

# **TARTU ÜLIKOO**

**Haridus- ja Teadusministeeriumi ning Tartu Ülikooli koostöölepingu  
nr 14 – 11/740 aruanne**

## **RIIGIEKSAMITULEMUSTE VÕRRELDAVUSE LOOMINE**

**Aire Raidvee (Tartu Ülikool)  
Anti Teepere (Riiklik Eksami ja Kvalifikatsioonikeskus)  
Olev Must (Tartu Ülikool)**

**Tartu 2008**

## Sisukord

<b>1. Sissejuhatus.....</b>	<b>3</b>
<b>2. Riigieksamitele seatavad eesmärgid .....</b>	<b>3</b>
<b>2.1. Konkreetseid uurimisküsimused.....</b>	<b>3</b>
<b>3. Riigieksamitele seatavate eesmärkide täidetud käesoleval ajal.....</b>	<b>4</b>
<b>3.1. Kas ja millised on hindamissüsteemide ühispõhimõtted erinevates ainetes? .....</b>	<b>4</b>
<b>3.2. Kas riigieksamihinnete palliomistus on ühetähenduslik? .....</b>	<b>5</b>
<b>3.3. Kas ühe aine piires eri aegadel saadud tulemused on samatähenduslikud? .....</b>	<b>7</b>
<b>3.4. Kas on võimalik hindamisel kasutatavad arvud seostada kvalitatiivsete hinnangutega, fikseerimaks hinnete samatähenduslikkust?.....</b>	<b>8</b>
<b>3.5. Kas riigieksamite tulemused saavad olla üliõpilaskandidaatide usaldusväärse võrdluse aluseks?.....</b>	<b>8</b>
<b>4. Võimalikud lahendused .....</b>	<b>8</b>
<b>4.1. Lahendused, mis püüavad võrreldavust parandada toortulemusi jaotuspõhiselt teisendades. ....</b>	<b>9</b>
<b>4.2. Tulemuste sisulisele konsistentsusele suunatud lahendused .....</b>	<b>13</b>
<b>4.2.1 Ülevaade modernse testiteooria pakutavatest võimalustest.....</b>	<b>13</b>
<b>4.3. Eksamitöö kvaliteedikontrollile suunatud lahendused.....</b>	<b>14</b>
<b>4.4. Läbi viidud eksamitööde hindamisele suunatud lahendus.....</b>	<b>14</b>
<b>4.5. Erinevate riigieksamiainete tulemuste võrreldavus.....</b>	<b>15</b>
<b>4.6. Riigieksamite referentspopulatsioonile suunatud lahendus.....</b>	<b>15</b>
<b>4.6.1. Millised peaks olema kompetentsi jaotusparameetrid referentspopulatsioonis? .....</b>	<b>16</b>
<b>4.6.2. Kas tuleks kehtestada referentsaasta kõigi eksamiainete puhul?.....</b>	<b>17</b>
<b>5. Ettepanekud.....</b>	<b>17</b>
<b>Lisa 1.....</b>	<b>19</b>
<b>Lisa 2.....</b>	<b>22</b>

## 1. Sissejuhatus

Analoogsete uurimuste või mõõtmiste suur arv on muutnud aktuaalseks küsimuse erinevate mõõtmistulemuste omavahelisest võrreldavusest. Suured rahvusvahelised uurimisprojektid nagu TIMSS, PISA, Euroopa sotsiaaluuring jt ei oleks mõeldavad, kui nad ei käsitleks tulemuste võrreldavuse probleemi näiteks erinevate riikide vahel. Ka erinevatest ajaetappidest pärinevate ja esmapilgul samade andmete kasutamine tõstatab samuti küsimuse sellest, kas andmed ikkagi on võrreldavad.

Koolihindeid on alati käsitletud ka võrdluse eesmärgil. Kui koolihindeid kasutatakse valikuvahendina nt kõrgkooli vastuvõtul, siis ongi esiplaanil võrdlus erinevate hinnete vahel ja siin on võrdluse täpsus ülioluline.

## 2. Riigieksamatele seatavad eesmärgid

Seadusandlikult on Eesti gümnaasiumi riigieksamil kolm põhifunktsiooni: 1) hinnata ainekava läbitust, st õppetöö tulemlikkust; 2) toimida kõrgkoolide vastuvõtukriteeriumina; 3) **olla aluseks õppetöö tulemuslikkuse võrdlemisel.**

Seega on riigieksamatele seatavad eesmärgid järgmised:

1. sama aine eksamite tulemused peaksid erinevate aastate lõikes olema võrreldavad (s.t samatähenduslikud);
2. erinevates ainetes sooritatud eksamite tulemused peaksid olema palliarvestuse mõttes võrreldavad;
3. eksamite tulemused peaksid hindama ainekava läbituse taset;
4. eksamite tulemused peaksid eksaminande piisavalt hästi ja usaldusväärsetl kõrgkoolide jaoks eristama.

### 2.1. Konkreetsed uurimisküsimused

Käsitledes Eesti haridussüsteemis riigieksamate hinnete võrreldavust, peame vastama vähemalt järgmistele küsimustele:

- A. Kas ja millised on hindamissüsteemide ühispõhimõtted erinevates ainetes?**
- B. Kas erinevate riigieksamate tulemuste palliomistused on samatähenduslikud?**
- C. Kas ühe aine piires eri aegadel saadud tulemused on samatähenduslikud? Kas võib ühe eksamiaine piires välja joonistada keskmiste hinnete ajalise trendi, teades, et võrdleme analoogseid, samatähenduslikke tulemusi?**
- D. Kas on võimalik hindamisel kasutatavad arvud seostada kvalitatiivsete hinnangutega, fikseerimaks hinnete samatähenduslikkust, (hinde arvväärus võib olla sama, aga kas tema tähendus on sama)?**
- E. Kas riigieksamate tulemused saavad olla üliõpilaskandidaatide usaldusväärse võrdluse aluseks, (kas eristavad eksaminande piisavalt hästi kõrgkoolide vastuvõtukriteeriumidena)?**

### 3. Riigieksamatele seatavate eesmärkide täidetud käesoleval ajal

#### 3.1. Kas ja millised on hindamissüsteemide ühispõhimõtted erinevates ainetes?

Riigieksamate hinde omistamist korraldab seadusandlikus plaanis haridusministri määrused nr 75 (24.12.2001) ja nr. 40 (16.11. 2006), mille kohaselt toimib 100-palli süsteem; positiivne eksamisooritus (alates 2003. aastast) on 20 palli. Hindamiskomisjonide pädevuses on otsustus, millise kaaluga on ühe või teise ülesande sooritus koondtulemuse moodustumisel. Palliomistuse ühtlustamiseks võivad hindamiskomisjonid läbi viia standardiseerimisprotseduure.

Kui lähtuda ülalnimetatud määrustest, siis tuleks hindamisel kasutada 100-t erinevat eristust/hindepalli, kuid määrus ei osuta täpselt sellele, milliseid arvväärtsi hindamistel kasutada, välja arvatud juhtum, kui tuvastatakse eksamieeskirjade rikkumine. Sel juhul osutatakse võimalusele omistada eksaminandile 0 palli. Konkreetsete ainete hindamiseeskirjad peavad aga võimalikuks 0 palli omistamise ka muudel juhtudel. Seega on 0-pallise hinde interpretatsioon ambivalentne, sest ta tähistab kas eeskirjade rikkumist või tulemust.

On ka ilmne, et positiivne soorituslävi 20 palli on ambivalentne ning formaalne, sest ta ei seo nimetatud tulemuse omistamist ei ainekava ega sisulise interpretatsiooniga sellest, mida antud nivoo ületaja peab konkreetsetelt teadma või oskama.

Erinevate ainete hindamissüsteemide kõige olulisemaks ühiseks tunnusjooneks on püüdlus kasutada 100 hindepalli süsteemi.

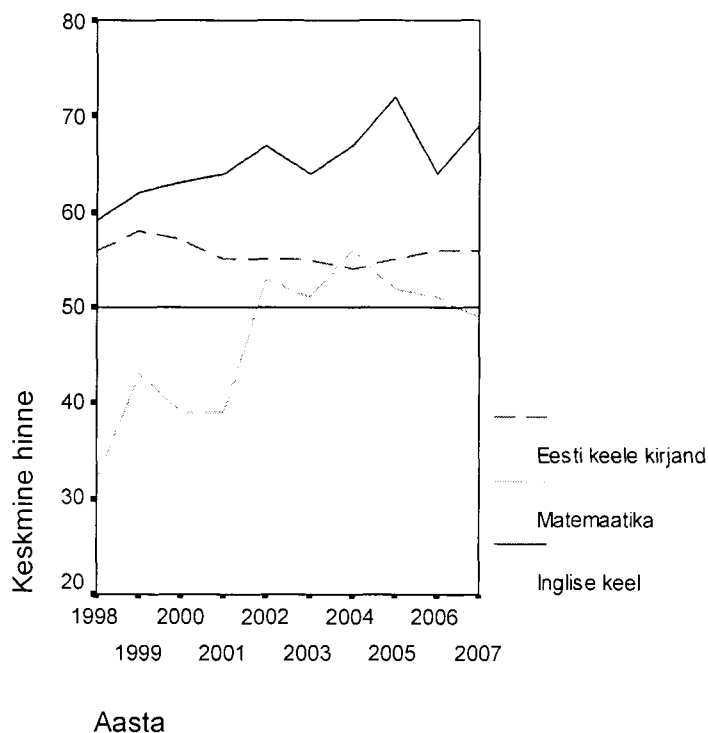
Kuigi kirjandi hindamisel kuni aastani 2005 kasutati sisuliselt 10-palli süsteemi ja kirjandi kui terviku hindamist, siis teiste ainete hindamistulemustega näiliku võrreldavuse saavutamiseks korrutati kirjandihinnet 10-ga. Selline lineaarteisendus ei muuda aga 10-t eristusastet 100-ks astmeks. Alates 2006. aastast rakendatakse kirjandi tulemuse hindamisel 20 eristusastet (neli erinevat aspekti, igaühe puhul 5 võimalikku tulemust). Lõpptulemuse arvutusvalem [ $aspekt1*9+aspekt2*7+aspekt3*3+aspekt4 = tulemus$  100-pallisel skaalal] ei muuda 20-t eristusastet 100-ks tasemeks.

Erinevate ainete hindamissüsteemide teiseks olulisemaks ühisosaks on hinde palliomistuses selge ja range mõõtmismeetrika puudumine. Põhiliselt lähtutakse ettekujutusest, et mingi ainekava osa teadmist või mitteteadmist võib hinnata pallisüsteemis, omamata selget ja kontrollitavat analüütilist alust erinevate hindepallide samaväärsuse hindamiseks. Lähtutakse ettekujutusest, et üksikute eksamiülesannete täitmise eest antavate pallide summeerimine annab täpse ja võrreldava eksamihinde/tulemuse. Metodoloogilises mõttes toimub üksikülesannete soorituste eest omistatud astakute summeerimine oletusega, et nii moodustub võrdsetele intervallidele rajanev koondskaala.

Hindamine 100-pallisel skaalal loob põhimõtteliselt suhteliselt suured hinnete diferentseerimise võimalused, kuid samas suurendab ka vastutust hinnete võrreldavuse tagamisel. REKKi tasandil on hindamispõhimõtted (hindamiste eristuskirjad, hindamisjuhendid) ainekesksed ning nad ei osuta sellele, kuidas mingi aine tulemust kõrvutada teistega (s.t koondtulemuse väärtustega pole seotud kindlaid kvalitatiivseid-sisulisi määratlusi) peale 3 ülalnimetatud põhimõtte (100-palli skaala, positiivne sooritus alates 20 pallist; 0 palli omistamine eeskirjade rikkumisel või palliarvestuses null-tulemuse

saavutamisel). Siit tuleneb, et erinevate ainete hindamiste ühtlustamise detailsemad põhimõtted praktiliselt puuduvad. Ühtlustatud põhimõtete puudumisest võib tuleneda ka eri ainete keskmiste eksamitulemuste suur varieeruvus. Selgusetu on näiteks see, kas tulemus 70 inglise keeles 2005. aastal on midagi sisuliselt sama või erinevat matemaatika 40 pallililise tulemustest 2001. aastal. Formaalselt on nende arvude kõrvutamine põhjendatud sellega, et mõlemas aines antud aastatel olid need tulemused väga lähedased vastavatele keskmistele sooritustele.

Joonisel 1 on kujutatud kolme suurima sooritajate arvuga gümnaasiumi riigieksami aine keskmised tulemused. Kolme vaadeldava riigieksami hinded on aastate lõikes väga kõikumavad. Vaadeldes 3 aine keskmiste tulemuste dünaamikat, on selge, et erinevate aastate toortulemuste võrdlemise põhjendatus nt kõrgkooli vastuvõtul on äärmiselt kaheldav.



Joonis 1. Kolme gümnaasiumi riigieksamiaine keskmiste tulemuste dünaamika 1998 – 2007

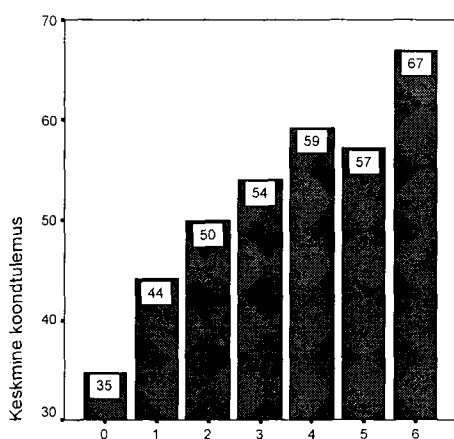
### 3.2. Kas riigieksamihinnete palliomistus on ühetähenduslik?

Hindamine oma olemuselt on mõõtmine. REKKi poolt läbiviidav riigieksamite tulemuste hindamine on olemuselt klassikalise testiteooria (KTT) rakendus. KTT-l rajanev mõõtmise ideoloogia tugineb seisukohale, et mingi omaduse mõõtmiseks on loodud kindlaid mõõtetalone kasutatav mõõtmisinstrument (test, ülesannete kogum) ning et mõõtmise kõige olemuslikum sisu on uuritava objekti/omaduse võrdlemine selle mõõtmisinstrumentiga. Mõõtmise olemuseks on sisuliselt järelalusprotsess, mis tuleneb konstantsete mõõtmisetalonide rakendamisest mõõdetavale objektile. Esmapilgul võiks arvata, et riigieksamite hindamisel mõõtetalonideks on pall, mis omistatakse mingi elementaarõppimise või -õskuse eest.

Riigieksameid reguleerivad seadusandlikud aktid ei selgita, mida tähendab pall või mille eest võib eksamikomisjon palli omistada, st selgusetu on, kuidas tuleb mõõtetaloni hindamisel

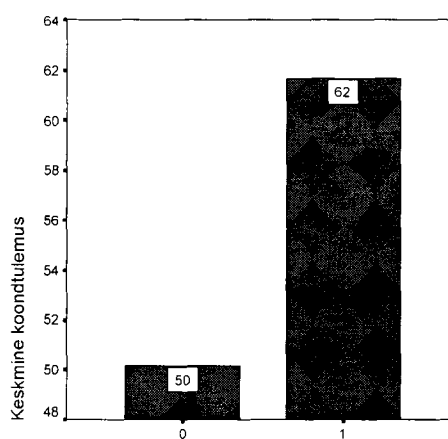
rakendada. Vastavates määrustes osutatakse, et riigieksamiitoid hinnatakse „100-palli süsteemis”, ent avamata jäetakse see, mida taoline süsteem tähendab. Eelkõige jääb lahtiseks küsimus sellest, mis on hindamissüsteemi etalonmõõtühiku – „palli” – tähendus, sisu. Hindamisjuhistest ja eristuskirjadest selgub, palliga tähistatakse ainekavas ette nähtud mingi konkreetse teadmise või oskuse esitamist ning nõ elementaar-pallide kaalumise ja summeerimine annab lõppkokkuvõtteks riigieksami tulemuse. Mõningates ainetes (näit matemaatika) võib üks tervikülesanne koosneda kuni 20 elementaar-palli eristusest. Oluline on rõhutada, et taoline elementaar-palli omistus on oma olemuselt eksamikomisjoni hinnang eksamiülesande tähtsusele/olulisusele. Kui eksamikomisjoni palli ekspert-hinnangut võrrelda sellega, kui raske on eksamineeritavate jaoks ühe või teise palli saamine, siis selgub, et erinevused on küllaltki suured. Joonistel 2 ja 3 on kujutatud kahe erineva 2005. a. geograafiaeksami ülesande palliomistust ning joonisel 5 ühe 2007. a. matemaatikaeksami ülesande palliomistust lähtuvalt sellest, milline oli palli saanud eksamineeritava koondtulemus. (Ühtlustamise mõttes on geograafia eksami palliomistus esitatud täispallides.) Vaatleme kõigepealt esimese palli raskust. Jooniselt 2 selgub, et palli 1 said eksamineeritavad, kelle üldtulemus oli 44; joonisel 3 kujutatud esimese (ja ainsa) palli saajate keskmine tulemus oli 66 ja joonisel 4 kujutatud ülesande esimese palli saajate keskmine tulemus oli 30. Toodud numbrite kõrvutamisel selgub, et kõige raskem oli saada 1 pall 14-nda ülesande eest, kõige kergem aga 5-nda ülesande eest. Kui vaadelda antud 1 palli raskusena keskmist üldtulemust, siis peame kõrvutama arve 66 ja 30. Järelikult on hindepallide raskuse kaalud väga erinevad.

Järgmine kõrvutus: 21. ülesande eest oli võimalik saada maksimaalselt 6 palli, taolise punkti arvu saajate keskmine üldtulemus oli 67. Kui kõrvutame arvu 67 teise vaadeldava ülesande 1 palli saamise raskusastmega (62), siis leiame et arvud on väga lähedased. Ehk teisisõnu – esimeses ülesandes kuue palli saajad olid samade üldtulemustega kui teises ülesandes 1 palli saajad. Kolmanda näiteülesande korral vastab samale raskusele pallide 7...9 saamine. On selge, et riigieksami palliomistuste eeskirjad ei taga pallide võrdsust ega võrreldavust eksamineeritavate seisukohast. Hindamise elementaarühik - pall - ei ole ühetähenduslik. Õigete vastuste või nende alusel antavate hindepallide summeerimine ei taga automaatselt summaarsete tulemuste ühetähenduslikkust ja võrreldavust.



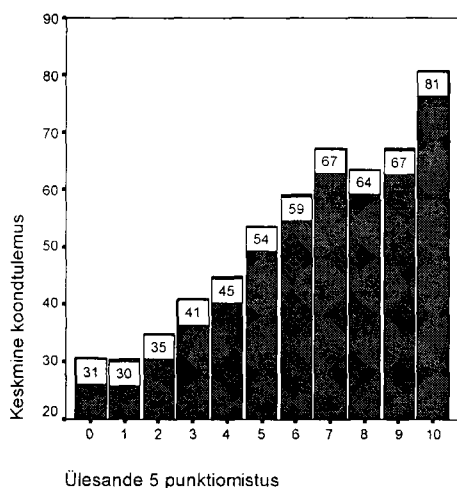
Ülesande 21 punktiomistus

Joonis 2. Ülesande 21 palliomistus.



Ülesande 14 punktiomistus

Joonis 3. Ülesande 14 palliomistus.



**Joonis 4.** Ülesande 5 palliomistus.

Näitlikustatud vastuolulised tulemused tulenevad sellest, et riigieksamite hindamis-eeskirjade metodoloogiline alus on küsitava väärtusega. Eeskirjad lähtuvad sellest, et on olemas mingid „objektiivsed”, ekspertide poolt mõistetavad ühtsed mõiste „pall” määratlused/suurused, ent mis eksamineeritavate jaoks oma raskusastmelt võivad oluliselt erineda.

### 3.3. Kas ühe aine piires eri aegadel saadud tulemused on samatähenduslikud?

Hinnete samatähenduslikkuse hindamise lähtepunktiks on hindamisjuhised. Eksamitööid käsitlevad materjalid (hindamisjuhised, eristuskirjad) toonitavad, et eksamitööd lähtuvad riiklikust õppekavast, kuid samas on erinevatel aastatel eksamite korralduses tehtud muutusi, mis võivad olla oluliselt mõjutanud eksamite tulemusi. Vastavate tulemuste samasust kontrollivaid uurimusi aga pole. Näiteks on muudetud:

1. eksamitöö sooritamise aega (matemaatikas on teise osa sooritamise aeg olnud tüüpiliselt 120 minutit, aga on ka 150 minutit aastail 2003 ja 2008);
2. eksamitööde ülesannete arvu ja üksikülesannete kaalu (eksamiülesannete arv ja nende eest omistatav maksimaalne pallide arv on muutunud matemaatika riigieksamil; vrld 2007. a. ja 2008.a. palliomistusi varasemate aastatega);
3. kogu eksami korraldust (inglise keele katseline riigieksam 2003);
4. kogu hindamissüsteemi (kirjandi hindamisel tervikhindamiselt üleminek komponentide kaalutud ja sõltumatule hindamisele; 2006. ja järgnevatel aastatel eksamid võrdluses varasematega);
5. eksami sooritusläve (2003. aastal kehtestati 20-palline soorituslävi, mida varasemalt polnud).

Kuivõrd eelpool osutatud muudatused puudutavad eksameid, mida sooritab kõige suurem arv isikuid, siis on selge, et ka vähemalt näidetes osutatud muudatused omavad mõjusid väga paljudele hinnete ja hindedaajatele. Sisuliselt on tegemist muutustega, mis muudavad ainesiseste hindamiste tulemused võrreldamatuteks, sest eksamiülesannete või –töö sisu ja hindamise olemus on muutunud.

### **3.4. Kas on võimalik hindamisel kasutatavad arvud seostada kvalitatiivsete hinnangutega, fikseerimaks hinnete samatähenduslikkust?**

Praegu kasutatav hindamissüsteem seostab ühese kvalitatiivse hinnanguga vaid 100-pallise hinde-skaala ühe punkti, so 20-pallise tulemuse, millest alates on ületatud sooritatuse piir. Samas pole enamiku õppeainete (v.a võõrkeeled) puhul määratletud, millist sisulist oskuste taset teatud tulemus peegeldab. Haridusalastes võrdlusuuringutes (TIMSS, PISA) on püütud tulemuste parema tõlgenduse huvides mõningad hineskaala väärtused seostada kvalitatiivsete määratlustega selle kohta, mida eksamineeritav peaks teadma või oskama, et saavutada mingi konkreetse tulemus.

### **3.5. Kas riigieksamite tulemused saavad olla üliõpilaskandidaatide usaldusväärse võrdluse aluseks?**

Mitmed kõrgkoolid on üle läinud lävendipõhisele vastuvõtule. Vastuvõtulävendite rakendamine on aga toonud kaasa mõningaid raskusi, mis on seotud sellega, et riigieksamite tulemusi kajastavaid numbrilisi näitajad on raske interpreteerida eelkõige sellest seisukohast, kui palju mingite hinnete saajaid ja potentsiaalseid üliõpilaskandidaate on realselt olemas. Üldreeglina on vastuvõtulävendid kehtestatud gümnaasiumi riigieksamite [toorskooriliste] tulemuste alusel. On ilmnenu, et eksamite tulemuste jaotusstatistikud varieeruvad oluliselt nii ainete kui aastate lõikes. Antud asjaolu osutab võimalusele, et [ka võrdse kompetentsitaseme puhul] on teatud aastatel/ainetes gümnaasiumi riigieksameid sooritanud isikuil kõrgkooli astumisel eelis teiste eksamineeritute ees. Näiteks on ebavõrdses olukorras tudengikandidaadid, kes on sooritanud eksamid erinevates võõrkeeletes, millede osakaalud sisseastumiskonkursil on võrdsed, kuid hinnete jaotuskarakteristikud eksaminandide populatsioonides erinevad. Konkreetselt võib välja tuua tulemused aastast 2005, kus vene keele võõrkeelena keskmine hinne oli 78,22, inglise keeles aga 71,93. Ilmselt olid kõrgkoolidesse vastuvõtul, (kus vastuvõtukriteeriumina kasutati võõrkeele hinnet), eelisolukorras need üliõpilaskandidaadid, kes olid sooritanud võõrkeele eksami vene keeles. Ilmselt raskused vastuvõtuotsustuste kujundamisel ajendasid prof Ene Tiidu ettepanekut lineaarselt standardiseerida riigieksamite tulemused.

#### **Järeldus:**

Riigieksamite hinnete adekvaatse võrdluse ja interpreteerimise kõige suurem raskus on selles, et puudub seletus, mida üks või teine hinne tähendab. Defineerimata on ühikud, mida riigieksamite hindamissüsteem kasutab, ja puudub teoreetiline selgus, kuidas peaksid erinevate ainete ja aastate riigieksamite hinded omavahel olema võrreldavad ning milline peaks olema hinnete jaotumise alus.

Riigieksamid praegusel kujul ei vasta seadusandluses sätestatud eesmärkidele tulemuste võrreldavusest.

## **4. Võimalikud lahendused**

**Lahenduse üldsuund on** hindamissüsteemi korrastamine ning ühtlustamine, mis seob üheks tervikuks nii ainekava koostamise, selle sooritustasandite kvalitatiivse määratlemise ja konkreetsete eksamiülesannete konstrueerimise kui ka eksamitöödele hinnete omistamise eeskirja. See on kahtlemata pikk ja aeganõudev protsess. Seetõttu võib lahendust käsitleda ka samm-sammulisena, alustades kõigepealt esmastest, kiiremini teostatavatest muudatustest, mis levendaksid riigieksamite tulemuste mittevõrreldavust. Järgnevas käsitletaksegi lahendusi lihtsamatest süsteemsemateni, vaagides võimalike variantide eeliseid ja puudusi.

#### 4.1. Lahendused, mis püüavad võrreldavust parandada toortulemusi jaotuspõhiselt teisendades.

Mõningaid lahendusi riigieksamite hinnete suurema võrreldavuse saavutamiseks on toodud tabelis 1.

**Tabel 1.** Mõningaid võimalusi suurendada gümnaasiumi riigieksamite tulemuste võrreldavust

Muutuse sisu	Muutuse matemaatiline olemus	Positiivsed efektid	Negatiivsed efektid
Hinnete keskmiste võrdsustamine (näit kõik keskmised = 50).	Lineaarteisendus: hinded teisendatakse konstandi lisamisega kujule, et keskmine hinne oleks kõikides ainetes võrdne (näit 50).	Sarnased keskmised tulemused. Näilik vastavus soovitava eksamite raskusvõrdsusele.	1. Mõõtiühik puudub endiselt. 2. Eksameid ei pruugi sooritada võrreldavad õpilaskontingendid ja seetõttu tulemus ebaõige. 3. Hindeskaala ei pruugi mahtuda 1 – 100 piiresse. 4. Säiluvad erinevused tulemuste hajuvuses.
Hinnete keskmiste ja hajuvusnäitajate võrdsustamine (näit kõik keskmised = 50 ja standardhälbed = 16,67).	Lineaarteisendus: hinded teisendatakse hajuvusühikutele ja jaotusele antakse soovitav kuju.	Sarnased keskmised, sarnased hajuvusnäitajad.	1.-3. Vt eelmine 1. - 3. 4. Säiluvad erinevused jaotuste asümmeetrias; hinnete hajuvused on vaid näivalt võrdsed.
Empiiriliste jaotuste normaliseerimine.	Mittelineaarne teisendus: empiirilised andmed asendatakse normaaljaotuslikega eesmärgiga saavutada soovitud jaotusparameetrid (näit keskmine = 50; Sd = 16.67; normaaljaotus)	Sarnased keskmised, sarnased hajuvusnäitajad, jaotuste sümmeetrilisus, ühtlustatud (hajuvus)ühik (SD =16.67)	1. Empiirilise jaotuse asümmeetria kõrvaldamisega jäetakse kõrvale võimalus, et valikulise eksami on sooritanud paremad õpilased. 2. Hindeskaala ei pruugi mahtuda 1-100 piiridesse.
Hinde esitamine protsentskaalal.	Monotoonne teisendus: hindepalli väärtuse asemel kasutatakse tema kumulatiivset protsentiili.	1. Väga selge tõlgendus; ühik = protsentiil. 2. Kergesti rakendatav kõrgkooli vastuvõtu otsustustes. 3. Erinevate ainete hinnete kirjeldavad statistikud on suhteliselt konstantsed. 4. Protsentiilhinne on esialgse hindega otseses funktsionaalses seoses. 5. Hindevahemiku 1 – 100 maksimaalne kasutamine.	Eelkõige äärmuslike väärtuste lähedal ei ole lähtehinde ja protsentiilhinne vastavus 1:1.
Normaalkõvera ekvivalendi (NKE) kasutamine	Hinne normaliseeritakse kujule: $21.06z + 50$ ja ümardatakse täisarvuks.	Intervalliskaala teke.	Äärmuslike väärtuste lähedal surub teisendus algväärtused kokku

Kõik esitatud lahendid on suuremal või väiksemal määral seotud normaaljaotusest tulenevate ideede ja ettekujutustega. Juba üksnes standardhälbe kui olulise hindekoore iseloomustava statistiku kasutamine osutab normaaljaotusele. Normaaljaotusliku muutuja standardhälve geomeetriselises interpretatsioonis osutab normaalkõvera käänupunktile, ehk kohale, mis eristab erinevaid muutuste (sageduste) astmeid. Mida enam käsitletava muutuja jaotus erineb normaaljaotuslikust, seda enam kaotab standardhälve oma sisulist mõtet. Kõikide ainete suhtes ühtne lineaarse transformatsiooni eeskiri – keskmine hinne 50 ja standardhälve 16,67 – peab samuti silmas olukorda, kus lähtetulemused on suhteliselt normaaljaotuse lähedased. Standardhälve väärtus 16,67 tuleb sellest, et kui hinnete vahemikud 0-st kuni 50-ni ja 50-st 100-ni on 50 palli ja kui tahta need 50 palli jaotada nii, et enam kui 99% hindeid mahuksid vahemikku 0 - 100, siis hinnete standardhälve peaks olema  $50:3 = 16,67$ . (Keskmisest  $\pm 3$

standardhälvet haarab 99,7% juhtumitest normaaljaotusliku muutuja korral.) Kui rakendada taolist lineaarse transformeerimise (**standardiseerimise**) eeskirja 2007. a gümnaasiumi füüsika riigieksami tulemustele, siis mitte ühelgi juhul ei tekiks hinnet 100 ja paljudel juhtudel peaksid hinded olema miinusmärgiga. Füüsika riigieksami maksimaalne hinne oleks vaid 75, st keskmisest vaid 1,5 ( $25/16,67 = 1,5$ ) standardhälbe kaugusel. Mitte-ootuspärased lineaarse standardiseerimise tulemused on põhjustatud sellest, kuivõrd lähteandmestiku jaotused erinevad normaaljaotuslikest ja kuivõrd soovitud keskmine tulemus erineb reaalselt olemasolevast keskmisest tulemusest (vt lisa 1). Hinnete lineaarteisendus eesmärgiga saavutada nende ühesugune keskmine ja hajuvus ei loo tegelikult praegusest paremat alust hinnete võrreldavusele.

Statistilised **normaliseerimisalgoritmid** (Blom, Van der Waerden jt), omistades ümber lähteväärtusi (piltlikult: venitades pikemaks või kokku surudes originaal-skaalaväärtusi, tagavad lahenduse, kus tõepoolest keskmise tulemuse 50 ja standardhälbe 16,67 korral vahemik 1...100 täitub, ent veelgi enam, vahemik 1...100 ületatakse, sest normaaljaotus peab silmas eelkõige pidevaid suursi, millel puuduvad ülemised või alumised piirid. Kui eksamisooritajaid on 10 000 ja nende tulemused peaksid jaotuma normaaljaotuslikult keskmisega 50, siis tõenäoliselt peaks nullist väiksemate või 100-st suuremate väärtustega hinnete saajaid olema umbes 30. Teisalt pole normaliseerimisalgoritmid diskreetsete ja arvult väikeste mitte-normaalsete algandmete teisendamiseks parimad, sest mõningate diskreetsete skaalapunktide vahele võivad tekkida kunstlikult suured vahemikud. Näide jällegi füüsikaeksami hinnetest: Blom'i meetodil normaalskooride maksimumtulemus oleks 96, ent sellele eelmine väärtus vaid 89. Samuti, tugevalt asümmeetrilise jaotusega eksamite puhul ning eeldusel, et eksaminandid olid kõik aine valdamises tugevad, teeksimine normaliseerimisega eksami tugevalt sooritanutele liiga, viies nende tulemused normaliseerimisteisendusega soovitavale ühtlustatud keskmisele tulemusele (50) lähemale. Normaliseerimisprotseduuri kõige suuremaks puuduseks on asjaolu, et ta muudab oluliselt lähteandmeid suhteliselt kunstlike eesmärkide saavutamiseks (normaaljaotus läbi väärtuste ümberskaleerimise).

**Tulemuste protsentiilskaala** keskendub hinnete järjestustele ja nende sagedustele mingil hindamisperioodil (aastal).

Protsentiilskaala rakendab riigieksami tulemustele teisendust, mille käigus igale toortulemusele seatakse vastavusse tema protsentiilastak. Terminoloogilise selguse huvides olgu märgitud, et

- protsentiilastak (ik *percentile rank*) on konkreetse elemendi väärtusega võrdsete või sellest väiksemate väärtustega elementide proportsioon kogu jaotusest, arvatatuna protsentides.
- protsentiil on kvantiil, mis jagab jaotuse 100-ks (elementide arvu mõttes) võrdseks osaks, (samas võib igast konkreetsest jaotuspunktist mõlemal pool olla võrdse väärtusega elemente), olles seega hajuvuskarakteristik ning riigieksami tulemusena mitterakendatav.

Tulemus protsentiilskaalal (e konkreetsele toortulemusele vastav protsentiilastak) on väljendatav järgmise valemiga:

$$\frac{cf_i}{N} \times 100\%$$

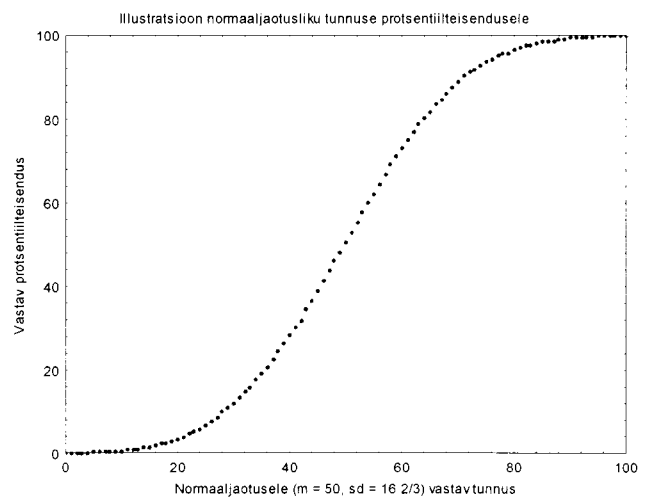
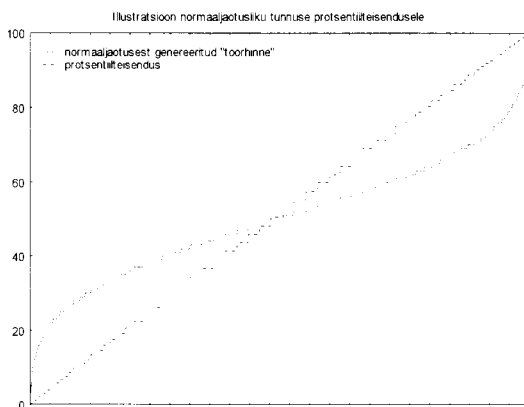
kus  $cf_i$  on konkreetse tulemusega võrdsete või sellest madalamate tulemuste hulk ning  $N$  on valimi suurus.

Näide. Tulemus 85 protsentiilskaalal ütleb, et antud tulemus on samaväärne või tugevam võrreldes 85%-ga kogu vaatlusalusest jaotusest, ning 15% on sellest tugevamad.

Antud lähenemise puhul vastab protsentiiltulemusele jaotuskõvera alune pindala kuni antud järjestuspunktini.

Protsentiilskaala puhul on oluliseks probleemiks see, et toortulemuste teisendamine protsentiilastakuteks ei ole eelkõige jaotuse äärmuslike väärtuste (s.t nii madalaimate kui kõrgeimate pallide) osas üks-ühene (vt joonis 4). Näiteks, eeldades, et eksaminande on ligi 10 000 või isegi vaid ca 5000, ning toortulemused (vahemikus 0...100) jaotuvad normaaljaotuse kohaselt (keskmisega 50 ja standardhällbega  $16\frac{2}{3}$ ), vastab kõigile toortulemustele alates 91...93-st pallist protsentiilhinne 100, kaotades seega täielikult eristuse kõigi eksamineeritute vahel, kes said palle vahemikus 91/93...100. Selline olukord kehtib teoreetilise normaaljaotuse puhul. Seniste riigieksamite tulemused ei ole üldjoontes väga täpselt normaaljaotusele vastanud. Samas, eeldusel, et reaalne andmestik vastaks normaaljaotusele, on oluline teadvustada, et kui valimi suurus on 10 000, siis (keskmise 50 ja standardhällbe  $16\frac{2}{3}$  puhul) jääks toortulemuse vahemikku 91...100 ca 50 isikut, mis on ~0,5% kogu valimist.

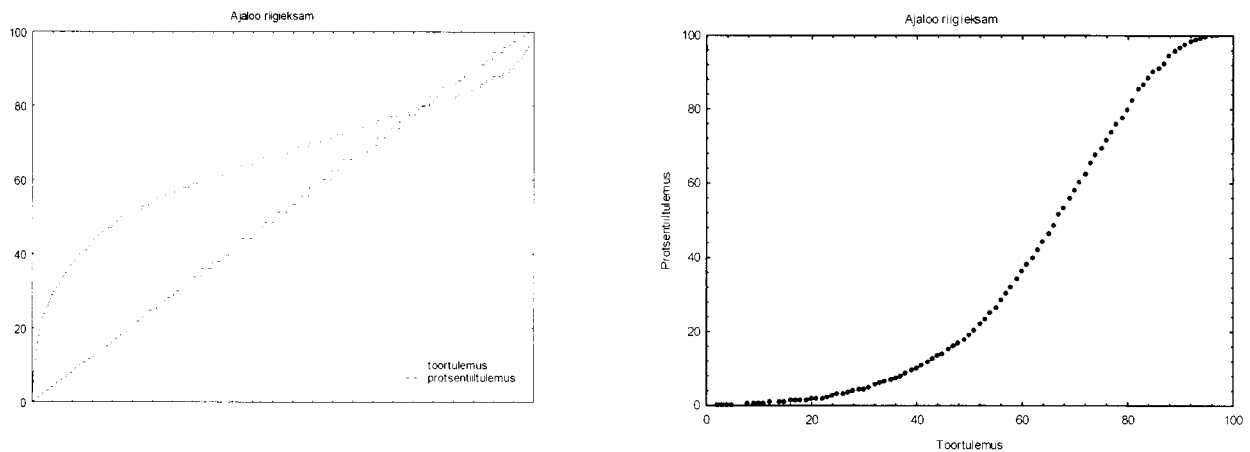
2007. a. andmetel põhinev väljavõte sellest, mil määral reaalsete riigieksamite toortulemuste viimine protsentiilskaalale toortulemusi kokku surub, on toodud lisa 2. Küsimus sellest, kas taoline kokkusurumine jääb mõõtmisvea piiresse, eeldab juba iga konkreetse eksamitöö eraldi analüüsi. See, kuivõrd taoline kokkusurumine kõrgkoolide vastuvõtupraktikas probleemiks osutub, sõltub konkreetsetest vastuvõtutingimustest; sellest kuidas jagunevad konkreetseid riigieksameid sooritanud üliõpilaskandidaadid erinevate kõrgkoolide ja erialade vahel.



#### Joonis 4 Normaaljaotusliku muutuja esitamine protsentiilskaalal

Joonise 4 vasakul paneelil on kujutatud normaaljaotusliku tunnuse (vahemikus 0...100, keskmisega 50 ja standardhällbega  $16\frac{2}{3}$ ) protsentiilteisendus, kus x-teljel on konkreetse tulemuse saanud individid, y-teljel aga tema toortulemus (sinine joon) ning sellele vastav protsentiilhinne (punane joon). Joonise 4 paremal paneelil on täpselt sama andmestik, kuid kujutatud selliselt, et x-teljel on algne toortulemus ning y-teljel sellele vastav protsentiilhinne.

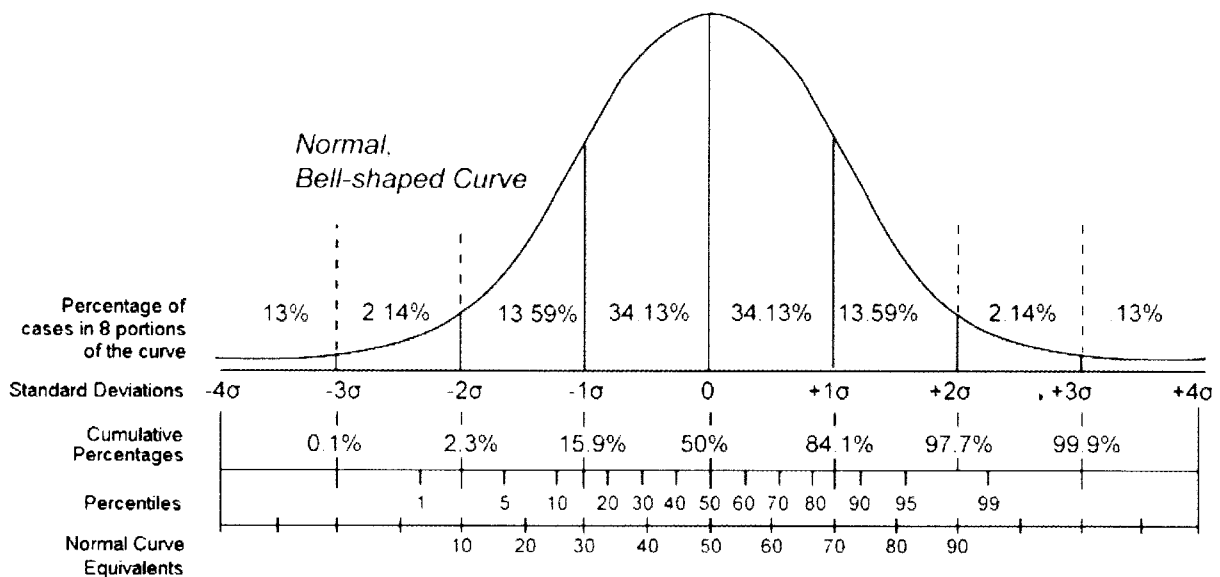
Konkreetne näide protsentiilhinde rakendusest ajaloo riigeksami tulemustele on kujutatud joonisel 5.



**Joonis 5. Ajaloo riigeksami (2007) tulemuste esitamine protsentiilskaalal.**

Joonise 5 vasakul paneelil x-teljel on kujutatud ajaloo riigeksami konkreetse tulemuse saanud individ, y-teljel aga tema toortulemus (sinine joon) ning sellele vastav protsentiilhinne (punane joon). Joonise 5 paremal paneelil on täpselt sama andmestik, kuid kujutatud selliselt, et x-teljel on algne toortulemus ning y-teljel sellele vastav protsentiilhinne.

**Normaalkõvera ekvivalent (NKE) (joon 6)** on transformatsioon, mida kasutatakse psühhomeetrias standardiseeritud testitulemuste parema interpreteeritavuse huvides. Normaalkõvera ekvivalent jagab tulemuste pingerea skaala vahemikus  $\pm 3$  standardhälvet 99-ks võrdsete intervallidega sagedusteks. Väärtusi, mis jäävad väljaspoole  $\pm 3$  standardhälvet, vaadeldakse kui ekstreemseid, mida saab vastavalt vajadusele käsitleda kui iseseisvaid kategooriaid või liita eelnevate väärtustega. Arvutuslikult on NKE normaaljaotuse lineaarteisendus kujule:  $NKE = 21,06z + 50$ .



**Joonis 6. Normaalkõver, protsentiilid, normaalkõvera ekvivalendid**

## 4.2. Tulemuste sisulisele konsistentsusele suunatud lahendused

Kui toorhinnete sisulise konsistentsusega on probleeme (dimensionaalsus pole kontrollitud, küsimustes esineb hälbeid mingi isikutunnuse osas, tulemuste jaotus on tugevalt asümmeetriline, palliomistuse seos mõõdetava oskuse tasemega on kontrollimata), suudavad ülalkirjeldatud transformatsioonid olukorda vaid "kosmeetiliselt" parandada, pakkumata sisulist lahendust, mis peitub riigieksamite süsteemi põhjalikus läbivaatamises.

### 4.2.1 Ülevaade modernse testiteooria pakutavatest võimalustest

Üheks vahendiks sisulisel eksamitööde kvaliteedikontrollil on erinevad modernse testiteooria (ik *item response theory*, edaspidises *IRT*) mudelid, mille kõigi ühendavaks alustalaks (ja fundamentaalseimaks erinevuseks võrreldes klassikalise testiteooriaga) on tõenäosuslik lähenemine isiku uuritava omaduse taseme ja tema konkreetsele küsimusele õigesti vastamise tõenäosuse seostamisel, mis vastandub klassikalise testiteooria põhinemisele valimist lähtuval kvantiilhinnangul. Sellest erinevusest lähtuvad *IRT* mitmed olulised eelised klassikalise testiteooria ees.

Peamised *IRT* mudelid põhinevad logilistel regressioonil, millega defineeritud seos on väga hea lähend paljude looduses toimuvate protsesside, ja nagu ilmneb, ka konkreetse isiku ja testiküsimuse interaktsiooni väljundi, kirjeldamiseks.

Tõenäosuslik lähenemine tagab mitmete *IRT* mudelite (täpsemalt, Rasch'i mudelite) raames spetsiifilise objektiivsuse – mõõtmistulemuste valimist sõltumatuse, s.t võimaluse hinnata testiküsimuste ja isikuparameetreid teineteisest sõltumatult, mis on üks peamine *IRT* eelis klassikalise testiteooria ees.

Teine peamine eelis seisneb selles, et kui klassikalises testiteoorias opereeritakse järjestusskaalal, on spetsiifilise objektiivsuse tagatuse korral (Rasch'i mudelite) puhul võimalik mõõtmiste tasemeks saavutada intervallskaala. Samas, kui klassikalise testiteooria raames liidetakse ja keskmistatakse toorskoore, mis on sisult järjestustunnused, ning ei arvestata, et skaala lõikes võib erinevatel küsimustel olla üldskoori jaoks sisuliselt erinev kaal, erinev eristusvõime, ja/või erinev õigesti arvamise tõenäosus, siis *IRT* raames on kõiki neid parameetreid võimalik kvantifitseerida ja arvesse võtta.

Näiteks, klassikalise testiteooria raames käsitletakse polütoomsete vastustega ülesandeid vahemikaskaalal, mis pole aga teoreetiliselt õigustatud, kuna üksikküsimuste siseselt võivad lävede vahed olla varieeruva pikkusega, (rääkimata võimalusest, et lävede järjekord võib olla pööratud või reaalsete vastuste jaoks kollabeerunud, – mõned lävevahemikud võivad olla kasutamata) – *IRT* empiirika põhjal praktiliselt ei esine olukorda, kus reaalsed vahemikud oleksid vastuskategooriate tõenäosusjaotuste mõttes võrdsete pikkustega.

Lisaks kahele kirjeldatud fundamentaalsele eelisele annab *IRT* mitmeid olulisi praktilisi lisavõimalusi –

- juba mainitud võimalus hinnata õige vastuse tõenäosusfunktsiooni alumist asümptooti konkreetse küsimuse puhul, mis näiteks valikvastustega küsimuse puhul ei tarvitse võrduda valikvariantide arvu pöördväärtusega; samuti hinnata erinevate küsimuste erinevat eristusvõimet konkreetse uuritava omaduse suhtes; raskusastmest rääkimata;
- võimalus uurida ja kompenseerida küsimuse hälvet, s.t olukorda, kus sarnase uuritava omaduse tasemega inimeste õige vastuse tõenäosus konkreetsele küsimusele erineb sõltuvalt

nende rühmakuuluvusest. (Näiteks on võimalik kontrollida, kas eesti- ja venekeelsed eksamitööd on oma raskusastmetelt võrreldavad);

- võimalus uurida skaala dimensionaalsust, nt võimalus koostada mitmemõõtmelisi eksamitööd (arvestades, et ilmselt ei ole riigieksami-ainet, mis oma olemuselt oleks ühemõõtmeline), ning ühtlasi hinnata eksaminandi taset erinevatel uuritavatel dimensioonidel (erinevates ainevaldkondades, nt matemaatika puhul – trigonomeetria, algebra, ...);
- võimalus seostada erinevaid skaalasisid – viia erinevatel skaaladel saadud tulemused sisuliselt võrreldavateks: konkreetseks väljundiks oleks erinevate aastate riigieksamitööde raskusastmete võrreldavuse uurimine.

- kuna *IRT* raames on mõõtmisvea jaotus uuritava omaduse tasemete lõikes muuthajus, (klassikalise testiteooria püsihajusaga võrreldes), mis on sisevaliidsuse mõttes adekvaatsem lähenemine, aga mõõtmisviga on pöördvõrdeline informatiivsusega, mis omakorda allub skaala omadusi muutes kontrollile, on võimalus disainida soovitud parameetritele vastav skaala – soovitud dimensionaalsusega. maksimaalse mõõtmistäpsusega uuritava omaduse jaotuse teatud piirkonnas, nt kui on vaja eristada vastajaid suurema täpsusega teatud kindlas jaotuse alas – riigieksamite puhul on ainekava läbituse hindamise kõrval ühtlasi tarvilik tagada ka eksaminandide võimalikult hea võrreldavus, hea eristatus jaotuse parempoolses otsas.

### 4.3. Eksamitöö kvaliteedikontrollile suunatud lahendused

Kvaliteetne eksamitöö on võrreldavate ja tõlgendatavate riigieksamitulemuste alustala. On loomulik, et eksamitöö panevad kirja oma ala asjatundjad ainespetsialistid. Samas on inimvõimete piire ületav ülesanne hinnata pelgalt peale vaatamise teel eksamitöö konstruktivaliidsuse kehtivust konkreetset populatsioonil. Siin saavad appi tulla psühhomeetria meetodid, mille abil on võimalik väga konkreetset uurida, milline on eksamitöö dimensionaalsus (mitut faktorit või osaoskust töö "mõõdab"), kas test on aus/õiglane nt sugude lõikes, erinevate eksamikeelte lõikes, millised on erinevate ülesannete panused lõpptulemusse (intervallskaala kontroll), kas polütoomsete ülesannete puhul toimib palliomistus loogiliselt (pööratud lävede kontroll). Ühtlasi on võimalik uurida, kui võrd vastab eksamitöö raskus populatsiooni teadmiste tasemele, ning vältida olukordi, kus väga suur osa eksamitööst on suunatud väga väikese osa eksaminandide eristamisele (nt 10% palliarvestuse sajast võimalikust pallist eristab alla 1% eksaminandidest), sest on võimalik täpselt kalibreerida konkreetsete ülesannete raskusaste konkreetse populatsiooni jaoks ning valida optimaalne arv sobiva raskusega ülesandeid.

Siin oleks väga kasulik ainespetsialistide jõud ühendada psühhomeetrilise kvaliteedikontrolliga, tulemusena oleks lihtsam nii eksamitööde hindamine kui tulemuste võrdlemine ja tõlgendamine.

### 4.4. Läbi viidud eksamitööde hindamisele suunatud lahendus

Nagu ka praegu, jälgivad eksperdid iga eksamitöö puhul selle vastavust eelnevalt seatud kriteeriumidele/ainekavale. Kuna teiste riikide, aga ka meie endi kohalikud kogemused on näidanud, et inimvõimete kohane ei ole siiski väga täpselt ja järjekindlalt kvalitatiivsetele kriteeriumidele tuginedes hindadeid omistada. Seega oleks tulemusi kindlasti vajalik ka statistiliste meetodite abil saadavatega kõrvutada ning leida ekspertide ja statistiliste meetodite hinnangute pealt "kuldne kesktee". Sealjuures ei ole primaarne, kas lõpptulemuste esitamine toimub toortulemustena, protsentiilskaalal, või *IRT* meetodite abil leitavate hinnangutena. Tähtis on, et tulemuste sisu ja tähendus oleks tõlgendatavad ning võrreldavad eksaminandide, erinevate eksamiaastate, ja ideaalis ka ainete lõikes.

#### 4.5. Erinevate riigieksamiainetes tulemuste võrreldavus

Kas erinevate riigieksamiainetes tulemused saavad olla võrreldavad? See on riigieksamite käsitlemisel üks kompleksseim probleem, millele üheselt vastata on äärmiselt keeruline. Erinevate ainetes võrdlemine eeldab ühismõõtmise olemasolu üle vastavate ainetes. Ilmselt ei saa selleks ühismõõtmeks olla muu, kui üldine akadeemiline edukus/võimekus. Viimase mõõtmiseks on kahtlemata paremaid vahendeid kui riigieksamid.

Siiski on olemas kvantitatiivsed meetodid erinevate eksamiainetes suhteliste raskusastmete võrdlemiseks, neist levinumad baseeruvad taaskord *IRT* (täpsemalt, Rasch'i analüüsi) meetoditel; samuti paarikaupa kattuvate, erinevates ainetes eksameid sooritanute valimite võrdluseks; erinevate eksamiainetes tulemuste keskmiste võrdlemisel.

#### 4.6. Riigieksamite referentspopulatsioonile suunatud lahendus

Kõige üldisemas plaanis võivad riigieksamite tulemusi mõjutada eksamitööde erinev raskus, eksamisooritajate kontingendi erinevad kompetentsitasemed ning loomulikult juhuslikkus.

Et sedastada, millistest faktoritest on tingitud erinevate aastate eksamitulemuste sarnasused-erinevused, pakub asendamatu võimalusi referentspopulatsiooni kasutamine.

Kõik gümnaasiumi riigieksamite ained on õppekavas kohustuslikud, ent reeglina on eksamiainetes valik vabatahtlik (va emakeel). Nii on tekkinud olukord, et riigieksameid sooritab mingis aines ainult teatud osa õpilasi, kõikide gümnaasiumi lõpetavate isikute teadmiste tase eksami ainetes jääb aga teadmata. On vähemalt kaks võimalikku lahendit: lähtumine kõikide gümnaasiumilõpetajate üldpopulatsioonist ( $N$ ) või ainult antud aineseksami sooritajate populatsioonist ( $n$ ). Teine lahend on senini teostunud ning ei nõua mingeid täiendavaid töid. Ent raskused tekivad eelkõige erinevate aastate hinnete võrdlemisel ja erinevuste interpreteerimisel. Võtame näitena füüsika riigieksamite hindet. Aastal 2001 oli füüsika riigieksami keskmine hinne madalaim (43,14). Eksami sooritas 1280 isikut. Aastal 2005 oli füüsika keskmine hinne kõrgeim (67,07) aga sooritajate arv madalaim (546). Erinevuste võimalik seletus on selles, et füüsika kompetentsi tase gümnaasiumi lõpetajate seas pole muutunud, erinevused aastate lõikes on tingitud vaid sellest, et aastal 2005 tulid füüsika riigieksamile eelkõige need õpilased, kes valdasid füüsikat paremini. See on ainult hüpoteetiline ja raskesti tõestatav seletus, ent referentspopulatsiooni määratlemine aitab siin luua suuremat selgust. Referentspopulatsiooni määratlemine kõikide õpilastega, kes antud ainet õppisid, loob antud valdkonnas selguse. Kas ja kuidas võrd on erinevate aastate tulemused võrreldavad. Ent küsimus pole üksnes referentspopulatsiooni määratlemises, vaid ka üldpopulatsiooni sooritusaste kohta informatsiooni kogumises. Taoline töö on aga **valikuurimustena** tehtav ja informatsioon kogutav ja kindlasti lisandub nii väärtuslikku informatsiooni lisaks riigieksamite hinnete adekvaatsele interpreteerimisele ka selleks, kuidas üldse hinnata vastava õppekava täitmist. Sisuliselt tähendaks see, et sama-aegselt  $N$ . aasta riigieksamiülesannete koostamisega kogutakse varjatud valikuuringutega informatsiooni referentspopulatsiooni kompetentsuse kohta, et kalibreerida reaalses eksamiülesannetes raskus ja tuletada referentspopulatsiooni jaotusparameetrid.

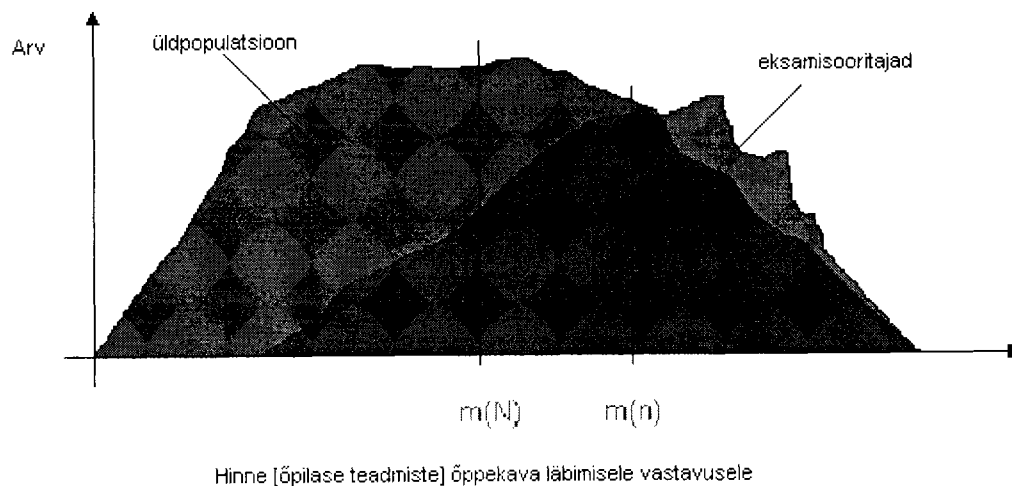
Referentspopulatsiooni määratlemine aitab riigieksamitel täita ka teisi olulisi funktsioone – hinnangut õppekava läbitusele. Praegune valikutel baseeruv süsteem ei anna adekvaatset teavet selle kohta, milline on teadmiste ja oskuste tase riigieksamiainetes kogu gümnaasiumi

lõpetaval kontingendil (väljaarvatud kirjandi abil emakeele oskuse hindamine), ent selle teadmine võiks olla haridussüsteemi edukuse hindamisel olulise väärtusega. Lahenduse miinuseks on kahtlemata lisatöö referentspopulatsioonide parameetrite kogumisel.

#### 4.6.1. Millised peaks olema kompetentsi jaotusparameetrid referentspopulatsioonis?

Ilmselt lahendusi võib siin olla mitmeid. Üks võimalik lahendus on järgmine: gümnaasiumi lõpetajate **üldpopulatsioonil** mistahes gümnaasiumi riigieksami aine tulemused transformeeritakse vastavaks normaaljaotusele keskväärtusega 50 ja standardhälbega 16,67. Sellisel juhul jääks põhiline osa (99,7%) tulemustest vahemikku 1 – 100.

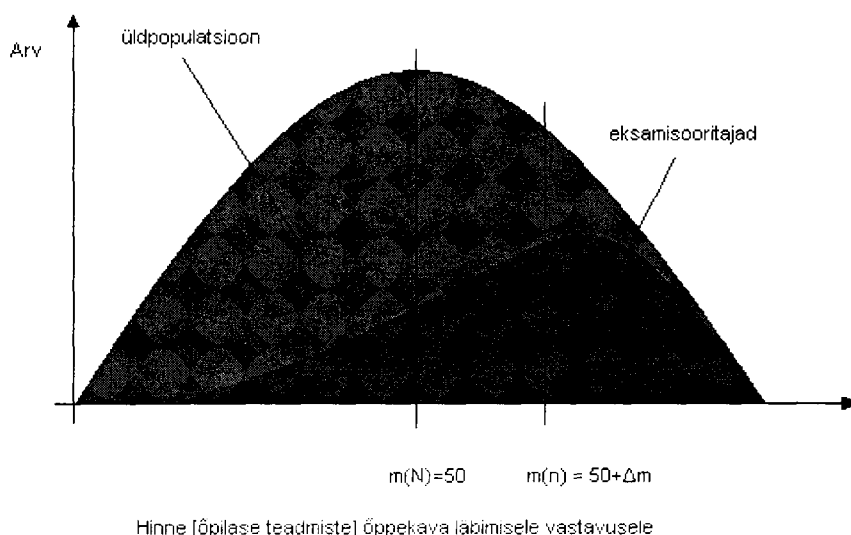
Referentspopulatsiooni ja jaotusparameetrite olemust aitavad avada joonised 7 ja 8.



**Joonis 7.** Eksamihinnete toorskooride jaotuvus enne normaliseerimist.

$m(N)$  – eksamihinde (teoreetiline) keskväärtus üldpopulatsioonis, tuletatud peidetud testi abil, mis viiakse pisteliselt läbi enne riigieksamit gümnaasiumilõpetajatel;  
 $m(n)$  – eksamihinde empiiriline aritmeetiline keskmine antud eksami sooritanute seas.

Empiiriline hinnete jaotus eksami sooritanute seas võib joonisel toodust meelevaldselt erineda. Samas, kuna antud eksamit realselt mittesooritavate gümnaasiumilõpetajate eksamihinnete teoreetiline jaotus ekstrapoleeritakse pisteliselt läbiviidava peidetud testi tulemustelt, on see eeldatavalt lähedane normaalkujule.



Joonis 8. Riigieksami hinnete jaotuvus pärast normaliseerimist.

#### 4.6.2. Kas tuleks kehtestada referentsaasta kõigi eksamiainete puhul?

Teadaolevalt praeguste riigieksamite mõnede ainete puhul sisuliselt referentsaastat juba rakendatakse, kasutades eksamitöodes ülesandeid mitmel järjestikusel aastal. Referentsaasta on aasta, mille tulemustega hakatakse täpsustama/kalibreerima järgnevate aastate eksamitöid ja tulemuste hindamist. On täiesti võimalik, et teatud aja möödudes selgub, et referentsaastat on vaja vahetada. Referentsaasta ja referentspopulatsioon koos aitavad tagada hinnete stabiilsuse ja võrreldavuse.

Referentsaastast tulenevalt on vajalik rakendada protseduuri, kus kahel järjestikusel aastal viiakse antud aineeksamisisse sisse ühisosa, nn „ankurülesanded“, mis võimaldavad võrrelda kahe aastaksamipopulatsioonide kompetentside tasemeid, tagavad baasi erinevatel järjestikutel aastatel eksamit sooritanud kontingentide võrdlemiseks. See tagab eksamite raskusastmete püsivuse ajas. Kui sellist „ankurdamist“ kasutada mitmete järjestikuste aastate lõikes, on mistahes antud ajavahemikku jäävate aastateksamipopulatsioonide kompetentsid võrreldavad, mis viiks hindamisprintsipi võimalikult objektiivseks ja aastast sõltumatuks ning tagaks äärmiselt olulise informatsiooni nii hariduspoliitika arendamiseks kui ka teadusuuringuteks.

Referentsaastat rakendavad mitmed tuntud testimissüsteemid: NAEP, SAT, aga ka rahvusvahelised haridusmõõtmiste projektid nagu PISA ja TIMSS.

### 5. Ettepanekud

- A. Kõigi eksamiainete jaoks oleks tarvilik välja töötada ainekaval põhinev kvalitatiivsete määratluste süsteem, mis oleks abiks nii eksamitöö koostamisel kui selle hindamisel.
- B. Kõigi ainete eksamitööde koostamisel oleks väga vajalik, et ainespetsialistidele pakuksid eksamitööde kvaliteedikontrolli juures abi statistikud ja psühhomeetria spetsialistid. Eksamitööde palliomistus tuleks summeeritud hinnangutel põhinevalt järjestusskaalalt viia (metodoloogiliselt põhjendatud) intervallskaalale ning iga-külgselt jälgida eksamitööde kvaliteeti üksikülesannete tasandil.

- C.** Tagamaks eksamineeritud kontingentide, aga ka eksamitööde raskusastmete võrreldavus aastate lõikes, tuleks kõigis eksamiainetes rakendada „ankurülesandeid“. Tuleks kaaluda referentspopulatsiooni rakendamise võimalusi, mis tooks kaasa täiendava töö erinevates eksamiainetes kogu populatsiooni kompetentsi hindamisele suunatud valikuuringute/tasemetööde näol.
- D.** Eksamitööde hindamisel tagada ekspertide hinnangutele lisaks ka statistilise analüüsi andmed, et kergendada ekspertidel tööde järjekindlat ja võimalikult objektiivset hindamist.
- E.** Seniks, kuni uut hindamissüsteemi kujundatakse, jätkata hindamist seni toiminud süsteemi kohaselt. Suuremat tähelepanu aga tuleks pöörata tulemustega kaasneva informatsiooni esitamisele. Eelkõige tuleks täiendavalt esitada informatsiooni, mis hõlbustaks tulemuste tõlgendatavust ja kasutamist. Haridus- ja Teadusministri määrus nr 40 õpitulemuste hindamise põhimõtetest sätestab küll terve rea analüüsi, mida riigieksamite tulemuste alusel tuleb teha ja esitada Haridus- ja Teadusministeeriumile, aga määrus ei osuta võimalustele, mida teha tulemuste paremaks tõlgenduseks tulemuste kasutajate poolt. Üheks võimaluseks on esile tuua informatsioon tulemuste pingereast. Tulemuse esitamine protsentiilskaalal on siin üheks võimalikuks lahenduseks.



## Lisa 1

2007. aasta gümnaasiumi riigeksamite hinnete mõned võimalikud transformatsioonid.  
(Tulemus protsentiilskaalal on esitatud täisarvudes, skaala väärtused <1 on ümardatud üheks)

## Ajaloos gümnaasiumi riigeksami hinded 2007

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	2 171	2	97	64,72	17,55	-0,69	0,23
Normaliseerimisteisendus (Van der Waerden)	2 171	-5	102	49,97	16,60	-0,00	-0,06
Normaliseerimis- teisendus (Blom)	2 171	-7	103	49,97	16,61	-0,00	-0,04
Lineaarteisendus (M=50; Sd=16,67)	2 171	-10	81	50,00	16,67	-0,70	0,24
Tulemus protsentiilskaalal	2 171	1	100	50,84	29,07	-0,02	-1,21
Normaalkövera ekvivalent	2 171	-24	117	49,99	21,07	-0,01	-0,04

## Bioloogia gümnaasiumi riigeksami hinded 2007

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	3 669	11	99	58,56	17,83	-0,02	-0,74
Normaliseerimisteisendus (Van der Waerden )	3 669	-6	108	49,99	16,61	-0,02	-0,03
Normaliseerimis- teisendus (Blom)	3 669	-7	110	49,99	16,63	-0,01	-0,01
Lineaarteisendus (M=50; Sd=16,67)	3 669	6	88	50,00	16,66	-0,02	-0,74
Tulemus protsentiilskaalal	3 669	1	100	50,82	28,89	-0,03	-1,20
Normaalkövera ekvivalent	3 669	-23	127	50,05	21,08	-0,01	-0,03

## Eesti keele kirjandi gümnaasiumi riigeksami hinded 2007

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	10 225	0	100	56,44	21,50	-0,11	-0,67
Normaliseerimisteisendus (Van der Waerden)	10 225	0	94	49,98	16,55	-0,01	-0,10
Normaliseerimis- teisendus (Blom)	10 225	0	95	49,99	16,58	-0,00	-0,08
Lineaarteisendus (M=50; Sd=16,67)	10 225	6	84	50,01	16,67	-0,11	-0,67
Tulemus protsentiilskaalal	10 225	1	100	50,65	28,91	-0,02	-1,21
Normaalkövera ekvivalent	10 225	-13	106	49,98	20,95	-0,01	-0,09

## Vene keele kirjandi gümnaasiumi riigeksami hinded 2007

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	3 581	0	100	58,08	18,70	-0,01	-0,72
Normaliseerimisteisendus (Van der Waerden )	3 581	-4	102	49,99	16,58	0,00	-0,05
Normaliseerimis- teisendus (Blom)	3 581	-5	102	49,98	16,61	-0,01,	-0,03
Lineaarteisendus (M=50; Sd=16,67)	3 581	-2	87	50,01	16,67	-0,01	-0,72
Tulemus protsentiilskaalal	3 581	1	100	51,13	29,30	-0,01	-1,21
Normaalkövera ekvivalent	3 581	-20	116	50,03	21,01	-0,00	-0,04

**Geograafia gümnaasiumi riigeksami hinded 2007**

Muutuse sisu	arv	min	Max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	7 132	8	98	56,36	14,79	0,01	-0,50
Normaliseerimisteisendus (Van der Waerden )	7 132	-11	109	50,04	16,62	-0,01	-0,01
Normaliseerimis- teisendus (Blom)	7 132	-13	110	50,03	16,63	-0,01	0,00
Lineaarteisendus (M=50; Sd=16,67)	7 132	-5	97	50,00	16,67	0,01	-0,50
Tulemus protsentiilskaalal	7 132	1	100	50,99	28,83	-0,03	-1,20
Normaalkövera ekvivalent	7 132	-30	127	49,98	21,06	0,01	-0,01

**Inglise keele gümnaasiumi riigeksami hinded 2007**

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	9 696	5	99	68,80	15,97	-0,48	-0,31
Normaliseerimisteisendus (Van der Waerden )	9 696	-12	110	49,99	16,63	0,01	-0,02
Normaliseerimis- teisendus (Blom)	9 696	-14	111	49,98	16,65	0,01	-0,01
Lineaarteisendus (M=50; Sd=16,67)	9 696	-17	82	50,01	16,66	-0,48	-0,32
Tulemus protsentiilskaalal	9 696	1	100	50,84	28,98	-0,02	-1,20
Normaalkövera ekvivalent	9 696	-32	128	50,03	21,06	0,00	-0,01

**Saksa keele gümnaasiumi riigeksami hinded 2007**

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	808	15	100	70,50	19,49	-0,43	-0,69
Normaliseerimisteisendus (Van der Waerden )	808	2	93	49,94	16,53	-0,02	-0,17
Normaliseerimis- teisendus (Blom)	808	0	94	49,95	16,60	-0,02	-0,12
Lineaarteisendus (M=50; Sd=16,67)	808	3	75	50,00	16,69	-0,42	-0,71
Tulemus protsentiilskaalal	808	1	100	50,86	29,02	-0,01	-1,19
Normaalkövera ekvivalent	808	-14	105	50,00	21,01	-0,03	-0,11

**Prantsuse keele gümnaasiumi riigeksami hinded 2007**

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	36	44	97	77,50	13,35	-1,14	0,58
Normaliseerimisteisendus (Van der Waerden )	36	18	82	49,94	15,41	0,01	-0,46
Normaliseerimis- teisendus (Blom)	36	15	85	50,03	16,12	-0,01	-0,29
Lineaarteisendus (M=50 Sd=16,67)	36	8	74	49,94	16,65	-1,14	0,57
Tulemus protsentiilskaalal	36	3	100	52,33	29,62	-0,05	-1,24
Normaalkövera ekvivalent	36	4	96	49,89	20,81	0,01	-0,20

**Keemia gümnaasiumi riigeksami hinded 2007**

Muutuse sisu	arv	min	max	keskmine	SD	Asümmeetria	järsakus
Esialgne tulemus	2 276	6	100	61,96	21,24	-0,28	-0,80
Normaliseerimisteisendus (Van der Waerden )	2 276	-5	100	49,96	16,57	-0,01	-0,07
Normaliseerimisteisendus (Blom)	2 276	-8	101	49,97	16,59	-0,00	-0,04
Lineaarteisendus (M=50; Sd=16,67)	2 276	6	80	50,01	16,69	-0,28	-0,80
Tulemus protsentiilskaalal	2 276	1	100	50,60	28,98	-0,01	-1,20
Normaalkövera ekvivalent	2 276	-24	114	49,99	21,08	-0,01	-0,05

**Matemaatika gümnaasiumi riigieksami hinded 2007**

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	5 513	0	100	49,26	19,09	-0,05	-0,25
Normaliseerimisteisendus (Van der Waerden )	5 513	-1	98	49,98	16,59	-0,01	-0,07
Normaliseerimisteisendus (Blom)	5 513	-1	99	49,99	16,60	-0,01	-0,06
Lineaarteisendus (M=50; Sd=16,67)	5 513	7	94	50,00	16,67	-0,06	-0,25
Tulemus protsentiilskaalal	5 513	1	100	50,72	28,88	-0,03	-1,20
Normaalkövera ekvivalent	5 513	-15	112	49,94	21,04	0,01	-0,06

**Riigikeele gümnaasiumi riigieksami hinded 2007**

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	4 834	0	100	63,23	23,41	-0,35	-1,03
Normaliseerimisteisendus (Van der Waerden )	4 834	-9	104	50,00	16,61	-0,02	-0,08
Normaliseerimisteisendus (Blom)	4 834	-11	104	50,00	16,64	-0,02	-0,05
Lineaarteisendus (M=50; Sd=16,67)	4 834	5	76	50,01	16,69	-0,36	-1,04
Tulemus protsentiilskaalal	4 834	1	100	50,62	29,01	-0,00	-1,20
Normaalkövera ekvivalent	4 834	-28	119	49,95	21,05	-0,01	-0,04

**Ühiskonnaõpetuse gümnaasiumi riigieksami hinded 2007**

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	5 287	3	97	58,24	14,70	-0,34	-0,03
Normaliseerimisteisendus (Van der Waerden )	5 287	-9	109	49,99	16,65	-0,01	-0,04
Normaliseerimisteisendus (Blom)	5 287	-11	111	49,99	16,67	-0,01	-0,02
Lineaarteisendus (M=50; Sd=16,67)	5 287	-13	94	50,01	16,66	-0,34	-0,03
Tulemus protsentiilskaalal	5 287	1	100	51,01	28,96	-0,03	-1,20
Normaalkövera ekvivalent	5 287	-29	129	50,03	21,08	0,00	-0,02

**Vene keele (võrkeelena) gümnaasiumi riigieksami hinded 2007**

Muutuse sisu	arv	min	max	keskmine	SD	asümmeetria	järsakus
Esialgne tulemus	471	22	100	75,37	16,51	-0,77	-0,14
Normaliseerimisteisendus (Van der Waerden )	471	2	98	49,96	16,53	-0,01	-0,17
Normaliseerimisteisendus (Blom)	471	0	100	49,95	16,59	-0,01	-0,12
Lineaarteisendus (M=50; Sd=16,67)	471	-4	75	50,17	16,83	-0,80	-0,12
Tulemus protsentiilskaalal	471	1	100	51,13	29,30	-0,01	-1,21
Normaalkövera ekvivalent	471	-15	115	49,95	20,98	-0,01	-0,03

**Lisa 2**

2007. aasta riigieksamate toortulemuste kokkusurumine jaotuse teisendamisel protsentiilskaalale. Kuna küsimuse all on riigieksamate võime olla üliõpilaskandidaatide usaldusväärse võrdluse aluseks, on vaatluse all vaid jaotuse parempoolne osa nende toortulemuste osas, mille puhul toimub protsentiilteisenduse käigus „kokku surumine“ e võimalik infokadu. Vastus küsimusele, kas infokadu on tõeline või vaid näiline, eeldab iga konkreetse eksamitöö puhul sedastamist, kui suur on mõõtmisviga jaotuse erinevates osades.

Ajalugu (N = 2171)

11 isikut (0,51%) - 100. protsentiilastak (toortulemused 96, 97)

28 isikut (1,29%) - 99. protsentiilastak (toortulemused 93, 94, 95, 96)

Bioloogia (N = 3669)

21 isikut (0,57%) - 100. protsentiilastak (toortulemused 95, 96, 97, 98, 99)

62 isikut (1,69%) - 99. protsentiilastak (toortulemused 92, 93, 94)

Eesti keele kirjand (N = 10225)

77 isikut (0,75%) - 100. protsentiilastak (toortulemus 100)

97 isikut (0,95%) - 99. protsentiilastak (toortulemused 97, 98, 99)

91 isikut (0,89%) - 98. protsentiilastak (toortulemused 95, 96)

134 isikut (1,31%) - 97. protsentiilastak (toortulemused 93, 94)

Füüsika, (N = 592)

3 isikut (0,51%) - 100. protsentiilastak (toortulemus 100)

16 isikut (2,70%) - 99. protsentiilastak (toortulemused 98, 99)

Geograafia, (N = 7132)

40 isikut (0,56%) - 100. protsentiilastak (toortulemused 90, 91, 92, 93, 94, 95, 96, 98)

72 isikut (1%) - 99. protsentiilastak (toortulemused 87, 88, 89)

96 isikut (1,35%) - 98. protsentiilastak (toortulemused 84, 85, 86)

Inglise keel (N = 9696)

91 isikut (0,94%) - 100. protsentiilastak (toortulemused 96, 97, 98, 99)

Keemia (N = 2276)

12 isikut (0,53%) - 100. protsentiilastak (toortulemused 99, 100)

29 isikut (1,27%) - 99. protsentiilastak (toortulemused 97, 98)

Matemaatika, (N = 5513)

28 isikut (0,51%) - 100. protsentiilastak (toortulemused 98, 99, 100)

66 isikut (1,2%) - 99. protsentiilastak (toortulemused 89, 90, 91, 92, 93, 94, 95, 96, 97)

46 isikut (0,83%) - 98. protsentiilastak (toortulemused 86, 87, 88)

68 isikut (1,23%) - 97. protsentiilastak (toortulemused 83, 84, 85)

50 isikut (0,91%) - 96. protsentiilastak (toortulemused 81, 82)

70 isikut (1,27%) - 95. protsentiilastak (toortulemused 79, 80)

Prantsuse keel, (N = 36)

1:1 vastavus toortulemuse ja protsentiilastaku vahel üle kogu jaotuse

Riigikeel, (N = 4834)

38 isikut (0,79%) - 100. protsentiilastak (toortulemused 99, 100)

Saksa keel, (N = 808)

1:1 vastavus toortulemuse ja protsentiilastaku vahel üle kogu jaotuse

Vene keele kirjand (N = 3581)

25 isikut (0,7%) - 100. protsentiilastak (toortulemused 98, 99, 100)

30 isikut (0,84%) - 99. protsentiilastak (toortulemused 95, 96, 97)

47 isikut (1,31%) - 98. protsentiilastak (toortulemused 92, 93, 94)

Vene keel võõrkeelena (N = 471)

5 isikut (1,06%) - 100. protsentiilastak (toortulemused 98, 99, 100)

Ühiskonnaõpetus (N = 5287)

29 isikut (0,55%) - 100. protsentiilastak (toortulemused 90, 91, 92, 93, 94, 95, 96, 97)

59 isikut (1,12%) - 99. protsentiilastak (toortulemused 87, 88, 89)

46 isikut (0,87%) - 98. protsentiilastak (toortulemused 85, 86)

61 isikut (1,15%) - 97. protsentiilastak (toortulemused 83, 84)

93 isikut (1,76%) - 96. protsentiilastak (toortulemused 81, 82)