

**Tartu Ülikool  
Psühholoogia instituut**

**Eve Kanarik**

**WAIS-III Aritmeetika alltesti kohandamine Eesti oludele -  
üksikülesannete ja esialgse skaala psühhomeetrilised omadused**

**Magistriprojekt**

**Juhendaja:** Olev Must, PhD

**Kaasjuhendaja:** Aire Raidvee, MSc

**Läbiv pealkiri:** WAIS-III Aritmeetika alltest

**Tartu 2010**

## Sisukord

Sisukord.....	2
Kokkuvõte.....	4
Abstract.....	5
1. Sissejuhatus.....	6
1.1. Wechsleri skaala.....	6
1.2. Aritmeetika alltest ja töömälu.....	6
1.3. Aritmeetika allskaala kohandamine.....	9
Kohandamise etapid.....	13
1.4. Töö eesmärgid.....	15
2. Meetod.....	16
2.1. Valim.....	16
2.2. Mõõtmisvahendid.....	16
Aritmeetika.....	16
Arvumälu.....	18
Sarnasused.....	18
Taibukus.....	18
2.3. Testijad ja testimisprotseduur.....	19
2.4. Andmeanalüüs.....	19
Analüüsi etapid.....	19
Rasch`i mudelist lähtuv analüüs.....	20
Klassikalisest testiteooriast lähtuv analüüs.....	21
3. Tulemused.....	22
3.1. Rasch`i mudelist lähtuv analüüs.....	22
Mudeli eeldused.....	22
Ülesannete sobitumine Rasch`i mudeliga.....	23
Ülesannete raskusjärjestus.....	23
Lävede järjestus.....	24
Ülesannete hälve.....	24
Skaala sobitumine Rasch`i mudeliga.....	24

3.2. Klassikalisest testiteooriast lähtuv analüüs.....	24
Ülesannete raskusjärjestus.....	24
Ülesannete eristusvõime.....	26
Skaala raskus.....	26
Skaalaskoorid.....	26
Sisereliaablus.....	27
Valiidsus.....	27
4. Arutelu.....	27
Viited.....	34

**Kokkuvõte**  
**WAIS-III Aritmeetika alltesti kohandamine Eesti oludele – üksikülesannete ja esialgse skaala psühhomeetrilised omadused**

Magistriprojekti eesmärgiks oli luua eestikeelsesesse WAIS-III Aritmeetika alltesti heade psühhomeetriliste omadustega ülesanded, mis moodustaksid originaaltestiga sisuliselt samaväärse skaala. Andmeid koguti Aritmeetika, Arvumälu, Sarnasuste ja Taibukuse testidega 191 normaalvalimi ning 20 vaimse alaarenguga inimeselt. Üksikülesannete ja skaala hindamiseks kasutati üksikvastuste ja klassikalise testiteooria meetodeid. Analüüsi käigus hinnati Aritmeetika üksikülesannete ja skaala sobitumist Rasch'i mudeliga; ülesannete raskust; ülesannete hälvet soo, vanuse ja hariduse osas; lävede raskuse järjekorda ning skaala dimensionaalsust, reliaablust, valiidsust, raskust ja esialgsete skaalaskooride jaotust. Eestikeelsesesse Aritmeetikasse ei sobinud 4 originaalülesannet, mis asendati lisaülesannetega. Eestikeelse Aritmeetika ülesanded olid sobivate psühhomeetriliste omadustega, v.a. üks ülesanne, mis kergelt eelistas madalama haridusega inimesi. Skaala toorskoorides ilmnes kerge negatiivne asümmeetria ja esialgsed skaalaskoorid varieerusid 4 standardhälbe ulatuses, mille tõttu on testi vaja täiendada raskemate ülesannetega. Skaala sisemine reliaablus osutus heaks, poolitustesti reliaablus ning seosed teiste alltestidega olid sarnased originaaltesti vastavate näitajatega.

**Märksõnad:** *WAIS-III, Aritmeetika, kohandamine*

### **Abstract**

#### **Adaptation of WAIS-III Arithmetic subtest into Estonian – psychometric properties of items and the preliminary scale**

The aim of this study was to create into WAIS-III Arithmetic subtest items with good psychometric properties which would form an equivalent scale to the original. Data was collected using WAIS-III Arithmetic, Digit Span, Similarities and Comprehension subtests from 191 normgroup people and 20 clinical group people (mental retardation). Item response theory and classical test theory methods were used to evaluate items' and scale fit with Rasch model, items' difficulty; differential functioning; thresholds positions and scale dimensionality; range of preliminary scale scores; internal and split-half reliability and validity. Four original items from Arithmetic subtest were replaced with appropriate additional items. Estonian Arithmetic items had appropriate psychometrical properties, except one item which was slightly biased toward people with lower education. Distribution of subscale raw scores was slightly negatively asymmetrical and preliminary scale scores varied between 4 standard deviations, therefore the scale needs to be complemented with more difficult items. Subscale internal reliabilities were good, split-half reliabilities and correlations with other subtests were similar to the original.

**Keywords:** *WAIS-III, Arithmetic, adaptation*

## 1. Sissejuhatus

### 1.1. Wechsleri skaala

WAIS-III (*Wechsler Adult Intelligence Scale – Third Edition*, Wechsler, 1997) on individuaaltest, mis on mõeldud täiskasvanute (16- kuni 89-aastaste) üldvõimekuse ja spetsiifiliste võimete hindamiseks. Wechsler defineeris intelligentsuse kui „indiviidi võime käituda eesmärgipäraselt, mõelda ratsionaalselt ja saada efektiivselt hakkama oma keskkonnas“ (Groth-Marnat, 1997). WAIS on üks kõige sagedamini kasutatavaid instrumente kliinilises töös (Rabin, Barr & Burton, 2005).

WAIS-III on hierarhilise struktuuriga, võimaldades skoorid välja arvutada neljal tasandil – koguskaala; sõnaline ja soorituslik skaala ning indekskaalad (sõnaline taibutus, tajupõhine töötlus, töömälu ja töötluskiirus). Neljanda tasandi moodustavad üksikud alltestid. Skaalaskooride (koguskaala, sõnaline/soorituslik ja indekskaalad) leidmiseks summeeritakse vastavate alltestide skaleeritud skoorid ja leitakse neile vastavus skaalal, mille keskmine on 100 ja standardhälve 15. Alltestide skaleeritud skoorid (keskmine 10, standardhälve 3) leitakse iga vanusegrupi jaoks eraldi.

WAIS-III alltestide jaotamine sõnalisse ja soorituslikku skaalasse põhineb Wechsleri praktilisel kogemusel ja ei ole toetatud faktoranalüütiliste uuringute poolt, samas on empiirilist kinnitust leidnud WAIS-III alltestide jagamine kolmetasandiliseks hierarhiliseks struktuuriks (koguskaala, indekskaalad ning üksikud alltestid) (Taub, Witta & McGrew, 2004; Saklofske, Hildebrand & Gorsuch, 2000).

### 1.2. Aritmeetika alltest ja töömälu

Aritmeetika alltesti (edaspidi Aritmeetika; lähtuvalt koostatava eestikeelse WAIS-III käsiraamatu kokkuleppelisele terminite kasutamisele kirjutatakse alltestide nimed suure algustähega) ülesannete lahendamine toetub mitmetele võimetele nagu arvutamisoskus, auditoorne mälu, järjestikune infotöötlus, keskendumisvõime ja tähelepanu, varastest kooliaastatest omandatud teadmised, loogiline mõtlemine, üldistusvõime, arvuliste probleemide lahendamise ning analüüsi- võime (Groth-Marnat, 1997). Aritmeetika on WAIS-III alltestidest üks kõige enam pingutust nõudev ja väga tundlik häirimise suhtes, kuna piiratud aja jooksul tuleb suuliselt esitatud ülesanded peast

lahendada. Testi kõrged skoorid näitavad head matemaatilist taiplikkust, keskendumisvõimet ja head auditoorset lühimälu.

Aritmeetika kuulub WAIS-III sõnalisse skaalasse ja töömälu indeksskaalasse ning on oluliselt seotud nii teiste sõnalise skaala alltestidega (Sõnavara, Arvumälu, Taibukus, Sarnasused, Informeeritus) kui ka töömälu indeksskaala alltestidega (Arvumälu, Tähtnumber järjestus). Kuna empiirilisel on kinnitust leidnud pigem WAIS-III kolmetasandiline struktuur (jagunemine koguskaalaks, indeksskaaladeks ja üksikuteks alltestideks) (Taub jt., 2004; Saklofske jt., 2000), siis Aritmeetika, kuuludes töömälu indeksskaalasse, peaks mõõtma töömälu.

Daneman ja Carpenter (1980) on defineerinud töömälu kui aktiivset süsteemi, mis salvestab ja samal ajal manipuleerib informatsiooniga. Näiteks ülesanne: *Kolmel riiulil on kokku 66 raamatut. Alumisel riiulil on kolm korda ja keskmisel kaks korda rohkem raamatuid kui ülemisel riiulil. Mitu raamatut on ülemisel riiulil?* Selle lahendamiseks on tarvis teha üks arvutus, hoida tulemust mälus samal ajal kui tehakse uusi arvutusi ning seejärel kombineerida kõikide oluliste arvutuste tulemused, et leida vastus küsimusele. Baddeley ja Hitch (1974) käsitlevad töömälu kolmeosalise süsteemina, millest allsüsteemid fonoloogiline ring (ik *fonological loop*) ja visuaalruumiline visandiplokk (ik *visuospatial sketchpad*) salvestavad ja säilitavad vastavalt sõnalis-auditoorset ning visuaalruumilist informatsiooni. Kogu süsteemi kontrollib piiratud mahuga keskne-täidesaatev tähelepanusüsteem, mille ülesandeks on tähelepanu keskendamine ja jaotamine, ebaoluliste stiimulite allasurumine ning tegevuse planeerimine. Hiljem kirjeldas Baddeley veel neljandat süsteemi, episoodilist puhvrit (ik *episodic buffer*), mis on modaalsusvaba ja tegeleb sündmuste abstraktsete representatsioonidega ning integreerib kahe allsüsteemi infot episoodilise mälu infoga (Baddeley, 2000).

Kuna Aritmeetika ülesanded esitatakse sõnaliselt, siis on nende ülesannete lahendamine seotud töömälu fonoloogilise ringi aktiivsusega. Baddeley (2003) on leidnud, et sõnaliselt esitatud töömälu ülesannete korral, mille puhul aktiveerub fonoloogiline ring, avaldab töömälule mõju fonoloogilise sarnasuse efekt (sarnaselt kõlavaid häälikute ridu mäletatakse vähem kui erinevalt kõlavaid häälikute ridu) ning sõna pikkuse efekt (töömälu maht väheneb häälikute arvu kasvades). Baddeley eristab fonoloogilises ringis kahte allsüsteemi – hoidla, kus säilitatakse mälujäljed mõneks

sekundiks, ning artikuloorset kordamise protsessi, mis aitab mälujälgi värskendada. Töömälu maht on piiratud, kuna kordamine toimub reaajas – meeldejäetavate ühikute arvu suurenedes jõutakse punkti, kus esimene mälujalg on kadunud enne, kui selle kordamiseni jõutakse. Artikuloorse kordamise protsessi olemasolule viitab see, et fonoloogilise sarnasuse ja sõna pikkuse mõju kaovad visuaalselt esitatud sõnade puhul.

Mitmed uurimused toetavad töömälu valdkonnaspetsiifiliste komponentide olemasolu, kuna töömälu on tugevemalt seotud samas valdkonnas võimekust mõõtvate testidega (Süss & Oberauer 2002; Haavisto & Lehto 2004; Mackintosh & Bennett, 2003). WAIS-III töömälu testid on esitatud suuliselt ning mõõdavad seega vaid sõnalis-audioorset töömälu.

Jensen arvab, et individuaalsed erinevused intelligentsustestide tulemustes on seotud just erinevustega töömälu võimekuses (Jensen, 1998). Kyllonen ja Christal (1990) leidsid, et töömälu ja taibukus on omavahel väga tugevalt seotud ( $r=,80\dots,90$ ). Mitmed hilisemad uurimused kinnitavad, et töömälu on tõepoolest seotud  $g$ -faktoriga, kuid pole siiski päris sama, mis  $g$  – töömälu seletab umbes 30 kuni 50%  $g$  hajuvusest (Conway, Kane & Engle, 2003; Conway, Cowan, Bunting, Thierrault & Minkoff, 2002). Ülesanded, mis ei eelda ainult info meeldejätmist, vaid ka sellega manipuleerimist, seostuvad  $g$ -faktoriga tugevamalt (Ackerman, Beier & Boyle, 2002; Conway jt., 2002). Arvatakse, et töömälu ja  $g$ -faktori seost vahendab keskne täidesaatev tähelepanusüsteem, kuna just nendes ülesannetes, mis nõuavad täidesaatva tähelepanusüsteemi kontrolli segavate tegurite (proaktiivne interferents; vastuste konkureerimine; harjumuslik, kuid ebasobiv vastamine) eiramiseks, on sooritus seotud töömäluga (Conway jt., 2003). Täidesaatva tähelepanu-süsteemi roll töömälu ülesannetes on toetada aktiivset eesmärgipärast informatsiooni säilitamist häirivate tegurite korral.

Mitmed uurimused näitavad, et töömälu ja  $g$  on omavahel seotud lühimälu kaudu. Colom, Flores-Mendoza, Quiroga ja Privado (2005) leidsid, et töömälu (info meeldejätmise ja samaaegne manipuleerimine) ja lühimälu (ainult info meeldejätmise) olid omavahel tugevalt seotud, kuid töömälu ennustas  $g$ -faktorit paremini kui lühimälu. Samas, kui lühiajalise mälu mõju hoiti statistiliselt kontrolli all, siis oli töömälu ennustavus madal, mis võiks viidata sellele, et töömälu lühimälu komponent vahendab seost töömälu ja  $g$  vahel. Colom, Abad, Quiroga, Shih ja Flores-Mendoza (2008) leidsid,



et lühimälu koos töötluskiirusega kirjeldas 88% töömälu testiskooride hajuvusest (lühimälu 79% ja töötluskiirus 9%), kusjuures kui nende osade mõju kontrolliti, siis töömälu ei ennustanud enam individuaalseid erinevusi vaimses võimekuses.

Üldiselt ollakse seisukohal, et töömälu ei ole päris ühtne süsteem, ning arvatakse, et erinevatel valdkonnaspetsiifilistel töömälu osadel võib olla lisaks passiivsele spetsiifilise info salvestamisele ka aktiivne roll, mis tegeleb samaaegselt selle info töötlemisega. Töömälu ja *g*-faktori seotuse osas ollakse üsna ühel meelel, kuid seose seletamisel jagunevad uurijad peamiselt kahte leeri – need kes arvavad, et töömälu ja *g* on seotud tähelepanuprotsesside kaudu, ja need, kes arvavad, et seda seost vahendab lühimälu.

Aritmeetika allskaala (edaspidi Aritmeetika) kuulub WAIS-III kohustuslike testide hulka ja pakub koos teiste koguskaala testidega võimaluse üldise vaimse võimekuse taseme (k.a. vaimse alaarengu) hindamiseks. Lisaks sellele võimaldab Aritmeetika hinnata sõnalist töömälu. Uuringud on näidanud, et häired töömälus võivad olla seotud mitmete psüühikahäirete (skisofreenia, posttraumaatilise stressihäire, aktiivsus- ja tähelepanuhäire, õpivilumuste spetsiifilised häired) ja neuroloogiliste häiretega (kerge kognitiivne kahjustus, frontotemporaalne ja vaskulaarne dementsus) (Piskulic, Olver, Norman & Maruff, 2007; Morey jt., 2009; Biederman jt., 2009; Carretti, Borella, Cornoldi & DeBeni 2009). Töömälu häirituse hindamine aitab kaasa kliiniliste hüpoteeside püstitamisele ja diferentsiaal-diagnoosimisele keerulisemate juhtumite korral.

### **1.3. Aritmeetika allskaala kohandamine**

Lähtuvalt rahvusvahelistest testi adapteerimise juhistest (*International Test Commission [ITC], 2010*) peaks adapteerimise tulemusena kohandatud test olema samaväärne originaaltestiga. Selle saavutamiseks tuleb testiülesannete ja instruksioonide kohandamisel arvestada populatsioonide lingvistiliste ja kultuuriliste erinevustega, tagada testimaterjalide ja protseduuride sobivus, kindlustada, et andmete kogumine võimaldab kasutada sobivaid statistilisi meetodeid testi erinevate keelevariantide võrreldavuse tagamiseks. Oluline on tagada testi valiidsus ja reliaablus. Kuna test adapteeritakse ja normid kogutakse sihtgrupi suhtes, tuleb andmeid koguda populatsiooni esindavalt

valimilt (*American Educational Research Association [AERA], National Council of Measurement in Education [NCME], American Psychological Association [APA], 1999*).

Selleks, et alltestid oleks vanusegrupiti ja alltestide lõikes võrreldavad, on WAIS-III alltestide toorskoorid teisendatud skaalaskoorideks (keskmine 10, standardhälve 3). See on saavutatud toorskooride kumulatiivse sagedusjaotuse normaliseerimisega ja neist skaalaskooride teisendamisega igas vanusegrupis eraldi. Skaalade (koguskaala, sõnaline/soorituslik ja indeksskaalad) skoorid on leitud vastavate alltestide skaalaskooride summeerimisel ning nende summa teisendamisel skaalale, mille keskmine on 100 ja standardhälve 15. Seega on adapteerimisel oluline, et Aritmeetika ülesannete raskuse ja inimeste võimekuse taseme varieeruvus oleks piisav, et võimaldada toorskooride teisendamist originaaliga samaväärsele kujule.

Testi eri keelevariantide võrdväärsuse tagamiseks tuleb välja selgitada ülesanded, mis ei sobi teises kultuuris kasutamiseks. Selleks uuritakse ülesannete hälvet (ik *differential item functioning*, DIF) ehk olukorda, kus uuritava võimekusdimensiooni võrdse taseme korral erineb testiküsimusele õigesti vastamise tõenäosus teatud isikutunnuste (nt soo, vanuse, hariduse) alusel eristuvates gruppides (Lord, 1980). Ülesannete hälbe vältimiseks, tuleb ülesannete loomisel ja tõlkimisel arvestada ülesannete ülesehitust, vältida stereotüüpset ja diskrimineerivat keelekasutust vähemusgruppe silmas pidades. Elosua ja López-Jauregui (2007) eristasid nelja DIF-i allikat: kultuuriline sobivus, tõlkeprobleemid, morfoloogilised-süntaktilised ja semantilised erinevused.

Testi tõlkimise teooriast tulenevalt on tõlke viga mitmemõõtmeline – tõlke viga pole ainult vigase tõlke tagajärg, vaid tuleneb mitmetest tõlkevea dimensioonidest (nt teksti esitamise stiil, semantika, sõnakasutus ja kontekst) (Solano-Flores, Backhoff, Contreras-Niño, 2009). Teooria autorite arvates pole tõlkeviga välditav, kuid seda on võimalik vähendada vigasid süstemaatiliselt kodeerides ja tõlkimisse mitmete erialade esindajaid kaasates. Samas ei pruugi ka võimalikult vähete tõlkevigadega ülesanded tagada ülesannete sobivust kultuuriliste erinevuste tõttu – mõõdetavad konstruktid ja konstruktiga seotud käitumised erinevad kultuuriti, tõlkimise käigus võib test muutuda mitmemõõtmeliseks (Van de Vijver & Leung, 1997). Seega on oluline testi adapteerida, mitte ainult tõlkida. Adapteerimisel tõlgitakse ülesanded, kuid muudetakse sõnastust või

ülesannete sisu nii, et need sobiksid kultuurilise kontekstiga (Hambleton, 2005).

Adapteerimisel lisatakse testi ka kultuuriliselt sobivamaid kuid originaalülesannetega sisuliselt sarnaseid ülesandeid juurde, et oleks võimalik ebasobivad originaalülesanded välja vahetada nii, et säiliks testi struktuur, reliaablus ja valiidsus.

WAIS-III on tõlgitud ja kohandatud erinevatesse keeltesse. Kohandamise protsess on erinevates riikides toimunud erineva põhjalikkusega ning kohalike testivariantidega on esile tulnud mitmeid probleeme.

Väiksemates riikides nagu Rootsi ja Taani tõlgiti testi instruksioonid ja ülesanded, kuid kohalikke norme ei loodud. Iverson, Lange ja Viljoen (2006) uurisid Ameerika ja Kanada normidest tulenevaid erinevusi psühhiaatriliste patsientide valimil ning leidsid, et Kanada normide kasutamisel saadi süstemaatiliselt madalamad tulemused testiskoorides võrreldes Ameerika normide kasutamisega. Autorid ei kahtle Kanada normide usaldusväärsuses ja selline erinevus Ameerika normidest lubab pidada Kanada testi raskemaks võrreldes Ameerika testiga. Seega omab testipunkt tähendust vaid kindlas kultuurikontekstis ning normide ülevõtmine ei ole enamasti õigustatud.

Inglise variandi loomisel anglitseeriti WAIS-III ülesanded ning loodeti kasutada Ameerika norme. 332 inimesest koosneva valimi analüüsimisel leiti, et koguskoori, indeksskooride ning kuue alltesti skaleeritud skooride keskmised olid oluliselt kõrgemad Ameerika valimi omadest (Wycherley, Benjamin, Crawford & Mockler, 1999). Seda tulemust seletati Flynn'i efekti (Inglise andmed koguti Ameerika omadest 2 aastat hiljem) ja valimiga, mis oli kallutatud vaimselt võimekamate inimeste suunas (tudengitest admekogujatel oli lihtsam uuringusse värvata oma tuttavaid, kes tõenäoliselt olid samuti kõrgema haridusega). Seejärel täiendati valimit, et see sisaldaks ka madalama vaimse võimekusega inimesi ja esindaks paremini populatsiooni hariduse ja elukutse osas (Wycherley, Lavender, Holttum, Crawford & Mockler, 2005). Seejärel leiti taas, et mitmete skaalade ja alltestide keskmised (välja arvatud töötluskiiruse indeksskoor) olid oluliselt kõrgemad Ameerika omadest. Autorid oletasid valimi skooriprofiilide põhjal, et kallutatud valim tuleneb inimeste vabatahtlikkuse alusel uuringusse värbamise protseduurist. On leitud, et vabatahtlikud on paremate vaimsete võimetega (Rosenthal, & Rosnow, 1975) ning seetõttu võis juhtuda, et vanematesse eagruppidesse saadi küll madalama haridusega, kuid vaimselt siiski võimekamad inimesed. Seejärel korrigeeriti

skooride ja alltestide keskmisi eeldusel, et töötluskiiruse indeks näitab valimi mittekallutatud võimekuse taset ning leiti, et Inglise variandile võib rakendada Ameerika norme.

Kallutatud valim tekitas suuri probleeme Mehhikos WAIS-III adapteerimisel, kus normide loomine toimus mitme aasta vältel neid järk-järgult täiustades (2001, 2003, 2004 aasta normid). Mehhiko WAIS-III normide valiidsust hinnates leidsid Suen ja Greenspan (2009), et Mehhiko keskmine koguskaala skoor oli 12 punkti kõrgem originaalist. Seega võiks Mehhiko testi Ameerika omast lihtsamaks pidada ning seletada erinevust teatud kultuuriliste erinevustega. Kuid esines ka muid olulisi probleeme nagu madal reliaablus, ebamäärane referentspopulatsioon, skooride normaliseerimise puudumine, teatud gruppide väljajäämine standardiseerimisvalimist, ebasobivate statistikameetodite kasutamine. Probleemid Mehhiko normidega kerkisid esile siis, kui kohus soovis surma mõista kurjategijat, kuid seaduses polnud surmanuhtlus lubatud vaimse alaarengu puhul. Mehhiko normide järgi osutus kurjategija intellekt piirialaseks, kuid Ameerika normide järgi ulatus vaimse alaarengu tasemele. Mehhiko WAIS-III autorid tunnistavad, et ressursside puudus ei võimaldanud WAIS-III standardiseerimist Mehhiko populatsiooni juhusliku stratifitseeritud valimi alusel ning soovivad vaimse alaarengu hindamisel kasutada Ameerika norme (Escobedo & Hollingworth, 2009). Mehhiko juhtum näitab küll, et testide adapteerimisel on oluline andmeid koguda populatsiooni esindava valimi põhjal ning kasutada sobivaid analüüsimeetodeid, kuid samas meenutab ka, et ühegi üksiku testitulemus põhjal ei tohi langetada otsust inimese üle. Tõenäoliselt pole ükskõik kui ideaalse adapteerimisprotsessi tulemusena siiski võimalik saavutada olukorda, kus ühe testi tulemus võiks õigustatult määrata inimese saatuse.

Foxcroft ja Aston (2006) andsid kriitilise hinnangu Lõuna-Aafrika WAIS-III adapteerimisele, kus lisaks inglise keelt emakeelena rääkivatele inimestele oli valimisse kaasatud ka inglise keelt teise keelena rääkivaid Lõuna-Aafrika rahvusgruppe. Viimased ei olnud aga oma keelegrupi suhtes sobivalt esindatud. Lisaks valimi puudustele oli inglise keelt teise keelena rääkijate grupis instruksioonide ja ülesannete kohandamine grupi eripärasid arvestades ebapiisav. Selle tulemusena ei suudetud luua testi, mis oleks võrdselt kasutatav erinevates keelegruppides.

Roivainen (2009) võrdles WAIS-III euroopa keeltesse (Saksa, Hispaania, Prantsuse) kohandatud sooritusteste Ameerika normidega, ning leidis, et Ameerika inimesed said võrreldes eurooplastega oluliselt madalamaid tulemusi tajupõhise töötuse indeksskaalal, kusjuures töötuskiiruse indeksskaalal olid tulemused sarnased. Nooremates vanusegruppides oli eurooplastel oluliselt kõrgem soorituslik IQ, kuid üsna võrdne tulemus töötuskiiruse indeksskaalal võrreldes Ameerika inimestega. Vanemates vanusegruppides olid sooritusliku IQ tulemused ligilähedased, kuid eurooplased said madalama tulemuse töötuskiiruse indeksskaalal. Roivanen seletab neid tulemusi kultuuriliste erinevustega.

Lisaks erinevustele kultuuride vahel võivad ka ühe kultuuriruumi sees eri aegadel saadud sama testi tulemused kaotada Flynn'i efekti (Flynn, 1999) tõttu võrreldavuse. Must, Nijenhuis, Must ja Vianen (2009) leidsid, et eri aegadel saadud sama IQ testi skoorid ei ole võrreldavad, kuna alltestide ja g-skoori tähendused on ajas muutuvad.

**Kohandamise etapid.** Aritmeetika kohandamine on mitmeetapiline protsess, mille eesmärk on luua eestikeelses testis sellised üksikülesanded, mille psühhomeetriselised omadused sobiksid originaalsete WAIS-III ülesannetega, moodustades ühemõõtmelise skaala, mille mõõtmise usaldusväärsus ja valiidsus oleks samaväärne originaaltestiga.

Käesolev uuring oli üks etapp Aritmeetika kohandamise protsessis. Selle eesmärk oli luua ja valida välja eestikeelses Aritmeetikasse üksikülesanded, millega oleks võimalik jätkata andmete kogumist suuremal valimil WAIS-III standardiseerimiseks, normeerimiseks ja valideerimiseks ning võimaldaksid testi kasutada kliinilises töös. Käesolev uurimus koosnes omakorda mitmetest etappidest:

- 1) Eestikeelse Aritmeetika alltesti ülesannete kogumi koostamine ja esialgne hindamine väiksemal valimil.
- 2) Ülesannete individuaalne testimine valimil, mille sooline, vanuseline ja hariduslik jaotus sarnaneb Eesti eestikeelse rahvastiku vastavate tunnuste jaotusega 2000. aasta rahvaloenduse põhjal.
- 3) Üksikülesannete analüüs, mille eesmärk oli valida välja sobivad ülesanded ja reastada ülesanded raskuse järjekorda. Selleks, et säiliks WAIS-III

Aritmeetika originaaltesti struktuur ja sisu, lähtuti ülesannete valikul järgmistest kriteeriumitest:

- üksikülesandele õigesti vastamise tõenäosus suureneb võimekuse kasvades;
- üksikülesannete eristusvõime on võimalikult hea;
- üksikülesannete skoorimine on võimalikult lihtne;
- üksikülesannetele vastamise muster võimaldab kolmeastmelist skoorimist 0, 1, 2;
- võimalus testimine katkestada teatud arvu valede vastuste järel
- üksikülesanded peavad olema erineva raskusega, et oleks võimalik hinnata võimeid võimalikult laiaades piirides.

4) Skaala analüüs, mille eesmärgiks oli hinnata eestikeelse Aritmeetika reliaablust ja valiidsust, raskust ning esialgsete skaalaskooride jaotust.

Uute ülesannete loomisel püüti ette näha võimalikke probleeme originaalülesannetega. Kuna Aritmeetika kuulub töömälu indekskaalasse, siis peaksid ülesanded pigem koormama töömälu kui eeldama valemite tundmist ja koolis õpitud teadmisi. Groth-Marnat (1997) järgi mõõdab Aritmeetika muuhulgas ka koolis omandatud teadmisi, kuid nagu mitmed uuringud on kinnitanud, kuulub Aritmeetika koos Arvumälu ja Tähtnumber järjestusega töömälu indekskaalasse (Taub jt., 2004; Saklofske jt., 2000). Samas ei välista hilisem töömälu käsitus faktiteadmiste kasutamist, kuna episoodilise puhvri kaudu on need töömälule kättesaadavad (Baddeley, 2000).

Mitmed originaalülesanded, mis toetuvad protsendi ja tõenäosusteooria tundmisele, eeldavad spetsiifilisi teadmisi. Eestis on protsentarvutus matemaatika õppekavas juba väga pikka aega ning kõik põhiharidusega inimesed peaksid olema seda õppinud. Tõenäosus on põhikooli õppekavas alles alates 1996. aastast ning selle tõttu võiks ülesanne eelistada nooremaid inimesi. Vastavate originaalülesannete mittedobivuse korral peaks olema võimalik need lisaülesannetega asendada. Seetõttu lisati Aritmeetika testi juurde ka erineva raskusega ülesandeid, mis koormaks pigem töömälu ega eeldaks valemite tundmist. Näiteks: *Latist, mille pikkus on 12 m, lõigatakse 3 tükki pikkustega 2 m, 2m 30 cm ja 4 m. Kui pikk tükk jääb alles?*

Erineva raskusega ülesannete loomisel lähtuti kahest põhimõttest – ülesande keerukusest ja raskusest. Keerukus on seotud ülesande omadustega ja sõltub ülesande lahendamiseks vajaminevast ajast (tehete arv) ja mälu mahust, mis on ülesande lahendamisega koormatud – mida rohkem tehteid ülesanne sisaldab ja mida rohkem vahetulemusi on tarvis mälus säilitada, seda keerukam on ülesanne. Ülesande raskus sõltub aga inimese omadustest, eelkõige kognitiivsest võimekusest see ülesanne lahendada. Ülesande raskus arvutatakse ülesannetele vastamiskäitumise põhjal. On leitud, et ülesande raskust mõjutab selle esitamise keeleline struktuur ja kontekst, sõnade arv, eelneva ülesande sisu, tehete hulk (Loftus & Suppes, 1972; Judd & Hickson Bilsky, 1989) ja sõnade pikkus, millega ülesanded esitatakse (Ellis & Hennely, 1980). Baddeley (2004) on leidnud, et sõnaliste töömälu-ülesannete puhul on olulised eelpoolkirjeldatud fonoloogilise sarnasuse ja sõna pikkuse efekt. Näiteks ülesanne *Kunstimuuseumi pilet maksab 29 krooni, ajaloomuuseumi pilet 35 krooni, loodusloomuuseumi pilet 30 krooni. Mitut erinevat muuseumi saaks turist külastada, kui tal on 89 krooni?* võiks olla raske, kuna koormab töömälu, sest meeldejäetavaid ühikuid on palju, ülesanne on esitatud pika teksti ja pikkade sõnadega, kusjuures mitmed sõnad sarnanevad fonoloogiliselt.

Aritmeetika ülesanded on esitatud probleemolukordade kaudu, millega inimesed igapäevaselt kokku puutuvad – keskmise palga ja kiiruse arvutamine, sisseostude tegemine jne. Uute ülesannete loomisel püüti taoline kontekst säilitada nii, et see oleks võimalikult neutraalne ega looks vastamiseeliseid teatud soolisse, vanuselisse või hariduslikku gruppi kuuluvatele inimestele. Näiteks: *Kaup maksis 400 krooni. Seda hinda aladati 10% võrra, mõne aja pärast alandati uut hinda veel 10 % võrra. Kui palju maksis kaup pärast teistkordset hinnaalandust?*

#### **1.4. Töö eesmärgid**

Magistrip projekti eesmärk oli luua ja valida välja eestikeelsesse Aritmeetikasse ülesanded, mis moodustaksid originaaltestiga sisuliselt samaväärsse skaala. Oluline oli, et testi oleks võimalik kasutada ka kliinilises töös. Selleks tuli eestikeelsesse Aritmeetikasse luua ülesanded, mille psühhomeetrilised omadused sobiksid originaalsete WAIS-III ülesannetega, moodustades ühemõõtmelise skaala, mille mõõtmise usaldusväärsus ja valiidsus oleks samaväärsed originaaltestiga. Selleks tuli hinnata

Aritmeetika üksikülesannete sobitumist Rasch`i mudeliga; ülesannete raskusjärjestust; ülesandesiseste lävede raskusjärjestust; soolist, vanuselist ja hariduslikku hälvet ning eristusvõimet. Skaala osas tuli hinnata selle mõõtmelisust; sobitumist Rasch`i mudeliga; skaala sisemist ja poolitustesti reliaablust; valiidsust; raskust, esialgsete skaalaskooride jaotust võrreldes originaaliga.

## 2. Meetod

### 2.1. Valim

Valim moodustus normgrupist ( $N=191$ ; 85 naist ja 106 meest) vanuses 16...89 aastat (keskmise vanus 44,5 a). Valimis olid esindatud 13 vanusegruppi: 16...17 ( $n=7$ ), 18...19 ( $n=7$ ), 20...24 ( $n=19$ ), 25...29 ( $n=19$ ), 30...34 ( $n=18$ ), 35...44 ( $n=29$ ), 45...54 ( $n=31$ ), 55...64 ( $n=26$ ), 65...69 ( $n=10$ ), 70...74 ( $n=10$ ), 75...79 ( $n=7$ ), 80...84 ( $n=4$ ), 85...89 ( $n=4$ ). Hariduse osas olid esindatud 5 kategooriat: algharidus ( $n=21$ ), põhiharidus ( $n=44$ ), keskharidus ( $n=57$ ), kutseharidus ( $n=40$ ), kõrgharidus ( $n=29$ ). Normgrupil ei esinenud häireid ega haigusi, mis oleks võinud mõjutada kognitiivseid võimeid ja seeläbi testitulemusi. 2000.a rahvaloenduse andmete põhjal moodustati normvalimi plaan (Burk, 2007), mis esindaks Eesti eestikeelset elanikkonda soo, vanuse ja hariduse osas. Katseisikud värvati uuringusse vastavalt valimiplaanile. Testitud valimi üksiktunnuste vastavust etteantud valimiplaanile kontrolliti  $\chi^2$ -testiga, mis näitas, et tegeliku ja planeeritud valimi jaotus ei erine statistiliselt oluliselt soo [ $\chi^2(1)=,02$ ;  $p=,88$ ], vanuse [ $\chi^2(12)=1,43$ ;  $p=,10$ ] ja hariduse [ $\chi^2(4)=,13$ ;  $p=,10$ ] poolest.

Lisaks tervetele inimestele testiti Aritmeetikaga ka väike kliiniline valim ( $N=20$ ), mis koosnes kerge ( $n=4$ ) ja mõõduka ( $n=16$ ) vaimse alaarengu diagnoosiga inimestest.

### 2.2. Mõõtmisvahendid

**Aritmeetika.** Käesolevas uuringus kasutati Aritmeetika eestikeelset prooviversiooni, mille koostamisel säilitati kõik 20 originaalülesannet. Originaal-ülesanded ja testi instruksioon tõlgiti eesti keelde ning lisati 20 uut ülesannet (edaspidi nimetatud lisaülesanded). Originaal-ülesannete tõlkimisel säilitati ülesannete arvulised



ühikud, muudeti mõõtühikuid (naelad tõlgiti kroonideks, miilid kilomeetriteks) ning muudeti esemeid nii, et need sobiksid kokku arvuliste väärtustega.

Näiteks: *A coat that normally sells for 60 pounds is reduced by 15 per cent during a sale.*

*What is the price of the coat during the sale?*

*Tavaliselt 60 krooni maksva särgi hinda alandatakse 15% võrra. Kui palju maksab särk pärast allahindlust?*

Lisaülesannete loomisel lähtuti erineva raskusega ülesannete tagamiseks järgmistest põhimõtetest:

- liitmine ja lahutamine on lihtsamad tehted kui korrutamine ja jagamine;
- mida rohkem ühikuid ülesandes esitatakse, seda keerukam ja raskem on ülesanne;
- mida rohkem operatsioone on vaja teha ülesande lahendamiseks, seda keerukam ja raskem on ülesanne;
- täisarvudega arvutamine on vähem keerukas kui kümnendmurdudega;
- pikad ja fonoloogiliselt sarnased sõnad muudavad ülesanded raskemaks;
- ülesannete lahendamine ei tohi eeldada spetsiifiliste teadmiste olemasolu (valemite tundmist);
- kontekst, milles ülesanded esitatakse, peavad olema seotud igapäevaste olukordadega.

Seejärel paluti ülesandeid ja instruksiooni hinnata 6 eksperdil, kes kuulusid WAIS-III kohandamise töörühma (1 doktorikraadiga matemaatik, 2 doktorikraadiga psühholoogi, 1 magistrikraadiga psühholoog ja 2 psühholoogiamagistranti). Ülesandeid korrigeeriti lähtuvalt ekspertide kommentaaridest (parandati sõnastust, lisati raskemaid ülesandeid).

Järgmiseks hinnati ülesannete esialgseid raskusi normvalimil. Selleks koguti andmeid 34 isikult (keskmine vanus 40,5 a; vanusevahemik 17...75 a, 17 meest ja 17 naist), kelle seas olid esindatud kõik haridustasemed (alg-, põhi-, kesk-, kutse-, ja kõrgharidus). Ülesannete raskusindeksid (ülesandele õigesti vastanute arv jagatud kõigi vastanute arvuga) varieerusid vahemikust ,15...1, mis lubas ülesandeid pidada sobilikuks, et jätkata andmete kogumist.

Aritmeetikas tuli katseisikul peast lahendada aritmeetilisi ülesandeid. Aritmeetika eestikeelses prooviversioonis oli kokku 40 ülesannet, millest esitati katseisikule suuliselt

36. Esimest 4 loendamisülesannet normgrupile ei esitatud, kuna need olid mõeldud kliinilisele grupile ja normgrupi testimisel loendamises vigu ei eeldatud. Seetõttu alustati normgrupi testimist viiendast ning kliinilise grupi testimist esimesest ülesandest. Katseisik pidi ülesandeid lahendama peast, abivahendeid kasutamata. Peale ülesande ettelugemist hakati võtma aega. Vajadusel võis ülesande ettelugemist korrata, kuid selle käigus aega ei katkestatud. Aritmeetika testis oli kõigi ülesannete ajalimiidiks 2 minutit. Kui testitav ei suutnud selle aja jooksul ülesandele vastata, mindi edasi järgmise ülesandega. Vastuseid skooriti kolme punkti skaalal, kus 0 punkti sai ebaõige vastuse eest, 1 punkti sai ajalimiidi sees antud õige vastuse eest ning 2 punkti kahe raskeima ülesande eest, millele vastati õigesti 10 sekundi jooksul.

**Arvumälu.** Arvumälu alltest (edaspidi Arvumälu) kuulub WAIS-III sõnalisse skaalasse ja töömälu indeksskaalasse ning viidi läbi Aritmeetika valiidsuse hindamiseks. Arvumälu koosnes kahest osast, millest esimene osa mõõtis sõnalist lühimälu ning teine osa sõnalist töömälu. Esimene osa koosnes 16 ülesandest, milles esitati katseisikule suuliselt numbrite ridu edaspidises järjekorras. Katseisik pidi numbreid samas järjekorras kordama. Teises osas oli 14 ülesannet, milles esitati samuti numbrite ridu, kuid katseisik pidi neid kordama tagurpidises järjekorras. Iga õigesti vastatud ülesanne andis 1 punkti, kokku oli võimalik saada 30 punkti.

**Sarnasused.** Sarnasuste alltest (edaspidi Sarnasused) kuulub WAIS-III sõnalisse skaalasse ning viidi läbi Aritmeetika valiidsuse hindamiseks. Sarnasustes esitati testitavale suuliselt kaks sõna, millele ta pidi leidma selgituse, mille poolest nende sõnadega tähistatavad mõisted on sarnased. Näiteks: *Mille poolest on sarnased maal ja kuju?* Eesti keelde kohandatud Sarnasuste vastuseid skooriti lähtuvalt skoorimisjuhendist kolmepunktilisel (0,1,2) skaalal, kolm kõige lihtsamat ülesannet skooriti kahepunktilisel skaalal (0,1).

**Taibukus.** Taibukuse alltest (edaspidi Taibukus) kuulub WAIS-III sõnalisse skaalasse ja seda kasutati samuti Aritmeetika valiidsuse hindamiseks. Taibukuses esitati testitavale suuliselt küsimusi sotsiaalsete reeglite, ühiskonna toimimise ja igapäevaeluliste olukordade lahendamise kohta. Näiteks: *Milleks kasutatakse raha?* Vastuseid skooriti skoorimisjuhendi abil, enamuses ülesannetes kolmepunktilisel skaalal (0,1,2), kolme kõige lihtsama ülesande puhul kasutati kahepunktilist skaalat (0,1).

Kuna andmete kogumise etapis ei olnud Sarnasused ja Taibukus veel eesti keelde kohandatud, siis koguti andmeid nende skaalade proovivariantidega (Sarnasused sisaldas 40 ja Taibukus 35 ülesannet). Aritmeetika valiidsuse analüüsis kasutati eesti keelde kohandatud Sarnasusi ja Taibukust (Burk, 2007), mis sisaldasid vastavalt 19 ja 18 ülesannet.

### **2.3. Testijad ja testimisprotseduur**

Andmeid koguti 2004. – 2008.a. katseisikuid individuaalselt testides. Testijateks olid psühholoogiatudengid ja praktiseerivad psühholoogid (kokku 22 testijat), kellest igaüks läbis eelnevalt testimistreeningu, kus harjutati WAIS alltestide nõuetekohast läbiviimist ning analüüsiti ilmnenu vigu. Testijad otsisid katseisikuid vastavalt valimiplaanile enda ja oma tuttavate tutvusringkonnast, tingimusel, et rohkem kui 1/3 katseisikutest ei oleks testija sugulased ega isiklikud tuttavad.

Katseisikutega viidi läbi neli sõnalist alltesti järjekorras: Arvumälu, Aritmeetika, Sarnasused ja Taibukus. Võimalusel viidi testid läbi ühe testimiskorraga pikemate pausideta, kuid vajadusel ka kahe testimiskorraga. Alltestide läbiviimisel originaaltesti alustamise ja katkestamise reegleid ei rakendatud, igale katseisikule esitati kõik ülesanded. Enamasti kulus nelja alltesti läbiviimiseks 90...120 minutit.

### **2.4. Andmeanalüüs**

Andmeanalüüsil kasutati üksikvastuste teooria ja klassikalise testiteooria meetodeid. Analüüsil oli kaks peamist eesmärki:

- 1) Aritmeetika üksikülesannete hindamine. Analüüsi käigus hinnati ülesannete sobitumist Rasch`i mudeliga; ülesannete raskusjärjestust; ülesandesiseste lävede raskusjärjestust; eristusvõimet; soolist, vanuselist ja hariduslikku hälvet.
- 2) Aritmeetika skaala hindamine. Analüüsi käigus hinnati skaala mõõtmelisust; sobitumist Rasch`i mudeliga; sisemist ja poolitustesti reliaablust; valiidsust. raskust, skaalaskooride ulatust.

**Analüüsi etapid.** Analüüs teostati järk-järgult peamiselt nelja etapina:

- 1) Esimeses etapis kontrolliti, kas andmestik on ühemõõtmeline ning seejärel analüüsiti iga testiküsimuse nelja erineva ajalimiidiga varianti (15, 30, 60 ja

120 sekundit), mis skooriti dihhotoomselt 0 ja 1 punktiga. Seega sisaldas andmestik 160 ülesannet. Analüüsi käigus eemaldati ebasobivad ülesanded ning loodi kolmest raskeimast ülesandest mitmeid polütoomseid ülesandeid erinevate ajalimiitidega.

- 2) Teises etapis korrati järelejäänud ülesannetega Rasch`i analüüsi ja valiti välja kaks polütoomset ülesannet.
- 3) Kolmandas etapis tehti uuesti Rasch`i analüüs järelejäänud Aritmeetika ülesannetele ja skaalale, arvutati skaala reliaabluskordajad, hinnati Aritmeetika raskust ja skaalaskooride jaotust.
- 4) Neljas etapp hõlmas Aritmeetika valiidsuse hindamist.

**Rasch`i mudelist lähtuv analüüs.** Käesolevas töös kasutati analüüsiks programmi RUMM2020 (*Rasch Unidimensional Measurement Model*, Andrich, Sheridan & Luo, 2003), mis rakendab Rasch`i lihtsat logistilist mudelit (Rasch, 1960). Rasch`i mudeli eelised klassikalise testiteooria ees on, et Rasch`i mudel võimaldab hinnata vastaja omaduse taset ja ülesande omadusi üksteisest sõltumatult, pakkudes võimalusi üksikülesannete täpsemaks hindamiseks, ega eelda selleks populatsiooni suhtes esindusliku valimi olemasolu (Embretson & Reise, 2000). Rasch`i mudel on unikaalne teiste IRT- mudelite seas, kuna võimaldab viia isikute ja ülesannete parameetrid ühisele intervallskaalale. Samas tuleb Rasch`i tulemusi hinnata kriitiliselt koos teiste meetoditega (klassikaline ja kvalitatiivne analüüs), mitte eelistada ühte meetodit pimesi teisele. Eesti keeles võib üksikvastuste teooria kohta lugeda Aire Raidvee bakalaureusetööst (Raidvee, 2003 ) ja Kristo Burki magistriprojektist (Burk, 2007).

Rasch`i mudeli rakendamiseks kontrolliti andmestiku ühemõõtmelisuse eeldust (Wright & Stone, 1999). Andmeid võis pidada ühemõõtmeliseks, kui jääkide ükski peakomponent ei seletanud üle 30% hajuvusest (Hair & Anderson, 1998) ja isikute eraldusindeks (ik *person separation index*, PSI) oli võimalikult 1 lähedal.

Ülesannete mudeliga sobitumise hindamiseks kasutati standardiseeritud jääkide suuruse määramist ja  $\chi^2$ -testi. Ülesanne sobitus mudeliga, kui standardiseeritud jäägi absoluutväärtus ei ületanud 2,5. Originaalülesandeid, mille standardiseeritud jääk oli lubatust suurem, kuid mille dispersioon oli väga väike (vähemalt 90% testititest vastas

õigesti) ainult jäägi väärtuste põhjal ebasobivaks ei peetud.  $\chi^2$ -testis korrigeeriti valepositiivsete tulemuste vältimiseks usalduse nivood 0,05 Bonferroni järgi väärtusele ,003 ( $=0,05/n$ , kus  $n = 17$  on ülesannete arv alltestis).

Ülesannete raskuse hindamiseks arvutas programm RUMM2020 vastavad logaritmilise šansisuhte ühikud ehk logitid.

Lävede ülesande-sisese raskusjärjestuse hindamiseks vaadati lävede järjekorra õiget asetust ehk esimene lävi pidi olema kergem kui teine lävi.

Ülesannete hälvete (ik *differential item functioning*, DIF) visuaalsel hindamisel võrreldi gruppide testiküsimuste kõveraid ja statistilisel hindamisel rakendati standardiseeritud jääkidele dispersioonanalüüsi. ANOVA F-statistiku olulisuse nivoo 0,05 korrigeeriti Bonferroni järgi väärtusele ,003 ( $0,05/n$ ,  $n$  on ülesannete arv alltestis). DIFi hindamiseks moodustati 4 intrevallklassi. Haridusliku DIF-i uurimiseks moodustati võrreldavatesse gruppidesse piisava arvu isikute tagamiseks kaks gruppi – madalama haridusega (sisaldas alg- ja põhiharidusega inimesi) ning kõrgema haridusega (sisaldas kesk-, kutse-, ja kõrgharidusega inimesi). Vanuselise DIF-i uurimiseks jagati inimesed 4 vanusegruppi (16-29, 30-44, 45-64, 65-89). Vanusegrupid moodustati nii, et igas vanusegrupis oleks piisavalt erineva võimekusetasemega inimesi.

Eestikeelse Aritmeetika skaala sobitumist mudeliga kontrolliti  $\chi^2$ -testiga, milles korrigeeriti valepositiivsete tulemuste vältimiseks usalduse nivood ,05 Bonferroni järgi väärtusele ,003 ( $,05/n$ , kus  $n$  on ülesannete arv alltestis).

**Klassikalisest testiteooriast lähtuv analüüs.** Igale küsimusele arvutati normvalimi põhjal raskusindeksid (ülesandele õigesti vastanute arv jagatud kõigi vastanute arvuga). Lihtsamate ülesannete hindamiseks, mille raskusindeks normgrupi põhjal tuli 1, kasutati kliinilise grupi õigesti vastamise sagedust.

Eristusvõime hindamiseks arvutati eristusindeks  $D$  ja hinnati testiküsimuse korrelatsiooni koguskooriga. Ülesandeid, mille  $D$  väärtus oli madal ja ka variatiivsus oli väike (90% või enam inimesi vastas õigesti) ainult madala  $D$  väärtuse tõttu ebasobivaks ei peetud. Testiküsimuse korrelatsiooni hindamiseks koguskooriga kasutati dihhotoomsete ülesannete puhul punkt-biseriaalset korrelatsiooni, kahe polütoomse

ülesande puhul Spearmani korrelatsiooni. Madala korrelatsiooni ja variatiivsusega ülesandeid esialgu testist välja ei jäetud.

Eestikeelse Aritmeetika raskuse võrdlemiseks originaaliga, leiti eestikeelse Aritmeetika toorskooridele Ameerika normide järgi skaalaskoorid ja arvutati nende skaalaskooride keskmine.

Aritmeetika esialgsete skaalaskooride leidmiseks hinnati kõigepealt eestikeelse Aritmeetika toorskooride normaaljaotusele vastavust Kolmogorov-Smirnovi testiga. Seejärel normaliseeriti toorskooride protsentiiljaotused nelja vanusegrupi kaupa ning arvutati normaliseeritud skooridele skaalaskoorid.

Aritmeetika sisereliaabluse hindamiseks kasutati Cronbach`i alfati, mis arvutati nii kõigi katseisikute kui nelja vanusegrupi jaoks. Poolitustesti reliaabluse hindamiseks reastati lõpliku Aritmeetika ülesanded raskuse järjekorda ning jagati pooleks nii, et ühe poole moodustasid paarisarvulise järjekorranumbriga ülesanded ja teise poole paaritu järjekorranumbriga ülesanded. Seejärel leiti Pearsoni korrelatsioon kahe poole summaskooride vahel kõigi katseisikute ja nelja vanusegrupi jaoks.

Aritmeetika valiidsuse hindamiseks arvutati Aritmeetika ning Arvumälu koguskoori, edaspidise ja tagurpidise osa vahel, Sarnasuste ja Taibukuse vahel Pearsoni korrelatsioonikordajad. Originaal-alltestide ja eestikeelsete alltestide korrelatsioonide erinevusi hinnati korrelatsioonide Fisheri z-skooride erinevuste standardvea ühikutele viimise kaudu.

### 3. Tulemused

#### 3.1. Rasch`i mudelist lähtuv analüüs

**Mudeli eeldused.** Analüüsi I etapis kaasati analüüsi Aritmeetika alltesti 141 küsimust 160st. Välja jäid need, millele vastamise varieeruvus oli 0. Standardiseeritud sobitusjääkide peakomponent-analüüs näitas, et esimene komponent seletas ära 5,03% varieeruvusest. Isikute eraldusindeks (ik *person separation index*) oli ,98.

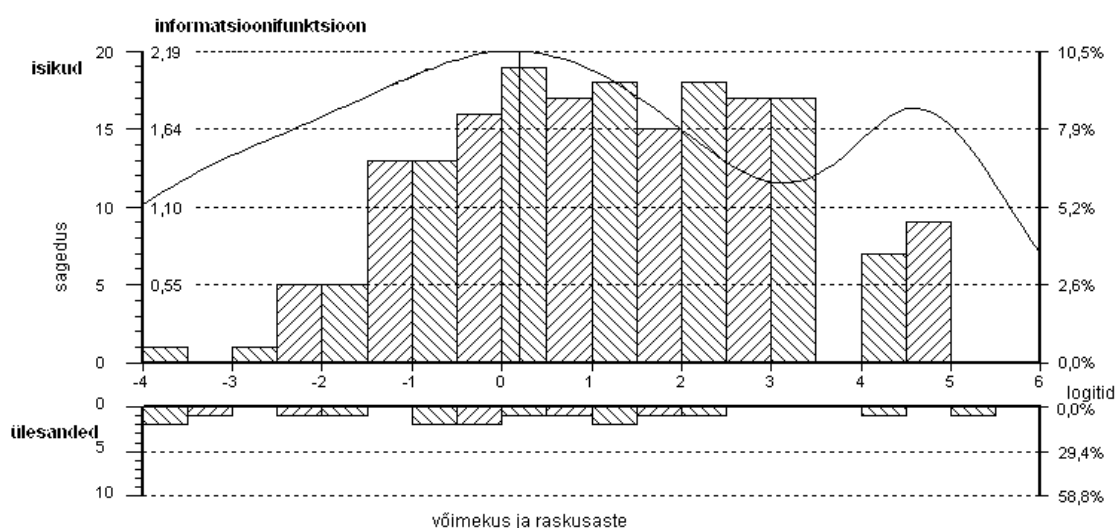
Analüüsi III etapis viidi uuesti läbi analüüs väljavalitud testiülesannete standardiseeritud sobitusjääkidele peakomponentide meetodil ning siis seletas esimene komponent ära

10,34% testiküsimuste varieeruvusest. Isikute eraldusindeks oli ,98. Seega võib nii algset kui väljavalitud ülesannete andmestikku pidada ühemõõtmeliseks.

**Ülesannete sobitumine Rasch`i mudeliga.** Analüüsi I etapis oli ühe originaalülesande sobitusjääk piirialane ( $-2,42, p = ,04$ ) ning jäeti seetõttu Aritmeetikast välja (*Kui te ostate seitse 20 sendist kommi ja annate müüjale 5 krooni, siis kui palju raha saate te tagasi?*). Hilisematel analüüsi etappidel sobitusid kõik originaalülesanded rahuldavalt Rasch`i mudeliga. Aritmeetika väljavalitud ülesannetekogu sobitusjäägid on toodud tabelis 1.

**Ülesannete raskusjärjestus.** Analüüsi I etapi tulemustest ilmnes, et eesti keelde tõlgitud originaalülesanded ei erinenud Eesti valimil ühtlaselt raskusastmete poolest. Seetõttu jäeti välja neli originaalülesannet. Aritmeetika testi lisati kuus lisäülesannet eesmärgiga tagada ülesannete raskusastmete võrdse sammuga jaotumine, asendada väljajäetud originaalülesandeid ning luua kolm polütoomset ülesannet. Aritmeetika väljavalitud ülesannete raskusastmed on toodud tabelis 1.

Aritmeetika originaalülesannete eestikeelse variandi suhteline raskusjärjekord ei ühtinud täielikult originaaltesti järjekorraga. Eestikeelse Aritmeetika lihtsamad ülesanded olid originaaltestiga võrreldes samas järjekorras, 6 ülesannet osutusid Eesti inimestele suhteliselt raskemaks ning neli ülesannet suhteliselt kergemaks.



Joonis 1. Eestikeelse Aritmeetika ülesannete raskuse ja isikute võimekuse jaotus

Analüüsi III etapis ilmnes, et väljavalitud Aritmeetika ülesannete seas puudusid sobiva raskusega ülesanded teatud võimekuse tasemel inimeste eristamiseks (joonis 1). Joonisel 1 on näha ülesannete informatsioonifunktsiooni kõver, mis näitab ülesannete eristusvõimet erinevatel võimekuse tasemetel.

**Lävede järjestus.** Analüüsi II etapis ilmnemise kolmest polütoomsest ülesandest ühel pööratud läved ehk esimese läve raskus oli raskem kui teise läve raskus (*6 nõela maksab 2 krooni ja 10 senti. Kui palju maksab 13 nõela?*). Seega otsustati see ülesanne edasisest analüüsist välja jätta. Ülejäänud kahe polütoomse ülesande lävede raskusjärjekord oli sobiv.

**Ülesannete hälve.** Analüüsi I etapis ei ilmutanud ükski originaalülesanne statistiliselt olulist soolist, hariduslikku ega vanuselist hälvet. Samuti ei esinenud hälvet lisaülesannetes, mis valiti väljajäetud originaalülesandeid asendama ning raskusastmelist jaotuvust parandama.

Analüüsi III etapis, kus arvutati uuesti statistikud Aritmeetika väljavalitud ülesannete kogumile, ilmnes väike ühtlane hariduslik hälve ühe ülesande puhul [ $F(df)=10,52; p = ,001$ ], mis eelistas madalama haridusega inimesi. Ülesanne tuli siiski säilitada, kuna polnud ühtegi sarnase raskusastmega ülesannet, mis oleks võinud seda asendada.

**Skaala sobitumine Rasch`i mudeliga.** Eestikeelse Aritmeetika skaala sobitus mudeliga rahuldavalt [ $\chi^2(34)=38,84; p=,26; \alpha=,003$ ].

### 3.2. Klassikalisest testiteooriast lähtuv analüüs

**Ülesannete raskusjärjestus.** Ülesannete normvalimil leitud raskusindeksid varieerusid eestikeelse Aritmeetika väljavalitud ülesannetel vahemikus ,15...1 ning olid jaotunud raskuse kasvamise järjekorras üsna ühtlaste vahedega.

Ülesannetele õigesti vastamise sagedus kliinilises valimis järgis samuti valdavalt ülesannete raskuse kasvamise järjekorda. See tähendab, et Rasch`i ja klassikalise analüüsi põhjal raskematele ülesannetele vastas kliinilises grupis järjest vähem inimesi õigesti. Erandiks olid kaks ülesannet (ülesanded 4 ja 5), mille normaalvalimi Rasch`i mudeli raskusastmed ei sobinud kliinilise valimi vastamissagedustega (tabel 1). Ülesannete järjestamisel lähtuti nende ülesannete puhul kliinilise valimi vastamissagedustest. Kaks



originaalülesannet, millele normaalvalimi isikud 100%-liselt õigesti vastasid, olid kliinilises valimis sama vastamise sagedusega ning üks neist jäeti eestikeelse Aritmeetika ülesannetekogust välja. (Asetage seitse kuupi testitava ette ja öelge: Kui teil on 7 kuupi ja 2 kuupi võetakse ära (eemaldage 2 kuupi), siis mitu kuupi jääb teile alles?).

**Tabel 1.** Eestikeelse Aritmeetika ülesannete statistikud

Ül	Kategoriatele vastamise protsent			Rasch'i mudeli statistikud				Klassikalise testi-teooria statistikud		
	Kat 0	Kat 1	Kat 2	rask.	jääk	$\chi^2$	p	rask.	D	r-pb
1	0	100 (100)						1		
2	0 (13,6)	100 (86,4)						1		
3	0 (27,3)	100 (72,7)						1		
4	1 (59,1)	99 (40,9)		-3,83	,41	3,76	,15	,99	<b>,02</b>	<b>,01</b>
5	2 (81,8)	98 (18,2)		-3,99	-,42	,65	,73	,98	<b>,20</b>	<b>,17</b>
6	2,6 (90,9)	97,4 (9,1)		-3,26	-,02	4,81	,09	,97	<b>,20</b>	<b>,14</b>
7	6,3 (77,3)	93,7 (22,7)		-2,44	,03	2,97	,23	,94	<b>,15</b>	<b>,27</b>
8	11,5 (90,9)	88,5 (9,1)		-1,88	-,90	,76	,69	,88	,35	,48
9	19,4	80,6		-,96	-,37	1,17	,56	,81	,50	,52
10	24,6	75,4		-,60	,25	1,66	,44	,75	,63	,51
11	28,3	71,7		-,39	-1,12	2,13	,35	,72	,72	,62
12	34,0	66,0		-,10	-1,28	4,60	,10	,66	,78	,65
13	38,7	61,3		,36	,19	1,02	,60	,61	,60	,58
14	47,1	52,9		,82	,29	3,05	,22	,53	,86	,65
15	50,3	49,7		1,00	-,80	4,50	,11	,50	,73	,66
16	57,6	42,4		1,46	-,19	1,25	,54	,42	,73	,61
17	62,8	37,2		1,84	-,31	1,67	,43	,37	,68	,61
18	70,2	29,8		2,35	-,10	,42	,81	,30	,67	,58
19	84,8	13,6	1,60	4,46	-,98	1,12	,57	,15	,42	,52
20	84,3	15,2	,50	5,15	-1,10	3,30	,19	,16	,44	,56

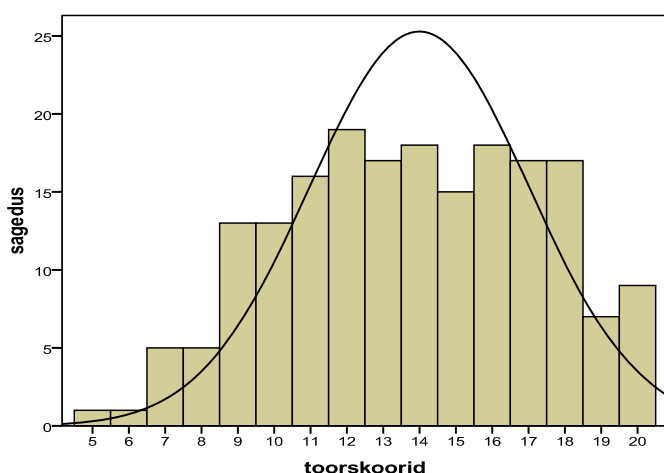
Märkused: N=191, sulgudes on kliinilise valimi vastuskategoriate jaotus protsentides, tumedalt on trükitud ebarahuldava suurusega statistikud (<0,3), Kat – vastuskategooria, rask. – ülesande raskus., D – eristusindeks, r-pb – ülesande-testi korrelatsioonikordaja (ülesannete 4-20 punkt-biseriaalne; ülesannete 19-20 korral Spearmani rho)

Normvalimi raskusindeksid ja kliinilise valimi ülesannetele vastamise sagedused on toodud tabelis 1.

**Ülesannete eristusvõime.** Üksikülesannete eristusindeks  $D$  näitas enamuse ülesannete puhul piisavat eristusvõimet varieerudes vahemikus ,02... ,86 (tabel 1), olles rahuldav kuni väga hea nende ülesannete puhul, mis olid piisava varieeruvusega. Mõned kergemad ülesanded, millele vastas enamusi inimesi (90% või enam normaalvalimis) õigesti, olid madala eristusvõimega. Punkt-biseriaalne ja Spearmani korrelatsioonikordaja varieerusid vahemikus ,01... ,66 (tabel 1) ning olid rahuldavad kuni väga head ülesannete puhul, mille varieeruvus oli piisav. Nelja madala varieeruvusega ülesande puhul oli korrelatsioon ebapiisav.

**Skaala raskus.** Eestikeelse Aritmeetika toorskooridele Ameerika normide järgi leitud skaalaskooride keskmine oli 10,41 (95% usalduspiirid 10,04...10,83) ja standardhälve 2,63 (95% usalduspiirid 2,38...2,85).

**Skaalaskoorid.** Kolmogorov-Smirnovi testi kohaselt ei vastanud eestikeelse Aritmeetika toorskoorid normaaljaotusele ( $D=,093$ ;  $p=,000$ ; asümmeetrianäitaja -,121; järsakus -,80; toorskooride keskmine 13,7 ja mediaan 14,0). Toorskooride normaliseeritud protsentiiljaotusele leitud skaalaskoorid varieerusid järgmiselt: 16- kuni 29- aastaste grupis 4...14, 30- kuni 44- aastaste grupis 4...16, 45- kuni 64- aastaste grupis 5...16, üle 65- aastaste grupis 4...16.



**Joonis 2.** Eestikeelse Aritmeetika toorskooride jaotus

**Sisereliaablus.** Eestikeelse Aritmeetika sisereliaablusnäitaja Cronbach'i alfa oli ,83, varieerudes erinevates vanusegruppides vahemikus ,74...84 (tabel 2). Poolitustesti reliaablusnäitaja oli ,84 varieerudes erinevates vanusegruppides vahemikus ,77...87 (tabel 2).

**Tabel 2:** Eestikeelse Aritmeetika toorskooride keskmised, standardhälbed, sisereliaablus ja poolitustesti reliaablus ning originaaltesti poolitustesti reliaablus erinevates vanusegruppides ja kogu grupis

	N	M	SD	$\alpha$	r-pt	Orig. r-pt
A (16-29)	52	15,5	3,37	,84	,86	,87-,89
A (30-44)	47	14,1	3,30	,81	,77	,88-,90
A (45-64)	57	13,7	2,99	,74	,78	,90-,91
A (>65)	35	11,0	3,12	,81	,87	,77-,91
A kokku	191	13,8	3,51	,83	,84	,88

Märkused: A-Aritmeetika, N-inimeste arv grupis, M-keskmine Aritmeetika toorskoor, SD-standardhälve,  $\alpha$ -Cronbachi alfa, r-pt – poolitustesti reliaablus, Orig. r-pt – originaaltesti poolitustesti reliaablus

**Valiidsus.** Eestikeelse Aritmeetika Pearsoni korrelatsioonid Arvumälu, Sarnasuste ja Taibukusega on toodud tabelis 3. Kõik korrelatsioonid olid statistiliselt olulised ( $p=,01$ ). Originaal-alltestide ja eestikeelsete alltestide korrelatsioonide erinevused ei osutunud statistiliselt olulisteks ( $p=,05$ ).

**Tabel 3:** Eestikeelsete ja originaal-alltestide omavahelised Pearsoni korrelatsioonid

	A	AM	AM-e	AM-t	S
AM	,58 (,52)				
AM-E	,48	,90			
AM-T	,56	,90	,63		
S	,61 (,57)	,40 (,40)	,35	,36	
T	,59 (,57)	,42 (,39)	,39	,38	,64 (,70)

Märkused: N=191;  $p=,01$ ; A-Aritmeetika, AM - Arvumälu, AM-e – Arvumälu edaspidised arvuread, AM-t – Arvumälu tagurpidised arvuread, S – Sarnasused, T – Taibukus, sulgudes on originaal-alltestide vastavad korrelatsioonid

#### 4. Arutelu

Eestikeelsesesse Aritmeetikasse õnnestus enamjaolt leida sobivad ülesanded, millest moodustuva skaala reliaablus ja valiidsus osutusid sarnasteks originaaltestiga. Siiski ilmnes ülesannetes mitmeid probleeme, mille tõttu tuleb käesolevat eestikeelse Aritmeetika skaalat täiendada.

Aritmeetikas ei sobitunud Rasch`i mudeliga vaid üks originaalülesanne.

Lisaülesannetest sobitusid mudeliga valdav enamus, pakkudes head valikut ka muude tunnuste poolest ebasobivate originaalülesannete asendamiseks. Eestikeelse Aritmeetika skaala sobitus Rasch`i mudeliga rahuldavalt.

Originaalülesannete raskusastmeline varieeruvus oli rahuldav, ulatudes väga kergetest ülesannetest rasketeni. Kõige raskemaks ülesandeks osutus originaalülesanne (tõenäosuse ülesanne), mis originaaltestis oli raskuselt eelviimane. See ülesanne võis raskeks osutada selle tõttu, et eeldas tõenäosuse arvutamise valemi tundmist. Kuna tõenäosus on põhikooli õppekavas alles aastast 1996, siis võimalik, et suur osa inimesi ei vastanud ülesandele õigesti just selle tõttu, et ei tundnud või mäletanud vastavat valemit. Seda arvamust toetavad ka testitavate kommentaarid ülesandele („*Ma ei tea tõenäosust.*“, „*Ma juba koolis tõenäosust ei osanud*“, „*Tõenäosus, mis see on?*“). Seega osutusid lisaülesannete loomise etapis tekkinud kahtlused tõenäosusülesande sobilikkuse osas eesti keelt kõnelevatele inimestele osaliselt õigeks, arvestades kommentaare, mida ülesanne testitavates esile kutsus. Teisalt ei toeta Rasch`i analüüs ülesande sobimatust ning ei näita vanuselist ega hariduslikku erinevat toimet. Sellel ülesandel võib isegi teatud väärtus olla teiste ülesannet seas – olles kõige raskem ülesanne, võimaldab see eristada kõrgema võimekusega inimesi. Võiks arvata, et taibukamad suudavad tõenäosuse probleemi ka ise loogiliselt lahendada ega pea ilmtingimata koolis õpitut meenutama. Samas ei muuda seda ülesannet raskeks ainult vastava valemi mittetundmine, kuna ülesanne koormab oluliselt ka töömälu. Ülesandes on palju ühikuid, see on esitatud pikkade sõnadega ja esineb fonoloogilist sarnasust (Baddeley, 2003) – sõna “kirjaklamber” kordub 5 korda. Ülesande lahendamiseks vajalikud detailid võisid meelest minna juba enne kui ülesande lahendamiseks vajalikule valemile mõtlema hakati. Seda oletust toetavad samuti testitavate kommentaarid („*Läks meelest*“, „*Ei suuda keskenduda*“). Hoolimata ülesande sobivuse vaieldavusest testi eestikeelse versiooni jaoks, otsustati see siiski säilitada, kuna asemele polnud pakkuda piisavalt rasket alternatiivi.

Aritmeetika alltestis esines erineva raskusega ülesandeid, mis võimaldavad mõõta nii madalama kui ka kõrgema võimekusega inimesi (joonis 1). Esines ka piisavalt lihtsaid ülesandeid vaimse alaarengu hindamiseks. Üks võimekuse ala esines siiski, mille täpsemaks mõõtmiseks sobivad ülesanded puudusid – seega võib sellesse vahemikku

kuuluva võimekusega inimene saada madalama tulemuse kui tema tegelik võimekus lubaks. Ühest küljest ei ole see testi kliinilise rakendamise seisukohalt väga suur probleem, kuna puuduvad ülesanded just võimekamate inimeste mõõtmiseks. Kliinilises töös kasutatakse testi pigem madalama võimekusega inimeste hindamiseks. Teisalt oleks vaja lisada eestikeelsesesse Aritmeetikasse vastava raskusega ülesandeid ning koguda andmeid juurde. Lõplikult selgub selle probleemi ulatus siis, kui WAIS-III adapteerimise järgmises etapis kogutakse andmeid suuremalt valimilt kogu testipaketiga ning arvutatakse Aritmeetika alltesti normtabelid koos toorskooridele vastavate skaalaskooridega. Kui skaalaskoorid progresseeruvad 1 punkti kaupa, siis mõjutab ülalpool mainitud puudus testitulemusi vähe, kuid progresseerudes rohkem kui 1 punkti kaupa jäävad võimekusevahemikud, mida test ei mõõda ning need tuleks täita sobivate ülesannetega.

Toorskooride kergelt negatiivse asümmeetrilise jaotuse põhjal võiks eestikeelsele Aritmeetikale ette heita seda, et raskeid ülesandeid on vähe. Siiski peegeldab Rasch'i mudeli informatsioonifunktsiooni tipp, et skaala ülemises osas on eristusvõime rahuldav. Käesoleva valimi põhjal eestikeelses Aritmeetikas laaeefekti ei ilmnenud (maksimaalset skoori ei saanud ükski inimene). Üksikutele ülesannetele vastamise mustrit vaadates võib laaeefekti risk siiski esineda kui andmeid kogutakse suuremal valimil. Nimelt leidis käesolevas valimis 2 inimest, kes said kahe raskeima ülesande eest ka boonuspunkte, kuid eksisid mõnes lihtsamal ülesandes. Lihtsamates ülesannetes eksimine võis nende inimeste puhul olla juhuslik ning suuremas valimis võiks leiduda inimesi, kes saavad maksimumpunktid. Seega on oluline edaspidise adapteerimise jaoks luua juurde raskemaid ülesandeid.

Originaalülesannete raskused ei olnud jaotunud ühtlaste vahedega ning ühtlasema jaotumise eesmärgil tuli eestikeelsest Aritmeetikast välja jätta neli originaalülesannet, mis asendati sobivate lisaülesannetega. Ülesannete raskusjärjekorda seadmisel lähtuti valdavalt Rasch'i raskusastmest ja klassikalise testiteooria raskusindeksist, mis enamasti olid kooskõlas ning kergemate ülesannete puhul sobisid ka kliinilise valimi ülesannetele vastamise sagedustega. Kahe ülesande (ülesanded 4 ja 5) puhul ei olnud normaalvalimi Rasch'i raskusastmed kliinilise valimi vastamissagedustega kooskõlas. Ülesannete järjestamisel lähtuti nende ülesannete puhul kliinilise valimi õigesti vastamise sagedustest,

kuna need ülesanded on mõeldud pigem vähem võimekamate (vaimse alaarenguga) inimeste eristamiseks (normgrupis on need ülesanded väga madala/puuduva eristusvõimega). Seda valikut toetas ka see, et Rasch'i raskusastmete erinevus oli väga väike, kuid kliinilise grupi õigesti vastamise sageduste erinevus üsna suur ning klassikalise testiteooria raskusindeksid olid kooskõlas kliinilise grupi ülesannetele õigesti vastamise sagedustega. Originaalülesannetest otsustati säilitada kõige lihtsam ülesanne, millele ka kõik kliinilise valimi isikud õigesti vastasid. Kuna kliiniline valim koosnes vaid kerge ja mõõduka vaimse alaarenguga inimestest, siis võiks selle ülesande säilitamine jätta võimaluse raske vaimse alaarenguga inimeste eristamiseks mõõduka vaimse alaarenguga inimestest. Eestikeelse Aritmeetika kliinilise kasutatavuse hindamiseks tuleks edaspidi andmeid koguda erinevate häiretega inimestelt.

Kuigi puudub info originaalülesannete raskusastmete arvuliste väärtuste kohta, saab siiski võrrelda nende suhtelist järjestust eestikeelses Aritmeetikas ja originaaltestis. Aritmeetika eestikeelse variandi originaalülesannete suhteline raskusjärjekord ei ühtinud täielikult originaaltesti vastava järjekorraga. Eestikeelse Aritmeetika lihtsamad originaalülesanded olid originaaltestiga võrreldes samas järjekorras, 6 originaalülesannet osutusid Eesti inimestele suhteliselt raskemaks ning neli suhteliselt kergemaks. Raskemaks osutunud originaalülesannete suhteliste asukohtade erinevused olid suuremad (erinevus originaalülesandega kuni 5 kohta) kui kergemaks osutunud ülesannete puhul (erinevus originaalülesannetega kuni 2 kohta). Raskemaks osutunud originaalülesanded olid pigem tõenäoliselt koormamisele suunatud, nende lahendamiseks piisas aritmeetilistest tehetest ega olnud tarvis valemite tundmist. Originaaltesti kaks kõige raskemat ülesannet vahetasid omavahel asukoha Aritmeetika eestikeelses variandis nii, et tõenäoliselt rohkem koormav ülesanne (kirjaklambrate ülesanne – sõna pikkuse ja fonoloogilise sarnasuse efekt, palju ühikuid) jäi eestikeelses variandis kõige raskemaks. Kergemaks osutunud originaalülesanded toetusid küll natuke ka koolis õpitud teadmistele (keskmise kiiruse arvutamine, kolme arvu keskmise leidmine), kuid ei koormanud oluliselt tõenäoliselt teksti pikkuse osas, võrreldes raskeimate ülesannetega. Seega võiks arvata, et eestlaste koolis õpitud teadmised on natuke paremad kui Ameerika inimestel, kuid tundub, et suurem koormus tõenäoliselt muudab ülesanded eestlaste jaoks raskemaks. Võimalik, et eesti keele omapäral võrreldes inglise keelega on siin oma osa kuna eestikeelsetel sõnadega on

rohkem silpe kui inglisekeelsetel, ning seetõttu koormab samasuguste probleemülesannete esitamine eesti keeles rohkem töömälu.

Analüüsi I etapis ei ilmnenu ühegi ülesande puhul soolist, vanuselist ega hariduslikku hälvet, kuid III etapis teostatud analüüs eestikeelse Aritmeetika ülesannetele näitas, et üks ülesanne (kasutatud mööbli ostmine) eelistab madalama haridusega inimesi. Selle ülesande hälvet võiks pidada väikeseks, sest vastasel juhul oleks see ilmnenu juba analüüsi I etapis. Kahjuks ei olnud võimalik seda ülesannet asendada, kuna puudus sobiva raskusega ülesanne, mis oleks sobinud kokku testi struktuuriga – leidis küll sobiva raskusega ülesandeid, kuid nende ajalimiit ei sobinud ülesande asukohaga. Üks võimalik seletus selle ülesande hälbele võiks olla seotud ülesande kontekstiga. Kasutatud mööblit ostavad tavaliselt madalama sissetulekuga inimesed, kes on tavaliselt ka madalama haridusega ning neile võib ülesande esitamise kontekst luua eelise. Selle ülesande hälvet tuleks kindlasti uuesti hinnata suuremal valimil.

Üldiselt näitas nii ülesande-testi korrelatsioon kui eristusindeks  $D$ , et ülesannete eristusvõime on rahuldav. Mitterahuldava eristusvõimega olid ülesanded, mille variatiivsus oli väga väike. Kuna Aritmeetika on mõeldud inimeste hindamiseks laialdsete võimete skaalal, siis peab test sisaldama ka väga lihtsaid ülesandeid, millele enamus inimesi vastabki õigesti. Seetõttu ei saagi nende ülesannete eristusvõime normgrupis olla hea ning ülesanded säilitati hoolimata madalast eristusvõimest.

Eestikeelse Aritmeetika toorskooridele Ameerika normide järgi leitud skaalaskooride keskmise ja standardhälbe järgi võib eestikeelse Aritmeetika raskust pidada originaaltesti raskusega võrdväärseks (originaaltesti keskmine 10 ja standardhälve 3).

Kuna Aritmeetika toorskooride jaotus ei vastanud normaaljaotusele (esines kerge negatiivne asümmeetria), siis esialgsete skaalaskooride leidmiseks tuli toorskooride protsentiiljaotused vanusegrupiti normaliseerida, mis võis andmeid skaalaskooride osas moonutada. Eestikeelse Aritmeetika esialgsed skaalaskoorid jaotusid 4 standardhälbe piirides (2 standardhälvet üle ja 2 standardhälvet alla keskmise), erinevalt originaaltesti jaotusest, mis jääb 6 standardhälbe (3 standardhälvet üle ja 3 standardhälvet alla keskmise) piiridesse. Kuna käesolevas uuringus leiti esialgsed skaalaskoorid vaid normgrupile, siis ei saagi skaalaskoor ulatuda 3 standardhälvet madalamale keskmisest.

Keskmisest kõrgemate skaalaskooride jäämine 2 standardhälbe piiridesse võis tuleneda sellest, et keegi testitavatest ei saanud maksimaalset võimalikku toorskoori. Arvesse võttes käesoleva uurimuse valimi mahtu (191 inimest), siis kolmanda standardhälbe piiridesse (2,3% jaotusest) oleks tõenäoliselt võinud mahtuda vaid 4...5 inimest. On võimalik, et need inimesed juhuslikult ei sattunud käesolevasse uuringusse ning suurema valimi korral õnnestuks uuringusse leida inimesi, kes võimaldavad laiendada skaalaskooride jaotust 3 standardhälvet üle keskmise. Kuna käesoleva uurimuse valim on väike ning skaalaskooride leidmine on sisukas vaid kõikide alltestide lõikes, siis lõplike skaalaskooride leidmise viis ja jaotus selguvad WAIS-III edaspidise adapteerimise käigus kui Aritmeetikasse on lisatud raskemaid ülesandeid ja andmeid kohutud suuremalt valimilt.

Eestikeelse Aritmeetika sisemine reliaablus oli nii vanusegrupiti kui kogu valimis heade näitajatega. Poolitustesti reliaablusnäitajad olid nii vanusegrupiti kui kogu valimis sarnased originaaltesti vastavate reliaablusnäitajatega (tabel 2). Natuke madalamad olid eestikeelse Aritmeetika poolitustesti reliaablusnäitajad 30...44 ja 45...64- aastaste gruppides, kuid siiski piisavate väärtustega, et peegeldada skaala usaldusväärsust.

Eestikeelse Aritmeetika valiidsuse hindamiseks leitud korrelatsioonikordajad Aritmeetika ja teiste WAIS-III sõnaliste alltestide (Arvumälu, Sarnasused ja Taibukus) vahel olid kooskõlas originaaltesti vastavate näitajatega (tabel 3) ning ei erinenud statistiliselt oluliselt. Sarnaselt originaaltestile oli eestikeelne Aritmeetika võrdväärselt seotud nii sõnalise skaala alltestide (Sarnasused, Taibukus) kui ka töömälu indeksskaala alltestiga (Arvumälu). Sellist Aritmeetika võrdväärsset jagunemist on täheldatud ka Wechsleri lastele mõeldud testis (Wechsler, 1991). Natuke tugevam seos ilmnes eestikeelse Aritmeetika ja Arvumälu tagurpidise osa vahel võrreldes seosega Aritmeetika ja Arvumälu edaspidise osa vahel. Kuna Arvumälu tagurpidine osa on suunatud sõnalise töömälu mõõtmisele (eeldab samaaegset numbrite meeldejätmist ja manipuleerimist) ja edaspidine osa vaid sõnalise lühimälu mõõtmisele (eeldab ainult numbrite meeldejätmist), siis suurem seos Arvumälu tagurpidise osaga võiks viidata Aritmeetika seosele töömäluga. Kuna Sarnasused ja Taibukus omasid suhteliselt madalaid korrelatsioone Arvumälu ja selle kahe erineva osaga, siis võiks arvata, et Arvumälu (sõnaline lühi- ja töömälu) on sõnalise võimekuse osakaal väiksem võrreldes Aritmeetikaga ning



Arvumälu ja Aritmeetika mõõdukat korrelatsiooni võiks osaliselt omistada ka töömälule. Kuna kõik alltestid olid omavahel statistiliselt oluliselt korreleeritud võib neid seoseid omistada ka *g*-faktorile. Ka WAIS-III kõik originaal-alltestid olid omavahel oluliselt seotud ning seda seost seletati samuti *g*-faktoriga (Wechsler, 1997). Lisaks *g*-faktorile võib statistiliselt olulisi seoseid omistada ka sõnalistele võimetele kuna käesolevas uuringus kasutati vaid sõnalise skaala teste. Aritmeetika valiidsusele viitab kaudsetl ka keskmiste testiskooride kahanemine vanuse kasvades (tabel 2). Waters ja Caplan (2005) leidsid, et vanemad inimesed said võrreldes noorematega kehvemaid tulemusi töömälu ülesannetes. Aritmeetika valiidsust toetab ka andmestiku ühemõõtmelisus. Kuna käesolevas uurimuses oli testide läbiviimine üsna ajamahukas, siis valiidsuse hindamiseks kasutatavate mõõdikute arv oli piiratud vaid WAIS-III sõnalise skaala alltestidega, millest enamusega viidi samaaegselt läbi sarnast uurimust (Burk, 2007). Edaspidi tuleb Aritmeetika valiidsust täpsemalt uurida suuremal inimeste grupil, viies läbi kogu WAIS-III testipakett ja teised vajalikud testid Aritmeetika divergentse ja konvergentse valiidsuse hindamiseks.

Kokkuvõtteks võiks öelda, et käesoleva uurimuse abil õnnestus eestikeelsest Aritmeetikast kõrvaldada valdav osa probleemsetest originaalülesannetest ning asendada need sobivamate lisaülesannetega. Esialgne eestikeelne Aritmeetika skaala mõõdab võimekust üsna laiaades piirides, kuid testi oleks soovitatav täiendada raskemate ülesannetega. Esialgse eestikeelse Aritmeetika mõõtmise usaldusväärsus ja valiidsus on sarnased originaaltestiga. Järgnevalt võiks Aritmeetikasse lisada raskemaid ülesandeid ja andmeid koguda kogu WAIS-III testipaketiga suuremal valimil eesmärgiga uurida käesolevast uurimusest ilmnunud probleeme eestikeelse Aritmeetika ülesannetega (üks hälbega ülesanne, vähe raskeid ülesandeid, puuduvad ülesanded teatud võimekuse tasemel inimeste täpsemaks eristamiseks) ning võimalusel uute ülesannetega need probleemid kõrvaldada. Seejärel võiks eestikeelse WAIS -III standardiseerida ja normeerida, uurida täpsemalt WAIS-III struktuuri Eesti valimil, hinnata WAIS-III alltestide divergentset ja konvergentset valiidsust ning nende sensitiivsust ja spetsiifilisust erinevate häirete osas.

### Viited

- Ackerman, P.L., Beier, M.E., Boyle, M.O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology*, 131, 567-589.
- American Educational Research Association [AERA], National Council of Measurement in Education [NCME], American Psychological Association [APA], (1999). *The Standards for Educational and Psychological Testing*.
- Andrich D., Sheridan B. & Luo, G. (2003). Rasch unidimensional measurement model 2020 (version 4.0 for Windows). Perth: RUMM Laboratory Pty Ltd.
- Baddeley, A.D., Hitch G.J. (1974). Working memory. Raamatus: G.A. Bower (Toim.), *The psychology of learning and motivation*, (47-89). New York: Academic Press.
- Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 11, 417-423.
- Baddeley, A.D. (2003). Working memory: looking back and looking forward. *Neuroscience*, 4, 829-839.
- Biederman, J., Petty, C.R., Ball, S.W., Fried, R., Doyle, A.E., Cohen, D., Henderson, C., Faraone, S.V. (2009). Are cognitive deficits in attention deficit/hyperactivity disorder related to the course of the disorder? A prospective controlled follow-up study of grown up boys with persistent and remitting course. *Psychiatry Research*, 170, 177-182.
- Burk, K. (2007). WAIS-III Sõnavara, Sarnasuste, Informeerituse ja Taibukuse alatestide kohandamine. Magistriprojekt. Tartu Ülikooli psühholoogia instituut.
- Carretti, B., Borella, E., Cornoldi, C., DeBeni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, 19, 246-251.
- Colom, R., Abad, F.J., Quiroga, M.A., Shih, P.C., Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, 36, 584-606.
- Colom, R., Flores-Mendoza, C., Quiroga, M.A., Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and individual*

- differences*, 39, 1005-1014.
- Conway, A.R.A., Cowan, N., Bunting, M.F., Thierrault, D.J., Minkoff, S.R.B. (2002). A latent variable analysis of working memory capacity, short term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163-183.
- Conway, A.R.A., Kane, M.J., Engle, R.W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive Sciences*, 7, 547-552.
- Daneman, M., Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, 19, 450-466.
- De Jong, P.F., Das Smaal, E.A. (1995). Attention and intelligence: The validity of the star counting test. *Journal of Educational Psychology*, 87, 80-92.
- Ellis, N.C., Hennely, R.A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, 71, 43-51.
- Elosua, P., López-Jauregui, A. (2007). Potential Sources of differential Item Functioning in the Adaptation of Tests. *International Journal of Testing*, 7, 39-52.
- Embretson, S.E., Reise, S.P. (2000). *Item response theory for psychologists*. London. Mahwah, New Jersey: Lawrence Erlbaum Associates, publishers.
- Escobedo, P.S., Hollingworth, L. (2009). Annotations on the Use of the Mexican Norms for the WAIS-III. *Applied Neuropsychology*, 16, 223-227.
- Flynn, J.R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, 54, 5-20.
- Foxcroft, C.D., Aston, S. (2006). Critically examining language bias in the South African adaptation of the WAIS-III. *SA Journal of Industrial Psychology*, 32, 97-102.
- Groth-Marnat, G. (1997). *Handbook of Psychological Assessment*. Wiley, 3.ed. New York, 132-204.
- Haavisto, M-L., Lehto, J. (2004). Fluid/spatial and crystallized intelligence in relation to domain-specific working memory: A latent-variable approach. *Learning and Individual Differences*, 15, 1-21.
- Hair, J.F., Anderson, R.E. (1998). *Multivariate data analysis with readings*. Upper Saddle River, N.J.: Prentice Hall.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into

- multiple languages and cultures. Raamatus: Hambleton R.K., Merenda P.F., Spilburger, C.D. (Toim.), *Adapting Educational and Psychological tests for Cross-Cultural Assessment*. New Jersey: Lawrence Erlbaum, 3-38.
- International Test Commission (2010). *International Test Commission Guidelines for Translating and Adapting Tests*. [<http://www.intestcom.org>]
- Iverson, G.L., Lange, R.T., Viljoen, H. (2006). Comparing the Canadian and American WAIS-III Normative Systems in Inpatient Neuropsychiatry and Forensic Psychiatry. *Canadian Journal of Behavioural Science*, 38, 348-353.
- Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Judd, T.P., Hickson Bilsky, L. (1989). Comprehension and Memory in the Solution of Verbal Arithmetic Problems by Mentally Retarded and Nonretarded Individuals. *Journal of Educational Psychology*, 81, 541-546.
- Loftus, E.F., Suppes, P.(1972). Structural variables that determine problem-solving difficulty in computer-assisted instruction. *Journal of Educational Psychology*, 63, 531-542.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mackintosh, N.J., Bennett, E.S. (2003). The fractionation of working memory maps onto different components of intelligence. *Intelligence*, 31, 519-531.
- Morey, R.A., Dolcos, F., Petty, C.M., Cooper, D.A., Hayes, J.P., LaBar, K.S., McCarthy, G. (2009). The role of trauma-related distractors on neural systems for working memory and emotion processing in posttraumatic stress disorder. *Journal of Psychiatric Research*, 43, 809-817.
- Must, O., Nijenhuis, J., Must, A., Vianen, A.E.M. (2009). Comparability of IQ scores over time. *Intelligence*, 37, 25-33.
- Piskulic, D., Olver, J.S., Norman, T.R., Maruff, P. (2007). Behavioural studies of spatial working memory dysfunction in schizophrenia: A quantitative literature review. *Psychiatry Research*, 150, 111-121.
- Rabin, L.A., Barr, W.B., Burton, L.A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: a survey of INS, NAN and APA division 40 members. *Archives of clinical neuropsychology*, 20, 33-65.

- Raidvee, A. (2003). Üksikvastuste teooria mudelite mõõtmisomadused Raveni maatriksite näitel. Lõputöö. Tartu Ülikooli psühholoogia osakond.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Roivainen, E. (2009). European and American WAIS III norms: Cross-national differences in performance subtest scores. *Intelligence*, doi:10.1016/j.intell.2009.10.001
- Rosenthal, R., Rosnow, R.L. (1975). *The volunteer subject*. New York: Wiley.
- Saklofske, D.H., Hildebrand, D.K., Gorsuch, R.L. (2000). Replication of the Factor Structure of the Wechsler Adult Intelligence Scale—Third Edition With a Canadian Sample. *Psychological Assessment*, 12, 436-439.
- Solano-Flores, G., Backhoff, E., Contreras-Niño, L.A. (2009). Theory of Test Translation Error. *International Journal of Testing*, 9, 78-91.
- Suen, H., Greenspan, S. (2009). Serious Problems with the Mexican Norms for the WAIS-III when Assessing Mental Retardation in Capital Cases. *Applied Neuropsychology*, 16, 214-22.
- Süss, H.-M., Oberauer, K., Wittmann, W.W., Wilhelm, O., Schulze, R. (2002). Working memory capacity explains reasoning ability – and a little bit more. *Intelligence*, 30, 261-288.
- Taub, G.E., Witta, E.L., McGrew, K.S. (2004). A Confirmatory Analysis of the Factor Structure and Cross-Age Invariance of the Wechsler Adult Intelligence Scale – Third Edition, *Psychological Assessment*, 16, 85-89.
- Van de Vijver, F., Leung, K. (1997). *Methods and Data Analysis for Cross-Cultural Research*. Thousand Oaks: Sage.
- Waters, G., Caplan, D. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory*, 13, 403-413.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler D. (1997). *Wechsler adult intelligence scale – third edition. Administration and scoring manual*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1997). *Wechsler adult intelligence scale – third edition. Technical Manual*.

- San Antonio, TX: The Psychological Corporation.
- Wycherley, R.J., Benjamin, L., Crawford, J., Mockler, D. (1999). Description of the UK study. Raamatus: D. Wechsler (Toim.), *Wechsler adult intelligence scale-third edition – UK. Administration and scoring manual*, (215-233). London: The Psychological Corporation.
- Wycherley, R., Lavender, A., Holtum, S., Crawford, J.R., Mockler, D. (2005). WAIS-III UK: An extension of the UK comparability Study. *British Journal of Clinical Psychology*, 44, 279-288.
- Wright, B., Stone, M. (1999). *Measurement Essentials*. Wide Range, Wilmington, Delaware.