

ARVUTITESTI VALIIDSUSE KONTROLLIMINE

Els Tasa, Heiti Pakk

Artikkel käsitleb vastloodud automaatse universaalse testi-missüsteemi psühholoogilise valiidsuse probleeme. Uurimustest selgub, et arvutil läbiviidava Eysencki ENL testi tulemused ei erine traditsioonilisest paberi-pliatsi-variantist. Katseisikute suhtumine arvutitestimisse on oluliselt positiivsem kui paberivariandisse. Lisaks ilmnes, et arvuti kasutamine vähendab tunduvalt testimise aega, rääkimata skoorimisest ja töötlusest.

Ülevaade probleemist

Tänapäeval kasutatakse pliatsi-paberitestide kõrval üsna rohkesti arvutiteste. Tuleb arvestada, et arvutitestimise tulemusi võivad mõjutada spetsiifilised faktorid. Nii ei ole sarnased kiirustesti arvuti- ja pabervariandid, tulemused arvutivariandil sõltuvad suurel määral trükkimisoskusest arvuti klaviatuuril. Seega testi arvutisse viimisel kaasnevad testi vormi muutused võivad oluliselt muuta testitulemusi. Teiseks variatiivsuse allikaks on testitav. Arvutikogemuse puudumine, arvutiga seotud ärevus või arvutikartus võivad mõjutada testitulemusi.

Seega võib arvuti- ja pabertesti valiidsus erineda ning vajalikuks osutub eri normide väljatöötamine kummalegi variandile (Moreland, 1985b). Järgnevalt neist probleemidest pisut lähemalt.

Arvutitestimise erinevus avaldub juba küsimuste esitusviisis. Arvuti esitab tavaliselt ühe või paar küsimust korraga. Pabervariandi puhul on aga võimalik vaadata eelnevaid küsimusi ja võrrelda vastuseid küsimustikku edasi-tagasi sirvides. Nii viisi kujuneb ühtne vastustemudel. Samuti põhjustab see vastamisel kiirustamist ja vahelejätmissi. Arvuti võib tõsta tähelepanelikkust iga küsimuse suhtes. Nii võib arvutitestimine olla psühhomeetriselt *parem* traditsioonilisest testimisviisist.

Sama testi arvuti- ja pabervariantidel on leitud olulisi lahknevusi olenevalt vastamise viisist. Tavaliselt tuleb arvutitestis aktiivselt reageerida igale küsimusele. Isegi siis, kui katseisik soovib vastamata jätta, peab ta ikka nupule vajutama. Pabervariandis (näiteks MMPI puhul) ei tehta vastamata küsimuse kohale mingit märget — küsimus jäetakse passiivselt vahele. Ja nii antakségi arvutil MMPI-le rohkem (?) vastuseid kui paberil [Briskin, Kolotkin, 1977].

Teiseks on erinevused tingitud sellest, et inimesed võivad an-

da arvutile rohkem ja täpsemat intiimset infot kui inimesest intervjueerijale või trükitud küsimustikule vastates. On leitud [Skinner, Pakula, 1986], et alkoholismihaiged eelistavad vastata arvutile, kui teemad lähevad intiimsemaks. Mõned uurimused on näidanud, et arvutiintervjuu puhul teatatakse kasutatavad alkoholiannused ligi 30 % suuremad ning mainitakse rohkem sotsiaalselt ebasoovitavaid tegusid [Moreland, 1985b]. Samuti usaldasid naised oma seksuaalprobleeme rohkem arvutile kui samast soost psühhiaatritele [Erdman jt., 1985]. A.C. Carr ja A. Ghosh (1983) leidsid, et arvuti psühhiaatriline intervjuu kogus keskmiselt 5.4 ravi jaoks olulist vastust rohkem kui psühhiaater just sellistest valdkondadest nagu kriminaalne karistus, joomine, impotentsus, töölt vallandamine ning enesetapukatse.

Kuigi võiks arvata, et paberile trükitud küsimustik pakub sama privaatsuse ja standardiseerituse kui arvuti, on kaks uurimust kindlalt tõestanud vastupidist. Samas on aga mitmeid vasturääkivate tulemustega uuringuid. Ühes uurimuses alkohoolikutega ja mõnes üldise populatsiooniga ei olnud mingit olulist vahet paberil, arvutil või inimese poolt läbiviidavate intervjuude vahel. Oletati, et uuringu eesmärk ja uuritava populatsioon on olulised faktorid, mis võivad mõjutada arvutiuringute valiidsust [Moreland, 1985a].

Põhiliseks psühholoogilise uuringu valiidsuse ohuks võib pida vastuste suundumust, mis on tingitud sotsiaalsest kontrollist ja meeldida püüdmisest. Kui oletada, et arvutitestimine tõstab uuritava usaldustunnet, võib ta anda vähem suundumuslikke vastuseid. W.M. Evans ja J.R. Miller [1969] ning Stockwell ja Jackson [1983] võrdlesid arvutil ja paberil läbiviidud testide kõrgelt personaalseid ja neutraalseid küsimusi ning leidsid väiksed erinevused skaalade keskmistes, kuigi faktorstruktuurid mõlema variandi puhul olid samad. Selle põhjal oletatigi, et arvuti suudab üleva hoida KI-de suuremat huvitatust, seega antakse vähem suundumuslikke vastuseid. Nii aga võivad paberitesti normid osutada arvutitesti jaoks sobimatuks.

Võib väita, et arvutitest paneb need inimgrupid, kellel puudub arvutikogemus, ebasoodsamasse olukorda kui arvutikogemusega inimesed. Üheks selliseks inimgrupiks peaksid olema vanemad inimesed. D.F. Johnson ja C.B. White [1980] leidsid, et vanemad inimesed, kellel oli võimalus enne testimist ühe tunni ulatuses arvutil treenida, said kõrgemad skoorid *Wonderlic Personnel Inventory*'ga kui treenimata testitavad. Sellest järeldati, et algne arvutikartus on lühiajaline ja kaob enamikul peale vajalikku treeningut.

Üheks probleemiks arvutitestimise ja intervjueerimise juures on see, et arvuti öeldakse olevat impersonaalne, ebainimlik ning uuritavate jaoks võõristust tekitav. Paljude uurijate ühine seisukoht on, et uuritavad suhtuvad arvutisse vägagi positiivselt, kuigi peaaegu alati on inimesi, kes ei soovi arvutit kasutada [M.J. Bresolin, 1984]. R.W. Lucas jt. [1977] ja A.C. Carr jt. [1983] leidsid, et üle 80 % uuritava-

test pidas arvutiintervjuud sama lihtsaks kui psühhiaatri intervjuud. J.H. Greist jt. [1973] uurisid intervjuuga enesetapuriski ja leidsid, et pooled eelistasid arvutiintervjuud. D.M. White [1983] leidis, et 80 % uuritud kolledži õpilastest eelistas MMPI arvutivarianti ja mitte ükski pabervarianti. A.C. Carr ja A. Gosh [1983] leidsid, et isegi foobiatega patsiendid ei ilmutanud mingit vastumeelsust arvutil käitumusliku küsimustiku täitmise suhtes.

Üldiselt on uuritavad hinnanud arvutiga suhtlemist vähemalt sama kergeks kui teisi variante [Bresolin, 1984; Skinner, Allen, 1983], kuigi arvuti oli algul rohkem ärevust tekitav kui trükitud küsimustik [Lushene jt., 1974]. H.A. Skinner jt. [1985] leidsid, et kui uuritav teeb arvutitesti läbi, muudab see tema eelistusi arvuti kasuks. Samuti leidis Newsted [1985], et arvutil testinud eelistasid rohkem arvutit kui paberil testinud paberit. Järelikult peaks arvutikogemuse suurenedes suhtumine arvutitestimisse veelgi paranema.

On kontrollitud, kas arvuti eelistamine on seotud põhiliselt tema uudsusega. Pärast 4 – 8 tundi kestnud arvutiküsimustiku täitmist oli 82 % uuritavatest valmis teist korda need tunnid arvutiga veetma [Angle jt., 1978]. Siin ei saa ilmselt põhjus olla ainult uudsuses: 8 tunniga peaks see faktor oma mõju kaotanud olema.

Tegelikult tulenevad paljud hädad suuresti ka sellest, et arvutikartust on märksa rohkem professionaalide (arstid, psühholoogid) kui uuritavate juures. Arstid, kellel endal puudub arvutikogemus, väidavad palju sagedamini, et patsiendid ei nõustu arvutitestimisega [Fowler, 1985].

Vaatamata eespool toodud kahtlustele, on enamik uurijaid näidanud, et mitteekvivalentsus isiksusetestide eri variantide vahel on tüüpiliselt väike (kui seda üldse esineb), et olla praktiliste tagajärgedega. Paljud on tõestanud oma arvutivariantide valiidsust: L. Katz ja J.T. Dalby [1981] Eysencki isiksuseküsimustiku, B. Briskin ja R.L. Kolotkin [1977] MMPI ning J.R. Miller [1986] oma Rorschachi kompuuterversioonide ekvivalentsust pabervariantidega.

H.J. Eysencki ENL testi paber- ja arvutivariandi võrdlus

Laiemalt oli antud töö eesmärgiks proovida uut universaalset automaattestimissüsteemi. Universaalne automaattestimissüsteem "TESTER" on esimene sellelaadne katsetus Eestis. Programmid valmisid Els Tasa ja Jüri Alliku koostööna ega ole oma praeguses seisus midagi lõpetatud. Arvutil läbiviimiseks ja skoorimiseks on ette valmistatud 1 test: Eysencki ENL vorm A eestindatud variant [Toim, 1983].

Teiseks eesmärgiks oli võrrelda katseisikute suhtumist paberpliatsi- ning arvutitestimisse (kumba varianti ja miks nad eelistavad ning kuidas kumbagi hindavad).

Töö viidi läbi 100 katseisikuga (47 meest ja 53 naist), kelleks olid TÜ psühholoogia kateedri töötajad, bioloogia eriala III kursuse statsionaarsed ja psühholoogia kaugõppe VI kursuse üliõpilased.

Katseisikud jaotati kahte võrdsesse gruppi: üks täitis enne testi pabervariandi ja siis vähemalt ühe nädala möödudes arvutivariandi, teine grupp täitis testid vastupidises järjekorras.

Arvutivariant viidi läbi arvutil DVK-2M. Pärast kummagi variandi lõpetamist täitsid kõik katseisikud ankeedi, mis sisaldas 5-semantilise diferentsiaali 7-pallist skaalat, millel tuli hinnata antud katseprotseduuri:

huvitav	-	tüütu
kerge	-	väsitav
nürimeelne	-	taibukas
meeldiv	-	ebameeldiv
mugav	-	vaevanõudev

Pärast mõlema testivariandi lõpetamist täideti ka teine ankeet, kus paluti väljendada ja põhjendada oma eelistust arvuti- või pabervariandi suhtes, kummagi variandi miinused ja plussid välja tuua, ning teha ettepanekuid testiprotseduuride parandamiseks.

Pabervariant täideti traditsioonilisel viisil: katseisikutele jaotati kätte küsimused, katsekorraldaja luges ette instruksiooni ja katseisikud märkisid oma vastused paberile.

Arvutivariandi puhul sisestas katsekorraldaja arvutisse katseisiku nime, sünniaasta, hariduse, ameti ja perekonnaseisu. Siis luges katseisik arvuti kuvarilt instruksiooni ning katsekorraldaja näitas täpselt 3 vastamiseks vajamineva nupu asukohta. Kui KI oli valmis, hakkas arvuti küsimusi esitama: kuvaril oli korraga üks küsimus, küsimuse järjekorranumber ning vastusevariandid koos nupu tähistega, mida tuli vastamisel vajutada (selles testis EI - 1, JAH - 2). Veel oli kuvaril näidatud eelmise küsimuse tagasikutsumise võimalus (vajutus klahvile "E"). Võttes tagasi eelmise küsimuse, oli võimalik sellele uus vastus anda. Kui küsimusele oli vastus antud, ilmus kuvarile järgmine küsimus. Kui kõik testi küsimused olid vastatud, salvestas arvuti automaatselt vastused andmefaili. Nii arvutil kui paberil saadud testitulemusi (E-, N-, L-skaalade väärtusi) võrreldi omavahel. Dispersioonanalüüs antud andmetega näitas, et arvuti- ja pabervariandi skaalad ei erine oluliselt üksteisest.

	Keskmised			
	paber	arvuti	F	P
	\bar{x}	\bar{x}		
L-skaala	3.56	3.55	.008	.93
E-skaala	12.04	11.98	.134	.72
N-skaala	10.62	10.44	1.026	.31

Dispersioonanalüüsis ei leitud olulisi erinevusi ka kahe testigrupi vahel.

Lisaks võrreldi eraldi igale testi küsimusele antud vastuseid arvuti- ja pabervariandil. Igale üksikule küsimusele 57-st võis vastata kummagi variandi puhul "JAH" või "EI". Teisel testimisel võisid need vastused esimese vastustega kokku langeda või mitte. Kokkulangevuse olulisuse määra kontrolliti McNemari muutuste olulisuse testiga [van Tassel, 1981], selgus, et mittekokkulangevused ei ole olulised ($p < 0.05$).

Nende tulemuste põhjal võib öelda, et Eysencki ENL testi arvuti- ja pabervariant on ekvivalentsed.

Testimisprotseduuri hinnangud

Kõigi katseisikute hinnanguid arvesse võttes ilmnes, et arvutiprotseduuri hinnatakse (Studenti T-kriteeriumi järgi) oluliselt huvitavamaks ($t = 1.971$, $p < 0.05$), kergemaks ($t = 1.920$, $p < 0.05$), taibukamaks ($t = 1.955$, $p < 0.05$), meeldivamaks ($t = 3.308$, $p < 0.001$) ning mugavamaks ($t = 6.121$, $p < 0.0005$) kui testimist paberil.

Gruppe eraldi vaadates ilmnesid hinnangutes erinevused. Esimene grupp, kes täitis enne pabervariandi, hindas arvutit oluliselt paremaks ainult mugavuse poolest ($t = 3.145$, $p < 0.005$). Teine grupp täitis enne arvutivariandi ning hindas seda märgatavalt huvitavamaks, taibukamaks, meeldivamaks ning mugavamaks kui pabervarianti. Seega need, kellel oli eelnev arvutitestimise kogemus, hindasid pabervarianti oluliselt halvemaks arvutivariandist.

Teine ankeet täideti, kui mõlemad testivariandid olid tehtud. 100 katseisikust eelistas arvutivarianti 93 ja pabervarianti 7 inimest, 3 esimesest ja 4 teisest testigrupist. Põhjuseks, miks valiti pabervariant, võib pidada arvutivõõristust — "*raske keskenduda, vastused panin kirja automaatselt*", "*paber on tuttavam, mõtted ei kandu mujale*", "*võib kogemata valele nupule vajutada*".

Enamik eelistas arvutivarianti, tuues põhjusena esile mugavust, lihtsust, huvitavust ja kiirust. Hinnati ka seda, et küsimused on ükshaaval loetavad — kogu tähelepanu saab pöörata küsimuse sisule, parem kontsentreerumise võimalus, vastused tulevad siiramad, impulsiivsemad. Samuti arvati, et arvutil tuleb vähem vigu, kuna ei pea pingsalt jälgima küsimuste järjekorda. Negatiivseks küljeks arvuti juures peeti müra ning soovitati parandada klahvide asetust.

Pabervariandi juures hinnati võimalust küsimustikus edasitagasi sirvida ning seega ennast kontrollida ja vastuseid parandada. Ainult üks vastaja pidas pabervariandi juures oluliseks vahetut kontakti testijaga. Pabervariandi miinuseks loeti seda, et pikk küsimustik sunnib tihtipeale asjatult kiirustama ja mõtlematult vastama.

Kokkuvõte

Saadud tulemuste põhjal võib öelda, et olemasoleval kujul [K. Toim, 1983] on Eysencki ENL testi arvuti- ja pabervariant võrdväärseid skaalade väärtuste põhjal kui ka eraldi testi küsimustele antud vastuseid võrreldes.

Uurimusest ilmnas aga, et mitmedki testi küsimused ei tööta korralikult: vastused on valdavalt ühesuunalised ning mõne korrelatsioon oma skaalaga liiga väike. Seega oleks vaja kontrollida testi valiidsust küsimuste kaupa esinduslikuma valimi peal ning viia uuesti läbi korralik standardiseerimise protseduur. See töö jääb aga väljapoole meie artikli eesmärke.

Katseisikud hindasid testimisprotseduuri arvutil paremaks kui paberil. Eriti ilmnas see tendents teisel katsegrupil, kes täitsid enne arvutivariandi. Seega mõjutab arvutikogemus suhtumist arvutitestimisse. Ja seda positiivses suunas.

KIRJANDUS

- Angle, H.V., Ellinwood, E.H., Carroll, J. Computer interview problem assessment of psychiatric patients // Orthner, F.H. Proceedings: The Second Annual Symposium on Computer Application in Medical Care. Washington. DC: Institute of Electrical and Electronics Engineers, 1978. P. 137 - 148.
- Bresolin, M.J. A comparative study of computer administration of the MMPI in an inpatient psychiatric setting: Unpublished doctoral dissertation. Loyola University of Chicago, 1984.
- Briskin, B., Kolotkin, R.L. Effects of computerized administration on scores of the MMPI // Applied Psychological Measurement, 1. 1977. P. 543 - 549.
- Carr, A.C., Ghosh, A., Ancill, R.J. Can a computer take a psychiatric history? // Psychol. Med. 1983. Vol. 13. P. 151 - 158.
- Carr, A.C., Ghosh, A. Accuracy of behavioral assessment by computer // Brit. J. Psychiatry. 1983. Vol. 142. P. 66 - 70.
- Erdman, H.P., Klein, M.H., Geist, J.H. Direct patient computer interviewing // J. Consult. Clin. Psychol. 1985. Vol. 6. P. 760 - 773.
- Evans, W.M., Miller, J.R. Differential effects on response bias of computer vs conventional administration of a social science questionnaire: An exploratory methodological experiment // Behav. Sci. 1969. Vol. 14. P. 216 - 227.
- Fowler, R.D. Landmarks in computer-assisted psychological assessment // J. Consult. Clin. Psychol. 1985. Vol. 6. P. 748 - 759.
- Geist, J.H., Van Cura, L.J., Kneppreth, N.P. A computer interview for emergency room patients // Computer Biomed. Res. 1973. Vol. 6. P. 257 - 265.
- Johnson, D.F., White, C.B. Effects of training on computerized test performance in the elderly // J. Appl. Psychol. 1980. Vol. 65. P. 357 - 358.

- Katz, L., Dalby, J.T. Computer and manual administration of the Eysenck Personality Inventory // *J. Clin. Psychol.* 1981. Vol. 37. P. 586 – 588.
- Lucas, R.W., Mullin, P.J., Luna, C.B.X., McInroy, D.C. Psychiatrists and a computer as interrogators of patients with alcohol-related illnesses: A comparison // *Brit. J. Psychiatry.* 1977. Vol. 131. P. 160 – 167.
- Lushene, R.E., O'Neal, O.F., Dunn, T. Equivalent validity of a completely computerized MMPI // *J. Person. Ass.* 1974. Vol. 38. P. 353 – 361.
- Moreland, K.L. Validation of computer-based test interpretations: Problems and prospects // *J. Consult. Clin. Psychol.* 1985a. Vol. 6. P. 816 – 825.
- Moreland, K.L. Computer-assisted psychological assessment in 1986: A practical guide // *Computers in Human Behavior.* 1985b. Vol. 1. P. 221 – 223.
- Newsted, P.R. Paper versus online presentations of subjective questionnaires // *Man-Machine Studies.* 1985. Vol. 23. P. 231 – 247.
- Skinner, H.A., Allen, B.A. Does the computer make a difference? Computerized versus face-to-face versus self-report assessment of alcohol, drug, and tobacco use // *J. Consult. Clin. Psychol.* 1983. Vol. 51. P. 267 – 275.
- Skinner, H.A., Allen, B.A., McIntosh, M.C., Palmer, W.H. Lifestyle assessment: Applying microcomputers in family practice // *Brit. Med. J.* 1985. Vol. 290. P. 212 – 214.
- Skinner, H.A., Pakula, A. Challenge of computers in psychological assessment // *Professional Psychology: Research and Practice.* 1986. Vol. 1. P. 44 – 50.
- Toim, K. Isiksuse psühodiagnostika. Tartu: TRÜ, 1983.
- Van Tassel, D. Basic-Pack Statistic Programs for Small Computers. New Jersey: Prentice-Hall. Inc., 1981.
- White, D.M. An assessment of the comparability of a computer administration and standard administration of the MMPI: Unpublished master's thesis. University of Alabama, 1983.

ПРОВЕРКА ВАЛИДНОСТИ ТЕСТОВ НА КОМПЬЮТЕРНОЙ СИСТЕМЕ

Э. Таса, Х. Пакк
Резюме

Разработана универсальная компьютерная система тестирования. Работа посвящена психологической валидации этой системы. Показано, что результаты теста Айзенка ЭНЛ не отличаются от традиционного бумажного варианта теста.

Отношение испытуемых к компьютерному варианту теста значительно более положительное, чем к бумажному. Выполнение компьютерного варианта занимает существенно меньше времени, не говоря уже о времени, уходящем на обработку данных.

THE PSYCHOLOGICAL VALIDITY OF THE UNIVERSAL COMPUTERIZED TESTING SYSTEM

E. Tasa, H. Pakk

Summary

The present article deals with the psychological validity of the universal computerized testing system. It is shown that the results of the computerized Eysenck ENL Test are not different from the traditional paper-pencil variant. The subjects' attitude towards the computerized variant was significantly more positive than towards the traditional one. In addition, the computerized system significantly reduces the time of the testing procedure, to say nothing of scoring.