

RECENZIJA KNJIGE "ZANESLJIVOST IN VELJAVNOST MERJENJA"

Zdenko Lapajne

**ANUŠKA FERLIGOJ, KARMEN LESKOŠEK IN TINA KOGOVŠEK (1995).
ZANESLJIVOST IN VELJAVNOST MERJENJA. LJUBLJANA: FAKULTETA ZA
DRUŽBENE VEDE (METODOLOŠKI ZVEZKI 11). ISBN 86-80227-47-1**

Kvaliteta merjenja ter skladnost teorije in stvarnosti sta osrednji vprašanji sleherne empirično utemeljene znanosti, ki ne zanimata več samo psihologov, temveč ju zlasti v zadnjih desetletjih vedno bolj poglobljeno raziskujejo tudi sociologi, komunikologi, politologi, pedagogi in raziskovalci drugih družbenih ved ob pomoči matematikov, metodologov in filozofov znanosti. V znani zbirki *Metodološki zvezki* je izšla knjiga *Zanesljivost in veljavnost merjenja*, ki ji je uspelo razmeroma zapleteno snov obdelati zgoščeno in pregledno na dobrih stosedemdesetih straneh. Namenjena je boljšemu študentu višjih letnikov ali podiplomcu različnih družbenih ved, ki obvlada osnovne pojme opisne in večrazsežne statistike ter se ne ustraši tudi kakega obrazca. Pri tem pa se knjiga precej spretno izogne matematično strogemu dokazovanju, kar družboslovce tako hitro odbije od podobnih, vendar bistveno obsežnejših in težje berljivih del. Poudarek knjige je na razumni uporabi metod, s katerimi lahko preverjamo kvaliteto merjenja v raziskovalni praksi. Računske podrobnosti avtorice namenoma prepuščajo priročnikom za rabo priljubljenih programskih paketov, kot sta SPSS in LISREL.

Snov je razdeljena v šest poglavij. V prvem se bralec podrobneje seznani z merjenjem v družbenih vedah, z napakami pri merjenju in s prizadevanji za višjo kakovost merjenja. Korake od teorije do merskega inštrumenta ilustrira nastajanje vprašalnika za merjenje teoretičnega konstrukta 'zadovoljstvo z delom'. Bralec, ki tega morda še ni vedel, na sedmi strani zve: »Meriti brez napak ni mogoče.« Napake lahko kvečjemu zmanjšamo na najmanjšo dosegljivo mero.

Drugo poglavje obravnava zanesljivost merjenja. Na kratko, vendar precej poglobljeno, prikaže klasično testno teorijo, opredeli zanesljivost ter podrobneje opiše kongenerični, tau-enakovredni in vzporedni model. Sledi kritika klasične testne teorije in razširitev merskega modela Sarisa in Andrewsa v obliki sistema enačb in vzročnega diagrama. V odstavku o zanesljivosti večkratnih meritev se psiholog ponovno sreča z obrazcem Spearmana in Browna. Sledi kratek prikaz posledic nezanesljivosti meritev na oceno korelacijskega in regresijskega koeficienta.

Tretje poglavje je posvečeno ocenjevanju zanesljivosti merjenja v praksi s ponavljanjem meritev na različne načine in z merami enakovrednosti, kamor sodita metodi razpolovitve in notranje konsistentnosti. Bralec, ki že pozna temeljne pojme faktorske analize, lahko zve, kaj nam o zanesljivosti povesta komponentna in faktorska analiza. Avtorice ilustrirajo poglavje s primerom merjenja stališč o spolnih vlogah pri študentih z vprašalnikom, ki ga sestavlja sedemnajst Likertovih lestvic. Na tem mestu bo psihološko izobraženi bralec najbrj pogrešal opredelitev in praktično uporabo različnih koeficientov diskriminativnosti. (V izpisu procedure RELIABILITY iz SPSS, denimo, najdemo kar dva v stolpcih z angleškima imenoma *corrected item-total correlation* in *squared multiple correlation*.) Zlasti prvi koeficient diskriminativnosti lahko bistveno olajša konstrukcijo sestavljene meritve v želji po čim višji notranji konsistentnosti, saj z eno številko povzame informacijo cele vrstice korelacijske matrike še zlasti, kadar je test dolg (denimo: iz več kot 30 postavk) in so zato manj pregledni drugi pristopi, ki so tudi računsko bistveno zahtevnejši.

Drugi del knjige, ki je nekoliko daljši, je posvečen veljavnosti. Algebre je manj, je pa zato snov vsebinsko bolj zapletena, saj so metodološko strožje obravnave nekaterih pogledov na veljavnost značilni šele v zadnjih treh desetletjih. Tudi terminologija v najboljši tuji literaturi je daleč od enotnosti: odlično načitane avtorice, ki v seznamu literature navajajo dobrih sto naslovov (od tega 35% iz tega desetletja), so, denimo, izpisale in večinoma

posrečeno poslovenile naslednje stalne zveze: *veljavnost analitičnega postopka*, *diagonala* ~i, *diskriminacijska* ~, *ekološka* ~, *empirična* ~, *faktorska* ~, *interpretativna* ~, *intrinzična* ~, ~ *izsledkov*, *kazalec* ~i, *koeficient* ~i, ~ *konstrukta*, *konvergentna* ~, *kriterijska* ~, *logična* ~, ~ *merjenja*, ~ *merskega postopka*, *multitestna* ~, *napovedna* ~, *nomološka* ~, *notranja* ~, *ocenjevanje* ~i, ~ *okolja*, ~ *podatkov*, *pojmovna* ~, *preverjanje* ~i, *pragmatična* ~, *razsežnostna* ~, ~ *raziskovanja*, *relativna* ~, *sočasna* ~, *statistična* ~, *teoretična* ~, *ugotavljanje* ~i, *validacija*, ~ *teorije*, *vindikacija*, *vsebinska* ~, *vzorčna* ~, *vzporedna* ~, *zdravorazumska* ~, ~ *znane skupine* in *naposled zunanja* ~. Priznati moram, da je moj osebni slovarček krajši in pogrešam samo še *veljavnost ključa/ključno veljavnost* (Hubbard in Schumacher, 1971). Še največ reda so v ta terminološki blodnjak vpeljali standardi ameriškega psihološkega društva, ki so se omejili na *vsebinsko*, *napovedno* in *sočasno veljavnost* ter *veljavnost konstrukta*. Kdo ve, zakaj avtorice te standarde dosledno navajajo po Splichalovi knjigi o analizi besedil? Bralce bo najbrž razveselila novica, da objava slovenskega prevoda tega vplivnega besedila iz leta 1985 čaka samo še na dovoljenje nosilca avtorskih pravic. Dotlej lahko ob izvorniku v sili navajamo tudi hrvaški Kulenovičev prevod (1992).

V četrtem poglavju knjige se bralec podrobneje seznanja z vsebinsko veljavnostjo ter obema oblikama kriterijske veljavnosti (napovedno in sočasno). Pojemovno trši oreh je veljavnost konstrukta: zanjo se mora bralec poglobiti v vlogo teoretičnih konstruktov v znanosti ter njihovo teoretično in empirično smiselnost. Na koncu poglavja ponovimo lekcijo, ki smo se je naučili že v prvem delu knjige: koeficient kakršnekoli kriterijske veljavnosti ne more preseči kvadratnega korena koeficienta zanesljivosti. »Nobena izmerjena spremenljivka ne more biti veljavna, ne da bi bila zanesljiva. Toda zanesljivo merjenje ni nujno veljavno.« (str. 103)

Peto poglavje je namenjeno bolj praktičnemu preizkušanju veljavnosti konstrukta, zlasti s pomočjo *večrazsežnega - večmetodnega pristopa*, kot sta ga predlagala Campbell in Fiske, ter z njegovimi dopolnitvami s konfirmatorno faktorško analizo. Poglavje zaključuje razmeroma podrobno analiziran primer uporabe tega pristopa pri analizi kvalitete podatkov o 'zadovoljstvu z ivljenjskimi razmerami' v anketi slovenskega javnega mnenja. Zadnje poglavje vsebuje pregleden in zgoščen pregled standardov kvalitete merjenja.

Delo opisuje sodobno stanje raziskovanja zanesljivosti in veljavnosti v večjih slučajnih vzorcih. Zaradi prizadevanja avtoric po zgoščenosti in zaradi navedb iz drugotnih virov deluje poznavalcu zgodovine psihologije nekaj

izjav nekoliko anahronistično. Letnica 1983 ob prvi omembi znanega obrazca Spearmana (1910) in Browna (1910) sicer pošteno vodi do vira, iz katerega so ga navedle avtorice (str. 26), vendar pravima avtorjema po krivici odzame kake tričetrst stoletja prvenstva. Lord in Novick (1968) prav gotovo nista »začetnika« klasične testne teorije (str. 12): do njune knjige se je že nabralo toliko novejših testnih teorij, da sta starejšo (prim. Gulliksen, 1950) poimenovala za »klasično« in poskrbela za njeno strogo aksiomatsko predstavitev.

Knjigo odlikuje sistematičnost, zgoščenost in obilica didaktično skrbno izbranih ter ne pretirano zapletenih praktičnih primerov. Pohvalna je tudi skrb za slovensko strokovno izražanje. Tu smo doslej pod premočnim vplivom angleščine ali kakega drugega jezika in se zato v okvirih tako različnih ved pogosto čisto po nepotrebnem te ko razumemo. Tako se mi zdijo izrazi *napovedna* in *sočasna veljavnost* ter *veljavnost konstrukta* stvarno in jezikovno ustrežnejši od drugih objavljenih predlogov, kot sta npr. *prognostična veljavnost*, ki prevladuje v slovenskem psihološkem in pedagoškem slovstvu (npr. Sagadin, 1968), ali *konstruktna veljavnost* pri Vodopivcu (1993) če ne omenim stvarno napačno prevedene *konkurenčne veljavnosti*. Nekaj pomislekov pa mi poraja uporaba matematičnega izraza *spremenljivka* v pomenih kar treh angleških strokovnih terminov *score*, *item* in *variable*. Pri pisanju o različnih psihodiagnostičnih sredstvih očitno potrebujemo skupni izraz za to, čemur pri različnih meritvah pravimo *naloga*, *(pod)vprašanje*, *problem*, *trditev*, *izjava*, *lestvica* in podobno; zato sam že poldrugo desetletje angl. *item* prevajam s *postavka*. Za angl. *score* se mi zdi *dosežek* vsebinsko ustrežnejši kot *rezultat* ali dolgovezno *število točk* še zlasti, kadar ni celo število. Matematik seveda upravičeno vidi *spremenljivke* (nikar: *variable* ali celo *varijable*!) na vseh ravneh pojmovne hierarhije, vendar smo pri pisanju o stvarnih merskih postopkih pogosto prisiljeni v istem stavku ločevati več ravni.

Kaj se bo prijelo v slovenskem strokovnem izrazju, bosta pokazala čas in usklajenost uredniških ali lektorskih posegov v različnih strokovnih revijah, testnih priročnikih, standardih in zlasti učbenikih, kakršen je ta.

Zanesljivost in veljavnost merjenja sodi na knjižno polico slehernega psihologa in študenta psihologije, prav gotovo pa med obvezno branje raziskovalca, pisca testnih priročnikov ali sodelavca v programih zunanjega preverjanja znanja.

NAVEDENO SLOVSTVO

1. Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296-322.
2. Gulliksen, H. (1950). *Theory of Mental Tests*. New York: Wiley.
3. Hubbard, J. P. in Schumacher, C. F. (1971). *Measuring Medical Education. The Tests and Test Procedures of the National Board of Medical Examination*. Philadelphia: Lea & Febiger.
4. Lord, F. M. in Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
5. Sagadin, J. (1968). *Prognostična veljavnost sedanjih meril za sprejem novincev v gimnazijo v SR Sloveniji*. Ljubljana: Zavod SRS za šolstvo.
6. Spearman, C. (1910). Correlation calculated with faulty data. *British Journal of Psychology*, 3, 271-295.
7. *Standardi za pedagoško i psihološko testiranje*. (1992, prev. A. Kulenović.) Zagreb: EDUCA.
8. Vodopivec, B. (1993). Sodobno pojmovanje konstruktne veljavnosti. *Psihološka obzorja*, 2(2), 37-56.

Freudova psihoanaliza je gotovo med tistimi smermi ("volani"), ki so najbolj zaznamovale psihološko znanost v našem iztekajočem se stoletju, prav gotovo pa je prav Freudova psihoanaliza tisti dosežek psihologije tega stoletja, ki je najbolj vplival na tokove sodobne miselnosti tudi zunaj ožje psihološke stroke. Brez Freudovega vpliva si npr. ne moremo predstavljati sodobne filozofije, še manj pa sodobne umetnosti, zlasti ne sodobne literature in filma. Morda se je tudi zato v sami psihološki stroki (in zunaj nje) že oblikovalo mnenje, da gre pri Freudovem delu že za bolj zgodovinsko dogajanje in da je za današnji čas primerno predvsem razumevanje Freuda v filozofskih in antropoloških razsežnostih. Zato tudi ni slučajno, da so se pri nas v novejšem času, vsekakor pa v obdobju, ki traja že svojih dvajset let, vsaj v znanstveni publicistiki uveljavili zlasti strukturalistični in postmoderni pogledi na Freuda. Kot da je bilo ob tem nekam pozabljeno dejstvo, da je Freudova psihoanaliza tudi kot tekoča znanost še vedno živa in se še vedno razvija. Posebno zanimivo pri tem je to, da je najmočnejši razvoj sodobne psihoanalize bil prav v tisto smer raziskovanj, ki jih je npr. strukturalna analiza v svoji polemiki z "ego psihologijo" najbolj zanemarila in potiskala v ozadje, to je v raziskovanje razvoja jaza in objektivnih odnosov.

Ta, v svetu vendarle prevladujoči koncept moderne psihoanalize gotovo ni bil odsoten v sami psihološki praksi, vse premo pa je bil prisoten v slovenski psihološki publicistiki. S knjigo "Razvojni analitična psihoterapija" pa smo