

ALTE eksamite hea tava põhimõtted
Redigeeritud variant
oktoober 2001

Koostanud ALTE tegevusjuhendi töörühm arutamiseks ALTE liikmetele

Sisukord

1.0. Sissejuhatus

- 1.1. ALTE tegevusjuhend.
- 1.2. Eksamiprotsessis osalejad.
- 1.3. Hea tava saavutamine.
- 1.4. Arusaam eksamite kasulikkusest.

2.0. Eksami kvaliteet: hea tava saavutamine

- 2.1. Valiidsus.
 - 2.1.1. Valiidsuse aspektid
 - 2.1.2. Andmed konstrukti kohta.
 - 2.1.3. Andmed sisu kohta.
 - 2.1.4. Andmed kriteeriumide kohta.
 - 2.1.5. Hea tava.
- 2.2. Reliaablus.
 - 2.2.1. Reliaabluse aspektid.
 - 2.2.2. Hea tava.
- 2.3. Mõju.
 - 2.3.1. Mõju aspektid.
 - 2.3.2. Hea tava.
- 2.4. Praktilisus.
 - 2.4.1. Praktilisuse aspektid.
 - 2.4.2. Hea tava.
- 2.5. Teenuse kvaliteet.
 - 2.5.1. Teenuse kvaliteedi aspektid.
 - 2.5.2. Hea tava.

Allikad ja viited.

1.0. Sissejuhatus

Kõigi hindamise aspektide olulisimaks mõisteks on õiglus, mis moodustabki alltoodud *hea tava* põhimõtete konteksti.

Alati kui eksameid hariduskontekstis laiaulatuslikult kasutatakse, mõjutavad need nii üksikisikuid kui ka institutsioone ja ühiskonda tervikuna.

Eksamite võimaliku laialdase mõju tõttu on tähtis, et nii *eksami koostajad* kui ka selle *kasutajad* järgiksid kõigile ühtset tegevusjuhendit, mis tagaks kvaliteetsed hindamisprotseduurid ning kõigi huvigruppide õiglase kohtlemise. Selline tegevusjuhend peab põhinema kindlatel hindamise hea tava põhimõtetel, mis võimaldavad saavutada hindamise kõrge kvaliteedi ja õigluse.

Arutelu selle üle, mis on hea tava antud dokumendi käsitluses, on püüe väljendada muret vastutuse pärast hindamise kõigis valdkondades, millega ALTE liikmed tegelevad.

See rõhutab andmete kogumise vajalikkust *valideerimise* eesmärgil ning *uurimis- ja arendustöö* osaksamiprotsessis. Seega arenevad allpool esitatud põhimõtted uurimis- ja arendusprogrammide laienedes tõenäoliselt tulevikus edasi.

1.1. ALTE tegevusjuhend (1994)

1.1.1. Psühholoogilises ja hariduslikus hindamises on USA-l pikaajaline standardite kehtestamise kogemus. 1999. aastal ilmus kõige uuem väljaanne dokumendist *Standards for educational and psychological testing* (AERA/APA/NCME), kuid varasemad väljaanded on pärit juba 1954. aastast. Nimetatud dokument on keeletestimisel mõjutanud nii testi koostajaid kui testi kasutajaid ning sellele viidatakse pidevalt keeletestimisalases kirjanduses (vt nt Bachman 1990, Bachman&Palmer 1996). Alates kaheksakümnendate aastate algusest on ETS (*Educational Testing Service*, USA) koostanud oma dokumendi *Standards for Quality and Fairness* (1981, 1987, 2000), toetudes sealjuures paljus AERA/APA/NCME dokumendile.

Hiljuti vaatas ILTA – *the International Language Testing Association* – läbi rahvusvahelised testimisstandardid (1995) ja aastal 2000 avaldas oma *Eetikakoodeksi (Code of Ethics)*. See dokument esitab üheksa annoteeritud põhimõtet, mis "*põhinevad moraali filosoofial ja mille eesmärgiks on kujundada professionaalset käitumist*".

1.1.2. Euroopa kontekstis on ALTE tegevus hakanud mõjutama professionaalsuse standardeid. 1994. aastal avaldas ALTE oma esimese tegevusjuhendi, kus kirjeldati kvaliteedistandardeid, mida assotsiatsiooni liikmed oma keeleeksameid koostades püüdsid järgida.

Enne tegevusjuhendi avaldamist arutasid ALTE liikmed aastatel 1991–1993 läbi selle algvariandi. Algvariant põhines dokumendil *The Code of Fair Testing Practices* (1998), mille oli Washington DC-s välja töötanud ühtse testimistegevuse komitee (*Joint Committee on Testing Practice* 1988) ja mille eesmärgiks oli esitada laiapõhjaline avaldus selle kohta, millised võiksid olla eksami kasutajate ootused ning huvigruppide kohustused ja vastutus *õigluse* poole püüeldes.

1.1.4. Tegevusjuhend eristab testimisprotsessis kolme põhilist huvigrupi:

- eksami koostajad (nt eksamikeskused ja muud institutsioonid, kes on ALTE liikmed);
- eksaminandid, kes sooritavad eksami omal valikul, kellegi suunamisel või mingist vajadusest tulenevalt;

- eksami kasutajad, kellel on vaja kasutada eksamitulemusi mingi otsuse langetamiseks või mõnel muul põhjusel.

1.1.5. Lisaks sellele määrab tegevusjuhend huvigruppide kohustused neljal erineval alal:

- eksamiarendus,
- eksamitulemuste tõlgendamine,
- õigluse poole püüdlamine,
- eksaminandide teavitamine.

On oluline, et koodeks rõhutab huvigruppide ühisvastutust ja keskendub nii eksami kasutajate kui ka eksami koostajate vastutusele õigluse taotlemisel.

1.1.6 Lisadokumendi *Principles of Good Practice for ALTE Examinations* koostasid Saville ja Milanovic 1991. ja 1993. aastal ning see arutati läbi ALTE koosolekutel Alcalá de Henaresis 1992 ning Pariisis ja Münchenis 1993. Eesmärgiks oli kehtestada üksikasjalikumalt sõnastatud reeglid, mille ALTE liikmed peaksid kõrgete professionaalsete standardite ellurakendamiseks omaks võtma.

Soov jõuda hea tavani tulenes paljudest ALTE liikmete töödelt (nt töö, mida tegi UCLES) ja hindamisalasest tegevusest üldse (nt Bachmani, Messicki uurimustest ja AERA/APA/NCME standarditest, 1985).

ALTE liikmed taotlesid dokumendile tuntud välisekspertidelt (Spolsky, Bachman, jaanuar-märts 1994) tagasisidet, mida arutati Arnhemis 1994. aastal. Kuigi dokumenti kogu ulatuses ei avaldatud, lisati osa sellest uude dokumenti nimega *Users Guide for Examiners*, mille ALTE Euroopa Nõukogu nimel koostas (1997).

1.1.7 Praegune redigeeritud variant dokumendist *ALTE eksamite hea tava põhimõtted* (2001) põhineb varasemal versioonil, kuid on põhjalikult nüüdisajastatud ja mitmes osas ümber töötatud. Detailsemalt käsitletakse *valiidsuse* ja *reliiabluse* küsimusi kui kõige olulisemaid ning nendega seonduvalt vaadeldakse ka eksamite mõju üksikisikule ja ühiskonnale.

1.1.8 See versioon nagu eelmisedki põhineb väga suures osas AERA/APA/NCME standardite dokumendi redigeeritud variandil (1999) – eriti valiidsuse ja reliiabluse osas – ning Bachmani 1990. ning Bachmani ja Palmeri 1996. aasta töödel.

1.2. Eksamiprotsessis osalejad

1.2.1. Eksamid mõjutavad mitte ainult eraisikuid, vaid ka institutsioone ja ühiskonda tervikuna, nagu ALTE tegevusjuhend märgib. Ometi ei ammenda tegevusjuhendis loetletud kolm huvigruppide kategooriat kogu eksamitegevuses osalejate ringi, kes on ALTE eksamitega seotud.

Eksamitega seotud üksikisikute hulka kuuluvad eksaminandid ja nende toetajad (õpilased, vanemad, õpetajad, töökohta taotlejad, töövõtjad jne). Lisaks kuuluvad nende hulka spetsialistid ja teadustöötajad, kes kavandavad, koostavad, viivad läbi ja valideerivad eksameid (st õpetajad, ülesannete koostajad, konsultandid, eksamineerijad, kooli omanikud, eksamikeskuste töötajad, inspektorid, õpikute koostajad jne).

Protsessiga seotud institutsioonide hulka kuuluvad koolid, ülikoolid ja kolledžid, valitsusasutused, kirjastajad, äri- ja tööstusettevõtted.

Üksikisikutele ja institutsioonidele tuleb kasuks, kui testimine aitab neil eesmärged saavutada ning ühiskonnale tuleb kasuks, kui nende eesmärkide saavutamine aitab kaasa üldise heaolu edendamisele.

ALTE liikmed							
Õppijad Õpetajad Kooli omanikud Testi koostajad Konsultandid Eksamineerijad/ Hindajad Eksamikeskuse töötajad Õppematerjalide autorid Kirjastajad (inter alia?)	<table border="1"> <tr> <td>Testi konstrukt</td> <td rowspan="4">Eksami tulemused</td> </tr> <tr> <td>Testi vorm</td> </tr> <tr> <td>Testi tingimused</td> </tr> <tr> <td>Testi hindamis- kriteeriumid</td> </tr> </table>	Testi konstrukt	Eksami tulemused	Testi vorm	Testi tingimused	Testi hindamis- kriteeriumid	Õppijad Vanemad/hooldajad Õpetajad Kooli omanikud Vastuvõtavad institutsioonid Riigiasutused Kutsekojad Tööandjad Teadustöötajad (inter alia?)
Testi konstrukt	Eksami tulemused						
Testi vorm							
Testi tingimused							
Testi hindamis- kriteeriumid							

(vt L. Taylor 1999)

See diagramm näitab, kui palju on huvigruppe, keda võib pidada *eksamiprotsessis osalejateks*. Nende hulgas on inimesi, kes aitavad kaasa ALTE eksamite koostamisele ja läbiviimisele ning ka neid, kes kasutavad eksamitulemusi ja tunnistusi.

ALTE liikmed peavad olema valmis jälgima kõikide huvigruppide vaateid ja suhtumist ning korrigeerima/muutma oma tegevust olenevalt sellest, kuidas huvigrupid eksameid kasutavad ja mida nendest arvavad.

1.3. Hea tava saavutamine

1.3.1. Huvigruppide (vt loetelu ülal) asjakohane kaasamine eksamiprotsessi on tähtis hea tava saavutamise põhimõte.

1.3.2. Hea tava saavutamine sõltub veel kahest põhimõttelisest printsiibist:

- a) eksamite koostamise, läbiviimise ja valideerimise ressursside ratsionaalne planeerimine ja juhtimine;
- b) kõiki eksamiprotsessi aspekte puudutavate andmete/info kogumine, säilitamine ja kasutamine.

Kui vajalikke andmeid ei suudeta koguda, siis ei ole nõutavate standardite saavutamist ja nendest kinnipidamist võimalik *tõestada* (nt kõik, mis puudutab valiidsust, reliaablust, mõju ja praktilisust).

1.3.3. Hea tava järgivaid ja seega ka kõrge kvaliteediga eksameid on võimalik koostada ainult siis, kui kogu eksamiprotsessi juhtimiseks kasutatakse nõuetekohaseid tegevusi. Seetõttu on vajalik välja töötada otstarbekas eksami koostamise, läbiviimise ja valideerimise protsessi kirjeldus (mis sisaldab korduvate tsüklite mõistet).

- 1.3.4. Selle käsitlusviisi järgi peab esimene faas olema *planeerimine*, mis sisaldab üksikasjalikku olukorra analüüsi (st teostatavuse uuring), mis vaatleks uue eksami tunnetatud vajadust hariduskontekstis. Eesmärgiks oleks tuvastada need *kaalutlemist nõudvad asjaolud ja piirangud*, mis on eksami arendusprotsessis olulised ja mis määravad selle, kuidas eksami kasulikkus saavutatakse.
- 1.3.5. Kui otsustatakse uut eksamit koostama hakata, on tarvis saavutada kokkulepped neis protseduurides, mis puudutavad vähemalt järgmisi valdkondi:
- arendusprotsessi juhtimisstruktuur;
 - selge ja lõimitud rollide ja vastutuse jaotus;
 - projekti edukuse tagamiseks määratud tähtaegade ja ressursside monitoorimise mehhanism;
 - kasutatava eksami koostamisprotsessi juhtimise metodoloogia (st ülesannete koostamine, läbivaatamine, toimetamine, eeltestimine, ülesandepanga loomine, eksamitöö koostamine).
- 1.3.6. Kui eksamit tegelikult kasutama hakatakse, tuleb koguda andmeid materjalide koostamise ja eksami läbiviimise kohta, et otsustada, kas tegevused vastavad eksami maksumust ja efektiivsust puudutavatele praktilistele ootustele. Tsüklilised protsessid, mis järgnevad algele planeerimisele, on seotud pidevalt toimuva eksamikoostamise protsessi monitoorimisega ja eksami tegeliku läbiviimisega. Et kõiki tegevusi monitoorida, tuleb need hoolikalt dokumenteerida ja andmeid koguda. Monitoorimismetoodika hõlmab igat sorti andmekandjaid, mis aitavad luua eksami dokumenteeritud ajaloo; see omakorda võimaldab nii *protsess- kui ka kokkuvõtvat hindamist*. Kogutavad andmed võivad olla nii püsiaandmed (kogutud faktid ja arvud) kui ka muutuvad andmed (tunded, muljed, suhtumised jne).
- 1.3.7. Sellise lähenemise põhiseisukoht on, et valideerimine moodustab protsessi lahutamatu osa. Et seda saavutada, tuleb eksami pidevas koostamise ja läbiviimise protsessis kasutatavad protseduurid planeerida nii, et adekvaatsed andmed oleksid kogutud, salvestatud ja vajadusel kättesaadavad.
- 1.3.8. Põhimõtteks peab olema eksami *üldise kasulikkuse* maksimaalne suurendamine. See tähendab, et eksami ühe aspekti kohta kogutud andmestik on vältimatult oluline ka teiste aspektide jaoks. Näiteks andmeid, mis on kogutud eksami reliaabluse kohta, ei kasutata kitsalt ainult reliaabluse määramiseks, vaid ka selleks, et tegelda valiidsuse ja mõju küsimustega (nii nagu neid allpool defineeritakse).

1.4. Arusaam eksami kasulikkusest

- 1.4.1. Hea tava põhimõtted, mida siin esitletakse, on koostatud selleks, et tagada ALTE eksamite vastavus konkreetsetele kriteeriumidele, mis iseloomustavad järgmisi *eksami kvaliteedi näitajaid*:
- valiidsus
 - reliaablus
 - mõju
 - praktilisus

 - teenuse kvaliteet.

Võttes põhimõttekindlalt nimetatud aspekte arvesse, tagavad ALTE liikmed selle, et täidetakse tegevusjuhendis omaksvõetud kohustused ja säilitatakse piisavalt kõrge kvaliteeditase.

- 1.4.2. Nii *valiidsust* kui *reliablust* on mõõtmist ja keeletestimist puudutavas kirjanduses põhjalikult käsitletud. Näiteks *The Standards for educational and psychological testing* (nii 1985. kui ka 1999. aasta väljaanne) pakub ulatusliku ülevaate ja Bachmani 1990. aastal ilmunud raamat pühendab kummalegi mõistele terve peatüki. Teised kaks kvaliteedinäitajat on eksamikoostajatele alati olulise tähendusega olnud, kuid on keeletestimiskirjandusse ilmunud alles viimasel ajal. Praeguseks on jõutud ühisele arusaamisele, et ühtegi eksami kvaliteedinäitajat ei saa hinnata eraldiseisvana ning et iga näitaja suhtelist tähtsust tuleb määratleda selleks, et maksimaalselt suurendada eksami *üleüldist kasulikkust* (vt Bachman ja Palmer, 1996).
- 1.4.3. *Eksami kasulikkuse* mõiste nõuab, et igas hindamissituatsioonis tuleb leida nelja kvaliteedinäitaja – valiidsuse, reliaabluse, mõju ja praktilisuse – sobiv *tasakaal*. Lisaks sellele on ALTE liikmetel kui eksamite koostajatel, kelle teeneid kasutatakse kogu maailmas, tähtis silmas pidada osutatava *teenuse kvaliteeti*. See paneb ALTE liikmetele kohustuse vastutada kõige eest, mis puudutab nende eksamite kasutamist; see tähendab osutada eksami kasutajatele kvaliteetselt teenust, mis oleks vastavuses hea tava põhimõtetega, nagu need selles dokumendis on esitatud.
- 1.4.4. Kõik eksamid on seotud kindla kontekstiga ja see tähendab, et praktilisi *kaalutlemist nõudvaid asjaolusid ja piiranguid* tuleb eksami koostamisel ja selle läbiviimisel arvestada viisil, et igas olukorras saavutataks eksami kvaliteedinäitajate vahel nõutav ja antud olukorrale vastav tasakaal (arvestades nt hariduskonteksti, eksaminandide vajadusi ja eksami eesmärki). Eksami *üldise kasulikkuse* saavutamiseks tuleb määrata kvaliteedinäitajate vastastikune tähtsus. Head eksamid ei ole aga võimalik koostada ilma kõiki näitajaid piisavalt arvestamata.
- 1.4.5. Mõeldes sellele, millises kontekstis eksamit hakatakse koostama, on tarvis arvestada spetsiifilisi *kaalutlemist nõudvaid asjaolusid ja piiranguid*, mis seda olukorda iseloomustavad. Need ei ole kõigile ALTE eksamitele ühised, need näitavad, kas eksam on teostatav ning olemasolevate *ressurssidega* koostatav ja läbiviidav. Mis ressurssidesse puutub, siis mõeldakse siin nii *sisemisi* ressursse, mis on ALTE institutsioonide käsutuses, kui ka neid, mis on kättesaadavad selles kontekstis, kus ALTE eksameid läbi viiakse. Eriti tuleb kontrollida ja juhtida eksami koostamise ja läbiviimise kulusid.

2.0. Eksami omadused.

ALTE liikmed peavad saavutama hea tava nelja ülalnimetatud eksami kvaliteedinäitaja – valiidsuse, reliaabluse, mõju ja praktilisuse – suhtes.

2.1. Valiidsus

2.1.1. Valiidsuse aspektid

Valiidsust peetakse üldiselt kõige tähtsamaks eksami kvaliteedinäitajaks; see puudutab eksami sobivust ja tähenduslikkust mingis kindlas hariduskontekstis ning järeltule, mida eksami tulemuste põhjal tehakse. Valideerimisprotsess on tõendusmaterjali kogumine nende järeltule põhjendamiseks.

Valiidsus tähendab tavaliselt seda, mil määral suudetakse tõestada, et saadud testitulemused on eksaminandi keeleoskuse taseme õige peegeldus. Kuigi tänapäeval defineeritakse valiidsust kui *ühtset kategooriat* (vrd Messicki peatükki Linni raamatus 1988), oleks otstarbekas kirjeldada eksami valiidsust mitme teise kaasneva mõiste kaudu, millest lähtuvalt valiidsust on võimalik tõestada:

- konstruktivaliidsus – mil määral on testi tulemused kooskõlas testi aluseks oleva kommunikatiivse keeleoskuse mudeliga;
- sisuvaliidsus – mil määral katab test kogu teadmiste ja oskuste kogumit, mis on oluline ja kasulik tõsielusituatsioonis, ning autentset keelekasutust;
- kriteeriumide valiidsus (prognoosiv ja võrdlev valiidsus) – mil määral korreleeruvad testi tulemused mingi teise tunnustatud välise kriteeriumiga, mis sama teadmist või oskust mõõdab (nt mingi teise tasemesüsteemiga nagu näiteks ALTE raamdokument).

2.1.2. Konstruktivaliidsus

Konstruktivaliidsuse fookuses on eelkõige eksami punktitulemus või hinne kui mingi iseloomustav tunnus või „konstrukt“. Eksami koostaja defineerib mõõtmiseks oskuse iseloomustavad tunnused ja just need definitsioonid moodustavad konstrukti.

Keeletestimise kontekstis tähistab konstrukti *kommunikatiivse keeleoskuse mudel* (vrd Canale ja Swain 1980, Canale 1983, Bachman 1990).

Konstruktivaliidsuse kohta andmete kogumine eksami valiidsuse tagamiseks algab eksami koostamisega ja jätkub, kui eksam kasutusele võetakse ning kõiki tegevusjuhiseid järgitakse.

Konstrukti kohta tehtavate järelduste kinnitamine on seotud paljude mõõtmise aspektidega, nagu näiteks eksami vorm, läbiviimise tingimused või mõõdetav oskuse tase, mis kõik nõuavad suurt tähelepanu ja võivad mõjutada eksami tähendust ja tõlgendamist.

Formatted: Estonian

Konstruktivaliidsust peetakse testi valideerimisel sageli '*ühendavaks kategooriaks*', mis liidab sisu ja kriteeriume puudutavad mõisted (Messick 1980). Konstrukti valideerimine kui protsess tugineb paljudest allikatest pärinevatele tõendusvahenditele, mis võimaldavad konstrukti avada. Millist meetodikat konstrukti avamiseks vajalike tõendusvahendite kogumisel rakendada, sõltub kõnealusest valideerimise probleemist ja mõõdetavate konstruktide osatähtsusest. Erialakirjanduses on kasutatud suurt hulka statistilisi võtteid, millest enamik tugineb korrelatsioonil ja mis sageli kasutavad eksperimentaalseid andmekogumisevõtteid.

2.1.3. Sisuvaliidsus

Sisuvaliidus uurib, mil määral eksami üksiküsimuste ja ülesannete valim kajastab määratletud *sisu valdkonda*. See on seotud nii *asjakohasuse* kui ka *kaetusega*.

Valimi esindatust mõõtvaid meetodeid on palju; valdkonna peamisi tunnuseid võib kirjeldada *mudeli* abil (nt *Waystage* ja *Threshold* taseme eristuskiri) ning vastava ala spetsialistidel võib paluda eksamiküsimused nende tunnuste järgi kategooriatesse jagada; sel moel võib sisu esindatuse üle otsustada.

Selle sisuvaldkonna kirjeldamine, mida eksam peaks eeldatavalt katma, on ALTE eksamite jaoks väga tähtis. See, mil määral üksikküsimuste või ülesannete formaat esindab vastavat valdkonda, on määrava tähtsusega ning huvigruppide ja vastavate ekspertide kaasamine on testi koostamise protsessis võtmeküsimuseks.

Sageli võib kontrollolukordade detailide ja iseloomustavate joonte kindlaksmääramiseks kasutada nende „reaalses elus“ toimimise süstemaatilist vaatlemist (vrd Bachman ja Palmer *Target Language Use – TLU – domain*, 1996, lk 44–45). Selliseid vaatlusi võib valdkonna sisu esindavat valimit koostades siduda eksperthinnangutega.

Erilise tähelepanu pööramine testi sisu ja ülesannete *autentsusele* ning suhtele *sisendi* ja eksamil eeldatava vastuse e *väljundi* vahel on üks sisuvaliidsuse tähtsatest tunnustest. ALTE eksamite ülesannete ja materjalide autentsust võib pidada selle hindamismetoodika peamiseks tugevuseks. Eksami sisu peab olema kavandatud nii, et eksaminandi vastus sisendile (ülesannetele, üksikküsimustele jne) annaks piisava tõendusmaterjali kontrollitava oskuse (st konstrukti) kohta. Eksaminandi vastused testi sisendile on eksaminandi ja testi sisu vahelise vastastikmõju tulemus. Testi sisu autentsus ja eksaminandi vastastikmõju autentsus sellele sisule on kõrget valiidsust taotleva eksami koostaja jaoks tähtsad kaalumist nõudvad asjaolud.

(Vt Widdowson 1978, 1983 *olukorra ja vastastikmõju autentsus* ning Bachman ja Palmer 1996 nende käsitluste arvestamine keeletestides.)

Kokkuvõtvalt võib öelda, et sisuvaliidsus on seotud nii eksami koostamisega kui ka valiidsust kinnitavate tõendusandmete kogumisega pärast seda, kui eksami koostamisetapp on lõppenud ja eksam on kasutuses.

2.1.4. Kriteeriumide valiidsus

Kriteeriumide valiidsus on ALTE-le eriti tähtis, sest ALTE eksamid on *kriteeriumipõhised*. See tähendab, et ALTE raamdokumendi viis taset kajastavad teatud hulka kriteeriumide tasemeid, millega eksam on seotud. Selline lähenemine mõjutab nii teisi valiidsuse aspekte (nt testi sisu) kui ka reliaablusele hinnangu andmist ja reliaabluse kohta aruandmist.

Kriteeriumide valideerimise eesmärgiks on näidata, et testi punkt tulemused või eksami hinded on süsteemipäraselt seotud mingi välise kriteeriumi või kriteeriumidega (nt mingi muu antud oskust peegeldava indikaatoriga). Seetõttu on just kriteerium keskse tähtsusega. Seda kriteeriumit võib defineerida erineval viisil, nt gruppi kuuluvuse, mingi teise sama oskust testiva eksami tulemuse või samal oskusel põhineva reaalse elu ülesande lahendamise kaudu.

Viietasemeline süsteem on kõikidele ALTE eksamitele väliseks kriteeriumiks ja annab aluse nende lahtimõtestamiseks.

Erialakirjanduses käsitletakse kahte tüüpi kriteeriumide valiidsust– *võrdlev ja prognoosiv*.

- Võrdlev valideerimine kujutab endast info hankimist selle kohta, kui võrd võib eksami tulemusi kasutada, et määratleda või ennustada kriteeriumile vastavat käitumist. Kõige tavalisem info põhineb erinevate võrreldavate näitajate korrelatsioonidel (nt kahe erineva

üht ja sama oskust mõõtvat eksami tulemusel). Sooritustesti puhul (nt kirjutamisel ja kõnelemisel) võib kvalitatiivselt võrrelda kriteeriumi norme ja testi sooritajate tulemusi.

- Prognoosival valideerimisel on sarnane eesmärk, kuid infot kogutakse oletatava tulevase tulemuse suhtes, nagu näiteks tulevase eksami tulemused, toimetulek kõrghariduses või tulevasel töökohal. Sellise valiidsuse tõestamine on eriti oluline, kui eksami või testi tulemusi kasutatakse valiku tegemisel või tasemerühmade moodustamisel.

2.1.5. Valiidsuse tõendamisel peaks hea tava sisaldama vähemalt järgmist:

- | |
|---|
| <p>a) mõõdetava konstrukti jaksamiga kaetava sisuvaldkonna kirjeldus;</p> <p>b) eksamitulemuste kasutamine, sh kirjeldus selle kohta, kuidas andmed on sobivad nende põhjal tehtavate järelduste ning edasiste sammude tegemiseks;</p> <p>c) kasutatud valideerimismetoodika kirjeldus ning valideerimise tulemused, mille hulgas on</p> <ul style="list-style-type: none">• eksamisoorituste aluseks olevate protsesside loogiline ja kogemuslik analüüs;• eksamitulemuste ja teiste muutujate vaheline seos, k.a sellise variatiivsuse allikad, mis ei ole konstruktiivselt seotud;• kuidas eksami küsimused/ülesanded koostati ja kuidas need on seotud selle valdkonna teadmiste või oskustega, mille kohta järeldusi tegema hakatakse;• loogiline ja kogemuslik andmestik, mis kinnitaks eristava valiidsuse osatulemusi;• nende ekspertide arv ja kvalifikatsioon, kes tegid valideerimisprotsessi jaoks olulisi otsuseid;• protseduurid, mida valideerimisprotsessi käigus oluliste otsuste langetamisel kasutati;• põhjendus ja metoodika, mida rakendati eksami eristuskirja koostamisel (k.a kasutatud materjalide valik jne);• põhjendus ja metoodika kriteeriumi olulisuse määramisel;• kvantitatiivse andmestiku tõlgendamiseks vajalik teave. <p>d) uute valiidsusuuringute läbiviimine, kui on olulised muutused toimunud eksamis, selle läbiviimises, võimalike eksaminandide tunnusoontes või kontrollitava sisu valdkonnas;</p> <p>e) teabe edastamine eksami kasutajatele selleks, et aidata neil tõlgendada valiidsusuuringuid niivõrd, kui need puudutavad kavandatava eksami tulemusi, nagu näiteks positiivse/negatiivse sooritamise otsust, valiku tegemist või tasemerühmadesse paigutamist.</p> |
|---|

(cf. AERA/APA/NCME Standards, 1999, lk 17–24)

2.2. Reliaablus

2.2.1. Reliaabluse aspektid

Reliaablus on igasuguse mõõtmise põhimõiste ja aitab kaasa üldise valiidsuse saavutamisele.

Keeletestimises tähendab reliaablus seda, mil määral testi tulemused on *stabiilsed, järjepidevad ja mõõtmisvigadeta*. Ehk teisisõnu, mil määral võib eksami hinnete põhjal eksaminandi kohta otsuseid langetada. Reliaabluse näitajad peaksid arvestama mitte ainult olulisi vea allikaid, vaid ka seda, milliseid otsuseid eksami hinnete põhjal võidakse tegema hakata.

On terve hulk põhjusi, miks inimene võib saada erineva tulemuse, olles sooritanud ühe ja sama eksami kaks eeldatavalt analoogset varianti. Kui tulemuste erinevust ei ole võimalik põhjendada, on tegemist mõõtmisveaga. Mõõtmisvead vähendavad üksikisiku eksamitulemuste reliaablust ja seega ka nende üldistatavust.

Reliaablust määratakse ja näidatakse enamasti *reliaablusindeksi* kaudu. Kuna see on üldmõiste, siis viga, mida see näitab, sõltub konkreetsest arvutamismeetodist ja kuna kõik vea allikad ei ole iga eksami jaoks olulised, on eksami koostaja ülesandeks otsustada, mil moel vea variatiivsust näidatakse. See võib tähendada, et esitatakse standardviga, usaldusvahemik, sõltuvusindeksid jne.

Enamik keeletestide reliaabluse arvutamise käsitlustest põhineb üldisemal haridusliku ja psühholoogilise testimise meetodikal, nt *APA Standards* 1954. kuni 1985. aastani. Uue köite peatükkides "Reliaablus ja mõõtmisvead" (*Standards* 1999, I osa, II alajaotus), eristatakse kolme üldist reliaabluse kategooriat, mille nimetused erialakirjanduses harilikult on:

- a) alternatiivsete variantide indeks (arvutatakse erineval ajal läbiviidud analoogsete eksamivariantide tulemuste abil);
- b) kordustesti indeks;
- c) sisemise ühtsuse indeks.

Nendest kolmest on just seesmise ühtsuse indeks, nagu Cronbach'i alfa või KR20, tavaliselt kasutusel selleks, et objektiivsete testide reliaablust määratleda (nt valikvastustega lugemis- või kuulamistest). Neid indekseid on suhteliselt lihtne arvutada, mille tõttu teisi ja võib-olla kohasemaid näitajaid eriti sageli ei kasutata (nt kordustesti näitajaid kasutatakse vähem, sest piisavat infot on realselt raske saada).

Kõnelemis- ja kirjutamistestide kohta ütleb *APA Standards*, et kui testi hindamiseks on vaja eksamineerija või hindaja tegevust, siis on tähtis vaadelda reliaablust hindamisotsuste täpsuse ja järjepidevuse seisukohalt. Paljud ALTE eksamite kõnelemis- ja kirjutamistestid kuuluvad just sellesse kategooriasse, sest hinnanguid langetavad eksamineerijad.

Subjektiivse hindamise (hindajate) reliaablust määratakse tavaliselt korrelatsioonide kaudu, nt hindaja usaldusvärsuse ja hindamise usaldusvärsuse korrelatsioon.

2.2.2. Reliaabluse tõendamisel peaks hea tava sisaldama vähemalt järgmist:

- a) tõsiseid jõupingutusi, et kindlaks määrata ja kvantifitseerida peamised mõõtmisvea allikad, sh
- hindepaaride eeldatav reliaabluse tase konkreetses kontekstis (nt hinded, mida eksaminand on saanud kahe erineva ülesande eest, mida peetakse raskusastmelt võrdseiks);
 - tulemuste üldistatavus ülesannete ja küsimuste, sama eksami erinevate variantide, eksamineerijate, erinevate eksami läbiviimiskordade jne puhul;
- b) kinnitus, et eksamihinded, kaasa arvatud osatulemused ja hinnete kombinatsioonid, on nende võimaliku kasutuse seisukohast piisavalt usaldusväärsed;
- c) teabe esitamine reliaabluse, mõõtmise standardvea vm sarnase kohta, nii et eksami kasutajad saaksid otsustada, kas eksamihinded on nende võimaliku kasutuse seisukohast piisavalt usaldusväärsed;
- d) teabe esitamine kõrvalekallete ning muude võimalike vigade allikate kohta, mida peetakse tulemuste tõlgendamisel oluliseks;
- e) eksamihinnete reliaabluse või järjepidevuse näitajate esitamine meetodite abil, mis vastavad eksamihinnete olemusele ja võimalikule kasutamisele ning võtavad arvesse tulemuste tõlgendamisel oluliseks peetavate kõrvalekallete allikaid;
- f) reliaabluse analüüsi andmed, k.a
- eksamihinnete reliaabluse või järjepidevuse hindamise meetodite kirjeldus ning nende meetodite kasutamise põhjendus, reliaabluse analüüsis ilmsiks tulnud kõrvalekallete peamised allikad ning kasutatud valemid ja/või vastavad viited;
 - reliaablusindeks, üldine mõõtmisviga, klassifitseerimise järjepidevuse indeks või muu vastav teave eksami hinnete järjepidevuse kohta;
 - mõõtmise standardvead või muud hinde järjepidevuse näitajad nendest eksamil saadud hinnete aspektidest, mille põhjal tehakse järeldusi üksikisikute kohta;
 - erinevate sõltumatute hindamiste ühtelangevuse määr, kui kasutatakse subjektiivhindamist;
 - ühe ja sama eksami raames esitatud osatulemuste või kogumtesti hinnete korrelatsioon.
- g) nende tingimuste kirjeldus, millistes reliaablid tulemused saadi, k.a
- eksami populatsiooni kirjeldus, nt demograafilised andmed, haridustase, amet;
 - analüüsivalimi selektsiooni ja sobivuse kirjeldus, k.a vaatluste arv, vahendid, standardhälve analüüsivalimis ja igas muus grupis, mille kohta reliaablus määratakse;
 - kui on tegemist subjektiivhindamisega, siis hindamise alus, k.a hindajate valimine ja väljaõpe ning eksamitööde jaotamise põhimõte hindajate vahel ning hindamiserinevuste korrigeerimine;
 - erinevate testivariantide ajalise vahe ja rakendamise järjekorra põhjendus, kui kasutati alternatiivvariandi või kordustesti meetodit.

(vrd AERA/APA/NCME Standards, 1999, lk 31 – 36)

2.3. Mõju

2.3.1. Mõju aspektid

On üldiselt teada, et eksamiteenuse pakkujana on ALTE liikmetel oluline mõju haridusprotsessidele ja ühiskonnale tervikuna. See mõju avaldub vähemalt kahel tasandil:

- a) *makrotasandil*, mõjuna üldistele haridusprotsessidele;
- b) *mikrotasandil*, mõjuna üksikisikutele (huvigruppidele), keda eksamitulemused mõjutavad.

Üks üldise mõju valdkond puudutab ALTE rolli hindamisprotsesside ja sellega seotud pedagoogiliste probleemide üldisele mõistmisele kaasaaitamises nii Euroopas kui ka kogu maailmas. Seda tehakse avalikku teavet, uurimistöid ja nõustamisteenuseid pakkudes. Sihiks peaks olema testimise eesmärkide ja meetodite parem mõistmine ning eksami kohta oleva teabe (tulemused, hinded jne) õige kasutamine.

Mis üksikisikutesse puutub, siis peab tagama, et eksam oleks õiglane ja mitte tendentslik.

Positiivne mõju *õpetamisele ja õppimisele* on oluline aspekt, mis ilmneb nii makro- kui ka mikrotasandil. Just niisuguses kontekstis käsitletakse selliseid mõisteid nagu *näivvaliidsus* (testi meeldivus) ja *tagasimõju*. On oluline uurida, milline mõju on eksamil selles kontekstis, milles neid kasutatakse. Eksamiarendajad peavad tegutsema printsiibil, et nende eksamil ei oleks negatiivne, vaid oleks võimalikult positiivne mõju.

2.3.2. Mõju tõendamisel peaks hea tava sisaldama vähemalt järgmist:

- a) eksami eristuskirjade ja detailse ainekava koostamine ja esitamine;
- b) professionaalsete tugiprogrammide pakkumine institutsioonidele ja igale õpetajale/õpilasele, kes eksamit kasutab.

Positiivset hariduslikku mõju on võimalik saavutada ka järgmiste tegevuste kaudu:

- sobivate ekspertide kaasamine kõikidel eksamiarenduse etappidel;
- eksamiülesannete koostajateks sobivate ekspertide väljaõpetamine ja rakendamine;
- eksamineerijateks sobivate ekspertide väljaõpetamine ja töölevõtmine.

Kui eksam võetakse kasutusele, tuleb kehtestada kindlad protseduurid, mille abil saaks koguda teavet, mis lubaks eksami mõju hinnata.

See tähendaks järgneva teabe kogumist :

- kes eksamit sooritab (s.o eksamisooritajate kirjeldus);
- kes eksami tulemusi kasutab ja mis eesmärgil; kes eksamisooritajaid eksamiks ette valmistab ja millistes tingimustes;
- milliseid kursusi ja materjale eksamiks ettevalmistamiseks koostatakse ja kasutatakse;
- millist mõju eksam üldiselt avalikkusele (nt haridusstandardile) avaldab;
- kuidas suhtuvad eksamisse need, kes on haridussüsteemi kaudu eksamitega otseselt seotud (nt õpilased, eksaminandid, õpetajad, lapsevanemad, jne);
- kuidas suhtuvad eksamisse need ühiskonnaliikmed, kes asuvad väljaspool haridussüsteemi (nt poliitikud, ettevõtjad jne).

Selline teave peaks organisatsiooni piires kõigile ALTE liikmetele kättesaadav olema, seda nt aruannetena. Nende aruannete kohandatud versioone peaks olema võimalik kasutada ka teistel huvigruppidel.

Kogutud teabe põhjal peab olema võimalik näidata, et eksam on piisavalt valideeritud ja reliaabel selles kontekstis, kus seda kasutatakse. See juba iseenesest aitab positiivset mõju saavutada.

2.4. Praktilisus

2.4.1. Praktilisuse aspektid

Praktilisus on teksti kasulikkuse mõiste lahutamatu koostisosa ja mõjutab eksami erinevaid aspekte.

Praktilisust saab defineerida kui ressursside mahtu, mis on vajalik eksami koostamiseks ja läbiviimiseks mingis kindlas kontekstis.

Iga eksami *praktilisus* hõlmab kaht tegurit:

- a) ressursid, mis on vajalikud sellise eksami koostamiseks, mille kvaliteedinäitajad (valiidsus, reliaablus ja mõju) on õiges tasakaalus;
- b) olemasolevad ressursid.

Praktiline eksam ei esita olemasolevatele ressurssidele põhjendamatuid nõudmisi. Kui olemasolev ressursimaht ületatakse, ei ole eksam praktiline. Sellisel juhul peab eksami koostaja kas muutma eksami vormi või põhjendama sellele eraldatavate ressursside suurendamise vajalikkust. Kui teine lahendusvariant praktilistel kaalutlustel võimatuks osutub, tuleb valida esimene lahendusvariant.

Enne eksamiarendusprotsessi jätkamist tuleb otsustada, kas eksam on praktiline ka juhul, kui muudetakse eksami eristuskirja. Kui mitte, siis ei peaks tööd jätkama.

2.4.2. Praktilisuse tagamise hea tava peaks sisaldama järgmist:

Head tava võib saavutada ainult juhul, kui *kogu eksamiarendusprotsessi juhitakse* nõuetele vastavate tegevuste kaudu, k.a eksami koostamine, läbiviimine ja valideerimine.

Praktiliste eksamite koostamine nõuab, et kasutusele võetaks *üksikasjalik eksamiarendusprotsessi mudel* – vt nt *User Guide for Examiners*.

Alati kui töötatakse välja uut või uuendatakse olemasolevat eksamit, tuleb järgida kindlaid protseduurireegleid, mis suunaks arendusprotsessi juhtimistegevust ning võimaldaks ülesandeid ja vastutust selgelt ja komplekselt jagada. Peavad olema loodud kontrollimehhanismid, milles on ära näidatud ajagraafikud ja ressursid, ning samuti meetodika selleks, et juhtida eksami koostamise protsessi siis, kui eksamit tegelikult koostama hakatakse (ülesannete koostamine, läbivaatamine, toimetamine, eeltestimine, ülesandepanga loomine, eksamitööde kokkupanemine).

Eksamiarendusprotsess peab algama *eksami praktilisuse analüüsiga*, mis peaks sisaldama vähemalt järgmist:

- uue eksami eesmärk
- raskusaste eeldatavate eksaminandide jaoks (nt ALTE raamdokumendile toetudes)
- välised tegurid
 - nõudlus
 - konkurents – teised samalaadsed kasutatavadksamid
 - ühiskonna esitatud nõudmised või tingimused (nt lapsevanemad, ministeerium jne)
- sisemised olulised tegurid
 - eksami konstrukti ja sisu puudutavad teooriad
 - tehnoloogilised saavutused
 - institutsionaalselt kehtestatud parameetrid
- uue eksami arvatav olulisus ja vastuvõetavus eeldatavatele eksaminandidele ja eksami kasutajatele;
- eksami osalustasu
- olemasolevad ressursid
 - eksami koostamiseks
 - läbiviimiseks
 - tulemuste esitamiseks
 - samasuguse eksami koostamiseks ja läbiviimiseks tulevikus.

Eksami kasulikkuse määratlemine on nii tsükliline kui ka korduv. *Praktilisuse* probleemid mõjutavad otsuseid kõikides eksamiarendamisefaasides. Eksami läbiviimisega seotud küsimusi arvesse võttes on tähtis läbi mõelda, mil määral on olemasolevate ressursidega (nt inimesed, tehnilised vahendid, aeg ja raha) sellise tasakaalu saavutamine (eksami koostamise protsessi erinevate aspektide vahel) võimalik.

2.5. Teenuse kvaliteet (täiendatakse koosolekul toimuvate arutelude järgi).

2.5.1. Teenuse kvaliteedi aspektid

Teenuse kvaliteet puudutab eksami koostaja võimet täita eksaminandide ja eksami kasutajate ees teatud kohustusi.

Sia hulka kuuluvad turvaliste eksamimaterjalide koostamine, eksami andmete ja tulemuste konfidentsiaalsus ning samuti eksamitulemuste kohta tekkinud küsimuste ja vaiete lahendamine.

2.5.2. Teenuse kvaliteedi tagamise hea tava peaks sisaldama järgmist:

- Uurimistulemuste kättesaadavaks tegemine ja uurimustulemuste vastastikune retsenseerimine.
- Üksikisikuid ja institutsioone puudutava teabe nõudmine ainult sellisel juhul, kui see on neile arvatavalt kasulik selle kaudu, et uurimistööd eksamiarenduse seisukohalt saaks jätkata ja tulemust parandada. Sellise teabe kogumise eesmärk peaks olema selgitatud kõigile, keda see puudutab.
- ALTE liikmete käsutuses olevate ja üksikisikuid või institutsioone puudutavate nii toor- kui töödeldud andmete konfidentsiaalsuse hoidmine. Samuti peab julgustama kõiki gruppe või institutsioone, kellelt andmeid saadakse või kellele neid antakse, sama reeglit järgima.

- Realistlike tähtaegadega üleandmiskohustuse võtmine ja sellest kohustusest kinnipidamine.
- Pädeva kvaliteedikontrolli kasutamine, et tagada kõigi ALTE liikmete pakutud toodete ja teenuste korrektsus ja üleandmine lubatud aja piires.
- Kohustuse võtmine teavitada neid, keda teabe avaldamise järel mõjutavad selles ilmnunud vead.
- Nende osapoolte teavitamine, kelle tegevust mõjutavad tõsised ajagraafikust kõrvalekaldumised.
- Eksamiküsimuste ja nendega seotud materjalide ülevaatamine ja toimetamine, et vältida võimalikku ebakorrektsust või keelt.
- Puuetega eksamisooritajatele kohandatud eksamivariantide koostamine või eksami läbiviimise protseduuri kohandamine, kui võimalik.
- Eksaminandidele vajaliku teabe andmine eksami sisu, küsimustüüpide, tööjuhiste ja vajalike eksamisooritamise strateegiate kohta. Püüda saavutada võrdne informatsiooni kättesaadavus kõikidele kandidaatidele.
- Eksaminandide teavitamine sellest, kui kaua nende eksamitulemusi andmebaasis säilitatakse ning kellele ja millistel tingimustel eksamitulemusi väljastatakse.
- Kaebuse või vaide esitamise ning lahendamise protseduuride kirjeldamine.

Allikad ja viited

AERA/APA/NCME (1985/1999). *Standards for educational and psychological testing*. Washington: AERA;

Formatted: Finnish

ALTE (1994) Code of Practice;

ALTE/Council of Europe (1997) *Users Guide for Examiners*;

ALTE (1998) *Handbook of Language Examinations and Examination Systems*;

Bachman L.F. (1990). *Fundamental considerations in language testing*. Oxford: OUP;

Bachman, L. F. and Palmer, A. (1996). *Language Testing in Practice*. Oxford: OUP;

Canale, M. (1983) *On some dimensions of language proficiency*, in Oller 1993;

Canale, M and Swain, M. (1980) *Theoretical bases of communicative approaches to second language teaching and testing*. *Applied Linguistics* 1,1;

Council of Europe:

Common European Framework of Reference (1997/2001),

Threshold Level (1990),

Waystage Level (1990);

ETS (1981, 1987, 2000) *Standards for Quality and Fairness*;

ILTA (1995) *Task Force on Testing Standards*;

ILTA (2000) *Code of Ethics*;

Linn, R.L. ed. (1989) *Educational Measurement*. Third edition. New York: ACE/Macmillan;

Messick, S.A. (1980) *Test validity and the ethics of assessment*. *American Psychologist*, 35;

Messick, S.A. (1989) *Validity*, in Linn 1989;

Saville, N. and Milanovic, M. (1991, 1993), *Principles of Good Practice for ALTE Examinations*, Unpublished documents;

Taylor, L. (1999) *Constituency matters*. Paper presented at Language Testing Forum, Edinburgh, November 1999;

Washington D.C. Joint Committee on Testing Practices (1988) *The Code of Fair Testing Practices in Education*;

Widdowson, H.G. (1978) *Teaching Language as Communication*. Oxford: OUP;

Widdowson, H.G. (1983) *Language Purpose and Language Use*. Oxford: OUP.