

Tartu Ülikool
Pedagoogika osakond

Jaan Mikk

AINETESTID

Loengukonspekt TÜ üliõpilastele

Tartu 2002

Teine, ümbertöötatud trükk

Küljendus Heino Heinla

© Jaan Mikk

Sisukord

Eessõna	5
Sissejuhatus	7
Milleks me vajame ainete?	7
Ainetesti mõiste ja liigid	9
Ainetestide kasutamise ajaloost	12
Testiülesannete liigid	15
Valikvastustega ülesanded	15
Testiülesannete teised liigid	27
Programmitest	37
Testiülesannete esinduslikkus	37
Testimise juhend	41
Testivastuste skoorimine	43
Programmitesti tulemuste teisendamine koolihinneteks	45
Ainekava optimaalse mahu määramine	47
Eristustest	51
Testiülesannete raskus	51
Testiülesannete diagnoosiv väärtus	55
Eristustesti tulemuste standardskaalad	59
Eristustesti tulemuste seostamine koolihinnetega	65
Testi kvaliteedi näitajad	67
Testi reliaablus	68
Testi valiidsus	73
Ainetestid tänapäeval	79
Pedagoogilise Testimise Teenistus USA-s	79
Testimine arvuti abil	83
Riigieksamid	85
Nõuanded testi täitjale	88
Lõppsõna	93
Kirjandus	96
Lisad	99
Testitulemuste standardskaalad	101
Raveni testi normid	105
SAT harjutustesti normid	106
Mõõteskaalad pedagoogikas	109
Mõistete selgitusi	110

Eessõna

T. Husén (1997, lk 254–255) loetleb kuus pedagoogilise uuringu suunda, mis on hariduspoliitikat mõjutanud kõige rohkem. Nende kuue seas on Binet uuringud õpilaste grupeerimisest testide abil ja programmistide alal tehtavad tööd. Testimine on avaldanud haridusele kogu maailmas tugevat mõju. Ka meil väidetakse, et riigieksamid on mõjutanud koolitööd rohkem kui teised uuendused viimase 10 aasta jooksul. Aga meil puudub testide alane uurimistöö ja kõrgkoolides pole ka ulatuslikumaid kursusi testide koostamisest ning hindamisest. Seetõttu võime me testide tegemisel eksida või osa nende võimalusi kasutamata jätta.

Testidel on nii pooldajaid kui vastaseid. Testide pooldajad väidavad, et nendega on võimalik inimese omadusi objektiivselt mõõta, võrrelda ja anda inimesele soovitusi tegevuste valikuks. Testide puudumisel on avatud tee suvalisteks otsustusteks inimese elutee kujundamisel. Eriti palju on kritiseeritud hindamist suuliste vastuste põhjal. Testide vastased väidavad, et testide täitmine traumeerib paljusid õpilasi ja et testide abil on võimalik mõõta vaid osa inimese omadustest, mistõttu mõõtmine viiks mõne omaduse ületähtsustamisele teiste, võib-olla olulisemate omaduste kahjuks. Testide kasutamine võib süvendada erinevusi inimeste vahel.

Raskused testide koostamisel ja kasutamisel ei tohiks põhjustada käegalöömist vastavale teadus- ja õppetööle. Pigem vastupidi, testimise keerukas valdkond nõuab palju mitmekülgset ettevalmistatud spetsialiste. Ilma nendeta võimegi jääda tasemele, kus me olime kümnekond aastat tagasi ja halvasti koostatud või vääralt kasutatud testidega diskrediteerida testimise idee.

Käesolev õppevahend TÜ üliõpilastele püüab anda ülevaate ainetestide teooriast: mis on ainetest, kuidas seda koostada, kasutada, kuidas testi ja selle tulemusi analüüsida. Lisaks teooriale on vajalik ka praktika, et saada hinnatud testijaks. Loodetavasti pakub õppevahend huvi ka õpetajatele ja teaduritele, kes alustavad ainetestide koostamist.

Tänapäevase abi eest käesoleva õppevahendi koostamisel retsensente Aleksander Pulverit ning Olev Musta, keeleteimetajat Leelo Jagot ja samuti oma õpetajaid Kalju Toime ning Inge Unti.

Tartus, 30 aprillil 2001.

Sissejuhatus

Milleks me vajame aineteste?

Ühiskonnad ja inimesed erinevad selle poolest, kui palju materiaalseid vahendeid neil on kasutada. Mõnes riigis on neid vahendeid palju, teises vähe. Ühiskonnad erinevad aga ka selle poolest, mida inimesed õigeks peavad. Mõnes ühiskonnas hinnatakse kõrgelt kristliku eetika reegleid, teises peetakse omakasu ellujäämise oluliseks vahendiks. Mida objektiivsemalt hinnatakse inimeste tööd, seda rohkem on inimesed motiveeritud töötama ja seda kõrgem on ühiskonna elatustase. Objektiivne hindamine on edasiliikumise oluline alus.

Õpilaste suhtumine hindamisse kujuneb ka koolis. Koolis on aga väga raske õpilasele objektiivset hinnet panna, sest segamas on mitmesugused tegurid. Vaatleme neid tegureid, et mõelda, kuidas nende mõju vähendada.

1. Õpilase hinnet mõjutavad tema eelmised hinded. Kui need on head, siis usuvad õpetajad, et ka käesolev vastus on hea ja hindavad seda kõrgemalt kui õpilase tausta tundmata. Seetõttu näiteks peavad üliõpilased halvaks tooniks, kui õppejõud õpinguraamatut sirvib, enne kui ta on oma hinde fikseerinud.
2. Õpilase käitumine võib samuti mõjutada tema hinnet. Üldiselt õpivad korraliku käitumisega õpilased paremini ja sellele reeglile alateadlikult toetudes võib õpetaja oodata head või halba vastust enne veel, kui õpilane on vastama hakanudki. See ootus paratamatult mõjutab hinnet.
3. Õpilased, kes on teatud määral omaks võtnud õpetaja vaated ja käitumise, saavad veidi paremaid hindeid. See omaksvõtt muudab vastavad õpilased õpetajale meeldivamaks ja see hoiak omakorda mõjutab hinnet. See mehhanism selgitab, miks poistel ja klassi uutel õpilastel on raskem head hinnet saada.
4. Tugevas klassis hinnatakse õpilasi rangemalt ja nõrgas klassis leebemalt. See on paratamatu õpilaste arendamise seisukohalt lähtudes, kuid samas muutuvad erinevate koolide hinded võrreldamatuks.

5. Noor õpetaja kardab välja panna äärmuslikke hindeid — tal ei ole veel kogemust, mida õpilased suudavad. Väga heade või väga halbade teadmistega õpilased saavad seetõttu ebaobjektiivse hinde.
6. Nõudliku õpetaja käe all õpitakse tavaliselt rohkem, kuid väga nõudlik õpetaja võib õpilastes hirmu tekitada ja seetõttu ei suuda õpilased talle vastata isegi seda, mida nad teavad.
7. Halb hinne tähendaks nii mõnigi kord õpilasele täiendavaid (ülejäu käivaid) pingutusi ja õpetajale täiendavat ajakulu. Nende vältimiseks halastab õpetaja õpilasele ja paneb välja positiivse hinde puudulike teadmiste juures.

Ülalloetletud tegurid puudutavad kõigepealt õpilaste suuliste vastuste hindamist. Kirjalike vastuste hindamise on objektiivsem, eriti siis, kui vastuste hindamiseks on välja töötatud kindlad reeglid. Ainetestides on need reeglid tavaliselt olemas, kuna seal on fikseeritud, milline vastus annab punkti, milline ei anna. Ainetestid võimaldavad hindaja subjektiivsust tunduvalt vähendada ja seetõttu on need arenenud maades üsna laialt kasutusel.

Ainetestide abil on võimalik ühtsetel alustel hinnata paljude koolide õpilasi. See võimaldab võrrelda koole, mis oleks võimatu, kui iga õpetaja ise koostab oma koolis kontrollivahendi. Ainetestide abil on võimalik võrrelda ka eri maade koole. Selleks on tarvis test võimalikult täpselt adapteerida teise keelde. Meil on selliseks näiteks Meeli Pandise (1996) poolt läbiviidud rahvusvaheline lugemisoskuse test.

Ainetestid võimaldavad kokku hoida õpetajate aega. See efekt on meil küll seni veel väike, kuna õpetajad peavad ise teste koostama. Kui aga kord koostatud testid hakkavad õpetajate seas levima, siis on õpetajatel lihtne nende seast valida sobiv ja sellega õpilastele arvestustöö teha. Suure tõuke selliseks arenguks annab arvutiprogramm APSTest 2 (Kelder, Nigul, Prank, 2001).

Liigne innustumine ainetestidest võib sundida õpilasi pühendama õppimisele lubamatult palju aega. See võib kahjustada kooli paljude oluliste eesmärkide saavutamist, näiteks õpilaste füüsilist, emotsionaalset või esteetilist arengut. Ainekavade maht koolis peaks vastama õpilaste õpivõimalustele. Momendil see nii ei ole — ainekavad on üle paisutatud faktilise materjaliga. Õppematerjali optimaalse mahu leidmise kõige usaldusväärsemaks teeks on programmitestide tulemused representatiivses õpilasgrupis. Seegi on üks põhjus, miks on kasulik aineteste koostada ja läbi viia.

Ainetesti mõiste ja liigid

Sõna *test* tuleb inglise keelest ja tähendab *katse, proov*, ka eksam, kontrolltöö. Teste tehakse elus väga mitmel alal: tehnikas, meditsiinis, psühholoogias, pedagoogikas jne. Meie räägime testidest pedagoogikas, aga arvestades pedagoogika tuginemist psühholoogiale, on sagedased ka võrdlused testidega pedagoogilises psühholoogias.

Test on vahend inimese omaduste mõõtmiseks. Inimese omaduste all on mõeldud tema teadmisi, oskusi ja isiksusomadusi. Testidega võib mõõta inimese teadmisi emakeelest, muusikast jne.

Mõõtmisi tehakse kõige enam füüsilikas. Seal on mõõdupuud üldaktsepteeritud (meeter, kilogramm jne) ja mõõtmistulemused seetõttu hästi mõistetavad. Pedagoogikas on mõõdupuuks testid. Need pole veel kuigi hästi tuntud ja mõõtmistulemuste üle vaieldakse. Tavaliselt eeldatakse, et test on väga hoolikalt koostatud, lähtudes alltoodud nõuetest. Seega standardiseeritus muudab ülesannete kogumi või küsimustiku testiks selle tavatähenduses.

Testi eripäraks on see, et tema ülesannetele on võimalik kiiresti fikseerida vastust. Vastamiseks on sageli tarvis ära märkida ülesande üks etteantud vastuse variantidest, kirjutada lünka üks sõna või arv jne. Vastuse kiire märkimise võimalus ei tähenda aga seda, et vastuse enda leidmine peaks samuti kiire olema. Nii mõnigi kord nõuab testiülesanne pingsat ja aeganõudvat mõttetööd, mille käigus see kiiresti märgitav vastus leitakse.

Testile lähedaseks mõisteks on *ankeet*, see on küsitlusleht mingite andmete saamiseks. Nii mõnigi kord nimetatakse ankeeti ka testiks. Eristame neid mõisteid järgmiselt. Ankeet koosneb küsimustest, millele lisatud kõik vastusevariandid on õiged. Näiteks küsimus "Millise erakonna vaateid Te kõige enam pooldate?" on ankeedi küsimus, sest vastaja osutamine igale legaalsele erakonnale väärib aktsepteerimist. Seevastu küsimus "Milline on täisnurkse kolmnurga pindala, kui tema hüpotenuus on 5 cm ja üks kaatet on 4 cm?" on testiküsimus, sest siin on vaid üks õige vastus: 6 cm^2 .

Testide kasutamine põhineb eeldusel, et inimese omadused ilmnevad tema tegevuses. Seega püütakse testiülesandega anda katseisikule tegevus, mille sooritamise edukuse järgi saab otsustada testijat huvitava omaduse arengutaseme üle.

Pedagoogilises psühholoogias jaotatakse testid tavaliselt kolme liiki: **ainetestid, võimetestid ja isiksustestid**. Selle jaotuse aluseks on inimese omaduse liik, mida testiga mõõdetakse. Ainetestidega mõõdetakse inimese emakeele, matemaatika jne teadmiste ja oskuste taset. Võimetestidega mõõdetakse verbaalsete võimete, kujutlusvõime jt võimete arengu taset. Isiksuse testid mõõdavad näiteks motive, ekstrasvertsust jm isiksusomadusi. Esitatud testide jaotus on meil üldiselt levinud ja me jääme selle juurde, kuigi sisuliselt võttes on ka võimed isiksusomadus ja kuuluvad seega isiksustestide alla. Ilmselt on võimetestide esiletoomine omaette liigina tingitud sellest, et võimetestid on hästi välja arendatud, ulatuslikult kasutusel ja mõneti erinevad paljudest teistest isiksustestidest. Õigupoolest on suur osa isiksustestide meie käsitluse järgi tegelikult isiksusküsimustikud.

Ainetest on küsimuste ja ülesannete kogum, millega mõõdetakse inimese õpitulemusi ehk õppeainepädevusi. Need on omandatud teadmised ja oskused. Testid võivad olla erinevates ainetes omandatud teadmiste mõõtmiseks erinevatel tasemetel: põhikooli, keskkooli, kõrgkooli tasemel. Teste võib koostada üksikute teemade omandatuse kontrollimiseks. Kontrollida saab mitte üksnes teadmisi vaid ka arutlusoskust.

Riiklikult kasutatavate standardiseeritud testide koostamine on ulatuslik töö, mille käigus testi esialgset varianti katsetatakse, analüüsitakse ja täiustatakse. Sellelt aluselt lähtudes liigitatakse ainetestid **standardiseeritud ja standardiseerimata ainetestideks**. Standardiseerimata aineteste nimetatakse ka õpetaja koostatud testideks, sest õpetaja tavaliselt ei tee oma testiga läbi kõiki neid protseduure, mis on vajalikud standardiseerimiseks. Õpetaja koostatud testidena võib vaadelda tunnikontrolle ja kontrolltöid. Samal ajal võib õpetaja koostada ka standardiseeritud ainetesti, mistõttu eelistame termineid standardiseeritud ja standardiseerimata ainetestid. Peab kohe ütlema, et Eestimaal on praegu raske leida kõikide nõuete järgi standardiseeritud aineteste. See ei tähenda, et olemasolevaid teste ei võiks kasutada, see tähendab, et nende tulemuste tõlgendamisel tuleb olla eriti ettevaatlik.

Eesti traditsioonidest lähtudes on oluline ainetestid jagada veel eristustestideks ja programmitestideks¹. **Eristustest** on ainetest, mille abil püütakse õpilasi eristada nende ainealaste teadmiste või oskuste järgi. Taoline test on vajalik olümpiaadidel ja kõrgkooli sisseastumistestina. Juhtus kord vabariiklikul füüsikaolümpiaadil, et 17 osavõtjat lahendas kõik ülesanded õigesti. Kuidas nüüd selgitada nende seast parim? Sisuliselt polegi see võimalik, kuna testiülesannete seas puudusid keerukamad ülesanded, mida vaid üks-kaks õpilast lahendada suutnuks. Eristustestides peaks olema nii kergeid kui raskeid ülesandeid, et testitulemuste põhjal oleks võimalik öelda, kes on kõige parem ainetundja, kes järgmine jne. Välismaal kasutatavad ainetestid on põhiliselt eristustestid.

Eesti traditsioonide kohaselt pannakse õpilasele hinne viis siis, kui ta kõik ülesanded õigesti lahendab. On loomulik, et mitu õpilast saab ainetesti järgi hinde viis ja paljud veel hinde neli. Kui õpilane lahendab alla poole ainetesti ülesannetest, siis on see ebageeldiv, sest hindeks tuleb kaks. Taoline ainetest ei ole eristustest, see on programmitest. **Programmitest** on test, mis sisuliselt vastab programmile (ainekavale). Programmitest kontrollib põhimõtteliselt kõikide eesmärkide saavutatust, mis vastava ainekavaga on seatud. Programmitestist ei jäeta välja ei liiga kergeid ega liiga raskeid ülesandeid. Testi koostamise ainsaks kriteeriumiks on ülesannete vastavus ainekavale. Programmitestiga kontrollitakse, mil määral on õpilased omandanud ainekava materjali. See on meil hindamise aluseks.

Eesti traditsioonid ainetestide kasutamisel on mõneti erinevad välismaistest, mistõttu tuleb meil vähemalt esialgu tunda ja kasutada nii programmi- kui eristusteste.

¹Samas tähenduses on kasutusel ka terminid normatiivne test ja kriteeriumtest (Lindgren & Suter, 1995) või normhindamine ja kriteeriumhindamine (Krull, 2000, lk 582). Need terminid viitavad otseselt vastavatele ingliskeelsetele sõnadele, kuid on sisuliselt väga lähedased: norm võib ju ka olla kriteeriumiks. Jääme seetõttu varem kasutusel olnud eestikeelsete terminite juurde, mis paremini avavad vastava mõiste sisu.

Ainetestide kasutamise ajaloost

Esimesed ainetestid võeti kasutusele üheksateistkümnenda sajandi keskel Ameerika Ühendriikides ja Inglismaal. Põhjuseks oli rahulolematuse hinnetega, mida õpilased said suuliste vastuste eest. Me nägime eespool, et need hindad sõltuvad väga paljudest faktoritest peale õpilase teadmiste taseme: küsimuste formuleerimine, hindaja poolt arvestatavad hindamisalused ja tähelepanelikkus, hindaja suhtumine õpilasesse jne. Paljude nende segavate faktorite mõju on võimalik hindamisel vähendada või koguni kõrvaldada, kui õpilane vastab kirjalikult varem hoolikalt koostatud küsimustele.

Ainetestid levisid üheksateistkümnendal sajandil aeglaselt. Sisulistele raskustele ainetestide koostamisel ja kasutamisel lisandus paljude inimeste vastuseis, kes ei olnud huvitatud teadmiste taseme objektiivsusest mõõtmisest. Nad rääkisid innukalt ainetestide tegelikust ja näivatest puudustest: ainetest ei võimalda mõõta kõikide aine õpetamisega taotletavate eesmärkide saavutatust, ainetest tähtsustab üle faktiliste teadmiste omandamise jne. Kriitikast ja vastuseisust hoolimata levisid ja täiustusid ainetestid möödunud sajandi esimesel poolel paljudes lääneriikides, sest ainetestide vored ületasid nende puudusi.

Esimesed standardiseeritud ainetestid võeti Ameerika Ühendriikides tarvitusele eelmise sajandi algul. Eriti hoogustus testide alane uurimistöö ja nende koostamine ning kasutamine peale Esimest maailmasõda. Kahekümnendail aastail trükiti Ameerika Ühendriikides kümneid testide kogumikke ja peaaegu pooled pedagoogika-alased uurimistööd olid pühendatud testidele. Neljakümnendail aastail muutus oluliselt testide sisu: faktiküsimused asendati suuresti ülesannetega, mis kontrollisid oskust rakendada teadmisi praktikas. Kuuekümnendail aastail trükiti Ameerika Ühendriikides üle kahe tuhande testi ja testi-kogumiku aastas.

Eesti Vabariigis koostas esimesena teste Juhan Tork. Ta töötas välja normid Eesti laste intelligentsuse kohta. Kui nõukogude väed 1944-ndal aastal Eestile uuesti lähenesid, siis oli Juhan Tork sunnitud tõdema: "Ma olen rumalaile õelnud, et nad on rumalad. Nad ei andesta seda mulle iialgi. Ma pean Eestist lahkuma". See tõdemus näitab ilmekalt tõsist probleemi, millega testide kasutajail tuleb arenevas riigis kokku puutuda.

Nõukogude ajal olid testid sisuliselt keelatud, kuid sellest hoolimata tegeldi Eestis kuuekümnendail-seitsmekümnendail aastail üsna intensiivselt testidega. Paljude ainete jaoks töötati välja standardiseeritud kontrolltööd, mis sisuliselt olid ainetestid. Vaatleme nendest testidest paari lähemalt.

G. Karu poolt koostatud füüsika redeltestid (Karu, 1974) on huvitavad selle poolest, et nendes on ülesanded jaotatud nelja omandamise taseme järgi: äratundmine, meenutamine, rakendamine tuntud situatsioonis ja rakendamine uues situatsioonis. Äratundmisülesanneteks on valikvastustega küsimused, meenutamisülesanneteks vabavastuselised küsimused ja kahel järgmisel tasemel esitatakse õpilasele lahendamiseks ülesandeid. Iga õppeveerandi materjali kohta on üks test kahes variandis. Teste on enne trükkimist koolis katsetatud ja nii on testidele lisatud katsetulemuste tabelid ning juhised testide läbiviimiseks ja vastuste hindamiseks.

O. Kärner koostas oma standardiseeritud kontrolltööd matemaatikas (Kärner, 1976) neljas võrdse raskusega variandis. Ühe aasta materjali kontrollimiseks oli kuus kontrolltööd. Ta katsetas oma kontrolltöid kolmel korral kokku peaaegu 2000 õpilasega. Enamus kontrolltööde ülesannetest nõuab vastuse kirjutamist arvuna või sõnana. Kontrolltööde kogumikus on juhend tööde läbiviimiseks, õiged vastused, iga vastuse eest saadav punktide arv, eeskiri punktide teisendamiseks hinneteks ja kontrolltöö katsetamisel saadud keskmised tulemused iga ülesande kohta. Viimaste põhjal saab iga õpetaja otsustada, kes tema õpilastest on standardgrupi keskmisest paremate teadmistega ja kes nõrgemate teadmistega.

Selliseid aineteste koostati seitsmekümnendate aastate algupoolel paljude ainete ja klasside jaoks. Nende abil oli võimalik ettekujutust saada vabariigi õpilaste teadmiste keskmisest tasemest. Hiljem see traditsioon kadus. Põhjustena võib oletada, et testide koostajate tööd ei tasustatud piisavalt, teadmiste taseme objektiivsete näitajate vastu ei tuntud kuigi suurt huvi, testide tulemused osutusid programiga seatud ideaalist tunduvalt madalamaks ja traumeerisid õpilasi, õpetajaid, koolijuhte. Nüüd me teame, et tolleaegne ideaal oli kõrgele seatud ja meil ei oleks olnud tarvis oma õpilaste teadmiste taseme pärast muretseda. Kuigi seitsmekümnendate aastate testimise tra-

ditsioon on meil tänapäevani jätkunud vaid üksikjuhtudel, oli see kogemus, millest meil on ka praegu õppida².

²Bibliograafilisi viiteid toleaeagsetele testidele Eestis võib leida näiteks I. Rapoorti jt. raamatust (Rapoort, Selg, Sotter, 1987).

Testiülesannete liigid

Valikvastustega ülesanded

Valikvastustega ülesanne või küsimus on selline, millele on lisatud võimalikud vastuse variandid ja vastaja peab nende seast leidma õige. Võimalike valikute arv on piiratud tavaliselt nelja-viiega, harva seitsmega.

Valikvastustega küsimustest on kõige lihtsamad **alternatiiv-vastuselised küsimused** ehk ei/jaa küsimused. Õpilasele esitatakse mingi väide ja õpilane peab märkima, kas see väide on õige või väär. Vaatame mõnd näidet.

1. *Õige või väär: Päikesevarjutus võib toimuda vaid noore kuu ajal.*
2. *Õige või väär: Ühe molekuli $C_{24}H_{50}$ põlemiseks kulub 74 molekuli hapnikku.*
3. *Temperatuuri tõustes vee molekulide kiirus*
 - a) *väheneb*
 - b) *suureneb.*
4. *Kas denaturatsioon on alati pöörduv?*

Alternatiivküsimuste oluliseks puuduseks on see, et õige vastuse juhusliku märkimise tõenäosus on suur: 50%. Seda puudust on aga võimalik oluliselt leevendada lahutades õigete vastuste arvust väärvastuste arvu. Teiseks puuduseks on see, et väär väide võib õpilase mälusse õigena püsima jääda. Selle puuduse leevendamiseks on otstarbekas hinnatava väite ette kirjutada sõnad "*Õige - Väär*", suunates nii õpilase kriitiliselt hindama järgnevat lauset. Võib-olla aitavadki alternatiivküsimused hajutada hoiakut, et kõik kirjutatu on õige.

Alternatiivküsimuste otstarbekusest räägib see, et paljud probleemid tegelikult eeldavad vaid ühte kahest võimalikust vastusest ja milleks hakata siis kunstlikult muid vastusevariante juurde mõtlema. Näiteks isegi ülaltoodud keemia-alane küsimus võib vabalt seista õpilase ees, kes teise õpilase vihikus näeb sellise ülesande nimetatud lahendust.

Esimesel pilgul võib paista, et alternatiivküsimusi on lihtne koostada: võta õpikust mingi lause, kirjuta ette "*Õige või Väär*" ja ongi

valmis. Nii lihtsalt saab vaid väga kehva testi. Alternatiivküsimustena ei soovitata kopeerida raamatu lauseid — need võivad õpilastele sõna-sõnalt meelde jääda ja võimaldavad nii kontrollida omandamist vaid madalaimal tasemel. Veelgi enam, kui õpilased kogevad, et testis tulevad neil õpiku laused, siis hakkavad nad õpikut sõna-sõnalt pähe õppima. See on väga ebasoovitav õpiharjumus. Järelikult on ka õige-väär küsimuste koostamine hoolikust ja oskust nõudev töö.

Valikvastusega ülesannetel on tavaliselt rohkem kui kaks vastusevarianti.

Kui abielupaari esimene laps on tüdruk, siis tõenäosus saada teise lapsena poiss on

- a) $1/8$
- b) $1/4$
- c) $1/2$
- d) $3/4$
- e) $1/1$

Niisugused ülesanded on välismaistes ainetestides peaaegu ainuvalitsevad. Neid kasutatakse, sest vastuse hindamine on lihtne ja praktiliselt vigadeta. Vastuse märgib õpilane sageli vastustelehele, tehes sellel vastava ovaali mustaks. Vastustelehti loeb arvuti. Kuna õpilase vastuste hindamine on täiesti täpne, siis nimetatakse valikvastustega teste ka objektiivseteks testideks. Tegelikult on täieliku objektiivsuse saavutamiseks oluline veel ülesannete valik, nende vormistus ja palju muud, aga siiski on see nimetus mõistetav, sest valikvastusega ülesanded annavad objektiivsema testitulemuse kui teised ülesannete liigid.

Valikvastusega ülesannetes on tavaliselt vaid üks õige vastus. Sellisel juhtumil nimetatakse neid ka mitmikvaliku küsimusteks. Samas on võimalik valikvastusega küsimusi koostada ka nii, et õige vastus on jaotatud kahe valikvastuse vahel. Siis nimetatakse neid küsimusi ka õigete vastusevariantide märkimise küsimusteks (APSTest 2, 2000). Meie räägime järgnevas põhiliselt valikvastusega küsimustest, millel on vaid üks õige vastus, sest need on rohkem levinud.

Valikvastustega ülesandeid võib **mitmeti vormistada**.
Vaatame mõnd näidet.

1. Valige sobiv eessõna!

Sie wohnt in Tartu ... ihrer Grossmutter

- a) zu
- b) zur
- c) bei der
- d) bei (Reppo, Rebane, 1993, lk 8).

2. Leidke valikvastus, mis on sisult kõige lähedasem esimese lause mõttele!

I don't mind if you leave right away

- a) *I don't think you'll leave*
- b) *I'm not sure if you take the right way*
- c) *I don't care if you stay or go*
- d) *I don't mean that you can go now (Peiker, 1993, lk 11).*

3. Milline väljend sobib paremini lausesse sulgudes oleva väljendi asemele?

Miss Universumi valijatekogu oli juba elimineerinud kõik peale viie kandidaadi ja nüüd oli nende ainus probleem leida võitja (viiele ülejäänule)

- a) viiest ülejäänust
- b) viieks ülejäänuks
- c) viie ülejäänu seast
- d) viie ülejäänu vahelt
- e) viie ülejäänu hulgast (Doster jt, 1979, lk 21).

4. Järgnevas näites on esitatud kaks lauset ja siis viis võimalikku otsustust nende kohta. Valige otsustuste seast välja kõige sobivam!

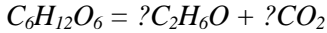
1. Kõik mehed on kollased.

2. Friedrich on mees.

Teine lause

- a) *esitab konkreetse juhtumi*
- b) *esitab erandi*
- c) *teeb üldistuse*
- d) *esitab järelduse esimesest lausest*
- e) *toetab esimeses lauses antud üldistust (Doster, 1979, lk 49).*

5. Millised koefitsiendid tuleb kirjutada alljärgnevasse reaktsiooni-võrrandisse, et see tasakaalustada?



a) 1;2

b) 2;2

c) 1;3

d) 1;1

e) 3;1 (Campbell, 1993, p. 39).

6. Which of the following is not thought to be ancestral to humans?

a) a reptile

b) a bony fish

c) a primate

d) an amphibian

e) a bird (Campbell, 1993, p. 669).

7. Esitatakse kaks suurust, üks tulbas A ja teine tulbas B. Otsustage, kumb nendest on suurem — kas A tulbas (variantvastus a), B tulbas (variantvastus b), või on nad ühesuurused (variantvastus c), või pole üldse võrreldavad (variantvastus d).

Nr	Tulp A	Tulp B	Valik
1.	0,16	1/6	a) b) c) d)
2.	4 protsenti 4805-st	190	a) b) c) d)
3.	Arvu 3/2 pöördarv	0,6	a) b) c) d)

Põhiline **kriitika** valikvastustega testiülesannete aadressil on see, et nendega ei saavat mõõta mõtlemisostkust. Ilmselt seetõttu kontrollitakse füüsika redeltestides mõtlemist vabavastuseliste küsimuste ja ülesannetega. Esimesel pilgul on see kriitika põhjendatud: kui vastused on antud, siis pole midagi muud, kui õige nende seast ära tunda. Tegelikult ei saa aga näiteks matemaatikaülesande vastust ära tunda enne, kui see ülesanne on lahendatud. Ja ülesanded pole ju tavaliselt nii lihtsad, et pilgu heitmisest ülesandele on selle vastus kohe käes. Peale selle võivad siin valikvastused anda õpetajale väärtuslikku informatsiooni: kui iga väärvalik on mõne lahenduseksituse tulemus, siis on õpetajal vastuste jaotusest näha, millist viga kui sageli tehti.

Keegi ei kahtle, et võimete testidega saab mõõta mõtlemisostkust. Samal ajal on nendes testides valdavalt valikvastusega küsimused. Näiteks võib vaadata kasvõi väljavõtet TÜ Sotsiaalteaduskonna

sisseastumistestist (Allik, 1997). See traditsioon näitab, et ka valikvastustega ülesanded võivad vastajatele tõsist pinget pakkuda. Kui küsimus nõuab pikemat mõttekäiku, siis tuleb see ka läbi teha, enne kui on võimalik õige variantvastus ära märkida. **Valikvastustega ülesandeid võib koostada aine omandatuse mõõtmiseks kõigil tasanditel.** Toome siin veel mõned näited.

1. Suhteliselt lihtne on koostada liigitusülesannet, mille lahendamine eeldab opereerimist objektide tunnustega. Järgneva näiteülesande on koostanud K. Jors algklasside loodusõpetuse testile.

Leia igale maakohale oma kodu

Saaremaa, Noarootsi, Hiiumaa, Viimsi, Muhu, Matsalu, Vormsi, Lahemaa.

<i>Saared</i>	<i>Poolsaared</i>	<i>Looduskaitsealad</i>
1.	1.	1.
2.	2.	2.
3.	3.	3.
4.	4.	4.

2. Kriitilise lugemise testis palutakse vastajal lugeda üht lõiku tummfilmist ja teist pantomiimist. Seejärel esitatakse näiteks järgmine küsimus.

Mõlema lõigu põhisisuks on

a) võtted vaatajate vapustamiseks

b) erisugused draamastiilid

c) visuaalsed elemendid draamas

d) auditooriumi vastupanu teatri võtetele

e) nostalgia meelelahutuse varaste vormide järele

(Taking the SAT I, 2000, lk 16).

Mõni nendest variantvastustest sobib hästi ühele lõigule, mõni teisele, üks kajastab mõlema lõigu ühist kõrvalideed ja vaid üks esitab ühise põhisisu.

3. Verbaalse mõtlemise mõõtmiseks kasutatakse sageli analoogiaülesannet.

Leidke esimese sõnapaari sõnade võimalikult täpne seos ja siis leidke järgneva viie vastusevariandi seast see sõnapaar, mille sõnad on samasugune seos!

Professionaalsus — ekspert

Tunnustus — võitja

Võit — atleet

Vaimustus — entusiast

Kohtumine — ametnik

Meditatsioon — doktor (Taking the SAT I, 2000, lk 52).

4. Analoogiaülesanne esitatakse sageli ka teisel viisil.

Leidke seos kahe esimese sõna vahel ja siis selle seose alusel otsige kolmandale sõnale paariline nelja valiku seast!

Kirjutuslaud — sahtel

Autotee — ...

a) auto

b) kraav

c) kruus

d) elektriliin.

5. Üldistamisoskuse mõõtmiseks võib kasutada järgmist ülesannet.

Leidke antud sõnadest üks, mis ei kuulu teistega samasse gruppi.

a) Anu

b) Kaja

c) Piret

d) Peeter

e) Salme.

Nii analoogia- kui üldistamisülesandeid võib esitada ka mitteverbaalsel materjalil.

6. Järgmine ülesanne näib sobivat mõtlemise meta-analüüsi oskuse mõõtmiseks.

"Kõik õunad siin korvis on hapud. Ma proovisin nelja õuna, ma tean."

Rääkija

- a) ignoreerib fakti, et õunad on korvis
- b) on proovinud kõiki õunu korvis
- c) on söönud oma elu jooksul palju õunu
- d) eeldab, et kõik õunad korvis on ühesugused
(Doster, 1979).

7. Hea näite ülesannetest mõtlemisvõime kontrollimiseks sotsiaalteaduste alal leiame H. C. Lindgreni ja W. N. Suteri (1995, lk 368) õpikust.

Lincolni Gettisburgi kõnet võib käsitleda kui

- a) sõja pidamise õigustust
- b) rahu sõlmimise õigustust
- c) demokraatia arendamist
- d) üleskutset kuulekusele.

Kõiki võimalusi valikvastustega ülesannete vormistamiseks pole siin võimalik esitada, sest testikoostajate loovus on suur. Vaatleme järgnevalt soovitusi **valikvastustega ülesannete koostamiseks**.

Ülesande sisu peaks olema kaasaegne ja õpilastele huvitav (Genesee & Upshur, 1996, lk 182). Samas peab see vastama õpetamise eesmärkidele — pole õige kontrollida seda, mida õpetamise eesmärgiks ei seatud.

Eristustestide puhul tuleb arvestada ka seda, mil määral on õpilased valmis kavandatavat ülesannet lahendama (Genesee & Upshur, 1996, lk 182). See puudutab nii ülesande sisu kui ka esitamise vormi.

Loomulikult peab küsimuse/ülesande tekst olema selge, lühike ja täpne. Seega tuleks sõnastatud küsimused üle vaadata veendumaks, et nendes pole liigseid sõnu ega raskestimõistetavaid väljendeid. Küsimuse mitmetimõistetavust on autoril endal sageli raske märgata, mistõttu on oluline kolleegi kriitiline pilk ja proovitesti tulemused.

Valikvastuste seas on tavaliselt vaid üks õige. Kui aga mitu valikvastust on õiged, siis tuleb instruksioonis märkida, et ühe küsimuse kaks või enam valikvastust võivad õiged olla. Või siis saab kasutada näiteks järgmist valikvastuste esitamise viisi.

Viie õpilase testitulemuste aritmeetiline keskmine oli 80, mediaan 75, madalaim tulemus 60 ja kõrgeim 100. Ülejäänud kahe õpilase testitulemusteks pakutakse

I. 66 ja 99

II. 70 ja 90

III. 76 ja 89

Õige pakkumine

a) puudub

b) ainult I

c) ainult II

d) I ja III

e) I, II ja III (Taking the SAT I, 2000, lk 27).

Peibutusvastuste valikust sõltub suuresti ülesande raskus ja efektiivsus. Kui valikuks esitatud väärvastused on õpilase jaoks ilmselgelt valed, siis on tal lihtne leida õige vastus. Kui aga peibutusvastused on õigele vastusele lähedased, näiteks osaliselt õiged, siis on valikvastustega ülesanne raskem kui sama küsimus ilma valikvastusteta. Heades valikvastustega ülesannetes on kõigil vastusevariantidel külgetõmbejõud. See on valikvastustega ülesannete kvaliteedi üks olulisi aluseid, mida testi koostaja peab alati hoolikalt jälgima. Heade vastusevariantide saamiseks soovitatakse esitada küsimus/ülesanne õpilastele eelkatses ilma vastusevariantideta ja sagedasemad väärvastused võtta peibutusvalikuteks.

Kuidas valikvastuseid konstrueerida ei tohi selle kohta üks näide. Testi koostaja kirjutab üheks valikvastuseks õige vastuse, teiseks selle detailse erijuhtumi, kolmandaks õige vastuse ülemäärase üldistuse ja neljandaks lisab veel midagi. Testile vastaja loeb nüüd ainult valikvastuseid ja arutleb järgmiselt. Üks valikvastus on teistest sisult oluliselt erinev — see ei sobi. Ülejäänud kolmest on üks väga üldine ja üks hästi konkreetne — ka need ei sobi. Õige vastus on sisult kokku kuuluvate vastusevariantide seast "keskmine". Taolise reegli olemasolu

näitab, et niisugusel põhimõttel koostati variantvastuseid ja juhtub, et koostatakse ka praegu.

Loomulikult on see testikoostaja praak ja teadmiste puudus. Heade valikvastuste saamiseks on vaja tunda hästi ainet ja koolikogemusest teada, kuidas õpilased võivad ühele või teisele küsimusele vastata. Õige vastuse asukoha võiks vastuste seas määrata täringu viskamisega või juhuslike arvude tabeli abil. Kui valikvastused on loogiliselt järjestatavad, siis tuleks seda järjestust ka kasutada. Näiteks arvud tuleks esitada kasvamise järjekorras.

Valikvastused võiksid ülesande põhiosaga grammatiliselt korrektselt seostuda, sest siis on lihtsam õpilasel vastata. Valikvastustega küsimus sõnastatakse jaatavas kõnes lünklausena, mille lünga täitmise võimalused on ette antud. Kui küsimuse põhiosa esitada küsilause ja seejärel valikvastused, siis peab vastaja kõigepealt transformeerima küsilause jaatavaks lauseks ja alles seejärel saab ta vaadata, milline valikvastus on õige. Selline transformatsioon ei ole ülesande mõtte ja selle võib ära jätta.

Esitatud soovitusel räägib vastu välismaine testimise praktika. Valikvastustega ülesannete põhiosa on sageli esitatud küsilause näiteks USA-s kasutatavates testides (Taking the SAT I, 2000; Taking the SAT II, 2000). Ilmselt ei ole nimetatud transformatsioon õpilaste jaoks suureks probleemiks, kui nad on sellega testides harjunud. Küsimuse vormis on valikvastusega ülesandeid lihtsam koostada. Aga ka siis peaksid ühe ülesande kõik valikvastused olema ühest semantilisest klassist ja samas grammatilises vormis (Genesee & Upshur, 1996, lk 187).

Valikvastuste koostamine on täiendav töö lisaks ülesannete endi valikule ja vormistamisele. Kui testi kasutatakse vähe ja sellega ei soovita saada eriti täpseid tulemusi, siis sageli loobutakse valikvastuste koostamisest ülesannetele, millel need ei ole ilmsed. Seevastu standardiseeritud testides on valikvastusega ülesanded laialt levinud, kuna nad võimaldavad väiksema vaevaga saada täpset informatsiooni suure hulga katseisikute ainealasest pädevusest.

Valikvastusega ülesannetele **saab märkida vastuseid ka juhuslikult** ilma ülesannet tegelikult lahendamata. Nii mõnigi juhuslik märkimine satub õigele vastusele ja nii on võimalik teenida lisapunkte. Võtame äärmusliku näite: õpilane ei tea ainest midagi ja õige-väär

küsimustega testis märgib vastuse huupi kõigile kümnele küsimusele. Tõenäosusteooria järgi satub ta viiel juhul õigele vastusele ja viiel juhul valele vastusele. Seega on õigeid vastuseid pool ja meie hindamis-traditsioonide kohaselt tuleks sellele õpilasele panna rahuldav hinne, kuigi ta midagi ei teadnud ja võib-olla küsimusigi ei lugenud. Kui valikvastuseid on rohkem kui kaks, on õigele vastusele huupi sattumise tõenäosus väiksem, kuid siiski võib see hinnet mõjutada.

Esitatud probleemi on võimalik lahendada kahel viisil: **mõjutada õpilasi, et nad vastuseid huupi ei märgiks**, või võtta huupi vastatud küsimuste eest punkte maha. Palve või nõue õpilastele mitte märkida vastuseid juhuslikult võib tõhusaks osutuda maades, kus on harjutud täpselt täitma eetikanorme. Peale selle peetakse siin veel oluliseks anda õpilastele piisavalt aega kõikidele küsimustele vastamiseks. Viimast soovitus testimise puhul tavaliselt ei täideta — testi täitmise aeg on rangelt piiratud. Inimene peab suutma ka vaimseid operatsioone sooritada piisavalt kiiresti, muidu jääb tema tööjõudlus madalaks. Ja eks olegi ju testimise eesmärgiks sageli inimeste eristamine töö taseme seega ka tööjõudluse järgi. Testimist on väga raske korraldada, kui lubada igal õpilasel vastata nii kaua, kui ta soovib. Üks õpilane vastab tunniga ja teine vajab samaks tööks viis tundi.

Väärvastused on pedagoogiliselt väga ebasoovitavad — nad võivad kujundada harjumuse vastata huupi. Oletagem, et lennuki-konstruktor ei tea, millist valemit parajasti rakendada ja valib ühe nendest juhuslikult. Kui valem oli vale, siis viib see vigadeni, mille avastamine ja parandamine katsetustel on kulukas. Õnneks tegelikkuses taoliste drastiliste tagajärgedeni ei lähe, aga mine tea. Hiljem me näeme, et mõnel juhtumil soovivad testispetsialistid õpilastel märkida vastuseid pooljuhuslikult. Ilmselt see parandab testitulemust. Aga milliseid harjumusi see kujundab? Kas me teame, kust saavad alguse kulukad inimlikud eksimused?

Välismaal on rakendusel teine moodus: **valesti valitud vastuste eest arvestatakse punkte maha** (Doster, 1979, lk 49; Taking the SAT I, 2000, lk 76). Meile võib tunduda taoline praktika õpilase nõokimisena — ta on õigesti vastanud, aga meie mõnd tema vastust ei arvesta. Näiteks, mis õigusega õpetaja paneb kahe, kui õpilane on pool küsimustest õigesti vastanud ja 50% õigete vastuste eest pannakse tavaliselt kolm. Õpilaste silmis võib siin õpetajale jääda vaid tugevama õigus. Tegelikult on ka õiglus tema poolel ja seda tuleb osata selgitada.

Eelmises näites me oletasime, et õpilane ei tea midagi ja ta märkis kümnest alternatiivvastusega küsimusest viis õigesti ja viis valesti. Oletame nüüd, et teine õpilane teab poolt materjalist ja vastab viiele küsimusele teadlikult õigesti, ülejäänud viiele ta ei vasta üldse. Teisel õpilasel on viis õiget vastust, pole valevastuseid ja on viis vastamata küsimust. See õpilane saab meie kriteeriumide järgi kolme. On suur erinevus, kas õpilane on või ei ole märkinud valevastuseid valikvastustega küsimustele. Kui väärvastuseid märgitud ei ole, on ta vastanud kindlate teadmiste järgi ja mingeid punkte ei arvata maha. Kui väärvastuseid on märgitud, siis on tõenäoliselt vastatud huupi, on nii satunud ka õigetele vastustele ja punkte tuleb maha arvata.

Punktide mahaarvamine välismaa testides toetub kahele eeldusele. Esiteks peavad ülesanded olema koostatud nii, et valikvastuste seast on õige ainult üks. Teiseks oletatakse, et valikvastuseid juhuslikult märkides satub õpilane igale vastusevariandile võrdse tõenäosusega. Seega näiteks nelja valikvastuse puhul satub õpilane esimesele väärvastusele 25% juhtudest, teisele samuti 25% juhtudest, kolmandale 25% juhtudest ja õigele vastusele 25% juhtudest. Kokku tuleb nelja valikvastuse puhul iga kolme ebaõnnestud juhuvaliku kohta keskmiselt üks õnnestunud juhuvalik. Juhuslikult märkides saab õpilane siin iga kolme väärvastuse kohta ühe õige vastuse.³

Eelnev arutelu näitab, et juhuslikult õigesti märgitud vastuste arv on väiksem, kui valede vastuste arv. Märki tabanud juhuslike valikute arv on võrdne valede valikute arv jagatud valikvastuste arv miinus üks. Seega saame teadlikult õigesti vastatud küsimuste arvu leida järgmise valemiga:

³Loomulikult on tegelikkus keerulisem kui siinkirjeldatud lihtne mudel. Väärvastus võib olla õpilase peas kindla teadmisenähtena, ta võib selleni jõuda eksliku arutluse tõttu jne. Lihtne matemaatiline mudel ei arvesta neid nüansse, ta arvestab kõik vead kokku. Aga ega väärteadmine pole inimesele parem kui juhuslik väärvalik. Väärteadmine on küll mõttetöö tulemus, kuid selle tagajärjed võivad olla elus raskemad, kui juhuslikult tehtud ja loodetavasti peagi ununenud väärvaliku tagajärjed.

$$R_t = R - \frac{W}{k-1} \quad (1)$$

kus R_t — teadlikult õigesti vastatud küsimuste arv,
 R — õigete vastuste arv,
 W — valede vastuste arv,
 k — ühe küsimuse valikvastuste arv.

Valemist näeme, et punktide mahaarvamine toimub vaid siis, kui on väärvastuseid. Eksimine on ohtlik elus ja sealhulgas testi täitmisel.

Kuna valikvastuste arv on tavaliselt 2–5, siis on võimalik valemit (1) esitada ka konkreetsemalt iga juhu jaoks eraldi.

Kui valikvastuseid on 5, siis tõeline tulemus = õigete vastuste arv miinus väärvastuste arv jagatud 4-ga.

Kui valikvastuseid on 4, siis tõeline tulemus = õigete vastuste arv miinus väärvastuste arv jagatud 3-ga.

Kui valikvastuseid on 3, siis tõeline tulemus = õigete vastuste arv miinus väärvastuste arv jagatud 2-ga.

Kui valikvastuseid on 2, siis tõeline tulemus = õigete vastuste arv miinus väärvastuste arv.

Nendest täpsustustest näeme, et väärvastuste arv tuleb enne lahutamist jagada seda väiksema arvuga, mida vähem on valikvastuseid. See on ka loomulik, sest väiksema arvu valikvastuste puhul on kergem sattuda õigele vastusele. Kui valikvastuseid on vaid kaks, siis tuleb õigete vastuste arvust lahutada väärvastuste arv. Rakendades seda reeglit ülaltoodud näitele, kus õpilane midagi ei teadnud ja andis 5 õiget ning 5 vale vastust, näeme, et tema tõeline tulemus on 0, nagu ta eeldatud teadmisedki. Et väärvalikutest tingitud parandust oleks kergem sisse viia, selleks grupeeritakse testis ülesanded tihti valikvastuste arvu järgi.

Kui ülesanded pole grupeeritud valikvastuste arvu järgi, siis tekib soov iga väärvastuse eest kohe punktid maha arvata. Seda mahaarvamist saab teha järgmiste reeglite järgi.

Kui valikvastuseid on 2, siis vale vastuse eest saab -1 punkti.

Kui valikvastuseid on 3, siis vale vastuse eest saab -0,5 punkti.

Kui valikvastuseid on 4, siis vale vastuse eest saab -0,33 punkti.

Kui valikvastuseid on 5, siis vale vastuse eest saab -0,25 punkti.

Toodud reeglid kehtivad, kui õige vastuse eest saab +1 punkti ja vaid üks valikvastus on õige. Kui õiget vastust hinnatakse +2 punktiga, siis tuleb eeltoodud arvud korrutada kahega jne.

Valem (1) on kooskõlas seisukohaga, et vabavastuseliste ülesannete puhul väärvastuste eest punkte maha ei arvata. Vabavastuseliste küsimuste ja ülesannete puhul on õpilase valikuvõimalused vastuse vormistamiseks väga suured ja seega tuleks nimetaja valemis (1) väga suur, mistõttu murd ise on väga väike ja lahutada pole õieti midagi.

Väärvalikutest tingitud parandust kasutatakse enamikus välismaistest ainetestidest. Eriti oluline on see siis, kui me tahame teada, mitme protsendi ulatuses on õpilane aine omandanud ja millist hinnet talle panna. Seega programmitestides on valemi (1) rakendamine obligatoorne. Seevastu mõnedes eristustestides väärvalikutest tingitud parandust ei kasutata, näiteks Raveni testis. Põhjenduseks on siin asjaolu, et eristustestide eesmärgiks on katseisikuid eristada ja selleks tuleb nad panna lihtsalt võrdsetesse tingimustesse. Sageli ei teatata nende testide täitmise aega õpilastele ette ja nad ei saa seega viimastel sekunditel lahendamata ülesannetele vastuseid huupi märkida. Seejuures on siiski oluline katseisikuid eelnevalt informeerida, kas väärvalikute eest punkte maha arvatakse või mitte.

Kokkuvõttena tahaksin soovitada valikvastustega ülesannete puhul väärvalikute eest punkte maha arvata. See on õiglasem, lähem elule ja vastab arenenud riikide tavale.

Testiülesannete teised liigid

Väga sageli kasutatakse meie ainetestides **vabavastuselisi küsimusi ja ülesandeid**. Tavalised matemaatika-, füüsika-, keemiaülesanded ongi vabavastuselised ülesanded, sest õpilane peab nende ise vastuse koostama. Vabavastuseliste küsimuste näiteks võivad olla järgmised küsimused.

- 1. Millal toimus laulev revolutsioon?*
- 2. Millised on pedagoogilise takti põhireeglid?*
- 3. Miks paistab päike vahel horisondil punasena?*

Vabavastuselisi ülesandeid on kergem koostada kui valikvastustega ülesandeid, sest viimaste puhul tuleb küsimusele veel võima-

likke vastuseid formuleerida. Eespool nägime ka, et valikvastuste saamiseks tuleks küsimus esitada õpilastele enne vabavastuselisena. Seega võib vabavastuselist küsimust vaadelda lähtekohana valikvastega küsimuse saamiseks. Vabavastuselise küsimuse teiseks vooruseks on see, et ta võimaldab õpilasel ise oma vastuse formuleerida. Seega areneb õpilase väljendusoskus ja hiljem vastuseid lugedes näeb õpetaja, kuidas mõtleb üks või teine õpilane. Kolmandaks tuleb märkida, et mõtlemisoscuse kontrollimiseks on kergem koostada vabavastuselisi küsimusi. Tõepoolest, divergentseid mõtlemisülesandeid on parem esitada vabavastuselistena ja loovülesandeid ei saagi valikvastustega anda. Sellegipoolest ei tasu arvata, et vabavastuselised küsimused on alati mõtlemisküsimused. Vabavastuselistena võib väga hästi esitada lihtsaid meenutamisküsimusi nagu esimene ülaltoodud näidetest. Seega sõltub küsimuse kvaliteet ikkagi rohkem tema sisust kui vormist.

Vabavastuseliste küsimuste põhiliseks puuduseks on see, et vastuseid on suhteliselt raske hinnata. Need raskused on tingitud mitmest asjaolust. 1. On tülikam ja aeganõudvam lugeda õpilase käsikirjalisi vastuseid, kui fikseerida, kas ta on märkinud õige valikvastuse. 2. Sageli on osaliselt õige vastuse puhul raske otsustada, kas lugeda see õigeks või vääraks. On muidugi võimalik anda sellise vastuse eest pool punkti või osaliselt punkte, aga üks seegi tee hindamise keerukamaks. 3. Vabavastuseline küsimus võib osutada teatud määral ebamääraseks ja siis tekivad jällegi hindamisprobleemid. Näiteks hindab õpetaja laulva revolutsiooni aastaarvu märkinud õpilase vastuse ebatäiuslikuks, kuna ta pidas silmas laulva revolutsiooni kulminatsiooni kuupäeva. Loomulikult on siin tegemist ebatäpse küsimusega, kuid see ilmneb alles vastuste hindamisel. Järelikult on ka vabavastuselistel küsimustel soovitatav eeltest. 4. Hea kirjamees võib oma stiiliga varjata vastuse sisulisi puudusi ja vastupidi. Ka käekirjast sõltub hinne vastusele väga tugevasti (Lindgren & Suter, 1995, lk 366). Seetõttu ei ole vabavastuse hindamine nii objektiivne kui etteantud vastustest valiku hindamine. Ilmselt mõjutavad siin hindamist kõik need subjektiivsed tegurid, mida tavaliselt suuliste vastuste hindamisega seostatakse.

Lühidalt öeldes on vabavastuselisi küsimusi kergem koostada, kuid vastuseid neile on raskem hinnata. Seetõttu kasutatakse vabavastuselisi küsimusi väikese katseisikute arvu korral ja testi algvariandis. Kui õpilasi on vähe, siis pole mõtet valikvastuste koostamisega vaeva näha

— lihtsam ja huvitavam on lugeda õpilaste koostatud vastuseid. Kui katseisikuid on palju, siis tasub valikvastuste koostamine end ära.

Kui testid on korrektselt koostatud, siis annavad valikvastuselised küsimused ja vabavastuselised küsimused sama häid tulemusi. Loomulikult tuleb hoolitseda selle eest, et küsimused oleksid kõikide nõuete järgi koostatud.

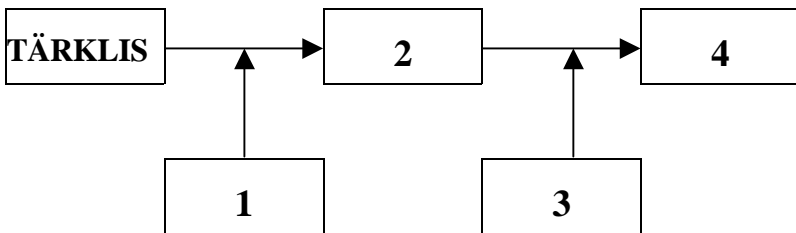
Vabavastuseliste küsimustele on lähedased **lünkülesanded**.

Lünkülesanded on tekstid, milles on mõne sõna või arvu asemel lünk ja õpilane peab selle lünga täitma. Kui tekst on lühike, näiteks lihtlause, siis on selles üks lünk; pikemas lauses või tekstis on mitu lünka. Lüngaks valitakse sõna või sõnaühend, mille taastamine võimaldab kontrollida soovitud mõtte omandamist.

1. Kui pool arvust x on kaheksa siis $x = \dots$
2. Kolmnurga kaatedid on 3 ja 4 cm. Selle kolmnurga hüpotenuus on cm.
3. Valgud ehk koosnevad ühest või mitmest, kus aminohappe jäägid on seotudsidemega.

Kuigi ühte lausesse võib teha mitu lünka, siiski peaks allesjäänud tekst olema piisavalt informatiivne, et oodatavat vastust üheselt määrata. Liiga sagedaste lünkade puhul hakkab lünga täitmine sõltuma sellest, kuidas õnnestus täita naaberlünka. Selline ülesannete kokkukulamine ei ole soovitatav — me saame mitme ülesande asemel ühe, mille täitmist hinnatakse mitmekordse punktide arvuga. Või siis sõltub esimese lünga täitmisest teiste täitmine ja me saame mitu õiget vastust nagu järgmises näites.

Täitke tühjad kastid



*Seda ülesannet on kontekstiväliselt raske mõista, mistõttu vaata-
tame ka autori (K. Mägi) ja lahendajate poolt pakutud vastuseid.*

Autor: 1. Amülaas

Lahendajad: 1. Vesi

2. Maltoos

2. Lahustatud tärklis

3. Maltaas

3. Suhkur

4. Glükoos

4. Kisell

Parempoolne vastus näitab lahendajate loovust ja seega sobiks see ülesanne loovustesti. Tavalises ainetestis soovitakse aga õigena näha vaid üht vastust. Selleks tuleks anda ülesandes piisavalt informatsiooni lõpptulemuse üheseks määramiseks.

Lüngad tekstis võiksid olla ühepikkused, et lünga pikkus ei saaks orientiiriks sobiva vastuse otsingul. Õpilastele tuleks teatada, kas lüngas on alati üks sõna või võib seal olla ka mitu sõna.

Lünkülesande sisuline lähedus küsimustele tuleneb sellest, et küsimuse saamiseks jäetakse lausest mõni sõna või väljend ära, asendatakse see küsisõnaga ja paigutatakse sõnad ümber vastavalt küsilause struktuurile. Siiski ei ole lünkülesanded küsimustega identsed. Lünkülesannete eelis on selles, et neid on lihtsam koostada ja nad on esitusviisilt ökonoomsemad. Lünkülesannete puuduseks on see, et nad suunavad teksti pähe õppima ja annavad vähem võimalusi mõtlemisülesannete esitamiseks. Eelnevast tuleneb, et lünkülesandeid ei tohiks mingil juhul koostada õpiku lausetest.

Lünkülesannetega testi võib koostada nii, et õpilased kirjutavad vastuse testivihikusse. Meil aga on oluline teste hiljemgi kasutada ja siis peaks õpilane oma vastuse kirjutama eraldi vastuselehele. Selleks tuleb testis lüngad nummerdada ja õpilane kirjutab vastuse eraldi lehele vastava numbri järele.

Vastuste hindamisel võib õigeks lugeda vaid selle sõna, mis maha oli kustutatud, või kõik sisuliselt lünka sobivad sõnad. Teine moodus on veidi täpsem ja kindlasti õiglasem, kui on oluline teada, mitu protsenti lünkadest õpilane täitis õigesti. Seega võiks õpetaja seda moodust kasutada omakoostatud kontrolltöodes. Samas võib tekkida probleeme õpilase mõne vastusega: kas lugeda see õigeks või mitte, kui sisuline kõrvalekalle lause esialgsest mõttest pole kuigi oluline. Siit tekkivad paratamatult mõningad hindamisvead ja loomulikult on tülikam kaaluda iga võimaliku vastuse õigsust, kui fikseerida, kas see langeb kokku etteantud vastusega või mitte. Nendel põhjustel võib

mõnikord õigeks lugeda vaid selle vastuse, mis langeb sõna-sõnalt kokku kustutatud sõnaga. Selline hindamismoodus on õigustatud kui võrreldakse õpilasi omavahel, sest õpilased on ühesugustes tingimustes: nad kõik kaotavad punkte seetõttu, et pole leidnud lünga täitmiseks algteksti sõna. Seega võib eristustestides, näiteks kõrgkooli sisseastumistestides õigeks lugeda vaid sõnasõnalised vastused.

Hindamisraskuste vähendamiseks kasutatakse valikvastuseid ka lünkülesannete puhul. Iga lünga täitmiseks võib ette anda teatud sõnad, mille seast õpilane valib sobiva. See on siis valikvastustega lünkülesanne.

Pourquoi es-tu debout si ... puisque tu dois te lever de bonne heure?

a) tard

b) loin

c) peu

d) haut (*Taking the SAT II, 2000, p. 53*).

Valikvastustega lünkülesannete voorused ja puudused on ilmselt analoogilised valikvastustega ülesannete vooruste ja puudustega.

Lünkülesannete üks eriliik — täiendtest — väärib eraldi käsitlemist. **Täiendtest** on lünktest, mille algtekstist on kustutatud iga seitsmes sõna. Sellist testi võib üsna lihtsalt koostada arvuti. Täiendtest on hea vahend tekstide raskuse võrdlemiseks. Kuna täiendtestis kustutakse algtekstist iga seitsmes sõna, siis ei mõjuta testi koostaja suva siin lünkade valikut ja kõikides võrreldavates tekstides valitakse lüngad ühtmoodi. Seetõttu ei sõltu täiendtesti tulemus lünkade valikust vaid ainult tekstide raskusest ja õpilaste võimetest. Kui õpilaste lugemis- oskused on võrdsed, siis saab täiendtestiga hästi võrrelda tekste nende raskuse järgi. Ainult igast tekstist tuleks teha kõik seitse täiendtesti varianti, sest mõneleheküljelise teksti puhul on selle täiendtesti variandid erineva raskusega.

Ingliskeelsetes täiendtestides kustutakse iga viies sõna. Uurimused eestikeelsete tekstidega näitasid, et meil sõltub nii sagedaste lünkade puhul ühe lünga täitmine teise lünga täitmise edukusest. Seetõttu on eesti keeles sobiv kustutada iga seitsmes sõna. Kustutada

võib sõnu muidugi veelgi harvemini, kuid siis on vaja pikemat teksti, et saada soovitud arvu lünki. See reegel lünkade sageduse kohta kehtib täiendtesti puhul ja võib olla teatud orientiiriks lünkülesannete puhul.

Täiendtestis kasutatav sõnade mehhaaniline mahakustusviis tekitab hulgaliselt lünkasid, millele keegi ei suuda vastata ja sellest aspektist ei ole täiendtest ökonoomne õpilaste võrdlemisel. Selle tõttu ei saa ka täiendtestis saadud õigete vastuste protsendile rakendada programmitesti hindenorme (vt Programmitesti tulemuste teisendamine koolihinneteks).

Täiendtesti vastandiks on **liigsete sõnadega tekst**. Õpilase ülesandeks on need liigsed sõnad maha kustutada.

Hobused haukusid sõid koplis, kui parm peremees nende poole jooksis.

Liigsete sõnadega tekst harjutab kindlasti kriitilist lugemist. Sellist teksti on kasutatud vaikse lugemise kiiruse mõõtmiseks. Luge-
miskiiruse mõõtmisel on mõnikord oluline, et loetaks iga sõna ja liigsete sõnade mahakustutamine on selle kontrollimiseks üks võimalus. Nende sõnade leidmine ei peaks olema eriti keeruline ja vastuse märkimine ei võta erilist aega. Muidugi võib ka keerulisemasse teksti sõnu lisada ja siis saame ainetesti, millega kontrollitakse materjali põhjalikumat omandamist.

Valikvastustega ülesande üheks edasiarenduseks on **kõrvutamisülesanded** ehk vastavusse seadmise küsimused. Need on ülesanded, millele võib anda ühed ja samad valikvastused. Kõrvutamisülesandes on tavaliselt kaks tulpa objekte ja nende vahel tuleb leida vastavus mingi reegli järgi.

1. Seadke vastavusse riigid ja nende pealinnad

- | | |
|-------------|--------------|
| A. Eesti | 1. Helsingi |
| B. Saksamaa | 2. Moskva |
| C. Leedu | 3. Stockholm |
| D. Soome | 4. London |
| E. Läti | 5. Vilnius |
| | 6. Tallinn |
| | 7. Berliin |
| | 8. Riia |

2. Seadke vastavusse mõisted ja nende olemus

<i>alus</i>	<i>näitab tegevuse kohta, aega, viisi</i>
<i>öeldis</i>	<i>näitab, millele tegevus on suunatud</i>
<i>sihitis</i>	<i>näitab tegevust või olemist</i>
<i>määrus</i>	

3. Seadke vastavusse füüsikas kasutatavad tähised

a) <i>F</i>	1) <i>N</i>
b) <i>m</i>	2) <i>m/s²</i>
c) <i>p</i>	3) <i>kg</i>
	4) <i>N/m²</i>

Viimane näide on teistest keerukam selle poolest, et siin pole öeldud, mille järgi objekte vastavusse seada. Õpilane peab kõigepealt taipama, et vasakus tulbas on füüsikaliste suuruste tähised ja paremas tulbas mõõtühikute tähised. Alles siis saab ta hakata neid vastavusse seadma.

Kõrvutamises ülesandes on oluline ühes tulbas anda rohkem objekte kui teises. Kui neid oleks ühepalju, siis tuleks viimase ülesande vastus automaatselt välja. Ühest täiendavast vastusevariandist jääb tegelikult isegi väheseks, sest kõrvutamises ülesannete puhul vastuste juhuslikku aimamist ei arvestata, kuigi see on siingi teatud määral võimalik.

Vastuste märkimiseks on kaks võimalust. Esimese näite puhul saab õpilane oma vastustelehel tähe A järel kirjutada näiteks numbriga 6 jne. Teise näite puhul tuleks õpilasel noolega ühendada esimese ja teise tulba kokkukuuluvad objektid. Siis pole test enam teistkordselt kasutatav, kuid nooremate õpilaste puhul on see lihtsam vastuse märkimise viis eelistatavam. Kolmandas näites võib vastuste märkimiseks anda eraldi vastuselehe, millel on järgnevad kirjed:

- a) 1) 2) 3) 4)
- b) 1) 2) 3) 4)
- c) 1) 2) 3) 4)

Siin saab vastaja nüüd igal real ära märkida õigeks peetava vastusevariandi. Selline vastusteleht sobib arvutiga hindamiseks ja arvutitestides.

Üheks võimaluseks on veel kasutada **ümberpaigutusülesandeid**. Nendes on antud mõned objektid, mis tuleb teatud tunnuse alusel õigesse järjekorda seada.

1. *Seadke reageerimisaktiivsuse alusel ritta järgmised keemilised elemendid*

A) Fe B) Na C) Cu D) Zn E) K

2. *Reastage riigid elanike arvu kasvamise järjekorras*

A) Inglismaa

B) Rootsi

C) Prantsusmaa

D) Saksamaa

E) Hispaania

3. *Järjestage alltoodud väited protsessi loogika järgi*

A) Vesi külmub

B) Vesi paisub külmumisel

C) Vesi satub kaljupragudesse

D) Kalju mureneb

Teatud kõhklustega võib ainetesti ülesandeks lugeda ka **essee**. Näiteks USA-s on emakeele testi üheks osaks essee. Selle kirjutamiseks pole tarvis üksikute ainete süvatundmist, küll aga võib kasutada oma kogemust ja teadmisi erinevatelt aladelt põhjendamaks oma väiteid. Essee teema on täpselt ette antud ja aega on 20 minutit. Õpilastele selgitatakse essee olemust järgmiselt. "Esseed kirjutatakse, et anda Teile võimalust näidata, kui hästi Te oskate kirjutada. Seepärast väljendage oma mõtteid teema kohta selgesti ja mõjusalt. See, kui hästi Te kirjutate, on hoopis tähtsam, kui see, kui palju Te kirjutate, aga teema täielikuks käsitlemiseks tuleb Teil ilmselt kirjutada rohkem kui üks lõik. Olge konkreetne" (Taking the SAT II, 2000, p. 11).

Essee eest võib USA-s saada 1–6 punkti. Eksperdid hindavad esseedes nende üldist meeldivust, struktuuri, sõnavalikut, lause ehitust ning punktuatsiooni, näidete sobivust, ideede loogikat (Taking the SAT II, 2000, p. 10).

Viimasel ajal on hakatud rohkem tähelepanu pöörama sooritus-ülesannetele. **Sooritusülesandes** tuleb õpilastel kasutada erinevaid teadmisi ning oskusi ja rakendada neid praktiliste probleemide lahendamiseks. Fikseeritakse ja hinnatakse mitte üksnes lõpptulemust vaid ka selleni jõudmise protsessi. Õpilaste töö hindamiseks peavad olema täpsed kriteeriumid ja kogenud hindajad, aga neilgi kulub rohkem aega ja nad ei hinda alati sama tööd ühtmoodi. Sooritustesti eeliseks on õpilase mõtlemisoskuse mõõtmine reaalses situatsioonis, kuid selle testi läbiviimine on keerukas.

Sooritustesti ülesande näitena vaatleme probleemi "Milline paberist käterätik on restoranis parim?". See sobib näiteks 8-nda klassi loodusõpetuses. Probleemi lahendamiseks antakse õpilasele katsevahendid (erinevad paberist käterätikud, mõõtevahendid jne) ja tööjuhend. Antud juhul võiks tööjuhendis olla järgmised soovitusel:

- 1) kavanda eksperiment vee neeldumise kohta,
- 2) teosta eksperiment,
- 3) esita eksperimendi tulemused aruandena,
- 4) analüüsi, mida tuleks veel arvestada paberist käterätiku valikul.

Iga etapi täitmist hinnatakse eraldi:

- 1) eksperimendi kava puhul hinnatakse selle loogilisust, detailsust,
- 2) mõõtmistel hinnatakse täpsust, tulemuste vormistamise korrektsust,
- 3) aruandes hinnatakse eksperimendi kirjelduse täielikkust, järelduste põhjendatust,
- 4) analüüsis hinnatakse, milliseid paberist käterätikute valiku faktoreid on vaadeldud ja käsitluse asjatundlikust (Capper, 1996, lk 149–150).

Ülesannete liigi valik sõltub paljudest asjaoludest. Ülalpool sai mainitud üht põhiteesi, mille kohaselt ulatuslikult kasutatavas testis õigustavad end valikvastustega ülesanded, aga ühes-kahes klassis tehtaval testil võivad olla vabavastuselised küsimused. Testide kasutamise algetapil võiks selles olla erinevat liiki ülesandeid, et pakkuda õpilastele vaheldust. Keskkooli lõpul peaks tingimata kasutama neid ülesande liike, mis esinevad riigieksamitestides. Ülesande sisu võib määrata sobivaima ülesande liigi. Ja lõpuks võib ülesande liik sõltuda ka testikoostaja eelistustest. Isiklikult eelistaksin valikvastustega

ülesandeid, mida maailmas kõige enam kasutatakse standardiseeritud testides. Üht liiki ülesanded paigutatakse testis tavaliselt kokku.

Ainetestide kursust üliõpilastele lugedes olen palunud neil koostada oma aine kohta test. Üliõpilased on testidesse võtnud väga huvitavaid ülesandeid, millest mõned ka ülalpool esinesid. Üliõpilased on märkinud ülesande juurde punktide arvu, mis selle ülesande lahenduse eest võib saada. Samas on **testides esinenud puudujääke**, millest olulisemad on järgmised.

1. Testile unustatakse kirjutada pealkiri: millise teema kohta, millises aines ja klassis see test sobib.
2. Ülesande või ülesannete rühma ette unustatakse kirjutada juhend, kuidas ülesannet lahendada.
3. Testis on liiga vähe ülesandeid.
4. Ühes ülesandes sisaldub vastus teisele ülesandele.
5. Mõtlemisülesandeid on liiga vähe.
6. Kõrvutamistestis unustatakse ühte tulpa lisada peibutusvastuseid.
7. Lünktestis kipuvad lüngad olema liiga sageli.
8. Hindamisjuhises kasutatakse liiga sageli murdarve. Kui ka ülesande osa lahendamise eest võib punkti saada, siis võiks terve ülesande lahendamist hinnata kahe või enama punktiga.

Heade testiülesannete koostamine eeldab ainetundmist, õpetamiskogemust ja ka testiteooria tundmist. Seetõttu on loomulik, et testide koostamine nõuab harjutamist ja koostatud testide kontrollimist praktikas. Eriti oluline on praktikas kontrollimine nende testide puhul, mida ulatuslikult rakendatakse või mille põhjal tehakse olulisi järeldusi, näiteks riigieksamitööde puhul.

Programmitest

Õpetajad on meil kogu aeg koostanud aineteste, mille abil saab kontrollida mõne tunni või mõne teema materjali omandatust. Selleks on nad valinud vastavast õppematerjalist mõned mõisted, seaduspärasused jne ja nende kohta koostanud küsimused. Õpetajad ei jäta õppematerjalist tavaliselt midagi olulist kontrollimata ega võta testi ka ülesandeid, millele vastamist pole nad õpetanud. Õpetajad koostavad tavaliselt programmiteste. Programmitesti ainus oluline tunnus on see, et selle ülesanded vastavad ainekavale. Ehk täpsemalt öelduna: programmitesti ülesanded on ainekava suhtes esinduslikud. Vaatleme lähemalt, kuidas selle nõude täitmist tagatakse teaduse reeglite järgi.

Testiülesannete esinduslikkus

Ainetestiga kontrollitav materjal on tavaliselt mahukas ja mitmepalgeline. Seal on palju mõisteid, fakte, seaduspärasusi ja neid tuleb omandada erineval tasemel: mõne mõiste puhul peetakse piisavaks lihtsalt selle äratundmist, teist tuleb aga osata rakendada praktiliste küsimuste lahendamisel jne. Siit tuleneb, et testiga ei õnnestu kontrollida materjali kõigi elementide omandatust kõigil soovitud tasanditel. Järelikult tuleb teha mingi valik, mis ei kahjustaks aga testitulemuse objektiivsust. Testitulemus võib muutuda ebatäpseks, kui testiga kontrollitakse põhjalikumalt õppematerjali mingite osade omandatust ja pealiskaudselt teiste osade omandatust. Siis määrab testitulemuse selle materjali omandamine, mida põhjalikumalt kontrollitakse, ja õpilasel pole kuigi palju kasu selle materjali omandamisest, mille kohta on vaid mõni üksik küsimus. Selliste eksituste vältimiseks peab ainetest olema **esinduslik** materjali suhtes, mille omandamist tema abil kontrollida tahetakse.

Esinduslikkus tähendab seda, et kõiki õppematerjali elemente kasutatakse testis sama sageli, kui nad selles materjalis esinevad. Kui näiteks Saksamaa loodust kirjeldatakse õpikus kümnel leheküljel ja Norra loodust ühel leheküljel, siis tuleks Saksamaa looduse kohta esitada kümme küsimust ja Norra looduse kohta üks küsimus. Kui testi mahub Saksamaa looduse kohta vaid kaks küsimust, siis jääb Norra

looduse teadmine selles testis kontrollimata. Toodud lihtsustatud näitest ilmneb muuseas, miks pikk test on lühikesest parem.

Õppematerjali elementidena on siin silmas peetud selle kõikvõimalikke osi: mõisted, seadused, faktid, omandamise tasemed, pädevused, õpetamise eesmärgid jne. Kui õpetamise eesmärgid on detailselt fikseeritud, siis on kõige lihtsam testi koostamisel lähtuda nendest. Selleks koostatakse iga eesmärgi omandatuse kontrollimiseks üks ülesanne. Detailselt fikseeritud õpetamise eesmärgid võivad aga olla väga mahukad, sest nendes on üles loetud kõik mõisted, seadused, faktid ja nende omandamise tasemed. Nii võib juhtuda, et testi ei mahu iga eesmärgi saavutatuse kontrollimiseks üht ülesannet. Siis tuleb nendest eesmärkidest teha valik, mis on esinduslik ja piisav.

Esindusliku kogumi valiku meetodid on hästi läbi töötatud sotsiaalteadustes (Mereste, 1975, lk 320–371; Tooding 1998, lk 154–170), kus küsitletavad isikud peavad esindama kogu populatsiooni. Kuna meil on programmitesti ülesannete valikul täpselt sama eesmärk, siis saame kasutada sotsiaalteadustes väljatöötatud meetodeid. Lühidalt kirjeldades on **ülesannete esindusliku kogumi saamiseks** järgmised võimalused.

1. Iga õpetuse eesmärk kirjutatakse omaette lehekesele, need asetatakse loosirattasse ja tõmmatakse sealt soovitud hulk õpetamise eesmäärke. Iga väljatõmmatud mõiste, seaduse, fakti jne kohta koostatakse ülesanne omandatuse kontrollimiseks ettenähtud tasemel. Selline ülesannete juhuslik valik annab hea tulemuse, kuid on raskesti teostatav.
2. Õpetuse eesmärgid on oma loetelus mingis järjekorras. Ainekavas on mõisted, faktid, seadused tavaliselt aineloojika järjekorras. Selles järjekorras võib õpetuse eesmärgid ka numereerida ja testi valida iga n-inda eesmärgi saavutatuse kontrollimise. Kui näiteks ainekavas on loetletud 600 mõistet ja testi soovitakse võtta 200 ülesannet, siis tuleks testi võtta ülesanne 3., 6., 9. jne mõiste kohta. Testi teise varianti võib siis võtta ülesanded 2., 5., 8. jne mõiste kohta. Õpieesmärkide fikseeritud loetelu puhul võib sellest elemente kontrolliks valida ka juhuslike arvude tabeli abil.
3. Testi võib koostada ka nii, et kõigepealt jaotatakse ainekava elemendid teatud olulise tunnuse järgi gruppidesse. Näiteks teadmised, mida tuleb a) ära tunda, b) meenutada, c) rakendada

tuntud situatsioonis, d) rakendada uues situatsioonis. Seejärel loetakse ära, mitu elementi on igas grupis ja testi võetakse elemente igast grupist võrdeliselt grupi suurusega. Kui näiteks ainekavas on elemente, mida tuleb ära tunda, kaks korda vähem kui elemente, mida tuleb meenutada, siis võetakse äratundmist nõudvate elementide seast kaks korda vähem elemente kui meenutamist nõudvate elementide seast. Igas elementide grupis tehakse omakorda valik rakendades üht ülaltoodud meetodeist. Loomulikult võib siinjuures rakendada elementide klassifitseerimiseks teisi aluseid — teema, elemendi liik jne — ja korraga kasutada elementide jaotust kahe või enama tunnuse järgi.

Praktiliselt kõigi nende meetodite puhul tõuseb probleem elementide ebavõrdsest kaalust: ainekursuse põhiline mõiste võib olla ühe sõnaga väljendatud niisama nagu vähemtähtis mõiste. Seega on nendel mõistetel võrdsed võimalused sattuda testi, kuid tegelikult oleks olulisem küsida põhimõistete omandamist. Selle probleemi olemasolul tuleks kontrollida, kas vaadeldav põhimõiste esineb ainekavas tõepoolest vaid ühel korral või tuleb ta kasutusele ja kontrollimisele ka teiste ainekava osade juures. Kui see nii ei ole, siis võiks ainekava põhielementidele anda suurema kaalu, et nende tõenäosus sattuda testi vastaks nende tähtsusele ainekavas. Piltlikult võiks seda ette kujutada nii, et põhimõiste kirjutatakse elementide loetellu kaks või kolm korda ja siis nummerdatakse elemendid, et leida neist iga n -is, mille kohta koostatakse testis küsimus. Loomulikult on nüüd põhielemendil suurem tõenäosus sattuda testi kui vähemtähtsal elemendil.

Ainekursuse **esinduslik test ei saa olla lühike**. Poole tunniga ei saa kontrollida, mil määral on omandatud mingi aine aastakursus. Lühikeses testis võivad olla juhuslikult ülekaalus näiteks kerged ülesanded ja siis saab õpilane parema testitulemuse, kui on tema teadmiste tegelik tase. Seetõttu on oluline, et ka testiülesannete esindusliku valiku korral oleks nende hulk küllaltki suur. USA-s kasutatavates koolijõudluse testides on iga aine kohta umbes 100 küsimust, millele lisanduvad mõtlemistestid emakeelest ning matemaatikast.

Mida rohkem on testis ülesandeid, seda väiksem on selle testi esindusviga. Samal ajal on aga pikemat testi raskem koostada ning läbi viia. Selle tõttu tuleks määrata igal juhul eraldi testi optimaalne pikkus, lähtudes lubatavast esindusveast ning kulutustest testile. Testi

ülesannete arvu määramisel lähtutakse väljavõtu piirvea valemist (Mereste, 1975, lk 336)

$$\Delta = \frac{t\delta}{\sqrt{n}} \quad (2)$$

milles Δ — väljavõtu piirviga,
 t — Studenti tõenäosuskordaja,
 δ — ülesannete raskuste standardhälve,
 n — ülesannete arv.

Avaldame sellest valemist ülesannete arvu n

$$n = \frac{t^2 \delta^2}{\Delta^2} \quad (3)$$

Kui nüüd eelkatses on teada ülesannete raskuste standardhälve, siis võib arvutada, mitu ülesannet on tarvis, et väljavõtu piirviga oleks etteantud suurusega.

Oletame näiteks, et ülesannete raskuste standardhälve on 0,2 ja me lubame väljavõtu piirveaks 0,02. Tõenäosuskordaja t leiame Studenti tabelist (Mereste, 1975, lk 466), pidades ülesannete arvu küllalt suureks, et seda antud tabelis lähendada lõpmatuseni, ja võttes lubatud eksimistõenäosuseks 0,05. Vajalik ülesannete arv on siis

$$n = \frac{1,96^2 * 0,2^2}{0,02^2} = 384$$

Valemiga (2) saab esindusviga prognoosida. Tegelik esindusvea saame teada ikkagi siis, kui me koostatud testi õpilastega läbi viime ja selle vea valemiga (2) uuesti arvutame.

Sama õppematerjali kohta saab koostada kaks või enam erinevatest ülesannetest esinduslikku testi. Kuna need testid on koostatud sama materjali kohta, siis on need võrdsed kõikide tunnuste osas, sealhulgas küsimuse raskuse osas, kuigi ülesandeid raskuse järgi testi ei valita. Näiteks vabariiklikes tasemetöodes või riigieksamitel peavad testi kaks varianti olema sama materjali suhtes esinduslikud.

Eelnevast tuleneb üks võimalus testide esinduslikkuse kontrollimiseks. Palume ühel õpilasgrupil (või mitmel võrdväärset grupil)

vastata testi eri variantidele ja vaatame kui palju erinevad tulemused eri variantide puhul. Ainult siis, kui see erinevus on küllaltki väike, võime me öelda, et selle testi ükskõik milline variant mõõdab tema aluseks olnud õppematerjali omandamist küllaltki täpselt. Kui eri variantide tulemused samadel õpilastel oluliselt erinevad, siis on kõige parem lähtuda keskmise tulemuse andnud variandist või kõik testivariandid uuesti koostada.

Testimise juhend

Testimise esimeseks sammuks on testitava materjali valik. Kuidas seda teha, vaatlesime eelmises osas. Seejärel tuleb iga testitava elemendi kohta koostada ülesanne, mille liike vaatlesime eelmises peatükis. Koostatud testi läbiviimisele esitatakse samuti nõudeid, milliste järgimiseta pole võimalik saada piisavalt täpseid tulemusi.

Test on mõõtmisvahend. Igasugune mõõtmine on võrdlemine mõõtühikuga. Et see võrdlemine annaks õigeid tulemusi, peab mõõtühik olema kogu aeg sama pikk. Testi puhul tähendab see, et testida tuleb alati ühtmoodi. Testimise ja tulemuste arvestamise meetodid tuleb täpselt fikseerida ja neid järgida. See on testi standardiseerimise oluline osa.

Testimise juhend esitatakse testi algul. Selles võiks olla kajastatud järgmised momendid.

1. Testi olemus: mida test mõõdab, millist õpilaste ettevalmistust ta eeldab ja kuidas kasutatakse testi tulemusi.
2. Lubatud abivahendid testi täitmisel. Nende loetelu võib olla väga erinev alates kustutuskummist ja lõpetades taskuarvuti ning teatmeteostega. Taskuarvuti kasutamisse suhtutakse viimasel ajal siiski ettevaatlikult, sest sellesse võib olla salvestatud informatsiooni, mille kasutamist testi koostajad ette ei näinud. Lubatud abivahendite loetelu peaks olema testi täitjale teada mõni päev enne testi täitmist, et ta saaks need vajadusel endale muretseda.
3. Kuhu ja kuidas kirjutada vastused: kas otse testi või eraldi vastustelehele? Kui on eraldi vastuste leht, kuidas seda siis täita. Kui vastustelehti hindab arvuti, siis on väga oluline teha valikut tähistav ovaal pliiaatsiga täielikult mustaks. Kas abiarvutusi ja muid märkmeid võib teha testivihikusse või lisalehele või pole need üldse lubatud?

4. Kui palju on testi täitmiseks aega ja kui palju on seal ülesandeid. Ajapiirangud on meile tulnud põhiliselt riigieksamitega, aga USA-s on need ammugi tarvilusel. Nad põhjendavad oma traditsiooni sellega, et kui õpilane ei suuda piisavalt kiiresti lahendada lihtsaimaid ülesandeid, siis ei jäägi tal aega keerulisemate ülesannetega tegelemiseks. Praktiliselt ei õnnestu peaaegu kunagi anda kõigile õpilastele piisavalt aega, sest mõni õpilane sooviks ikka veel edasi mõelda.
5. Valikvastustega ülesannete puhul arvestatakse välismaal vale vastuste eest punkte maha (Taking the SAT I, 2000, p. 76). Ka meil on otstarbekas seda teha ja vastav märkus lisada testi juhendisse. Välismaa testi juhendites näidatakse õpilastele koguni, kuidas seda reeglit enda kasuks tarvitada.

Lisaks sellisele üldinstruktsioonile tuuakse täpsemad selgitused ülesanneterühma kohta veel selle ees ja vajadusel tuuakse näidisülesanne koos lahendusega.

Testi läbiviija peab arvestama veel reegleid, mida testis tavaliselt ei trükita. Esitame nendest olulisemad.

1. Iga testi täitja istub omaette pingis, millede vahe on umbes üks meeter.
2. Igale testi täitjale antakse testi ülesanded ja leht vastamiseks.
3. Testija palub testi täitjatel kirjutada vastustelehele oma nime, kuupäeva, testi nimetuse, variandi numברי ja muud vajalikud andmed enne testitäitmise algust.
4. Vajadusel loeb testija ette testi instruktsiooni ja kirjutab tahvlile testi täitmise alguse ning lõpu aja.
5. Testija loob oma käitumisega rahuliku ja heasoovliku õhkkonna. Ta ei räägi kellegagi ega tegele kõrvaliste asjadega, väljendades nii lugupidamist testi täitjate töö vastu.
6. Õpilastele võib teatada, et testi täitmine lõpeb näiteks 5 min pärast.
7. Aja lõppedes palub testi korraldaja panna õpilastel kirjutusvahendid käest ja sulgeda testivihikud. Ta korjab kokku kõik testivihikud ja vastustelehed, teatab, millal õpilased saavad teada oma töö tulemused ja alles siis lubab õpilastel ruumist lahkuda.

Võimete testide puhul on reeglilik, et nende tulemusi teatatakse ainult katseisikule endale. Kas sama reeglit ei tuleks rakendada ka

ainetestide puhul, et vähendada testide võimalikke negatiivseid tagajärgi?

Testimise reegleid tuleb loomulikult täpselt täita. Ainult nii on võimalik ühtmoodi mõõta ja ainult siis saab test täita oma ülesandeid. Kui õpetaja annab oma õpilastele ettenähtust rohkem aega, siis võib ta mõnes õpilases äratada ootusi, mille realiseerimine osutub hiljem väga raskeks. Kui testija ei soovi märgata mittelubatud vahendite kasutamist testimisel, siis süvendab ta eelmisest riigist pärit mõtteviisi ja elulaadi. Test on objektiivne mõõtmisvahend ainult siis, kui seda tehakse ja hinnatakse täpselt nõuete kohaselt. Kes objektiivsust ei soovi, see teste ei tee ja astub isegi nende vastu välja.

Testivastuste skoorimine

Skoorimise all mõeldakse vastuste eest punktide andmist õpilasele. Kõikide õpilaste vastuste skoorimine ühesuguste põhjendatud reeglite järgi on testi objektiivsuse saavutamisel oluline samm. Reeglid punktide andmiseks peaks olema võimalikult täpsed ja selged. Kõige lihtsam on siin olukord valikvastustega ülesannetega: õige vastuse eest antakse üks punkt ja väärvastuse eest arvestatakse punktiosa maha. Mõeldavad on ka teistsugused reeglid, kuid neid ei kasutata. Skoorimise lihtsus ja täpsus on valikvastustega ülesannete oluline eelis.

Testiülesanded võivad olla erineva keerukusega ja siit tekib soov anda **keerukama ülesande eest rohkem punkte**. Skoorides taolise keerulise ülesande vastuseid ainult nulli ja ühega polegi võimalik märkida, et õpilane on mõne osaülesande lahendanud. Seega tundub olevat põhjendatud anda keerulisema ülesande lahenduse eest rohkem kui üks punkt. Millise ülesande osa eest punkti anda, peab olema täpselt fikseeritud skoorimisjuhendis. Näiteks sõnalises aritmeetikaülesandes võib iga õigesti tehtud tehte eest anda ühe punkti. Kui ajaloo testis on küsitud kolme kunstitegelast, kuid õpilane teab vaid kahte, saab ta kaks punkti kolmest jne.

Vabavastuseliste ülesannete ja küsimuste vastuste skoorimisjuhendit on raske koostada. Üks võimalus on lähtuda ideaalvastusest ja sellele iga osa eest anda üks punkt. Teine võimalus on testi katsetada ja ülesandele saadud vastused grupeerida. Iga konkreetse grupi tüüpvastus ja selle eest saadav punktide arv tuleks siis fikseerida skoorimisjuhendis.

Üks raskemini hinnatavaid testiülesandeid on ilmselt **essee**. Selle **hindamine** sarnaneb iluuisutamise hindamisega, kus vajatakse kogenud eksperte. Aga nendelgi on teatud reeglid tulemuse pallide fikseerimiseks. Ameerika Ühendriikide Testimise Teenistus kasutab esseede hindamisel järgmisi kriteeriume (Taking the SAT II, 2000, p. 12).

"Tulemus 6 punkti

Essee näitab selget ja järjekindlat kompetentsust, kuigi selles võivad olla juhuslikud eksimused. See essee

- käsitleb teemat efektiivselt ja ülevaatlilikult,
- on hästi organiseeritud ja täielikult välja arendatud, sisaldades täiesti sobivaid näiteid ideede toetamiseks,
- on kogu ulatuses kirjutatud ladiusas keeles, demonstreerides lausete erinevaid struktuure ja rikkalikku sõnavara.

Tulemus 5 punkti

Essee näitab järjekindlat kompetentsust, kuigi selles on juhuslikud eksimused või puudujäägid kvaliteedis. See essee

- käsitleb teemat efektiivselt,
- on üldiselt hästi organiseeritud ja välja arendatud, sisaldades sobivaid näiteid ideede toetamiseks,
- on kirjutatud ladiusas keeles, demonstreerides mõningat varieeruvust lausete süntaksis ja sõnavaras.

Tulemus 4 punkti

Essee näitab kompetentsust, kuigi selles on juhuslikud vead või puudujäägid kvaliteedis. See essee

- on teemakohane,
- on struktureeritud ja välja arendatud, sisaldades näiteid ideede toetamiseks,
- on kohati kirjutatud ladiusas keeles, sisaldades mõningaid grammatika- või stiilivigu,
- laseehituse varieeruvus on minimaalne.

Tulemus 3 punkti

Essee näitab arenevat kompetentsust. Selles essees on üks või rohkem järgmistest puudustest:

- mitteadekvaatne ülesehitus või teema arendus,
- ideid toetavad detailid on sobimatud või ebapiisavad,
- on mitu grammatika-, stiili- või lausestruktuuri viga,

Tulemus 2 punkti

Essee näitab mõningast ebakompetentsust. Seda esseed kahjustavad üks või mitu järgmistest puudustest:

- halb ülesehitus,
- nõrk teema arendus,
- ideid toetavad detaile on vähe või on nad sobimatud,
- grammatika-, stiili- või lausestruktuuri vead esinevad sageli,

Tulemus 1 punkt

Essee näitab ebakompetentsust. Seda esseed kahjustavad tõsiselt üks või mitu järgmistest puudustest:

- väga halb ülesehitus,
- väga nõrk teema arendus,
- keelekasutus ja süntaktilised vead on sellised, et mõtte taipamine on raskendatud".

Ilmselt võib skoorimise juhendit ka detailsemalt üles ehitada, andes teatud hulga punkte sisu, struktuuri, sõnastuse, näidete jne eest. Näiteks struktuuri eest võib anda kuni kolm punkti, jaotades siis hinnatavad esseed nõrga, rahuldava ja hea struktuuriga esseedeks. Sellise süsteemi puhul võib essee erinevate aspektide osakaal üldhinnangus olla erinev, näiteks sisu eest võib anda kuni 5 punkti. Igal juhul peab skoorimise juhend olema osa testi juhendist.

Programmitesti tulemuste teisendamine koolihinneteks

Aineteste viiakse läbi selleks, et selgitada õpitulemuste taset. Seda taset saab fikseerida millegi või kellegi suhtes. Meil on ainetestide puhul olnud hindamise aluseks kõige sagedamini ainekava. See, mil määral ainekavas fikseeritud eesmärgid on saavutatud, määrab ära õpilase hinde.

Programmitesti tulemuste põhjal saab panna õpilasele hindeid. Selleks tuleb vaid kokku leppida hinnete ja õigete vastuste protsendi vastavuses. Taolisi vastavusi on mitmeid, esitame neist kaks.

Eestimaa üldhariduskoolide õpilaste õpitulemuste ja hinnete vastavus on toodud tabelis 1 (Õpilaste ..., 2000). Nendes hinnete normides on õpetajal lubatud piire 5% võrra muuta sõltuvalt töö raskusest ja

mahust. Kolme alampiiriks on fikseeritud 50%, mis on üsna madal. See tähendab, et rahuldava hinde saab õpilane, kellel on pool materjalist omandamata. Sisuliselt on muidugi väga oluline, et kolmele vastava 50% sisse jääksid kõik põhiteadmised, mis on vajalikud aine edasiseks omandamiseks. Testide puhul seda nõuet aga kahjuks tavaliselt ei järgita.

Tabel 1

Programmitestile antud õigete vastuste protsendi ja koolihinnete vastavus (Õpilaste ..., 2000)

Hinne	Õigete vastuste protsent
5	90–100
4	70–89
3	50–69
2	25–49
1	0–24

Tartu Ülikoolis 1999 aastast kehtiv hindamissüsteem on toodud tabelis 2.

Tabel 2

Hindamissüsteem Tartu Ülikoolis

Hinde tähis	Hinde nimetus	Aineprogrammi omandatuse protsent
A	Suurepärase	91–100
B	Väga hea	81–90
C	Hea	71–80
D	Rahuldav	61–70
E	Kasin	51–60
F	Puudulik	0–50

Selleski süsteemis on rahuldava hinde alampiiriks seatud 51% õigeid vastuseid ja selle süsteemi rakendamine eeldab, et üliõpilasele esitatud küsimused võimaldavad saada ülevaadet aineprogrammi kui

terviku omandatusest. Positiivseid hindeid on selles skaalas viis, mis on kahe võrra rohkem kui üldhariduskoolis.

Meenutame, et enne programmitesti tulemuste järgi hinde fikseerimist on oluline õigete vastuste protsendist lahutada täppiläinud juhuslike vastuste protsent. Seda saab teha valemi (1) järgi.

Ainekava optimaalse mahu määramine

Programmitesti tulemuste järgi on võimalik hinnata mitte üksnes õpilasi vaid ka ainekavade jõukohasust. Peatume sellel probleemil pikemalt.

Kui programmitesti tulemused on kehvad, siis on tavaliselt süüdistatud õpetajat ja viimane süüdistab omakorda õpilasi ja lapsevanemaid. Niisugune lähenemine on aga poolik ja kohati ebaõiglane. Programmitesti ebarahuldavate tulemuste põhjuseks võivad olla ka liiga mahukad ainekavad⁴. Ainekavad peavad ju jõukohased olema reaalses tingimustes, see tähendab, antud õpilaste ja õpetajate puhul. Kui õpilaste esindusgrupp saab programmitestis liig palju mitterahuldavaid hindeid, siis tuleb üldkohustusliku ainekava mahtu vähendada. See on kõige kiirem tee õpilaste optimaalseks arendamiseks ning koolirõõmu suurendamiseks.

Programmitesti tulemuste põhjal võib arvutada, kui palju tuleks vähendada (või suurendada) ainekava mahtu. Arvutuste eeldusteks on rahuldava hinde alampiir (meil 50%), lubatav puudulike protsent (olgu see näiteks 5%) ja programmitesti tulemused õpilasgrupis. Nendest tulemustest leitakse 5% nõrgemaid õpilasi ning nende maksimaaltulemus punktides. See on tulemus, millest kõrgemaga peaksid õpilased saama kolme. See ongi siis kolme alampiir punktides ja optimaalne ainekava maht ületab seda kahekordselt.

Võtame näiteks geograafia ainetesti tulemused 854 üheksanda klassi õpilasel (V ja IX klassi ..., 1974) (Tabel 3).

⁴ Hiljuti konstateerisid ühe aine õpetajad, et nende aine tundide arvu on viimase kümne aasta jooksul vähendatud kolmandiku võrra, kuid aine maht pole muutunud. Paratamatult on see põhjustanud õpilaste suure ülekoormuse, sest ka 10 aastat tagasi olid aineprogrammid paljudele õpilastele ülejõukäivalt rasked.

Geograafia ainetesti tulemused

Punkte	Õpilaste arv	Õpilaste %	Summaarne %
1	1		100
2	3		100
3	8	1	100
4	6	1	99
5	22	3	98
6	24	3	95
7	31	4	92
8	44	5	88
9	59	7	83
10	73	9	76
11	80	9	67
12	67	8	58
13	103	12	50
14	80	9	38
15	67	8	29
16	59	7	21
17	53	6	14
18	30	3	8
19	29	3	5
20	15	2	2

Tabelist näeme, et 95% tugevamaid õpilasi sai tulemuseks vähemalt 6 punkti. Eelneva arutluse kohaselt tähendab see, et need õpilased oleksid pidanud saama vähemalt rahuldava hinde. Teisisõnu, 6 punkti peaks vastama poolele õppematerjalist. Kogu õppematerjali maht peaks siis vastama 12 punktile. Praegu on kogu õppematerjali maht vastavuses 20 punktiga, järelikult tuleks olemasolevat mahtu 8 punkti võrra vähendada. Kaheksa punkti moodustab 20-st 40%, seega tuleks selle testi andmetel õppematerjali mahtu vähendada 40% võrra.

Esimesel pilgul paistab, et arvutustes lubatud 5% puudulikke hindeid on liiga palju. Tegelikult võib arvata, et õppematerjali mahu sedavõrd ulatuslik vähendamine tõstab õpimotivatsiooni ka kõige nõrgematel õpilastel ja seetõttu ületab tulemuste tegelik tõus prognoosi.

Teiselt poolt on 95 protsendi tugevamate õpilaste minimaalse tulemuse määramine eelkirjeldataud viisil ebatäpne, mistõttu oleks hea, kui saaks jaotust selle punkti juures modelleerida. Mõnikord võib mudeliks sobida ka normaaljaotus.

Ainekava optimaalse mahu arvutusi tuleks tingimata teha selle koostamise käigus, aga neid arvutusi võivad teatud määral kasutada ka õpetajad oma klassis. Arvutustes on mõned kokkulepped (kolme alampiiir, lubatud puudulike %), mis võivad aja jooksul muutuda ja seda tuleb siis arvestada. Eriti tuleb aga silmas pidada, et arvutatav optimaalne maht tagab peaaegu täieliku õppeedukuse. Tugevamatele õpilastele on siinjuures ilmselt tarvis lisamaterjali, mis ületab selle mahu.

Eristustest

Programmitesti eesmärgiks on kindlaks teha, kui hästi õpilased on omandanud ainekavas ettenähtud õppematerjali. Selliste testide koostamine ja läbiviimine on meil aastakümnete pikkune traditsioon. Õpetaja poolt koostatud tunnikontrollid ja kontrolltööd on tavaliselt programmitesti funktsioonides.

Samas tõuseb üha jõulisemalt esile ainetestide kasutamine teisel eesmärgil — õpilaste eristamiseks ainealaste teadmiste ja oskuste järgi. Niisugune eristamine on vajalik näiteks kõrgkooli vastuvõtul.

Eristustestid on meil tegelikult juurdumas tänu riigieksamitele, mida me koostame välismaa testitraditsioonide järgi. Psühholoogias kasutatavad võimete testid on samuti eristustestid. Tutvume nüüd lähemalt eristustestide koostamise ja tulemuste tõlgendamisega.

Testiülesannete raskus

Kui kõik õpilased lahendavad ülesande, siis ei ole selle ülesande abil võimalik selgitada, kes nendest on aine paremini omandanud, kes halvemini. Samuti ei võimalda õpilasi eristada ülesanne, mida keegi ei suuda lahendada. Eristustesti on mõtet võtta ülesandeid, mida osa õpilasi lahendab, osa mitte. Samas pole ka mõtet võtta ainult neid ülesandeid, mida lahendab pool õpilastest, sest siis jääksid tugevamas pooles õpilased omavahel eristamata. Eristustesti ülesanded peaksid olema erineva raskusega. Näiteks mõnda ülesannet lahendab vaid 16% õpilastest, mõnda ülesannet pooled õpilastest, mõnd järele 84% õpilastest jne.

Ülesande raskus määratakse katseliselt, aga katses sõltub lahendamise tulemus oluliselt õpilaste võimekusest. Tugevama õpilasgrupi jaoks võib ülesanne olla kerge ja nõrgema grupi jaoks raske. Järelikult peame alati teadma, millise õpilasgrupi tulemuste järgi ülesannete raskus on arvatud. Alljärgnevas eeldame, et meil on fikseeritud õpilasgrupp, kelle jaoks ülesande raskust arvutame. Kõige loomulikum on selleks õpilasgrupiks võtta vastava klassi õpilaste esindusgrupp. Õigupoolest saamegi ülesande raskusest mingi klassi

õpilaste jaoks rääkida vaid siis, kui katse on tehtud selle klassi õpilaste esindusgrupiga.⁵

Testiülesande raskuse tavaliseks näitajaks on seda **mittelahendanud ja valeslilahendanud õpilaste protsent**. Selle näitaja saamiseks tuleb anda õpilastele piisavalt aega kõigi testiülesannete lahendamiseks. Vastasel juhtumil osa õpilasi ei jõua viimastena paigutatud ülesanneteni ja nende mittelahendamiste protsent tuleb suurem, kui on nende ülesannete tegelik raskus.

Testiülesande raskuse vastandmõisteks on ülesande jõukohasusmäär, mis arvutatakse ülesannet õigesti lahendanud õpilaste protsendina (Kärner, 1976, lk 19). Kui jõukohasusmäär lahutada sajast saame ülesande raskuse

$$D = \left(1 - \frac{R}{N}\right)100\% \quad (4)$$

kus D — ülesande raskus protsentides,
R — õigesti vastanud õpilaste arv,
N — õpilaste arv katses.

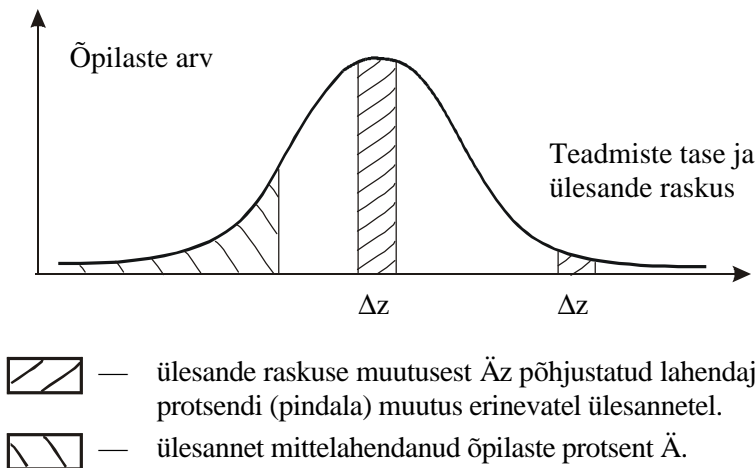
Selles valemis on otstarbekas kasutada ülesannet teadlikult õigesti lahendanud õpilaste arvu. Seega tuleks õigete vastuste näivast arvust lahutada parandusliige, mis oleneb väärvastuse valinud õpilaste arvust ja valikvastuste arvust ülesandes (vt valem 1).

Ülesande raskus protsentides võimaldab võrrelda, kumb kahest ülesandest on raskem. Samas ei saa protsentide järgi öelda, kui palju on üks ülesanne teisest raskem, sest ülesannet lahendanud õpilaste protsentide erinevus ei peegelda õigesti ülesannete raskuste erinevust. **Neid protsente pole võimalik liita ega arvutada nende aritmeetilist keskmist.** Vaatame seda probleemi lähemalt.

Sisuliselt on ülesande raskus õpilase teadmiste või oskuste tase, mis on vajalik selle ülesande lahendamiseks. Raskema ülesande lahendamiseks on tarvis rohkem teadmisi ja oskusi. Teadmiste tase ei

⁵ Esindusgrupi mõistest ja moodustamise viisidest võite lugeda näiteks L.-M. Toodingu (1998) raamatust.

kasva õpilasgrupis ühtlaselt: tavaliselt on keskmiste teadmistega õpilasi rohkem kui väga rumalaid või väga tarku õpilasi (joon. 1). Seetõttu kergete ülesannete raskuse suurenedes kasvab neid lahendanud õpilaste protsent aeglaselt (graafiku vasak osa). Keskmiste ülesannete raskuse kasvades kasvab neid lahendanud õpilaste protsent (joonealune pindala) kiiresti. Raskete ülesannete puhul muudab eelnevaga võrdne raskuse kasv jällegi vähe õigete vastuste protsenti. Seega ülesannet õigesti lahendanud õpilaste protsenti ja ülesande raskuse vahel pole lineaarset seost.



Joon. 1. Õpilaste tavaline jaotus nende teadmise ja oskuste järgi.

Selleks, et **ülesannet mittelahendanud (ja valesti lahendanud) õpilaste protsendilt üle minna ülesande raskuse näitajale**, tuleb kasutada normaaljaotuse jaotusfunktsiooni tabelit (lisa 1). Lisas 1 toodud tabeli kasutamisel tuleb protsentnormide tulbas kasutada ülesannet mittelahendanud õpilaste protsenti.

Oletame näiteks, et esindusgrupis 420-st õpilasest lahendas vaadeldud ülesande õigesti 120. Selle ülesande raskus protsentides on siis

$$D = \left(1 - \frac{120}{420}\right)100\% = 71,4$$

ja raskus z-skaalal 0,57 (Lisa 1). See on õpilastele keskmiselt sobivatest ülesannetest raskem.

Ülesannete raskus z-skaalal väljendub murdarvudes ja pooltel juhtudest negatiivsetes murdarvudes. Raskuse negatiivne näitav tähendab, et ülesanne on keskmisest kergem. Keskmise raskusega ülesande raskus väljendub z-skaalal nulliga. Raskuse positiivse näitavuga ülesanded on keskmisest raskemad.

z-skaalal väljendatud ülesannete raskustele võib leida aritmeetilise keskmise ja kasutada teisi intervalliskaala statistikuid. Õige ei ole aga arvutada, mitu korda on üks ülesanne teisest raskem, sest seegi skaala pole suhteskaala (vt lisa 4).

Eristustesti tuleb valida erineva raskusega ülesandeid. Nende ülesannete raskus võiks kõikuda vahemikus -2 kuni +2. Loomulikult peavad need ülesanded olema õpitud materjalist, kuid ülesannete sisuline sobivus pole eristustesti koostamisel ainsaks kriteeriumiks.

Testiülesannete esinduslikkus on ülesannete valiku ainsaks kriteeriumiks programmitesti koostamisel. Programmitesti koostamisel pole aga tarvis ülesannete raskust arvestada. Programmitest ei tarvitse õpilasi üldse eristada, näiteks keskkooli ainekava suhtes esinduslik test võib osutada paljudele sisseastujatele väga lihtsaks ja enamus õpilasi saab maksimaalse hinde. Eristustestis seda ei juhtu.

Ülesannete raskusega on Eestimaa testipraktikas seotud tõsine probleem. Ühelt poolt on testide koostamisel arvestatud ülesannete raskust, nii et testis ei oleks lihtsaid ülesandeid, mida kõik õpilased suudavad lahendada. Seega on lähtunud teatud määral eristustesti koostamise põhimõtetest. Teiselt poolt on sama testi tulemuste hindamisel kasutatud või püütud kasutada programmitesti tulemuste hindamise põhimõtteid. Ja siis on leitud, et õpilaste teadmised on väga madalad. Tegelikult on ebaseaduse oluliseks põhjuseks kokkusobimatute testimise põhimõtete samaaegne kasutamine. Niisuguste segaduste vältimiseks tuleb testi kavandamisest kuni selle tulemuste tõlgendamiseni täpselt silmas pidada, kas on tegemist programmitestiga või eristustestiga.

Testiülesannete diagnoosiv väärtus

Testid võib jagada kahte liiki: **homogeensed ja heterogeensed testid**. Homogeense testi kõik ülesanded mõõdavad sama oskust või võimet. Heterogeense testi erinevad ülesanded mõõdavad erinevate teadmiste ja oskuste omandamist; nende ülesannete lahendamisedukuse omavahelised korrelatsioonid pole väga kõrged. Ühe oskuse või võime testid on aga koostatud trepi põhimõttel: kui õpilane suudab vastata raskemale ülesandele, siis suudab ta vastata ka kõikidele kergematele ülesannetele. Võimete ja oskuste testid on sageli homogeensed testid. Kui oskuste testis mõõdaks üks ülesanne ühte oskust ja teine teist, millist oskust mõõdaks siis test tervikuna? Võiks ju vastata, et see test mõõdab mingit kompleksset oskust, kuid kompleksse oskuse mõõtmiseks on otstarbekam teha homogeensete testide pakett, millest igaiüks mõõdab ühte oskust ja kokku annavad nad selle kompleksse oskuse mõõtmisvahendi.

Kindlate oskuste ja võimete mõõtmiseks ette nähtud eristustestid peaksid olema homogeensed. **Kuidas homogeenset testi koostada?**

Homogeense testi saamiseks koostatakse või valitakse hulk ülesandeid, mille lahendamine näitab antud oskuse arengutaset. Nende ülesannete saamisel on otsustav asjatundjate arvamus ülesande sobivuse kohta oskuse mõõtmiseks. Ülesandeid võetakse vaatluse alla vähemalt 2–3 korda rohkem, kui neid on testis hiljem tarvis. Seejärel lastakse need ülesanded õpilaste esindusgrupil lahendada. Nii saab eelkatse tulemuste põhjal arvutada ülesannete raskust ja diagnoosivat väärtust. Testi valitakse ülesanded, mille diagnoosiv väärtus on kõrge.

Testiülesande diagnoosiv väärtus näitab, kui hästi eristab see ülesanne mõõdetava oskuse tasemeid. Kui mõned madalama oskuse tasemega õpilased lahendavad ülesande ja kõrgema oskuse tasemega ei lahenda, siis see ülesanne antud oskust hästi ei mõõda. Kui kõik kõrgema oskuse tasemega õpilased lahendavad ülesande ja madalama oskuse tasemega ei lahenda, siis see ülesanne mõõdab antud oskust hästi. Kui kõik testiülesanded mõõdavad sama oskust, siis on test homogeenne ja me teame, et see test mõõdab antud oskust. Selle testi ülesannete omavahelised korrelatsioonid on kõrged.

Eelnevast tulenebki põhimõtteline võimalus testiülesande diagnoosiva väärtuse arvutamiseks. Testiülesande diagnoosivat väärtust

näitab testiülesande korrelatsioon mõõdetava oskusega. Siin tõuseb aga kaks probleemi. 1. Mõõdetava oskuse tasemed katseisikutel ei tarvitse olla teada ja siis ei saagi korrelatsiooni arvutada. 2. Korrelatsioonikordaja suurus sõltub mitte ainult ülesande lahendamise ja oskuse taseme vahelise seose tugevusest, vaid ka ülesande raskusest ja seetõttu ei sobi hästi tavaline lineaarkorrelatsioon või point-biserial korrelatsioon.

Mõõdetava oskuse tasemete mitteteadmist katseisikutel ei peeta diagnoosiva väärtuse arutamisel suureks probleemiks: selleks tasemeks võetakse kõigi katsetamiseks koostatud ülesannete lahendamise üldtulemus. Eelduseks on asjaolu, et iga koostatud ülesanne peaks soovitud oskust mõõtma ja üheskoos mõõdavad nad seda oskust paremini kui iga üksikülesanne. Seega arvutatakse ülesande diagnoosiv väärtus selle ülesande ja testi üldtulemuse vahelise korrelatsioonina. Muidugi, kui mõõdetava oskuse tasemeid on katseisikutel võimalik määrata mingil muul viisil täpsemalt, siis kasutatakse seda moodust.

Korrelatsioonikordaja suurus sõltub sellest, kas korreleeritavate suuruste aritmeetilised keskmised asuvad skaala keskosas või mitte. Keskmise raskusega ülesande korrelatsioon testi üldtulemusega on kõrgem kui kerge või raske ülesande korrelatsioon. Seetõttu võib oskust täiesti õigesti diagnoosiva kuid äärmusliku raskusega ülesande korrelatsioon osutada madalamaks kui osaliselt valesti diagnoosiva, kuid keskmise raskusega ülesande korrelatsioon. Korrelatsioonikordajast lähtudes tuleks eelistada keskmise raskusega ülesannet, kuid tegelikult on antud juhul parem äärmusliku raskusega ülesanne.

Toodud põhjusel **on otstarbekas ülesande diagnoosivat väärtust iseloomustada spetsiaalse näitajaga**. Selle näitaja suurus ei peaks sõltuma ülesande raskusest, sest raskuse jaoks on omaette näitaja ja eristustesti on tarvis väga erineva raskusega ülesandeid.

Testiülesande diagnoosiva väärtuse näitajaks on otstarbekas võtta see, kui palju teeb see ülesanne õigeid õpilaste eristamisi rohkem kui vääraid eristamisi, arvutatuna osades eristamiste üldarvust. Esitatud idee mõistmiseks vaatame tabelit 4. Selles tabelis on õpilased ruudus A eristatud õpilastest ruudus B ülesande lahendamise järgi, kuid pole eristatud oskuse taseme järgi — kõigil on madal tase. Me ei saa öelda, kas ülesanne eristas neid õpilasi õigesti või valesti, ja seetõttu ei saa me ka nende ruutude võrdlust kasutada selle ülesande diagnoosiva väärtuse arutamisel. Õpilased ruudus A on madalama oskuse tasemega kui

õpilased ruudus C, kuid ka seda võrdlust ei saa me kasutada ülesande diagnoosiva väärtuse arvutamisel, sest need õpilased ei erine ülesande lahendamise järgi. Õpilased ruutudes A ja D erinevad nii oskuse taseme kui ka ülesande lahendamise edukuse järgi: tugevamad õpilased lahendasid ülesande õigesti ja nõrgemad ei lahendanud. Need on ülesande poolt tehtavad õiged eristamised. Iga õpilane ruudus A on õigesti eristatud igast õpilasest ruudus D. Kokku on selliseid õigeid eristamisi $A \times D$. Õpilased ruudus B on eristatud õpilastest ruudus C samuti nii oskuse taseme kui ülesande lahendamise edukuse järgi. Seekord aga on madalama oskuse tasemega õpilased ülesande lahendanud ja kõrgema oskuse tasemega õpilased ei ole seda lahendanud. Ilmselt on lahendus sõltunud millestki muust kui vaadeldavast oskusest. See ülesanne eristab vääralt iga õpilast ruudus B igast õpilasest ruudus C. Kokku on taolisi vääri eristamisi $B \times C$.

Tabel 4

Nelja välja tabel õpilaste jaotuseks
 ülesande lahendamise edukuse ja oskuse taseme järgi

Oskuse tase	Ülesande lahendus	
	Väär	Õige
Madal	A	B
Kõrge	C	D

Toetudes eeltoodud arutlustele näeme, et tabeli 4 andmetel teeb vaadeldav ülesanne $A \times D$ õiget eristamist ja $B \times C$ väärat eristamist. Õigete eristamiste arv ületab väärade eristamiste arvu $A \times D - B \times C$ võrra ja eristamisi on kokku $A \times D + B \times C$. Seega tabeli 4 tähistel puhul on ülesande diagnoosiv väärtus

$$d_v = \frac{AD - BC}{AD + BC} \quad (5)$$

Vastust ülesandele võib hinnata detailsemalt kui kahepalli-süsteemis ja ka oskuse tasemel on tavaliselt mitu eristatavat astet. Siis muutub tabel 4 suuremaks, kuid diagnoosiva väärtuse arvutuskeem

jääb samaks. Õigete eristamiste arvu saamiseks tuleb tabeli iga arv korrutada temast allpool ja paremal asuvate arvude summaga ja korrutised liita. Väärade eristamiste arvu saamiseks tuleb tabeli iga arv korrutada temast allpool ja vasakul asuvate arvude summaga ja korrutised liita. Õigete ja väärade eristamiste arvud tuleb panna valemisse (5) vastavalt $A \times D$ ja $B \times C$ asemele. Toodud reeglid kehtivad, kui tabeli peas ülesande lahendamise õigsuse näitajad suurenevad vasakult paremale ja tabeli vasakul serval oskuse tase tõuseb, liikudes ülalt alla.

Olgu näiteks mingi ülesande diagnoosiva väärtuse arvutamiseks saadud järgnev tabel 5.

Tabel 5

Erinevaid lahendustulemusi saanud õpilaste arvud erinevates oskusgruppides

Oskuse tase	Ülesande eest punkte		
	1	2	3
Madal	10	5	1
Keskmine	3	12	5
Kõrge	0	2	20

Selle tabeli andmetel on õigeid eristamisi

$$10 \times (12 + 5 + 2 + 20) + 5 \times (5 + 20) + 3 \times (2 + 20) + 12 \times 20 = 821.$$

Vääri eristamisi on

$$1 \times (3 + 12 + 2) + 5 \times 3 + 5 \times 2 = 42.$$

Ülesande diagnoosiv väärtus

$$d_v = \frac{821 - 42}{821 + 42} = 0,90$$

Toodud näites on ülesanne täiesti hea diagnoosiva väärtusega — ainult 5% eristamistest on väärad. Ilmselt võib testi võtta ka veidi väiksema diagnoosiva väärtusega ülesandeid. Kui diagnoosiv väärtus on null, siis teeb ülesanne sama palju vääri kui õigeid eristamisi ja see ülesanne testis ainult segab. Põhimõtteliselt võib diagnoosiv väärtus ka

negatiivne olla, kuid seda praktikas loodetavasti ei juhtu, sest see näitaks, et eksperdid on esialgsesse ülesannete kogumisse võtnud üsna sobimatuid ülesandeid.

Selle osa lõpetuseks pöördume tagasi algusesse: mitte iga eristustesti koostamisel pole tarvis ülesannete diagnoosivat väärtust arvutada ega arvestada. Riigieksamitööd haaravad ulatuslikku materjali, mille puhul me ei saa eeldada, et ühe osa omandamine on tingimata teise osa omandamise eelduseks. Riigieksamitöodes ja õpetaja koostatud eristustestides pole tavaliselt ülesannete diagnoosiv väärtus oluline, nende ülesannete valiidsus tagatakse eksperthinnangutega. Seevastu mõnda üksikut oskust, näiteks liitmisoskust, mõõtvates testides on ülesannete diagnoosiv väärtus oluline. Diagnostilisele väärtusele pööravad psühholoogid suurt tähelepanu võimete mõõtmisel.

Eristustesti tulemuste standardskaalad

Õpilase tulemust programmitestis hinnatakse selle järgi, mitu protsenti ülesannetest ta lahendas. See protsent näitab, kui hästi õppematerjal on omandatud ja nii saab õpilasele panna hinnet. Sellise praktika eelduseks on asjaolu, et programmitesti ülesanded esindavad õppematerjali. Eristustest ei vasta aga kuigi hästi ainekavale ja seetõttu ei saa tulemusi ka tõlgendada ainekavast lähtuvalt. Õpilase tulemust eristustestis tõlgendatakse selle järgi, kui palju punkte ta sai võrreldes teiste õpilastega.

Testi skoorimise tulemusena saadakse testi **toorpunktide arv** iga õpilase jaoks. Eristustesti toorpunktid on aga suhteliselt väheinformatiivsed. Nende põhjal saab küll öelda, kumb õpilane sai parema tulemuse, kuid kui palju parema, seda ei saa enam öelda. Asi on selles, et testis võib olla mõõdetava suuruse mõne vahemiku kohta palju ülesandeid ja teise vahemiku kohta vähe. Joonisel 2 on üht sellist situatsiooni illustreeritud: vahemikus FG on vaid üks ülesanne (tärn), sama pikas vahemikus GH on kuus ülesannet. Olgu kolm õpilast vastavalt tasemetel F, G ja H. Nende õpilaste oskuste tasemete vahed FG ja GH on võrdsed, kuid õpilane tasemel G lahendab vaid ühe ülesande rohkem kui õpilane tasemel F ja õpilane tasemel H lahendab 6 ülesannet rohkem kui õpilane tasemel G. Seetõttu ei näita toorpunktide erinevused õpilaste tõelisi erinevusi.

A B C D E F G H I

Joonis 2. Ülesannete võimalik jaotus testis. Iga tärn tähistab üht ülesannet, mõõdetav suurus kasvab vasakult paremale ja tähed näitavad võrdseid vahemikke selles kasvus.

Toorpunktid on täiesti kasutatud, kui me soovime võrrelda ühe õpilase kahe erineva aine teste omavahel. See, et õpilane sai matemaatika testis 70 punkti ja emakeele testis 60 punkti ei tähenda veel, et ta on matemaatikas tugevam. Võib-olla oli matemaatikatesti keskmine tulemus 55 punkti ja emakeeles 40 punkti. Siis ületab selle õpilase tulemus keskmist matemaatikas 15 punkti võrra ja emakeeles 20 punkti võrra. Tundub, justkui oleks see õpilane tugevam emakeeles, aga seegi järeldus on ennatlik — tuleb veel arvestada tulemuste standardhälbeid. Kui emakeeles on õpilaste tulemused väga erinevad ja matemaatikas ühtlaselt keskmise lähedal, siis võib see õpilane ikkagi suhteliselt tugevam olla matemaatikas.

Eelnevast on selge, et arvutused testi toorpunktidega on ilus, kuid kasutu mäng, sest toorpunktid moodustavad vaid järjestikaskaala. Toorpunktidelt tuleb üle minna standardpunktidele, mis võimaldavad tulemusi võrrelda. Enne kui me seda üleminekut kirjeldame aga veel teinegi põhjus, miks standardpunktid on kasulikud.

Eri testidel on maksimaalne võimalik tulemus erinev. Seetõttu ei ütle õpilase toorpunktide arv inimesele midagi, kui ta ei tea, milline oli õpilaste keskmine tulemus ja kui suur oli tulemuste varieeruvus. Selle asemel, et lisaks õpilase tulemusele teatada veel keskmine tulemus ja standardhälve, võib nende andmete põhjal väljendada õpilase tulemuse standardskaalal, mida kõik asjatundjad tunnevad ja ühtmoodi mõistavad. Testitulemuste standardskaalad on spetsialistidele tuntud nagu meetriga mõõtmise tavainimesele.

Konkreetselt testi tulemuste üleviimiseks mingile standardskaalale tuleb see test läbi viia esindusliku õpilasgrupiga.

Esinduslikke õpilasgruppe võib aga koostada erinevate õpilaste jaoks, näiteks 11 aastastele õpilastele, 12 aastastele õpilastele jne. Vas-tavalt sellele saame ka erinevad üleminekutabelid samale testile.

Näiteks võivad siin olla Raveni testi üleminekutabelid standardskaalale (e. normid) testi täitmisel grupis või individuaalselt ja sama testi normid 20-aastastele, 25-aastastele jne (lisa 2).

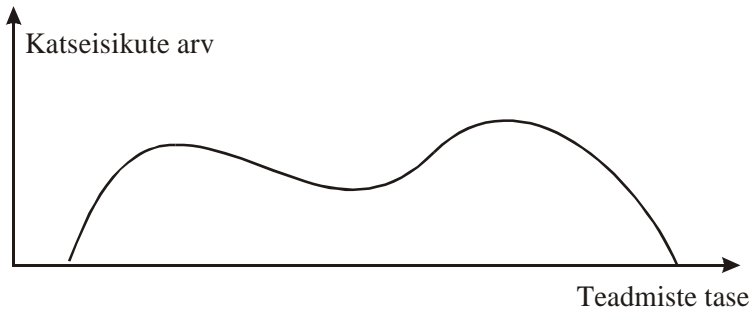
1. Põhimõtteliselt esimeseks testitulemuste standardskaalaks on **protsentskaala**. Õpilase tulemus protsentskaalal näitab, mitu protsenti õpilastest saab temast nõrgema tulemuse. Kui näiteks õpilase tulemus on 55 punkti ja selle testi protsentskaalal vastab 55-le punktile 95%, siis tähendab see, et 95% selle õpilase eakaaslastest saab temast selles testis nõrgema tulemuse. Seega on vaadeldava õpilase testitulemus parem kui enamusel tema eakaaslastest (vt lisa 2).

Ülaltoodud näites võrdlesime testitulemust kõigi eakaaslaste omaga, aga me võime seda veel võrrelda vastavate eriklasside õpilaste tulemustega ja leida näiteks, et seal 55 punkti vastab normile 60, mis tähendab, et selle õpilase tulemus on parem 60% eriklasside õpilaste tulemusest. Me võime lugeda õpilast kuuluvaks mitmesugustesse gruppidesse ja võrrelda tema tulemust nende gruppide tulemustega, saades niimoodi ühe testitulemuse jaoks erinevad näitajad erinevatest tabelitest. Ühel testil võib olla mitu standardskaalat sõltuvalt sellest, millisel katseisikute grupil need skaalad on välja töötatud.

Protsentnorm 50 langeb kokku keskmise tulemusega. Protsentnormides väljendatud testitulemustega pole aritmeetikatehted õigustatud. Seetõttu kasutatakse protsentsnorme harva kuigi neid on lihtne leida ja tõlgendada.

Järgnevale standardskaaladele üleminek põhineb eeldusel, et mõõdetav omadus jaotub esindusgrupis normaalselt. Kui testi toorpunktide jaotus ei ole normaaljaotus, siis tuleb toorpunktidele anda kaalud nii, et saadavad standardpunktid moodustaksid normaaljaotuse esindusgrupis.

Oletame näiteks, et testi toorpunktid annavad esindusgrupis "kahe küüruga kaameli" (joon. 3). Sellisel juhtumil tuleks osa keskmise raskusega ülesandeid paari kaupa kokku võtta ja paari eest anda üks punkt mõõdetava omaduse juurdekasvu. Piltlikult surume selle jaotuse keskelt kokku, nii et pindala jääks samaks. Loomulikult tuleb neid arvutusi teha täpsete matemaatiliste tabelite abil.



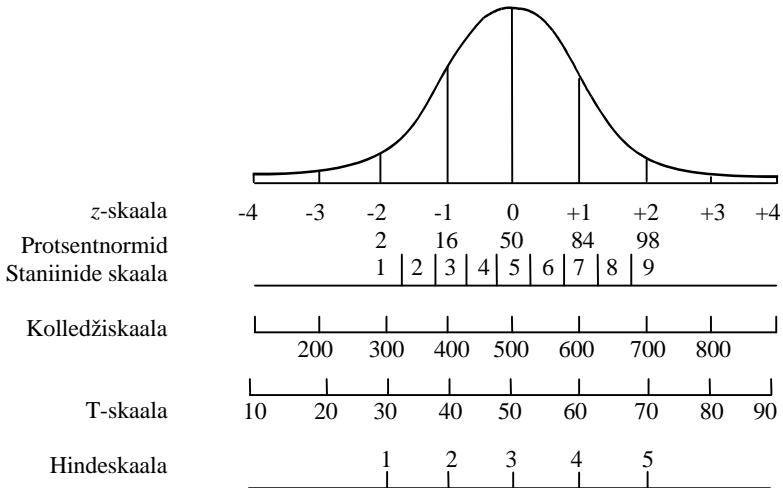
Joon. 3. Testi toorpunktide mittenormaaljaotus.

Joonis 1 ja 4 näitavad, kuidas testi toorpunktidelt minnakse üle standardpunktidele. Kõigepealt tuleb leida õpilaste protsent, kellel on vaadeldavast tulemusest nõrgem tulemus. Sellele protsendile tuleb liita vaadeldava tulemuse saanud õpilaste protsendist pool. Saadud protsendilt tuleb normaaljotuse tabeli järgi (lisa 1) üle minna tulemusele z -skaalas. Normaaljaotuse tabel annab seose punkti asukoha x -teljel ja sellest punktist vasakule jääva joonealuse pindala vahel. Pindala kujutab õpilaste protsenti, kes ülesannet ei lahendanud, ja punkti koordinaat vastab omaduse tasemele, millest allpool ülesannet ei lahendata. Saadavad standardpunktid x -teljel moodustavad intervalliskaala ja nendega on õigustatud liitmine ning lahutamine. Need tehted on seega õigustatud ka järgnevates standardskaalades.

2. Ülalnimetatud **z -skaala** ongi testi standardskaalade oluliseks liigiks, mis on iseseisvalt kasutusel ja ka paljude teiste skaalade aluseks. z -skaalal on keskmine tulemus null ja ühikuks standardhälve. Sisuliselt väljendatakse õpilase tulemust selles skaalas standardhälbe ühikuis. **Negatiivne tulemus z -skaalal tähendab, et õpilase tulemus on alla keskmise. Tulemus 0 ei tähenda siin sugugi nullilisi teadmisi, vaid hoopis keskmisi teadmisi vaadeldavas grupis. Tulemus 3 on z -skaalal väga heade teadmiste näitaja.**

Tulemused z -skaalal on täpsed ja sisuliselt lihtsad, kuid nad pole mugavad kasutada, sest pool tulemustest on negatiivsed ja enamus tulemusi murrud. Nendel põhjustel minnakse z -skaalalt üle teistele skaaladele korrutades z -skaala standardpunkte mingi konstandiga

(vältimaks murde) ja liites teise konstandi (vältimaks negatiivseid arve). Nii saadakse veel hulk erinevaid standardskaalasid.



Joon. 4. Üleminekul testitulemuste skaalade vahel (Lindgren, Suter, 1995, lk 379) (vt ka lisa 1).

3. **Staniinide skaala** on skaala, milles tulemuste keskväärtus esindusgrupis võrdub 5-ga ja standardhälve 2-ga. Tulemuse üleviimiseks z-skaalalt staniinide skaalale tuleb see korrutada 2-ga ja liita 5.

$$C = 2z + 5 \quad (6)$$

Tulemus 7 staniinide skaalal tähendab näiteks, et õpilane on ühe standardhälbe võrra esindusgrupi keskmisest parem ehk, teisiti öeldult, on tal tulemus parem kui 84%-il grupikaaslastest.

4. **Kolledži sisseastumistestide skaalal** on keskmine tulemus 500 ja standardhälve 100. Seos z-skaalaga on järgmine:

$$k = 100z + 500. \quad (7)$$

Seega teade, et USA keskkoolilõpetajate keskmine testitulemus emakeeles oli 444 punkti, ei tähenda, et keskkoolilõpetajad oleksid lahendanud keskmiselt 444 testiülesannet õigesti. See tähendab, et kollektiivne skaalale üleviidult said nad keskmiselt 444 punkti. Ülesandeid võis tegelikult olla 100 või 200 või mingi muu arv. USA keskmine tulemus 444 punkti kollektiivse skaalal on aga madal võrreldes skaala keskmisega 500 punkti.

5. **T-skaala**, mille keskmine on 50 ja standardhälve 10 punkti. Seos z-skaalaga on valemiga järgmine:

$$T = 10z + 50 \quad (8)$$

Tulemuste muutuvus T-skaalal on vahemikus 20–80 punkti, mis meenutab protsentnorme. Tegelikult on T-skaala intervalliskaala, aga protsentnormid mitte. Nende seos on näha jooniselt 4 ja tabelist lisas 1. Näiteks tulemusele 40 T-skaalal vastab protsentnorm 16. Kokkulangevus nendel normidel on ainult keskväärtuse osas — mõlemal 50.

6. **Hindeskaala**, mille keskväärtuseks on 3 ja standardhälveks 1. Sellise hineskaala pakkus Saksamaa koolisüsteemi jaoks välja Lienert. Meil tuleks üleminekuvaalem hineskaalale kirjutada järgmiselt

$$H = z + 3 \quad (9)$$

Seega tuleks koolihinde saamiseks z-skaala standardpunkte lihtsalt suurendada kolme võrra.

Piirdume nende näidetega testi standardpunktide skaaladest. Teooria illestamiseks on lisas 2 ja 3 toodud Raveni testi ja ühe SAT normid. Oluline on nende üleminekutabelite juures see, et iga skaala puhul on täpselt teada keskpunkt ning standardhälve, mistõttu on võimalik iga õpilase tulemust üldarusaadavalt tõlgendada.

Eristustesti tulemuste seostamine koolihinnetega

Eristustestide tulemused standardpunktides on asjatundjatele arusaadavad ja hästi tõlgendatavad. Mõnikord tekib siiski soov tulemuselt standardskaalas üle minna koolihinnetele. Siin ei saa aga enam kasutada Haridusministri poolt kinnitatud hindamise juhendit (vt tabel 1), kus õpilase tulemuse hindamise aluseks on tema poolt omandatud materjali protsent. Standardpunktide puhul on õpilase tulemuse hindamise aluseks õpilaste protsent, kelle tulemus jääb vaadeldava õpilase tulemusest allapoole. Õpilase tulemust hinnatakse võrreldes seda teiste õpilaste tulemustega. Seega on eristustesti tulemuste hindamine täiesti erinev programmitesti tulemuste hindamisest.

Eristustesti puhul on meil selgelt teada vaid õpilaste pingerida aine valdamise järgi. Me ei tea, millise osa õppematerjalist üks või teine õpilane selles reas on omandanud. Pingerea puhul ei jää üle muud, kui kokku leppida, milline protsent õpilastest saab puuduliku hinde, milline protsent õpilasi saab rahuldava hinde jne. Standardpunktide puhul ei saa mingil juhul panna puudulikku õpilasele, kelle tulemus jääb 50%-st allapoole, sest on ebaloomulik oletada, et pooled õpilased pole ainet omandanud. Eristustestide puhul peab olema teistsugune vastavus hinnete ja punktide vahel kui programmitestide puhul.

Eristustestide puhul kasutatavad hindenormid on tabelis 6. Need langevad kokku Lienerti poolt soovitatutega ja on esitatud ka meie asjatundjate poolt (Rapoport, Selg, Sotter, 1987, lk 132).

Tabel 6

Eristustesti tulemuste ja koolihinnete seos

Hinne USA-s	z-skaala	Stanii- nid	kolleđ i skaala	T-skaala	Hinne (Lienert)	Õpilaste %
A	> 1,5	>8	>650	>65	5	7
B	0,5 ...1,5	6...8	550...650	55...65	4	24
C	-0,5...0,5	4...6	450...550	45...55	3	38
D	-1,5...-0,5	2...4	350...450	35...45	2	24
E	<-1,5	<2	<350	<35	1	7

Tabelist 6 näeme, et hinded Lienerti järgi ei lange kokku meie koolihinnetega. Lienerti järgi saaks $24\% + 7\% = 31\%$ õpilastest puuduliku hinde, mida me ei saa aga lubada. Järelikult tuleb hinne "2" ehk "D" selles skaalas lugeda positiivseks hindeks. Siin ongi selgitus olulisele erinevusele meie hindedkaala (tabel 1) ja välismaa hindedkaala (tabel 6) vahel. Tabelis 6 on põhihinnete B, C ja D vahemikud kõik võrdsed ja järelikult on liitmine-lahutamine nendega õigustatud. Samas on siin tähtedena väljendatud hinnitel teine tähendus kui Tartu Ülikoolis kasutatavatel tähelistel hinnitel (tabel 2).

Testitulemuste tõlgendamine koolihinnete kaudu pakub testide kasutamise algperioodil suurt huvi. Tegelikult on hinded hoopis ebatäpsem õpitulemuste edastamisviis kui testitulemused. Seetõttu hakatakse edaspidi ilmselt rohkem kasutama testitulemusi, näiteks kolledži sisseastumisskaalas.

Testitulemustest saab kasutada mitte ainult summaarset resultaati vaid ka iga ülesande ja küsimuse vastamise tulemusi. Need näitavad, millised materjali osad on paremini ja millised halvemini omandatud. Õpetajale pakuvad huvi tema iga õpilase tulemused, et teada, kuidas just selle õpilase tööd edaspidi suunata.

Testi kvaliteedi näitajad

Testi kvaliteet sõltub otseselt sellest, kuidas seda testi on koostatud. Testi koostamise kirjeldus ja kasutamise eeskiri tuleb kirja panna. Seda kirjeldust nimetatakse **testi juhendiks** (ehk **testi manuaaliks**). Testi juhend annab ülevaate testi kvaliteedist. Juhendi puudumine näitab sageli kvaliteedi puudumist. Juhendi olemasolu ei tähenda aga veel, et test oleks tingimata kvaliteetne. Juhendist saab vaid informatsiooni testi kvaliteedi kohta.

Testi juhend peaks sisaldama järgmisi osi.

1. Testi eesmärk: millise ainevaldkonna või millise oskuse omandamist kontrollitakse, millise vanusega või millise ettevalmistusega õpilastele on test ettenähtud, kas testi kasutatakse õpilaste hindamiseks või nende järjestamiseks ainetundmise järgi, kuidas testi levitatakse.
2. Testi koostamise lühikirjeldus: kuidas valiti testiülesanded, kui ulatuslikud olid testi katsetused, keda esindas katsetel kasutatud õpilasgrupp ja milliseid tulemusi katsed andsid. Eristustestide puhul esitatakse veel tabelid, mille abil saab toorpunktidele üle minna standardpunktidele.
3. Testi läbiviimise eeskiri, milles peaks olema näidatud
 - a) nõuded õpilaste paigutusele testimise ajal,
 - b) ajalimiit testimisele,
 - c) lubatud abivahendid.

Näiteks matemaatika riigieksamil 2001 oli lubatud kasutada ainult teatud klaviatuuriga kalkulaatorit, ei olnud lubatud kasutada mobiiltelefoni, teatmikke, käsiraamatuid ega teiselt õpilaselt laenata kolmnurka või muid töövahendeid (Uudelepp, 2001).

4. Testitulemuste skoorimise eeskiri — millised õpilaste vastused annavad punkte ning kui palju.
5. Testitulemuste ja koelihinnete vaheline seos, kui testitulemuste järgi soovitakse hindeid panna või fikseerida, kes õpilastest tegi töö vähemalt rahuldavalt.
6. Testi reliaablus ja mõõtmisviga koos andmetega nende määramise meetodi ning õpilaskontingendi kohta.
7. Testi valiidsus koos andmetega selle määramisviisi kohta.
8. Testi koostamise aeg ja testi koostajate nimed ning kontaktaadress.

Eelnevas oleme rääkinud sellest, millised nõuded on testi eesmärgi fikseerimisele, testi koostamisele ja tulemuste skoorimisele ning tõlgendamisele. Nende nõuete täidetuse peaks selguma testi juhendist ja selle järgi saame otsustada testi kvaliteedi üle. Me ei ole veel rääkinud testi reliaablusest ja testi valiidsusest. Mõlemad nendest iseloomustavad testiga mõõtmise täpsust ja on seetõttu testitulemuste tõlgendamisel väga olulised.

Testi juhend on mõeldud selleks, et testi läbiviijad saaksid testi valida, selle tulemusi skoorida ja tõlgendada. E. Krull (2000, lk 575) refereerib aga õpilaste õiguste hartat, millest järgneb, et ka õpilasel on õigus teada, millise testi abil tema teadmisi ja oskusi mõõdetakse. Õpilased peaksid teadma, millise materjali omandamist kontrollitakse ja neile tuleks anda piisavalt aega, mis tähendab, et eelistada tuleks programmiteste. Õpilastel peaks olema õigus vaidlustada teadmiste kontrolliviise ja saadud hindeid, milleks oleks neil otstarbekas tunda testi juhendit.

Sissejuhatuses järgmisesse osasse võiks reliaabluse ja valiidsuse olemust iseloomustada järgmise kujundliku võrdlusega. Kui jäigalt kinnitatud püssist lastud kuulid tabavad kõik märki üksteise lähedal, siis on püss reliaabel. Kui tabamused on üksteisest kaugel, pole püss reliaabel. Kui iga lasuga tabatakse kümnet, siis on laskmine mitte ainult reliaabel vaid ka valiidne.

Testi reliaablus

Ükski mõõtmine pole täpne. Kui ma võtan mõõdulindi ja mõõdan oma kirjutuslaua laiust, siis võin kord saada tulemuseks 1001 mm, teine kord 999 mm, kolmas kord 1000 mm jne. Toodud näites on tegemist lihtsa mõõtmisega, aga inimese omaduste, sealhulgas teadmiste taseme mõõtmine on tunduvalt keerulisem. Võib ju õpilane praegu ülesande lahendada, kuid paari tunni pärast ei saa ta analoogilise ülesandega hakkama jne.

Õigem oleks siiski öelda: iga mõõtmine on täpne teatud piirini. Ülaltoodud näites on kõigil kolmel korral laua laiuseks saadud 100 cm, seega on mõõdetud sentimeetri täpsusega. Ka ainetestide puhul on tähtis teada, kui täpselt nad mõõdavad. Kas tegelik teadmiste tase võib erineda testitulemusest 1 punkti, 10 punkti või 100 punkti võrra?

Testi reliaablus näitabki, kui hästi langevad kokku kordustestimiste tulemused. Kui sama testi teist korda tehes saavad õpilased (peaaegu) sama tulemuse, siis on test reliaabel. Kui kordustesti tulemused on õpilastel oluliselt erinevad, siis pole test reliaabel. Testi reliaablus on testi truudus iseendale.

Eeltoodud testi reliaabluse määratlus on lihtne, kuid mitte päris täpne. Täpsem reliaabluse definitsioon toetub tõsiasjale, et testitulemuste erinevused on tingitud kahest tegurist: õpilaste erinevustest teadmistes ja mõõtmisvigadest. Kui mõõtmisvead on suured võrreldes õpilaste teadmiste erinevustega, siis see test ei sobi mõõtmiseks, sest selle reliaablus on madal. Mida väiksemad on mõõtmisvead võrreldes erinevustega mõõdetavas suurusel, seda kõrgem on testi reliaablus.

Testi reliaablus näitab, milline osa testitulemuste dispersioonist õpilasgrupis on tingitud mõõdetavast suurusel. Kui mõõdetava suuruse erinevused õpilasgrupis tunduvalt ületavad mõõtmisvigu, siis saadakse kordustestimisel praktiliselt samad tulemused, sest mõõdetava suurus lühikese ajaga ilmselt oluliselt ei muutu ja mõõtmisvead olid meie oletuse kohaselt väikesed. Seega langevad reliaabluse kaks toodud määratlust sisuliselt suuresti kokku.

Testi reliaabluse taset väljendatakse tavaliselt korrelatsioonikordajana⁶. Selle korrelatsioonikordaja leidmiseks on mitu moodust.

1. Ülaltoodud lihtsast definitsioonist tuleneb üks moodus testi reliaabluse määramiseks. Test viiakse läbi samal õpilasgrupil kaks korda ja arvutatakse tulemuste vaheline korrelatsioon. **Kordustesti** soovitatatakse korraldada järgmiselt: esimene kord nädala algul ja teine kord järgmise nädala lõpul, üks kord hommikupoolikul ja teine kord õhtupoolikul. Sellisel juhtumil on lühiajalised muutused õpilastes suurimad ja me ei hinda reliaablust üle.

See põhimõtteliselt lihtne moodus reliaabluse määramiseks pole praktikas siiski kuigi palju kasutusel, sest kordustesti tegemine on tülikas ja vahepeal võivad õpilased vastused küsimustele endale selgeks

⁶ Mõnikord iseloomustatakse testi reliaablust ka protsentarvuna. See reliaabluse näitav on kordustestimisel sama tulemuse saanud õpilaste protsent katseisikute seas (Genesee & Upshur, 1996, lk 247). Loomulikult tuleb siin iga kord täpsustada, mida mõistetakse sama tulemuse all.

teha kaaslaste või õpiku abil. Seega pole see reliaabluse näitaja ka päris täpne.

2. Tihti tehakse ainetest kahes või enamas variandis. Katsetuste käigus on otstarbekas siis lasta samadel õpilastel läbi teha mõlemad või kõik variandid. **Testi eri variantide vaheline korrelatsioon** on selle testi (mõlema variandi) reliaabluse üks näitaja. See on hea näitaja, sest ta arvestab enamikku testi reliaablust vähendavaid tegureid: testiülesannete ebaõnnestunud koostamine, eksitused testitulemuste hindamisel, õpilaste töövõime muutused jne.

See reliaabluse näitaja läheneb suurel määral juba valiidsusele, sest ta arvestab ka ülesannete ebaõnnestunud valikust tingitud erinevusi mõõtmistulemustes.

3. Reliaabluse leidmiseks kasutatakse sagedamini meetodeid, mis võimaldavad korrelatsioonikordaja arvutada ainult antud testi tulemuste põhjal ühel katsel. Lihtsam nendest meetoditest seisneb selles, et arvutustel jagatakse test pooleks: paaritu arvulise järjenumbriga ülesanded ühte poolde ja paarisarvulise järjenumbriga teise poolde. Seega saadakse arvutustes ühe testi asemel kaks pooltesti ja leitakse nende pooltestide tulemuste vaheline korrelatsioon r_1 . Terve test on pikem ja mida pikem on test, seda kõrgem on ta reliaablus. Terve testi reliaabluse r saab pooltestide vahelise korrelatsiooni põhjal arvutada **Spearman-Browni valemiga**

$$r = \frac{2r_1}{1 + r_1} \quad (10)$$

4. Tänapäeval on reliaabluse määramise valemitest kõige populaarsem **Cronbachi α** . Selles valemis vaadeldakse iga testiülesannet eraldi.

$$r = \frac{n}{n-1} \left(1 - \frac{\sum \ddot{a}_i^2}{\ddot{a}^2}\right) \quad (11)$$

kus n — ülesannete arv testis,
 \ddot{a}_i^2 — i -nda testiülesande dispersioon,
 \ddot{a}^2 — testi dispersioon.

Selle ja teiste reliaablusvalemite kasutusviisi kohta toob üksikasjalikke näiteid P. Kees (1984, lk 132–151) Asjasthuvitatud võivad kasutada ka teisi allikaid (Toim, 1987).

Kuna erinevad valemid annavad veidi erinevaid tulemusi, siis tuleks koos reliaabluskoefitsiendiga ära näidata ka selle arvutusvalem. Tänapäeval arvutatakse reliaablust elektronarvutitega, mille statistika-pakettides on programmeeritud ka reliaabluse valemid. Testi koostaja ülesandeks jääb siis vaid algandmete nõuetekohane sisestamine arvutisse.

Kui suur peaks olema testi reliaablus, et seda testi pidada piisavalt täpseks? Vastus sõltub sellest, mida tahetakse testiga teha. Kui testiga tahetakse mõõta üksiku õpilase teadmiste või oskuste taset, siis peab ta reliaablus olema väga kõrge: umbes 0,95 (Genesee & Upshur, 1996, lk 248). Sellisel juhtumil on testi viga umbes 5 korda väiksem kui testitulemuste standardhälve. Kui testiga tahetakse võrrelda kaht õpilasgruppi omavahel, siis ei pea testi reliaablus nii suur olema. Võib olla piisab isegi reliaablusest 0,7, sest grupi keskmine tulemus on alati väiksema veaga kui üksikmõõtmine.

Vaatame nüüd, **millest sõltub testi reliaablus**. Sellest näeme ka, mida teha, kui testi reliaablus on madal ja me tahame koostada täpsemat testi.

1. Lühike test ei anna täpset tulemust. Mitme ülesande lahendamise keskmine tulemus on täpsem kui ühe ülesande lahendustulemus. Mida pikem on test, seda reliaablum ta on. Ilmselt seetõttu ongi välismaal ainetestides sageli sada ülesannet. Kui meie test on lühike ja ebareliaabel, siis võime mõelda selle pikendamisele. Kui aga antud testi

ebareliaablusel on ka teised põhjused, siis pole testi pikendamisest olulist abi.

2. Testi reliaablus sõltub oskuse dispersioonist ehk hajuvusest katseisikute grupis. Kui katseisikud oluliselt erinevad oskuse tasemelt, siis on testiga mõõtmise viga suhteliselt väike oskuse dispersiooniga võrreldes ja reliaablus kõrge. Kui katseisikute rühmas on oskuse hajuvus väike, siis tuleb sama testi reliaablus väike.

See on mõneti ootamatu tulemus: objektiivse mõõtmisvahendi — testi — omadus sõltub sellest, millise katseisikute grupi abil selle omaduse väärtus kindlaks tehti. Aga nii see on, sest reliaablus on sisuliselt defineeritud testi mõõtmisvea ja mõõdetava suuruse dispersiooni suhte kaudu. Seetõttu on oluline koos reliaabluse koefitsiendiga näidata ka testitulemuste dispersioon sellel katsel. Teiselt poolt tähendab see, et suure hajuvusega grupis mõõtmisi tehes võime saada soovitud tulemusi ka madalama reliaablusega testi abil, kuid väikese hajuvusega grupis õpilaste eristamiseks peab testi reliaablus olema kõrge.

3. Loomulikult sõltub testi reliaablus testi kvaliteedist. Kui testiülesanded on raskesti mõistetavad või koguni mitmetähenduslikud, siis on õpilaste vastused neile ilmselt üsna juhuslikud.

4. Kui testi täitjat segatakse või saab ta kasutada naabrite abi, siis jällegi moonutub tulemus ja testi reliaablus väheneb. Olulised on ka katseisikute motivatsioon ja väsimus.

5. Kui õpilaste vastuste läbivaatamisel hinnatakse sisuliselt sama vastust kord õigeks kord vääraks, siis ei saa see test olla reliaabel. Seega sõltub testi reliaablus skoorimise juhendist ja selle täitmise täpsusest.

6. Üldiselt on homogeense testi reliaablus kõrgem, sest mitte-homogeenses testis võib üks õpilane teada üht osa ja teine teist osa materjalist. Testi homogeensuse tõstmine pole võimalik programmi-testis, küll aga saab siingi reliaabluse parandamiseks kriitiliselt üle vaadata ülesannete sõnastused ja hoolitseda testi korrektse läbiviimise ning hindamise eest.

Reliaabluskoeffitsient on laialt levinud ja hästi mõistetav, aga ülal me nägime (punkt 2) et tegelikult on reliaablusest olulisem teada testiga mõõtmise vea suurust ehk **testi viga**. Viimane on reliaablusega

otseselt seotud ja tuleneb reliaabluse definitsioonist. Vastavalt sellele definitsioonile

$$r = \frac{\ddot{a}^2}{\ddot{a}_t^2} \quad (12)$$

kus \ddot{a}^2 — oskuse dispersioon grupis
 \ddot{a}_t^2 — testitulemuste dispersioon grupis

Arvestades, et testitulemuste dispersioon on mõõdetava suuruse ja mõõtmisvea dispersioonide summa

$$\ddot{a}_t^2 = \ddot{a}^2 + \ddot{a}_e^2$$

saame

$$\ddot{a}_e = \ddot{a}_t \sqrt{1 - r} \quad (13)$$

Saadud valem on kooskõlas tõsiasjaga, et reliaabluse r suurenedes testi viga väheneb. Kui reliaablus oleks 1 siis oleks testi viga null. Samas me näeme, et testi viga on väiksem testitulemuste standardhälbest. See on ka loomulik, sest testitulemuste standardhälve kajastab peale testi vea ka katseisikute erinevusi.

Eelnevas nägime, et testi reliaablus sõltub mõõdetava suuruse dispersioonist katseisikute grupis. Testi veal sellist sõltuvust ei ole. Selle tõttu on testi viga isegi parem testi täpsuse näitaja kui testi reliaablus. Nii näiteks teatatakse USA Koolijõudluse testi kohta, et selle mõõtmisviga on 30 punkti kollektiivis skaalal ulatusega 200–800 punkti (Taking the SAT I, 2000, lk 5). Arvestades aga testi reliaabluse laialdast levikut tahaks soovitada testi mõõtmistäpsuse iseloomustamiseks kasutada nii reliaabluskoeffitsienti kui ka testi viga.

Testi valiidsus

Igapäevastes mõõtmistes valiidsuse probleemi ei teki. Kui meile poes kompvekke kaalutakse, siis pole mingit kahtlust, et me saa-

me nende hulga grammides teada. Need mõõtmisviisid on ammutuntud ja üldaktsepteeritud. Pedagoogikas juurdub aga mõõtmiste tava aegamööda. Me pole veel jõudnud ühisele arusaamale sellest, kuidas mõõta näiteks õpilase matemaatikalaste oskuste taset rääkimata tema matemaatikaalasest võimekusest. Imselt tuleb selleks anda talle lahendada matemaatikaülesandeid. Mida aga näitab see, kui ta need ülesanded edukalt lahendab: kas seda, et ta on antud teema hästi omandanud, või on tal hästiarenenud oskused ülesandeid lahendada või haarab ta matemaatikaülesannete lahendusi lennult? Kas nende ülesannete eduka lahendamise põhjal võime öelda, et ta lahendab ka teisi matemaatikaülesandeid hästi? Kui ta ülesannetega toime ei tulnud, mis võis siis olla selle põhjuseks: ta ei tundnud termineid, puudusid vajalikud oskused, ei saanud aru ülesannete keerukast esitusest või midagi muud.

Testide puhul tuleb uurida ja teada, mida näitab nende ülesannete edukas lahendamine, see tähendab, tuleb teada nende valiidsust. **Testi valiidsus** näitab, kui hästi test mõõdab seda, mille mõõtmiseks ta on ette nähtud. Definiitsioonist ilmneb, et valiidsus on testi kvaliteedi põhinäitaja. Teised testi karakteristikud on suuresti suunatud sellele, et saavutada kõrget valiidsust.

Testi valiidsus on seotud lähedalt tema reliaablusega, sest mõlemad näitavad mõõtmistäpsust. Erinevus on aga järgmises. Reliaablus näitab, kui hästi test mõõdab seda, mida ta mõõdab. Reliaabluse puhul on ükskõik, mida test mõõdab, peaasi, et mõõdab. Valiidsus näitab, kui hästi test mõõdab seda, mida temaga mõõta tahetakse. Loomulikult on just see tähtis. Test võib olla reliaabel, kuid ta ei tarvitse selle tõttu veel valiidne olla. Mittereliaabel test ei saa ka valiidne olla, sest kui test midagi ei mõõda, siis ei mõõda ta ka soovitud omadust. Valiidne test on ka reliaabel.

Testi valiidsusel on neli liiki, millest kõige olulisem on esimene — sisuline valiidsus. Sisulisest valiidsusest lähtutakse testi koostamisel ja seda hindavad eksperdid. Teised kolm valiidsuse liiki on määratavad siis, kui test on valmis ja nende suurust väljendatakse korrelatsioonikordajana.

1. Testi valiidsuse saavutamise põhiteeks on tema ülesannete valik sellisena, et nende lahendamine näitaks mõõdetava omaduse olemasolu õpilasel. See valik toetub eksperthinnangutele ja nii saavutatakse testi **sisuline valiidsus**. Lähtudes mõõdetava oskuse või tead-

misvaldkonna avaldumisvormidest, otsustavad asjatundjad, millised ülesanded testi algvarianti võtta. Sellest ülesannete valikust sõltub peaaegu kõik, sest edasine töö on vaid algvariandi lihvimine. Muidugi võib oskamatu töötusega hea algvariandi ära rikkuda, kuid kehva algvarianti lihvides lihvitakse vaid vigu.

Programmitesti sisulise valiidsuse saavutamise käik oli toodud eespool: programmitestis on oluline, et testiülesanded esindaksid ainekava. See oli programmitesti koostamise ainus nõue ja see viib valiidsusele. Siiski on siingi probleem: kas kontrollida kogu materjali, näiteks ka abimõisteid, nagu nad õpikus on, või piirduda ainekava eesmärkide saavutatuse kontrolliga nende omandamise kõrgeimal astmel.

Ainealaste eristustestide koostamine on keerulisem, eriti siis, kui tahetakse mõõta oskusi või koguni võimeid. Sellisel juhtumil tuleb vastavast ainekavast või ainevaldkonnast valida ülesandeid, mille lahendamine näitaks selle oskuse olemasolu eri tasanditel. Selles mõttes on ka ainealased eristustestid esinduslikud oma ainevaldkonna suhtes. Teiselt poolt aga ei saa eristustestis järgida ülesannete jaotust näiteks raskuste järgi selles ainevallas, sest selles testis peab olema kergeid ja raskeid ülesandeid sama palju, kui keskmise raskusega ülesandeid. Ainealaste eristustestide, näiteks mõtlemistestide, koostamine on meie jaoks alles uus ja veetlev valdkond.

2. Sisulisele valiidsusele lähedane on **ennustav valiidsus**, mis näitab, kui hästi võimaldab test prognoosida edukust mingis tegevusvaldkonnas. Näiteks võib testi eesmärgiks olla prognoosida edukust kõrgkoolis mingil erialal õppides. Sellisel juhul on mõeldav, et me oma teoreetiliste teadmiste põhjal otsustame, millised õpilase praegused omadused määravad kõrgkooliedukust, ja koostame nende omaduste mõõtmiseks testid nii, nagu sisulise valiidsuse nõuded seda ette näevad. Seejärel mõõdetakse õpilaste omadused ja vajaliku aja möödudes kontrollitakse, mil määral testitulemused korreleeruvad prognoositava omaduse tasemega.

Näiteks kõrgkooliedukuse prognoosimisel võib aluseks võtta riigieksamite tulemused. Õpilaste riigieksamite tulemuste korrelatsioon hilisemate õpingute edukusega on riigieksamite ennustava valiidsuse näitajaks. Selle korrelatsiooni leidmine võtab mitu aastat aega, kuni ilmneb vaadeldavate sisseastujate õppeedukus kõrgkoolis, aga see-eest

on ennustaval valiidsusel ka suur väärtus. Taolisi katseid on teinud meil A. Sukamägi (1994).

Kõrge ennustava valiidsusega testide põhjal võib prognoosida, millisel alal inimene saavutab kiiremini ja kergemini edu, ja nii võib anda inimesele soovitusi olulisteks otsusteks. Ennustava valiidsuse suure praktilise väärtuse tõttu soovivad mõned autorid võtta see peamiseks valiidsuse näitajaks. Enamik asjatundjaid eelistab siiski valiidsuse põhiliigina sisulist valiidsust, sest testi sisu on olulisem kui see, mida ta prognoosib.

3. Koostatud testi **võrdleva valiidsuse** leidmine on suhteliselt lihtne, kui on olemas teine sama omadust mõõtev test. Siis viiakse nii uus kui vana test läbi samade õpilastega ja leitakse nende tulemuste vaheline korrelatsioonikordaja, mis on mõlema testi võrdleva valiidsuse näitajaks teise suhtes. Taolised arvutused on eriti olulised siis, kui tahtakse vana testi asendada uuega, pikemat testi lühemaga või muud taolist.

Võrdlev valiidsus pole sisuliselt päris täpne: kui mõlema testi koostamisel on sisse lipsanud samasugune süstemaatiline viga, siis on võrdleva valiidsuse näitaja tegelikust valiidsusest suurem; kui testide koostamisel on juhtunud erinevad vead, siis on võrdlev valiidsus tegelikust väiksem.

Võrdleva valiidsuse arvutusest on näha, kui kitsas on piir valiidsuse ja reliaabluse vahel. Reliaabluse ühes arvutusviisis tuli test jagada kaheks pooleks ja nende vahel arvutada korrelatsioon. Kui aga need pooled nimetada kumbki omaette testiks, siis oleks see korrelatsioon juba valiidsus. Võimalik, et reliaabluse ja valiidsuse piiri võib tõmmata järgmiselt: kui on tegemist erinevate testidega (eri autorite poolt või eri aegadel koostatud testid), siis on nende vaheline korrelatsioon valiidsuse näitaja; sama testi eri variantide (sama autor, sama aeg, sama kontseptsioon) vaheline korrelatsioon on reliaabluse näitaja.

4. Valiidsuse kõige täpsema hinnangu annab **faktorvaliidsus**. Selle leidmiseks on vajalik mitu sama omadust mõõtvat testi, millega määratakse mõõdetava omaduse tase katseisikutel ja tehakse tulemuste faktoranalüüs. Faktoranalüüsi esimene faktor peaks kirjeldama ülddispersionist suurema osa, sest kõik need testid mõõdavad sama omadust. Esimest faktorit võib vaadelda mõõdetava omaduse kõige

täpsema näitajana. Kõigi kasutatud testide faktorlaadungid on nende valiidsuse näitajaks. Faktorvaliidsus on täpsem valiidsuse näitaja, sest tema aluseks olev esimene faktor on vaba võimalikest eksitustest üksiktestide koostamisel.

Faktoranalüüsil arvutatakse tavaliselt rohkem kui üks faktor. Kui teine faktor kirjeldab ülddispersionist olulise osa ja seda õnnestub tõlgendada, siis mõõdavad need testid veel õpilaste teistki omadust. Testide faktorlaadungid teisele faktorile on nende valiidsuseks teise omaduse mõõtmisel. Siit tuleneb, et ühel testil võib olla kaks valiidsust. Seda järeldust on veel hea illustreerida ennustava valiidsusega. Seal nägime, et test võidakse koostada teadmiste taseme mõõtmiseks ühes ainevallas ja selles ainevallas on testil ka valiidsus. Siis kasutatakse sama testi näiteks kõrgkooli edukuse prognoosimiseks ja sealt saab test omakorda kõrgkooli edukust prognoosiva valiidsuse. Järelikult on testil mitte valiidsus üldse, vaid valiidsus mingi omaduse mõõtmiseks, ja valiidsuskoeffitsiendi kasutamisel tuleb teada, millist omadust silmas peetakse. Ilmselt peavad need omadused olema üsna lähedased, mida ühe testiga mõõta saab, sest ühe testiga ei saa mõõta mitut erinevat omadust.

Eestis kasutataval ainetestidel ja küsimustikel puudub tavaliselt nii reliaabluse kui valiidsuse näitaja. Oleme harjunud lihtsameelselt uskuma seda, milleks testide ja küsimustike autorid soovivad oma mõõtvahendeid kasutada. Tegelikult tuleks aga huvi tunda nii reliaabluse kui valiidsuse määramisviiside ja -tulemuste vastu. Ka võõrkeeltest adapteeritud mõõtmisvahendite puhul tuleks nende kvaliteedi näitajad meil uuesti määrata. Kui ainetesti reliaablus ja valiidsus on nõutaval tasemel, ainult siis saame mõõtmistulemuste põhjal järeldusi teha. See puudutab ka riigieksamitöid.

Ainetestid tänapäeval

Pedagoogilise Testimise Teenistus USA-s

Eestis oli testide kasutamine pikka aega keelatud. Me ei tunne teste ega testimist praegugi veel kuigi hästi. Aga arenenud maailm kasutab teste intensiivselt. Toome selle kohta mõne näite.

Maailma suurim ja tuntuim testimise süsteem on Pedagoogilise Testimise Teenistus USA-s⁷ (<http://www.ets.org>). See eraettevõte on oma eesmärgiks seadnud teste koostada, läbi viia kõikide soovijatega ja tulemusi töödelda ning analüüsida. Pedagoogilise Testimise Teenistuses on umbes 2100 põhikohaga töötajat (What is ETS, 2001) ja tal on lisaks keskusele Princetonis veel umbes 2000 kohalikku testimise keskust eri maades. 1998/99 õppeaastal kasutati 12 testi, mida täitsid kokku enam kui 11 miljonit õpilast 181 riigis. Kõige populaarsem oli nendest testidest SAT I — 2,5 miljonit täitjat, aga ka noorte õpetajate professionaalsete oskuste testi täitis 0,8 miljonit inimest ja TOEFLI testi peaaegu 0,7 miljonit inimest (What is ETS, 2001). Pedagoogilise Testimise Teenistus pakub soovijaile ka võimalust testideks valmistuda — nad avaldavad paberil ja internetis näidisulesannete kogusid, soovitusi testide täitmiseks jne.

Miks kulutatakse testimisele nii palju raha? Sest riik tahab teada, millist efekti annavad kooli suunatud vahendid. Sest kool tahab teada, millisel tasemel on tal õppetöö. Enamus keskkoolilõpetajaid tahab näidata, et nende teadmiste tase vastab paremate ülikoolide sisseastumisnõuetele. Seda kõike on võimalik teada saada ühtse ja objektiivse testimissüsteemi abil.

Pedagoogilise Testimise Teenistuse kõige suuremaks ülesandeks on keskkoolilõpetajate testimine Koolijõudluse Testidega SAT I ja SAT II. See toimub seitsmel päeval aastas, näiteks 2000/01 õppeaastal 14-ndal oktoobril, 4-ndal novembril, 2 detsembril, 27-ndal jaanuaril, 31 märtsil, 5-ndal mail ja 2 juunil (Taking the SAT I, 2000, lk 80). Testidest osavõtuks kirjutavad soovijad sügisel avalduse ja tasuvad testimismaksu. Nad saavad osta testide näidiseid koos soovitustega testide edukaks täitmiseks ja neil on koolis õpetajad-konsultandid, kelle

⁷ College Entrance Examination Board and Educational Testing Service
P.O. Box 6200, Princeton NJ, USA.

poole pöörduda. Täidetavad ainetestid valib keskkoolilõpetaja lähtudes teaduskonnast, millisesse ta soovib astuda. Ettenähtud laupäevadel sõidavad õpilased kohalikku testimise keskusesse ja täidavad seal teste terve päeva. Tavaliselt on hommikupoolikul mõtlemistestid ja õhtupoolikul teadmiste testid. Kõik testid on ajaliselt piiratud ja ülesanded valikvastustega. Vastused töötatakse läbi Princetonis. Tulemustest teatatakse õpilasele endale ja kõrgkoolidesse, mis õpilane märkis oma avalduses võimalike edasiõppimiskohtadena.

Koolijõudluse test SAT I mõõdab verbaalset ja matemaatilist mõtlemisioskust. Selles testis on järgmised ülesanded (Taking the SAT I, 2000).

I. Verbaalse mõtlemise test:

- 1) analoogiaülesanded,
- 2) lause täiendamine (lünktest),
- 3) kriitilise lugemise test.

II. Matemaatilise mõtlemise test:

- 1) sõnalised ülesanded valikvastustega,
- 2) kahes tulbas antud suuruste võrdlemine,
- 3) vabavastuselised ülesanded.

Pedagoogilise Testimise Teenistuses koostatakse ja viiakse läbi veel SAT'i eeltesti PSAT, Arenenud Paigutusprogrammi AP, Haridusliku Progressi Rahvuslikku Testi NAEP, Kraadiõppe Teste GRE jne. Pedagoogilise Testimise Teenistus teeb ka testimise alast teaduslikku uurimistööd ning koostööd teiste riikidega.

Teste koostavad ning testivad teisedki keskused Ameerika Ühendriikides. Meil on võimalik USA testidest teatud ülevaadet saada TÜ Raamatukogus Põhja-Ameerika Õpingute Keskuses. Selleks et jätkata õpinguid Põhja-Ameerika ülikoolides, tuleb sageli eelnevalt täita teste. Tuntuim on nendest inglise keele test — TOEFL'i test.

Mida annab selline testimise süsteem? Pedagoogilise Testimise Teenistus kindlustab õpilaste teadmiste objektiivse ja ühetaolise mõõtmise üle kogu Põhja-Ameerika. Mõõtmise objektiivsus loob tugevad stiimulid õppimiseks ja soodustab õppijate ning õpetajate koostööd testideks valmistumisel. Testid ise koostatakse sellisena, et nad arvestaksid võimalikult paljude osariikide õppekavu ja osariigid omakorda püüavad õppekavad koostada nii, et nende õpilased suudaksid teste hästi täita. Pedagoogilise Testimise Teenistus asendab väga hästi kõrgkooli sisseastumiseksameid. Kuna testitulemused töötatakse läbi

kõik ühtses keskses, siis selgub terve riigi õpilaste teadmiste tase ja selle arengutendentsid.

USA Pedagoogilise Testimise Teenistusega analoogne asutus töötab Umeå Ülikoolis (<http://www.umu.se/edmeas/index.shtml>). Seal on välja töötatud Rootsi Koolijõudluse Test ning muid teste ja tehakse testide alast uurimis- ning õppetööd. Rootsi Koolijõudluse Test koosneb järgmistest osadest: sõnavara, kvantitatiivne mõtlemine, loetu mõistmine, diagrammide, tabelite ja kaartide interpreteerimine, ingliskeelse teksti mõistmine. Tähelepanuväärne on see, et Rootsi Koolijõudluse Testis mindi 1995 aastal eristustesti põhimõttelt üle programmitesti põhimõttele, kuna eristustest sundis õpilasi tervistkahjustavalt pingutama. Rootsi kõrgkoolidesse võetakse umbes pooled üliõpilased vastu Koolijõudluse Testi tulemuste järgi ja pooled koolihinnete järgi (Stage, 1998). Suhteliselt tugev on Umeå Pedagoogilise Mõõtmise Teaduskonnas kehaliste võimete mõõtmine ja ka liiklustestide koostamine.

Euroopas on Rahvusvaheline Õppejõudluse Mõõtmise keskus Madalmaades. Keskuse ülesanne on keerulisem selle tõttu, et Euroopas kasutatavad testid on eri keeltes. Eesti osales ühes selle keskuse poolt läbiviidud ulatuslikus rahvusvahelises uuringus. TÜ dotsendi Olev Musta eestvedamisel tõlgiti eesti keelde lugemisoskuse testid 3-nda ja 8-nda klassi õpilastele, testiti esindusliku õpilasgruppi ja töödeldi tulemused rahvusvaheliste standardite kohaselt. See on meil kõige korrektsemalt tehtud ainetest ja me võime nüüd öelda, kui hästi meie õpilased loevad võrrelduna maailma mitmekümne maa lastega (Pandis, 1996). TÜ matemaatikud on seotud analoogse Kolmanda Matemaatika ja Scienci Rahvusvahelise testiga.

Koolijõudlustesti tulemused pakuvad Ameerika Ühendriikides huvi kogu rahvale. Nende muutusi jälgitakse tähelepanelikult, sest see näitab hariduse taset terves riigis. Testitulemuste tõus on suursündmuseks nii rahvale kui riigijuhtidele. Nii näiteks avalikustas asepresident Al Gore Haridusliku Progressi Rahvuslikku Testi NAEP 1998 aasta tulemused 10. veebruaril 1999. Need tulemused on tabelis 7 (NAEP results ..., 1999). Tabelist näeme, et koolilaste lugemisoskus langes kümnendi esimesel poolel ja tõusis teisel poolel. See tõus pole

arvuliselt küll kuigi suur, kuid statistiliselt siiski usaldatav, nii et asepresident ise sellest esimesena teatas.

Tabel 7

USA kooliõpilaste lugemisoskus 1990-ndail aastail
(NAEP results ..., 1999)

Klass	1992	1994	1998
12	292	287	291
8	260	260	264
4	217	214	217

Kahel järjestikusel testimisel ei saa kasutada samu testi-ülesandeid, ometi võrreldakse USA-s erinevate testimiste tulemusi väga täpselt. **Kuidas on võimalik erinevate aastate testitulemusi viia ühisele skaalale?**

Võrdse raskusega ülesandeid kahe testivariandi jaoks ei suuda koostada vist ükski ekspert. Ülesannete raskus tuleb ikkagi määrata katsetel. Sama õppematerjali kohta koostatud erinevate testivariantide või testide raskuste võrdlemiseks on kolm meetodit (Kolen, 1999).

1. Võrreldavad testid antakse erinevatele, kuid sama üldpopulatsiooni suhtes esinduslikele katsegruppidele. Kuna need grupid on ühesugused, siis on erinevused gruppide tulemustes tingitud testivariantide raskuste erinevustest. Ühe grupi keskmine tulemus on tegelikult võrdne teise grupi keskmise tulemusega, ühe grupi kümne protsendi tugevamate õpilaste madalaim tulemus on võrdne teise grupi kümne protsendi tugevamate õpilaste madalaima tulemusega jne. Selle idee järgi saab ühe testi arvulised tulemused teisendada teise testi tulemusteks.

2. Võrreldavad testid antakse mõlemad täita samale õpilasgrupile. Et vältida testide esitusjärjekorrast tuleneda võivat viga, antakse pooltele õpilastest enne täita vorm A ja teisele poolele enne täita vorm B jne. Edasised arvutused on eelkirjeldatuga samad.

3. Osaliselt ühiste ülesannetega testivariandid antakse erinevatele õpilastele. Nende ühiste ülesannete lahendustulemuste põhjal saab siis võrrelda kahte varianti tervikuna omavahel. Need ühised ülesanded on teiste ülesannete vahele paigutatud.

USA koolijõudluse testi võrreldakse kahe eelmise testiga just kolmanda meetodi abil. Õpilastele esitatavas testis on umbes 20% ülesandeid kahest eelmisest testi versioonist, aga õpilased ei tea, millised ülesanded on samad erinevatel testimistel. Ühiste ülesannete lahendustulemusi ei arvestata õpilase hindamisel, neid on vaja vaid selleks, et koostada üleminekutabeleid ühe aasta testitulemustelt teise aasta testitulemustele.

Näeme, et eri aastate testide tulemusi saab võrrelda vaid eksperimentaalselt. Seejuures testid on küll erineva raskusega, kuid iga järgmise aasta testi tulemused võrdsustatakse esimese aasta testi tulemustega vastavate üleminekutabelite abil. Ilma taolise eksperimentaalse tööta on erinevate testide tulemuste võrdlused reeglina vigased.

Testimine arvuti abil

Pedagoogilise Testimise Teenistus võimaldab õpilastel teste täita arvutis. Arvuti esitab õpilasele ülesande ja õpilane vastab vajutades mõnele klahvile. Loomulikult tuleb Pedagoogilise Testimise Teenistuses testimiseks registreeruda ja testimise keskusesse tulla kindlal päeval. Testi täitja isik tehakse kindlaks kahe dokumendi järgi testimise algul ja ka peale vaheaega. Testimise ruumis pole lubatud süüa-juua ega teiste testitavatega rääkida. Lubatud on kasutada mustandi paberit, kuid seda ei tohi pärast endaga kaasa võtta. Seega, abinõud ebaausate võtete vältimiseks on sama ranged kui paberi-pliiatsi testidel. Kui testimisel saadakse ametlikke edukuse näitajaid, siis ei saa seda testi täita kodus.

Testi täitmine arvutis eeldab, et testitav enne tunneks testimise programmi — millised on juhtnupud, kuidas märkida vastuseid jne. Pedagoogilise Testimise Teenistus on selleks välja töötanud vastava õpiprogrammi, kus selgitatakse ja antakse võimalus harjutada juhtnuppude kasutamist. Õpiprogramm sisaldab ka näidisülesandeid, nii kvantitatiivseid, verbaalseid kui analüütilisi. Lisaks sellele on võimalik Pedagoogilise Testimise Teenistusest saada erinevaid harjutusteste elektroonsel kujul.

Eelnev lühikirjeldus näitab, kuidas testimine arvuti abil paistab välja õpilase poolt. Sellise tulemi saamiseks on aga tarvis ulatuslik eeltöö. Ainetest tuleb koostada ja siis spetsiaalse programmi abil õpilasele esitada. Selline programm tavaliselt võimaldab õpetajal teste

koostada, õpilasel testidele vastata ja vastuseid hinnata ning tulemusi salvestada.

Eestis on väga täiusliku **testimise programmisüsteemi APSTest 2** välja töötanud Tõnis Kelder ja Agu Nigul Rein Prank'i juhendamisel (APSTest 2, 2000; <http://www.ce.ut.ee/apstest2>). See programmisüsteem võimaldab:

- õpetajal koostada küsimusi kasutades seejuures ka pilte,
- kombineerida küsimustepangast erinevaid teste,
- õpilasel testidele vastata,
- arvutil hinnata õpilase vastuste õigsust,
- säilitada õpilase tulemusi tulemustepangas,
- õpilasel harjutada põhilistele küsimustüüpidele vastamist näitetesti abil.

APSTestis on võimalik kasutada järgmisi küsimuste liike:

- ei/jaa küsimus,
- mitmikvaliku küsimus,
- õigete variantide märkimise küsimus,
- vastavusse seadmise küsimus,
- liikide määramise küsimus,
- rühmitamisküsimus,
- järjestamisküsimus,
- lünkade täitmise küsimus,
- lühivastusega küsimus,
- arvulise vastusega küsimus,
- kaardiküsimus.

Vastuste õigsuse hindamiseks tuleb arvutile ette anda, milline on õige vastus. Seejuures on arvuliste vastuste puhul võimalus anda õige vastuse vahemik ja valikvastusega küsimuste puhul näidata, mitu punkti annab iga vastusevariant. Nii on võimalik osaliselt õige vastusevariandi valikul anda õpilasele pool punkti ja väära valiku puhul punkte maha arvata, nii nagu me valikvastusega küsimuste puhul rääkisime.

APSTest on nagu tühi vihik, millele tuleb testid sisse kirjutada, alles siis muutub ta koolis kasutatavaks. Neid teste APSTesti jaoks koostatakse aktiivselt (<http://www.opetaja.ee>; <http://www.tiigrihype.ee>).

Tartu Ülikooli psühholoogiaosakonnas kasutatakse õpi- ja eksamineerimisprogrammi EXAMINATOR 4.5 (1999). Selles programmis on võimalik kasutada ainult välismaa populaarseimat küsimuste liiki — valikvastustega küsimusi. Valikvastuste arv võib varieeruda kahest viieni ja nende juurde võib kirjutada kommentaari, miks vastus on õige või väär. Tänu nendele kommentaaridele võib programmi kasutada ka õppimiseks. Programm fikseerib vastamise aja, igale küsimusele antud vastuse ja arvutab õigete vastuste protsendi, mille järgi saab õpilase teadmistele vastavat hinnet leida.

Testide tegemisel arvutite abil on mitmed eelised. Õpetaja ei tarvitse õpilaste tööde läbivaatamisele aega kulutada, seda teeb arvuti. Arvuti võib ka tüüpilised eksimused üles otsida ja teha koondstatistikat. Arvuti hindab vastuseid kohe ja nii pole õpilasel tarvis kaua oma tulemusi oodata.

Eristustestide puhul on arvutil veel üks oluline eelis — ta võib kohaneda õpilasega ja esitada tema tasemele vastavaid ülesandeid. Algul esitab arvuti keskmise raskusega ülesandeid ja sõltuvalt nendele saadud vastuste õigsusest esitab järgnevalt kas kergemaid või raskemaid ülesandeid. Sellist testimist nimetatakse **adaptiivseks**. Tavalises eristustestis on õpilase taseme määramisel kõige olulisemad just need ülesanded, mida ta suudab lahendada tõsise pingutusega; kergemad ja raskemad ülesanded ei aita tema oskuste taset selgitada. Nii hoitakse kokku aega, ei traumeerita õpilast liiga raskete ülesannetega ja saadakse täpsem diagnoos, kuna õpilasele esitatakse tema tasemel rohkem ülesandeid.

Riigieksamid

Momendil Eestis kasutatavatest ainetestidest on kõige olulisemad riigieksamid. Need viidi EV Haridusministri määrusega sisse 1997 aastal ja on paljude ekspertide arvates muutnud koolielu rohkem kui mistahes teine koolireformi osa viimasel kümnendil. Riigieksamid motiveerivad suurt osa õpilastest väga intensiivselt õppima, aga mõned õpilased on loobunud taotlusest riigieksam hästi sooritada. Riigieksamid toimusid viimastel aastatel 11 aines (tabel 8). Lisaks riigieksamidele toimuvad ka põhikooli lõpueksamid.

Riigieksamite sooritajaid
(<http://www.ekk.edu.ee/statistika/riigieksamid/text.html>)

Aine	Eksami sooritajaid		
	1999	2000	2001
Ajalugu	6449	5866	5225
Biooloogia	5532	5786	4749
Emakeel	14395	14412	12806
Füüsika	1387	1273	1280
Inglise keel	9258	9461	8488
Keemia	3685	3578	3441
Matemaatika	8061	8023	7248
Prantsuse keel	58	57	67
Riigikeel	2171	3574	3682
Saksa keel	1809	1607	1407
Vene keel	408	481	509

Haridusminister on 1996. aasta määruses nr 26. riigieksamite kaks esimest eesmärki sõnastanud järgmiselt:

- 1) kontrollida riikliku õppekavaga määratletud õpitulemuste saavutatust eksamiainetes,
- 2) tagada gümnaasiumilõpetajate eksamihinnete võrreldavus ning gümnaasiumi lõpueksamite ühitamine kõrgkoolide sisseastumiseksamitega

(<http://www.ekk.edu.ee/Yldteave/maarused/text.html>).

Esimene nendest eesmärkides tähendab, et riigieksamid peaksid olema koostatud programmitesti põhimõttel, sest ainult sellega on võimalik kontrollida, millisel tasemel on omandatud õppekavaga määratletud õppematerjal. Teise eesmärgi teine osa on aga realiseeritav ainult siis, kui riigieksamid on eristustestid, sest muidu pole nende abil võimalik leida, kes on väärilised pääsema kõrgkooli riigieelarvelistele kohtadele. Seega on riigieksamil kaks erinevat eesmärki.

Riigieksamite kahest eesmärgist peab üks paratamatult prevaleerima. Meil on ülekaalus teine eesmärk — õppurite valik kõrgkooli. Selle väite kasuks räägib näiteks riigieksamite statistika, mis

väga üldiselt võttes on normaaljaotus keskväärtusega 50% juures. See on tüüpiline eristustesti tulemuste jaotus. Kui riigieksamid oleksid programmitestid, siis tuleks ju rakendada programmitesti hindenorme ja tulemuste eest alla 50% kirjutada puudulik hinne, nii nagu õpetajad omakoostatud tööde põhjal teevad.

Praktiliselt mitterealiseeritav taotlus gümnaasiumi lõpueksamite ühitamisest kõrgkoolide sisseastumiseksamitega on põhjustanud palju kriitikat: positiivse hinde alampiiir 1 punkt sajast soosib ilma õppimata riigieksami sooritamist, erinevate ainete ja erinevate aastate eksamitulemused pole võrreldavad jne (<http://www.ekk.edu.ee/index2.html>). Nende puuduste leevendamiseks tegi Õpetajate Liit hiljuti ettepaneku kehtestada riigieksami positiivse hinde künnisena 20–25 punkti. See madal künnis oleks siiski suureks sammuks edasi võrreldes praeguse künnisega 1 punkt. Õpetajate Liit soovib ka jaotada riigieksam kaheks osaks, millest üks kontrolliks kooli lõpetamiseks vajalikke oskusi ja teine oleks neile, kes soovivad astuda kõrgkooli (Muni, 2001). Paljud probleemid laheneksid ka siis, kui kõikide riigieksamite tegemine oleks õpilastele vabatahtlik (Kerstna 2001). Iga kool saaks siis oma tingimustele vastavalt õpilasi kõige otstarbekamalt arendada; haridusministeerium saaks riigieksamit teinute protsendi järgi teada iga kooli akadeemilise taseme selles aines; Riigieksamikeskus hoiaks raha kokku, kuna eksamid teeksid vaid kõrgkooli pürgijad.

Riigieksamitöid koostatakse Riikliku Eksami- ja Kvalifikatsioonikeskuse juhtimisel. Õpetajad saavad oma aine komisjonile ettepanekuid ülesannete kohta, mida võiks riigieksamil kasutada. Kitsas ring spetsialiste valib nendest välja ülesanded, mis riigieksamil tegelikult kasutusele tulevad. Ülesanded valitakse lähtudes ainekavast. Näiteks matemaatika riigieksamil on püütud haarata kõiki teemasid, mis kolmel viimasel aastal on õpitud (Uudelepp, 1999).

J. Allik ja O. Must (1999) on analüüsinud riigieksamite sobivust kõrgkooli vastuvõtukriteeriumina. Nad leiavad, et riigieksamid kontrollivad suurel määral faktilise materjali tundmist, mitte akadeemilist võimekust, kompetentsust ja probleemide lahendamise oskust, mis peaks olema kõrgkooli sissesaamise kriteeriumiks. Selle tõttu on kõrgkooli astumisel edukad äsja gümnaasiumi lõpetanud neiud, aga gümnaasiumi varem lõpetanud noormehed maapiirkondadest jäävad

kõrgkoolidest välja. Lähtudes sellest riigieksami puudusest soovivad J. Allik ja O. Must kasutada nende poolt välja töötatud akadeemilise võimekuse testi, mille eeskujuks on SWESat.

J. Allik ja O. Must (1999, lk 4) kirjutavad veel: "Psühhoomeetrilisest seisukohast on täiesti arusaamatu, miks opereeritakse riigieksami toorpunktidega, mitte aga normeeritud andmetega, mis lubaks hõlpsasti vältida testimise aastast ja keelest tingitud mõõtmisviigu". Tõepoolest, riigieksamitulemuste üleviimine standardskaaladele võimaldaks paremini võrrelda erinevate testivariantide ja erinevate aastate tulemusi, kui erinevaid teste või nende variante teeb küllalt suur ja esinduslik õpilaste kogum. J. Alliku ja O. Musta poolt toodud ajaloo näite puhul ilmselt annaks standardpunktide kasutamine õiglasemaid tulemusi kui toorpunktid. Mõnes aines on aga niivõrd vähe testi täitjaid, et ka standardpunktidele üleminek ei võimalda erinevate aastate tulemusi korrektselt võrrelda.

J. Allik ja O. Must (1999) kasutavad TÜ sotsiaalteaduskonnas erinevate aastate testitulemuste võrdlemiseks ühiseid ülesandeid nendel aastatel, nii nagu me nägime SAT'i puhul. Selline katseandmetel põhinev erinevate aastate ja erinevate testivariantide võrdsustamine annab kõige täpsemaid tulemusi. Riigieksamid on õpilaste teadmiste mõõtmise suhteliselt objektiivne vahend, kuid täielikuma õigluse saavutamiseks on oluline erinevate aastate ja erinevate testivariantide tulemused võrdsustada katseandmete põhjal.

Nõuanded testi täitjale

Ainetestid on meie õpilaste jaoks suhteliselt uus asi ja nendega esmakordselt kohtudes saavad nad testis halvema tulemuse, kui nende teadmiste tase tegelikult on. Testide täitmist tuleb harjutada. Välismaal on selleks mitmeid võimalusi. Trükitakse juba kasutatud aineteste, mille järgi on eelseisvateks testideks võimalik valmistuda, sest testi ülesannete liigid ja raskusaste ei muutu oluliselt. Trükitakse õppevahendeid, mis sisaldavad nõuandeid testi täitmiseks koos näidisülesannetega. Töötavad ka keskused, kus aidatakse soovijatel testideks valmistuda. Meilgi oleks kõike seda tarvis, et meie õpilased saaksid teiste riikide õpilastega võrdsetes tingimustes võistelda.

Allpool refereerime Pedagoogilise Testimise Teenistuse soovitusi õpilastele (Essential guidelines ..., 2001; Graduate ..., 1994;

Taking the SAT I, 2000; Taking the SAT II, 2000). Kõigepealt **kuus soovitus järgimiseks enne testi täitmise algust.**

1. Tutvuge hoolikalt testiülesannete liikidega, et Te mõistaksite kiiresti, mida Te peate tegema ja kuidas vastama! Need minutid, mis Te kulutate ülesannete juhendi mõtestamisele on ülesannete lahendamise ajast kaduma läinud. Teatud ettekujutuse ülesannete võimalikest liikidest saab ka käesoleva õppevahendi osast "Testiülesannete liigid".
2. Õppige tundma vastustelehte ja seda, kuidas sinna vastuseid märkida! Kui ühele ülesandele on märgitud kaks vastust, siis loetakse see vastus vääraks. Kui vastustelehti hindab arvuti, siis tuleb väärvastus täiesti maha kustutada ja õige vastuse ovaal teha täiesti mustaks.
3. Uurige ka seda, mille järgi Teie vastuseid hinnatakse ja kuidas punkte antakse!
4. Harjutage testiülesannete lahendamist! Näidisülesandeid leiate eelnevate aastate testivihikuist ja harjutuskogumikest. Juba harjutamisel piirake oma aega nii, nagu see toimub testimisel!
5. Kui testija selgitustest jäi midagi arusaamatuks, küsige kindlasti enne testi täitmise algust! Küsimised testi täitmise käigus viivad Teie aega ja mõnikord pole testi läbiviijail siis enam lubatud selgitusi anda.
6. Ja lõpuks kõige tähtsam: õppige aine hästi selgeks!

Testi täitmise käigus soovitatakse arvestada järgmist.

1. Lahendage kõigepealt ülesanded, mis tunduvad Teile kerge-tena! Raskemate ülesannete juurde võite tagasi pöörduda siis, kui Teil aega üle jääb. Märkige need vastamata küsimused testivihikus ära, et neid hiljem kiiresti üles leida. Ka valed valikvastused võite testivihikus maha tõmmata, et hiljem ainult allesjäänute seast õiget otsida. Selleks kõigepealt peab aga meil kindel olema, et lubatakse testivihikusse kirjutada nagu USA-s. Meil sageli veel ei lubata.

Toodud soovitus rõhutatakse välismaa testides, kus iga õige vastuse eest antakse tavaliselt üks punkt. Meil võidakse küll keerukamate ülesannete eest anda ka kaks või enam punkti, aga siiski ei tasu kogu aega mõnele raskele ülesandele kulutada. Kui ülesannete lahendus hinnatakse erineva punktide arvuga, siis peaks see punktide arv iga ülesande juures trükitud olema.

Mõnikord on ülesanded igas seerias reastatud raskuse kasvamise järjekorras. Siis tuleks kõigepealt lahendada seeriade esimesed ülesanded kuni raskete ülesanneteni. Sellele reeglile ei allu küsimused loetud lõigu kohta, mis paiknevad teksti sisu järjekorras. Programmitestis ja seni meil koostatud testides ei tarvitse ülesanded ka olla raskuse järjekorras, kuna meil sageli ülesannete raskust ei kontrollita.

Arvutitestides ei tarvitse olla võimalust ülesande juurde tagasi tulla. Seetõttu vastake, kui võimalik, neile kohe, ja jaotage oma aeg nii, et Te jõuaksite kõiki ülesandeid lahendada.

2. Kui Te teate, et mõni valikvastus on vale, siis võite ülejäänute seast valida vastuse ka juhuslikult. Täiesti huupi vastamisest pole kasu, kui valikvastustega ülesannetes ja alternatiivülesannetes arvestatakse valede valikute eest punkte maha. Kui sellist mahaarvamist pole, siis võib vastuse juhuslikult märkida kõigile ülesannetele, mida Te lahendada ei oska või mille lahendamiseks pole aega.

3. Valikvastustega küsimustele püüdke kõigepealt vastata valikuid lugemata. Pärast mõttes vastamist lugege valikvastuseid. Kui Teie vastus langeb kokku ainult ühe antud valikvastusega, siis on vastus leitud ja Te saate edasi minna. Kui kaks või enam vastust tunduvad sobivat, siis tuleb ülesannet uuesti lahendada, lähenedes sellele võimalusel teisest vaatepunktist.

4. Vabavastuseliste ülesannete puhul valede vastuste eest punkte maha ei arvestata. Seega võib neile anda ka vastuse, milles Te kindel ei ole.

5. Hea ja isegi väga hea tulemuse saamiseks ei pea Te vastama kõiki küsimusi eristustestis. Õiged vastused poolele küsimustest annavad tavaliselt keskmise tulemuse.

6. Püüdke vastata kõigile testiküsimustele! Praktika näitab, et need, kes vastavad ka viimastele küsimustele saavad paremaid tulemusi, kui need, kes testi lõpuni ei jõudnud.

Eristustestides on loomulik, et Te ei suuda osale küsimustest vastata. Ka need õpilased, kes saavad hea tulemuse, jätavad mõne küsimuse vastamata. See on tegelikult juhtnõör, mis peaks olema testide sissejuhatuses, et meie programmitestidega harjunud õpilased ei hakkaks liigselt närveerima, kui nad mõnele ülesandele vastata ei suuda, vaid läheksid edasi järgmise ülesande juurde.

Lisaks üldistele soovitudele antakse **juhtnööre ka üksikute ülesannete liikide kohta**. Toome mõned näited.

Analoogiaülesannete lahendamisel soovitatakse kõigepealt võimalikult täpselt sõnastada antud sõnapaariga väljendatud seos ja siis vaadata, milline valikvastuste sõnapaaridest väljendab sama seost.

Kaks antud sõna väljendavad teatud seost. Valige viiest antud valikust see sõnapaar, milline kõige lähedasemalt väljendab sama seost.

LEIVAPURU — LEIB

- a) unts — mōõtühik
- b) pilbas — puu
- c) vesi — pang
- d) nōör — köis
- e) võikreem — või

Mōttekäik võiks olla siin järgmine: leivapuru on leiva väga väike osa ja ka pilbas on puu väga väike osa, mistōttu õige vastus on b) (Taking the SAT I, 2000, p 8).

Valikvastustega lünktestis on oluline mõista täpselt lause tähendust. Ei piisa sellest, et sõna sobib lünga naabruses olevate sõnadega, ta peab sobima kogu lausega. Kui lauses on kaks lünka, siis võib algul vaadata, millised valikvastused sobivad ühte lünka ja ülejäänud vaatlusest välja jätta. Seejärel vaadata teise sõna järgi, milline valik allesjäänuiust on sobiv. Lõpuks tuleb kindlasti kontrollida, kas kogu lause on tehtud valiku puhul korrektne.

Lugemistestides võib olla otstarbekas lugeda enne küsimusi ja siis alles teksti, et teada, millele tekstis tähelepanu pōōrata. Kui küsimused on teksti sisu järjekorras, siis võib otsida tekstist vastuse esimesele küsimusele, siis teisele küsimusele jne. Tegelikult on küsimuste eelnev lugemine ajakulu ja kas see võte sobib konkreetsele isikule või mitte, tuleks enne testimist järele proovida.

Need on võtted, mis aitavad testi täitjal oma teadmisi täielikult kasutada. Testi tulemus sõltub teatud määral nende võtete valdamisest, kuid põhiliselt määrab testitulemuse siiski teadmiste tase. Seetōttu on põhiline soovitus väga lihtsalt väljendatav: testi edukaks täitmiseks omandage aine hästi. Kui ainet pole hiljuti koolis õpitud, siis tuleks

seda enne testi korrata. USA-s on leitud, et 20 tunnilised kursused tõstavad koolijõudlustesti tulemusi kollektiivis kaalal 10–15 punkti võrra. Tuupimisest eelmisel õhtul pole märkimisväärset kasu leitud, küll aga soovitatakse eelmisel päeval korrata põhivara ja testiülesannete liigid veelkord üle vaadata.

Lõppsõna

Kogemused näitavad, et ainetestid võimaldavad ühetaolisemalt ja objektiivsemalt mõõta õpilaste teadmiste, oskuste ja mõtlemise taset eri ainetes, vähendades nii võimalusi subjektiivseteks ja voluntaristlikeks otsustusteks hariduses. Diskussioonid ainetestide üle aitavad täpsustada hariduse eesmärgi. Ainetestid stimuleerivad õpilasi ning õpetajaid süstemaatiliselt ning eesmärgikindlalt töötama ja panevad kogu kooli personalile vastutuse õppetöö tulemuste eest (Lindgren, Suter, 1995, lk 385).

Meil on testide oskamatu ja hoolimatu kasutamine kord juba testid kompromiteerinud. Seda viga ei tasu korrata. Seetõttu on väga vajalik tundma õppida testide teooriat ja testide kasutamisel arvestada ka arenenud riikide praktikat.

Hea testi koostamine on suur töö. Algul ei jõuta võib-olla kõiki etappe läbida ja me saame osaliselt nõuetele vastavad testid. Loomulikult vastavad ka nende testidega saadud tulemused vaid osaliselt nõuetele.

Ainetesti koostamine algab testi eesmärgi fikseerimisest. Kui me tahame kontrollida mingi ainekava omandatust, siis peame koostama programmitesti. Kui me soovime õpilasi eristada ainealaste teadmiste ja oskuste järgi, siis peame koostama eristustesti (tabel 9).

Testi eesmärgi fikseerimisele järgneb ülesannete koostamine, mille puhul põhiliseks kriteeriumiks on nende ülesannete sisuline valiidsus: mil määral antud ülesanne võimaldab kontrollida seda, mida kontrollida tahetakse. Kontrollides ebaolulist suuname me õpetajate ja õpilaste pingutused edaspidi taolise ebaolulise materjali omandamisele.

Ülesannete valik programmitesti toimub ühe ainsa kriteeriumi järgi — ülesannete esinduslikkus kontrollitava materjali suhtes. Eristustestide puhul tõuseb selle kõrvale ülesannete raskus ja kui me soovime koostada mingi üksikokuse testi (homogeenset eristustesti), siis tuleb ülesannete valikul kolmanda alusena kasutada nende diagnoosivat väärtust.

Testi kvaliteedi põhinäitajad — reliaablus ja valiidsus — on väga olulised igal testi liigil. Nendest reliaablust saab määrata ainult katseliselt ja ka valiidsust oleks otstarbekas katsetel määrata. Igal korralikul testil peab olema juhend, milles on testi koostamise kirjelduse ja kvaliteedi näitajate kõrval toodud ka testi normid. Programmitestidel

pole küll normitabeleid ega normeerimist eristustestide mõttes, kuid neil on siiski eeskirjad tulemuste tõlgendamiseks kasvõi üleminekuna koolihinnetele.

Tabel 9

Ainetesti liikide võrdlus.

Testi liik	Eesmärk		Ülesannete valik			reliiablus	valiidsus	normid	juhend
	programmi oman-datavuse kontroll	õpilaste eristamine	esinduslikkus	raskus	diagn. väärtus				
Programmitest	++	+	++	-	-	++	++	+	++
Heterogeenne eristustest	+	++	+	+	-	++	++	++	++
Homogeenne eristustest	+	++	+	+	++	++	++	++	++

Testitulemuste tõlgendamisel tuleb arvestada väga paljusid asjaolusid: mida test mõõdab, kui täpselt ta mõõdab, millise õpilaskontingendi jaoks on test ette nähtud, kui võrd õnnestus täita testi läbiviimise nõudeid, milline oli õpilaste psühholoogiline ettevalmistatus testiks jne. Loomulikult tuleb sellele mõelda juba enne testimist ja loodetavasti saab kõikidele nendele küsimustele vastata positiivselt, kuid kui mõnes aspektis on puudujääke, siis tuleb seda arvestada testitulemuste tõlgendamisel. Dotsent Kalju Toim armastas korrata järgmist mõtet: kui õpilane on testi hästi teinud, siis me teame, et tal on head teadmised; kui õpilane on testi halvasti teinud, siis me ei tea, millest see on tingitud — võib-olla valutas tal pea, võib-olla oli tal enne testi suur pahandus, võib-olla polnud ta õppinud. See arutus illustreerib ettevaatlikkust, mis on vajalik testitulemuste tõlgendamisel.

Lõpetuseks tahaks korrata üht väga olulist hoiatust: **ainetestidega ei saa mõõta kooli kõigi eesmärkide saavutatust**. Näiteks kooli kasvatuslike eesmärkide saavutatuse hindamisel võib ainetestidest vaid kaudset kasu olla. Uued õpetamise eesmärgid ning meetodid võetakse algul kasutusele ja siis alles koostatakse nende jaoks testid, kui võimalik. Seega võib ainetestide ületähtsustamine takistada kooli uuendamistaotluste katsetamist. Lisaks ainetestidele on kooli tegevuse edukuse hindamiseks tarvis teisi vahendeid ja meetodeid, mis võivad teatud tingimustel osutada testidest olulisemaks.

Kirjandus

- Allik J. (1997).** Võimete test. Kogumikus: Tartu Ülikooli sisseastumistestid 1996. Tartu: TÜ, lk 94–98.
- Allik J., Must O. (1999).** Riigieksamite ja sotsiaalteaduskonna akadeemilise võimekuse testi (AVT) võrdlev analüüs. Käsikiri (vt ka <http://sys130.psych.ut.ee/~jyri/ee>)
- APSTest 2 (2000).** Phare ISE CD 5.
- Campbell N. A. (1993).** Biology. Redwood City, California a. o.: The Benjamin/Cummings Publishing Company. 1200 p.
- Capper J. (1996).** Testing to Learn, Learning to Test. Newark, Delaware: IRA & Washington: AED. 203 p.
- Doster W. C. a. o. (1979).** Barron's how to Prepare for the College Level Examination Program. Woodbury, New York. 568 p.
- Essential Guidelines for all Test Takers. 2001.** <http://www.ets.org/cbt>.
- Examinator 4.5. (1999).** Õpi- ja eksamineerimisprogramm valikvastutega testide sooritamiseks. Tartu: OÜ IP Tehnoloogia. 18 lk.
- Genesee F., Upshur J. A. (1996).** Classroom based Evaluation in Second Language Education. Cambridge University Press. 268 p.
- Graduate Record Examinations (1994).** Educational Testing Service. 48 p.
- Husén T. (1997).** Educational Research and Policy making. In John P. Keeves (ed.) Educational Research, Methodology, and Measurement: An International Handbook. Second edition. New York, etc.: Elsevier Science, 251–257.
- Karu G. (1974).** Füüsika redeltestid VII klassile. Tallinn: Valgus. 21 lk.
- Kees P. (1984).** Statistika pedagoogidele ja psühholoogidele. III osa. Tallinn: TPedI. 216 lk.
- Kelder T., Nigul A., Prank R. (2001).** Testide moodustamise ning testidele vastamise programmisüsteem APSTest 2. <http://www.ce.ut.ee/apstest2>
- Kerstna T. (2001).** Riigieksamid ja hindekünnis. Õpetajate Leht, 25 mai, lk 7.
- Kolen M. J. (1999).** Equating of Tests. In: G. N. Masters & J. P. Keeves. Advances in Measurement in Educational Research and Assessment. Amsterdam et al: Pergamon, pp. 164–175.
- Kärner O. (1976).** Standardiseeritud kontrolltöid matemaatikast. V klass. Tallinn. Eesti NSV Haridusministeerium. 68 lk.
- Lienert G. A. (1969).** Testaufbau und Testanalyse. Weinheim, Berlin, Basel. 599 S.
- Krull E. (2000).** Pedagoogilise psühholoogia käsiraamat. Tartu: TÜ Kirjastus. 637 lk.

- Lindgren H. C., Suter W. N. (1995).** Pedagoogiline psühholoogia kooli-
praktikas. Tartu: TÜ Kirjastus. 594 lk.
- Mereste U. (1975).** Statistika üldteooria. Tallinn: Valgus. 496 lk.
- Muni E. (2001).** Õppekava toimikond vaeb valupunkte. Õpetajate Leht, 18
mai, lk 8.
- NAEP results show heartening gains. (1999).** Reading Today, Vol 16, No 5,
p 1 and 4–5.
- Pandis M. (1996).** Suurim lugemisoskuse uurimus Eestis. Haridus, nr. 2, lk
20–27.
- Peiker S. (1993).** Practice Tests for School-Leavers. Tartu, 26 p.
- Rapoport I., Selg R., Sotter R. (1987).** Testõ v obuèenii inostrannõm
jazõkam v srednei škole. Tallinn: Valgus. 351 lk.
- Reppo K., Rebane M. (1993).** Teste dein Deutsch. Tartu. 56 S.
- Stage C. (1998).** Admission to Higher Education in Sweden and the Swedish
Scholastic Aptitude Test. Paper presented at the seminar on testing in
Tartu University. Manuscript, 10 p.
- Sukamägi A. (1994).** Kõrgkooli õpiedukuse prognoosimise võimalusi. Kogu-
mikis: Isiksuse omaduste mõjust noorte haridus- ja elutee kujune-
misele. Tartu: TÜ kirjastus, lk 99–107.
- Taking the SAT I. (2000).** Reasoning test. USA, CEEB & ETS. 80 p.
- Taking the SAT II. (2000).** Subject tests. USA, CEEB & ETS. 96 p
- Toim K. (1981).** Psühhomeetria teoreetilised alused. Õppevahend psüh-
holoogiaosakonna üliõpilastele. Tartu: TRÜ. 78 lk.
- Toim K. (1987).** Psühodiagnostika. Eksperimentaal-psühholoogia õppeme-
todiline materjal. Tartu: TRÜ. 84 lk.
- Tooding L.-M. (1998).** Andmeanalüüs sotsiaalteadustes. Tartu: TÜ. 175 lk.
- Unt I. (1971).** Ainetestide funktsioonid uurimistöös ja koolipraktikas. —
Nõukogude Kool, nr. 11, lk 808–815.
- Uudelepp H. (2001).** 21. Mail toimub matemaatika riigieksam. Õpetajate Leht
11. Mai, lk 15.
- Uudelepp H. (koostaja) (1999).** Matemaatika Riigieksam 1999. Tallinn:
Riiklik Eksami- ja kvalifikatsioonikeskus. 42 lk + lisad.
- What is ETS.** <http://www.ets.org> 2001. 2 lk.
- Õpilaste hindamise kord (2000).** Haridusministri 20. septembri 2000. a
määrus nr 33. HAMm RTL 2000, 102, 1601.
- V ja IX klassi geograafia 1973. a. kontrolltööde tulemused (1974).** Tallinn:
ENSV haridusministeerium. 50 lk.

Lisad

Testitulemuste standardskaalad

Protsent-normid	z-saala	Staniinid C	Kolledž i skaala	T-skaala	Hinne (Lienert)
0,135	-3,00	-1	200	20	
0,159	-2,95				
0,187	-2,90		210	21	
0,219	-2,85				
0,256	-2,80		220	22	
0,298	-2,75				
0,347	-2,70		230	23	
0,403	-2,65				
0,466	-2,60		240	24	
0,539	-2,55				
0,621	-2,50	0	250	25	
0,714	-2,45				
0,820	-2,40		260	26	
0,939	-2,35				
1,072	-2,30		270	27	
1,222	-2,25				
1,320	-2,20		280	28	
1,578	-2,15				
1,786	-2,10		290	29	
2,018	-2,05				
2,275	-2,00		300	30	1,0
2,56	-1,95				
2,87	-1,90	1	310	31	1,1
3,22	-1,85				
3,59	-1,80		320	32	1,2
4,01	-1,75				
4,46	-1,70		330	33	1,3
4,95	-1,65				
5,48	-1,60		340	34	1,4

Lisa 1 (järg)

6,06	-1,55				
6,68	-1,50	2	350	35	1,5
7,35	-1,45				
8,08	-1,40		360	36	1,6
8,85	-1,35				
9,68	-1,30		370	37	1,7
10,57	-1,25				
11,51	-1,20		380	38	1,8
12,51	-1,15				
13,57	-1,10		390	39	1,9
14,69	-1,05				
15,87	-1,00	3	400	40	2,0
17,11	-0,95				
18,41	-0,90		410	41	2,1
19,77	-0,85				
21,19	-0,80		420	42	2,2
22,66	-0,75				
24,20	-0,70		430	43	2,3
25,78	-0,65				
27,43	-0,60		440	44	2,4
29,11	-0,55				
30,86	-0,50	4	450	45	2,5
32,64	-0,45				
34,46	-0,40		460	46	2,6
36,32	-0,35				
38,21	-0,30		470	47	2,7
40,13	-0,25				
42,07	-0,20		480	48	2,8
44,04	-0,15				
46,01	-0,10		490	49	2,9
48,01	-0,05				
50,00	0,00	5	500	50	3,0
51,99	0,05				

Lisa 1 (järg)

53,98	0,10		510	51	3,1
55,96	0,15				
57,93	0,20		520	52	3,2
59,87	0,25				
61,79	0,30		530	53	3,3
63,68	0,35				
65,54	0,40		540	54	3,4
67,36	0,45				
69,15	0,50	6	550	55	3,5
70,88	0,55				
72,57	0,60		560	56	3,6
74,22	0,65				
75,80	0,70		570	57	3,7
77,34	0,75				
78,81	0,80		580	58	3,8
80,23	0,85				
81,59	0,90		590	59	3,9
82,89	0,95				
84,13	1,00	7	600	60	4,0
85,31	1,05				
86,43	1,10		610	61	4,1
87,49	1,15				
88,49	1,20		620	62	4,2
89,44	1,25				
90,32	1,30		630	63	4,3
91,15	1,35				
91,92	1,40		640	64	4,4
92,65	1,45				
93,32	1,50	8	650	65	4,5
93,94	1,55				
94,52	1,60		660	66	4,6
95,05	1,65				
95,54	1,70		670	67	4,7

Lisa 1 (järg)

95,99	1,75				
96,41	1,80		680	68	4,8
96,78	1,85				
97,13	1,90		690	69	4,9
97,44	1,95				
97,73	2,00	9	700	70	5,0
97,98	2,05				
98,21	2,10		710	71	5,1
98,42	2,15				
98,61	2,20		720	72	5,2
98,78	2,25				
98,93	2,30		730	73	5,3
99,06	2,35				
99,18	2,40		740	74	5,4
99,29	2,45				
99,38	2,50	10	750	75	5,5
99,46	2,55				
99,53	2,60		760	76	5,6
99,60	2,65				
99,65	2,70		770	77	5,7
99,70	2,75				
99,75	2,80		780	78	5,8
99,78	2,85				
99,81	2,90		790	79	5,9
99,84	2,95				
99,87	3,00	11	800	80	6,0

Raveni testi normid

V %	20	25	30	35	40	45	50	55	60	65
95	55	55	54	53	52	50	48	46	44	42
90	54	54	53	51	49	47	45	43	41	39
75	49	49	47	45	43	41	39	37	35	33
50	44	44	42	40	38	35	33	30	27	24
25	37	37	34	30	27	24	21	18	15	13
10	28	28	25	6	-	-	-	-	-	-
5	23	23	19	-	-	-	-	-	-	-

V — katseisikute vanus aastates.

% — katseisikuste protsent, kes saavad tabelis näidatust vähem punkte.

SAT harjutustesti normid
(Taking the SAT I, 2000, lk 78)

Toorpunktid	Verbaalne mõtlemine kolleđi skaalal	Matemaatiline mõtlemine kolleđi skaalal
78	800	
77	800	
76	800	
75	800	
74	790	
73	780	
72	760	
71	750	
70	740	
69	730	
68	720	
67	710	
66	700	
65	690	
64	680	
63	670	
62	660	
61	660	
60	650	800
59	640	790
58	630	770
57	630	750
56	620	730
55	610	720
54	610	700
53	600	690
52	600	680
51	590	670

Toorpunktid	Verbaalne mõtlemine kolleđi skaalal	Matemaatiline mõtlemine kolleđi skaalal
50	580	660
49	580	650
48	570	640
47	570	640
46	560	630
45	560	620
44	550	610
43	540	600
42	540	600
41	530	590
40	530	580
39	520	570
38	520	570
37	510	560
36	510	550
35	500	540
34	500	540
33	490	530
32	480	520
31	480	520
30	470	510
29	470	500
28	460	500
27	460	490
26	450	480
25	440	480
24	440	470
23	430	460
22	430	460
21	420	450

Lisa 3 (järg)

Toorpunktid	Verbaalne mõtlemine kolledži skaalal	Matemaatiline mõtlemine kolledži skaalal
20	410	440
19	410	430
18	400	430
17	390	420
16	390	410
15	380	410
14	370	400
13	360	390
12	360	380
11	350	370
10	340	370
9	330	360
8	320	350
7	310	340
6	300	320
5	290	310
4	280	300
3	270	280
2	260	270
1	240	250
0	230	230
-1	220	210
-2	200	200
-3	200	220
-4	200	210
-5	200	200

Mõõteskaalad pedagoogikas

Nr.	Nimetus	Olemus	Näited	Statistikud
1.	Nominaal-skaala e. nimetuste skaala	Objektide ekvivalentsus	Värvide nimed Riiete liigid Õpikute numbrid	Mood Sagedus protsentides Kooskõlakordaja
2.	Järjestik-skaala	Suurem-väiksem	Eksperthinnangud Koolihinded	Mediaan Protsentiil Astakkorrelatsioon
3.	Intervall-skaala e. vahemik-skaala	Suurem-väiksem kindla suuruse võrra	Celsiuse skaala Standardiseeritud testi tulemused	Aritmeetiline keskmine Liitmise-lahutamine Lineaar-korrelatsioon
4.	Suhteskaala	Nullpunkt sisulise tähendusega	Kelvini skaala Pikkuste mõõtmine Programmitesti tulemused	Variatsioonikordaja

Märkus: iga järgnev skaala sisaldab ka eelmise omadused.

Mõistete selgitusi

STATISTIKA — tulemuste matemaatilise töötlemise meetodid.

Dispersioon — tulemuste hajuvuse näitaja: mida suurem dispersioon, seda kaugemale kalduvad üksikmõõtmiste tulemused nende keskmisest väärtusest. Dispersiooni arvutusvalem:

$$\ddot{a}^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

kus \ddot{a} — dispersioon,
 \bar{x} — keskmine tulemus,
 x_i — üksikmõõtmise tulemus,
 n — vaadeldavate mõõtmiste arv.

Korrelatsioon — kahe muutuva suuruse vahelise seose tugevuse näitaja. Korrelatsioon muutub piires -1 kuni +1. Korrelatsioon 0 näitab seose puudumist, korrelatsioon 0,5 mõõdukat seost, korrelatsioon +1 täielikku seost. Negatiivne korelatsioon tähendab, et ühe suuruse suurenedes teine väheneb.

Standardhälve — tulemuste hajuvuse näitaja, ruutjuur dispersioonist.

TEST — lühivastuseliste ülesannete ja küsimuste kogum, mille lahendamise edukuse järgi otsustatakse inimese omaduste taseme üle.

Ainetest — test ainealaste teadmiste, oskuste, mõtlemisvõime mõõtmiseks.

Eristustest — ainetest, mille põhieesmärgiks on katseisikute eristamine nende ainealaste teadmiste, oskuste, mõtlemisvõime järgi.

Homogeenne test — ainetest, mis mõõdab inimese üht oskust.

Programmitest — programmi, ainekava või õpiku suhtes täielikult esinduslik ainetest.

Testi reliaablus — mõõdetava suuruse dispersiooni osa testitulemuste ülddispersioonis, väljendatakse tavaliselt korrelatsioonikordajana, mida võib lihtsustatult tõlgendada kui kordusmõõtmiste kokkulangevuse määra.

Testi valiidsus — mõõdetud suuruse ja mõõta soovitud suuruse kokkulangevuse määra, väljendatakse tavaliselt korrelatsioonikordajana.

TESTISKAALAD — standardsed punktide süsteemid, mille abil väljendatakse testitulemusi (vt. ka joonis 4),

z-skaala — testitulemuste skaala, millel keskmine tulemus võrdsustatakse nulliga ja tulemuste standardhälve ühega.

Staniinide skaala — testitulemuste skaala, millel keskmine tulemus võrdsustatakse viiega ja tulemuste standardhälve kahega.

Kolledž i skaala — testitulemuste skaala, millel keskmine tulemus võrdsustatakse viiesajaga ja tulemuste standardhälve sajaga.

T-skaala — testitulemuste skaala, millel keskmine tulemus võrdsustatakse viiekümnega ja tulemuste standardhälve kümmega.

TESTITULEMUSTE NORMEERIMINE — toorpunktide üleviimine standardskaaladele.

Standardpunktid — testitulemused mõnel standardsel testiskaalal.

Toorpunktid — testitulemused õigete vastuste eest saadud punktide summana.