkuvate õpilaste tulemused kõrvale jätta, eriti siis, kui võrreldakse kitsast ja laia eksamit. Kahe laia eksami võrdlemisel see ei mõju nii tugevalt. Pigem mõjutab tulemuse täpsust negatiivselt see, kui mingi hulk õpilaste tulemusi on välja arvatud.

Eeltesti arvutamisel kasutatigi kolme erinevat õpilaste valimit. Esimeses versioonis võeti arvesse kõik õpilased. Teises versioonis jäeti välja need õpilased, kes mõnda ülesannet üldse ei proovinudki lahendada. Kolmandas versioonis jäeti välja õpilased, kes saavutasid nii kitsa kui ka laia eksami ülesande puhul maksimaalse tulemuse. Kordame veel, et kolmandat versiooni on vaja ainult kitsa ja laia eksami omavaheliseks võrdlemiseks, mitte aga kahe aasta laia eksami tulemuste võrdlemiseks, sest seda maksimumtulemuste paarid ei moonuta. Samas on ka üldkogumis õpilasi, kes vastavad kõikidele ülesannetele õigesti ning nende protsent peaks ideaaljuhul olema valimis sama kui üldkogumis. Siis oleks skaala rikutusest tulenev moonutus valimi ja üldkogumi puhul sama. Analoogilise põhjendusega võib teatud arvutustes välja jätta õpilased, kes said kõikide ülesannete puhul 0 punkti.

Täpsus klassikalises testiteoorias

Täpsuse hindamiseks tuleb arvestada määratlust, et testiteoreetiliselt koosneb tulemus täpsest skoorist ja veast. Kui me hakkame üksikülesannetele põhinevalt kogu testi tulemust arvutama, siis täpsed skoorid liituvad aritmeetiliselt, vead aga geomeetriliselt. Teiste sõnadega - vea suhe täpsesse skoori liidetavate arvu kasvades väheneb.

Standardviga võrdub standardhälve (ca 20 punkti riigieksamite skaalal) jagatud ruutjuurega testitud õpilaste arvust (ca 18), mis on natuke üle 1 punkti, mis on meie eeltestimise täpsus ühe sigma usaldusnivool. Kuna eeltestis osalenud koolide kalendrist tulenevatel põhjustel jäi sellel aastal 10% ülesandeid (s-o 1 ülesanne) nimetatud moel testimata, siis lisab see veel täiendavalt umbes 10% vea.

Testitavate triplettidega tekkinud probleemid.

Praktilistel kaalutlustel anti igale eeltestis osalevale grupile lahendada kolm ülesannet: üks eelmise aasta laia eksami ülesanne, üks selle aasta kitsa eksami ülesanne ja üks selle aasta laia eksami ülesanne. Gruppe oli sama palju nagu igal aastal ülesandeid. Sellega tuvastati eelmise aasta laia ja selle aasta laia ülesannete raskuste vahe sooritusprotsentides. Tuvastati ka selle aasta laia ja kitsa eksami raskuste vahe sooritusprotsentides. Kõikide gruppide vastavate sooritusprotsentide vahede kaalutud keskmine (kaaluks ülesande maksimumpunktide arv) andis kogu eksami sooritusprotsentide vahed.

Täpsema tulemuse saamiseks oleks pidanud triplettide järjekorda roteerima, sest tavaliselt lahendatakse esimest ülesannet paremini kui teist või kolmandat, sest inimesed väsivad.

Võrreldes selle aasta laia kursuse matemaatikaga hinnatakse sellest tulenevalt kõige rohkem üleeelmise aasta matemaatika laia kursuse eksami tulemust ning veidi vähem selle aasta kitsa matemaatika eksami tulemust. Selle tõttu kahaneb veidi arvutatav kahe laia eksami vahe ning natuke ka selle aasta kitsa ja laia kursuse eksamite tulemuste vahe. Kokkuvõttes - mõlemad vahed võivad tegelikkuses olla veidi suuremad ning tegemist on justkui alumise piiri hinnanguga. Samas eeltestimise protsessi keerukamaks muutmisega võivad tekkida omakorda vead, mis on põhjustatud protseduuri keerukusest.

Mida me eeltestimise tulemusena teada saame

Eeltestiga saame ennustada, et kui eelmise aasta laia eksami kontingent oleks teinud selle aasta laia kursuse eksamit, mitu punkti nad oleksid siis saanud. Reaalne eksamil ilmnenud tulemus ilmestab siis võrrelduna selle prognoosiga, kui palju oli selle aasta eksamit teinud õpilased tugevamad või nõrgemad eelmise aasta õpilastest. See on see, mida muidu omistatakse tavaliselt kahe järjestikuse aasta eksami tulemuste võrdlusele, kui eeldatakse eksamite raskuse täielikku püsivust. Varasemad arvutused on näidanud, et reeglina kõigub eksami raskus märgatavalt vähem kui kontingentide tugevused ning võrdsete eksamite eeldus ei ole oluliselt rikutud.

Kuna nii kitsa kui ka laia eksami ülesannetes maksimumtulemusi saavutanud õpilased ei riku kahe laia eksami tulemuste võrdlust, siis saame väita järgmist.

- Kasutati esimest valimit ja saadi selle aasta laia eksami prognoosiks 0,54.
- Eelmisel aastal oli laia eksami tulemus 0,57. Seega oli sellel aastal eksam keskmiselt 3 punkti (täpsemates arvutustes 2,65) raskem. Märgime, et see on alampiir, tegelik vahe on väiksem, kuna eeltestimisel ei roteeritud ülesandeid.
- Tegelik selle aasta laia eksami tulemus ei olnud aga 0,54, vaid hoopis 0,52 ning seega oli sellel aastal laia kontingent kahe punkti võrra nõrgem (õpilased tegid eksamit 2 punkti nõrgemini kui eelmisel aastal). Märgime, et see on nüüd pigem ülempiir ja eeltestis ülesannete roteerimisel oleks see vahe vähenenud.
- Eelmise aasta laia kontingent oleks teinud selle aasta kitsast eksamit 66 punktile (arvutused valimi 3 alusel).
- Selle aasta lai kontingent oleks teinud kitsast eksamit isegi 64 punktile (olid 2 punkti nõrgemad eelneva põhjal).
- Kahe järjestikuse aasta kitsad kontingendid olid eeltestis otseselt sidumata.

Kui me täiendavalt arvestame, et eksamil avaldunud kitsa ja laia eksami tulemuste vahele (nende igaühe omal skaalal) juurde liites kitsa ja laia eksami skaalade omavahelise nihke, saame kitsast ja laia eksamit teinud kontingentide reaalse erinevuse. Teiste sõnadega viime nad ühisele skaalale.

- Selle aasta kitsa ja laia eksami skaalade nihe eeltesti põhjal oli 13,1 punkti (alampiir).
- Keskmiste tulemuste reaalne vahe 2017. aasta testides (lai – kitsas) oli $51,85 - 38,43 = 13,42$
- Kitsa ja laia eksami keskmiste vahe ja skaalade nihke summa annab kitsa ja laia kontingendi võimekusvahe, mis on hinnanguliselt 27 punkti.

Seega, eelmisel aastal täheldatud väiksem kitsa ja laia kursuse eksamite kontingentide tugevuste vahe ei tulenenud ainult etteennustatavate kitsa kursuse ankrute drillimisest laia kursuse eksamit teinud õpilaste poolt (sellest tõenäoliselt ka), vaid reaalsest kontingentide tugevuste erinevuse vähenemisest. Poliitiliselt on see muidugi hea uudis, et kitsast kursust õppinud õpilased liiguvad hoogsalt laia kursuse eksamit teinud õpilaste võimekusele järele.

Toome ära kontingentide tugevuste erinevuse muutuse ajas.

- 2017. aasta eeltestiga arvutatud skaalade nihke ja eksamis selgunud keskmiste tulemuste kaudu määratud kontingentide erinevus oli 27 punkti.
- 2016. aasta ühisosa põhjal oli kontingentide erinevus 24,73 punkti.
- 2015. aasta ühisosa põhjal oli kontingentide erinevus 31,13 punkti.
- 2014. aasta ühisosa põhjal oli kontingentide erinevus 32,88 punkti.
- Hüppeline muutus kontingentide erinevuses toimus kolmandal aastal kitsast eksamit tegeva kontingendi arvelt ühisosa järgi sooritusprotsendilt 30% 37,5% peale, kusjuures laia kursuse eksamit valinud õpilaste sooritusprotsent ühisosa järgi säilis 62% kandis.

Järeldused

- Kui laia eksami sooritajate kontingendi tugevus on langenud 2 punkti, siis peaks kitsa eksami kontingent olema langenud umbes sama palju ehk

hinnanguliselt umbes 4 punkti kandis, et kompenseerida veel lisaks kontingentide vahe erinevuse 2 punktist kasvu. Seega, 2016. aasta kitsa eksami tulemuse 7 punktisest tõusust jäi selleks aastaks alles ainult 3.

Eelneva informatsiooni põhjal saame ligikaudselt hinnata kitsa eksami enda raskuse kõikumist.

- Kitsa eksami tulemus 2017. aastal oli 38,43.
- Kitsa eksami tulemus 2016. aastal oli 39,9.
- Nähtav tulemus langes 1,5 punkti ning seega pidi kitsa eksami töö olema veel 2,5 punkti raskem, et 4 punkti nõrgem kontingent võiks näiliselt ainult 1,5 punkti langeda. See on parim hinnang, mida me eeltesti ja päris testi tulemuste abil suudame välja pakkuda. Samas tõdeme, et kitsas eksam oli nüüd tervikuna mõeldud vaid kitsast kursust õppinud õpilastele ning oleks võinud seega tõusta. Languse põhjus oli ilmselt formaadi muutus, sest tavapärase formaadi muutus tekitab reeglina tulemuste languse, mis mõne aasta jooksul taastub.

Mida me eksami tulemusena veel teada saime

Selle (2017) aasta matemaatika laiale kursusele vastava eksami histogrammis on näha, et 0 - 20 punkti saavate õpilaste arv on kasvanud. See võib olla ka peamine eksami raskuse 3 punktise muutuse põhjus, et kadus ära võimalus kitsa kursuse ettearvatavate ülesannete eest punkte saada. Nagu tõdesime eespool, kuni viie punktist eksami raskuse muutust peetakse aktsepteeritavaks. Arvestades muutuste suurust eksami formaadis on 3 punktine muutus pigem hea saavutus.

Laia ja kitsa eksami skaalade omavaheline nihe on liikunud 16 punkti pealt 13 peale, kuid 13 on alampiir. Kui triplete tsükleerida, võib tegelik tulla 16-le lähemale.

Kontingentide vahe on liikunud juba 2016. aastast 32 punkti asemel 27 punkti kanti, põhjuseks õpilaskonna formaadiga kohanemine. Tollasele 7 punktisele kasvule on järgnenud selle aasta hinnanguliselt 4 punktine langus, sest muutus formaat.

Kitsa eksami histogrammi on tekkinud 100 punktised tööd, ehkki neid pole palju. Nendel õpilastel oleks olnud potentsiaalselt võimeid psühhomeetria üldtuntud põhimõtetest tulenevalt ka laia kursuse eksamiks valmistudes üle 80 punkti saada. Kuid nad võivad olla laia kursust üldse mitte õppinud ning neil võivad olla teised väljakutsed, millele pühenduda. Vähemalt eksami valmise hetke seisuga.

Kokkuvõttes teeb ainult rõõmu, et uus ühisosata ja eeltestimisega eksamiformaadi köögipool paistab hetkel sama ilus ja kõnekas kui statistikaaruannete numbrid. Kindlasti on uue eksamiformaadi puhul tulemused täpsemad, sest eksami raskus on ulatuslikumalt vastav testitavate võimekusele.