

SODOBNO POJMOVANJE KONSTRUKTNE VELJAVNOSTI*

Blaž Vodopivec

1. UVOD

Koncept konstruktne veljavnosti je nastal v letih 1950-54, ko je poseben odbor Ameriškega psihološkega združenja (APA) skušal sistematizirati in dopolniti dotedanja spoznanja in postopke pri validaciji testov, za katere je menil, da so nepopolni in v veliki meri tudi neadekvatni. Morda ni naključje, da se je potreba po tem konceptu, ki je najpomembnejši prispevek in inovacija tega odbora, porodila v pododboru za projekcijske preizkuse.

Odbor je dotedanje prakse validacije sistematiziral v s-kriterijem-povezano (criterion-related) veljavnost, ki jo sestavljata sočasna (concurrent) in prediktivna (predictive) veljavnost, in vsebinsko (content) veljavnost. Po strokovni literaturi tistega časa se je širilo mnogo kritik o nezadostnosti teh dveh postopkov validacije pri dokazovanju veljavnosti testa (se pravi pri dokazovanju hipoteze, da test res meri tisto, kar naj bi meril). Zato se je odboru zdelo, da mora v kartotečni omarici, ki se imenuje veljavnost, odpreti še en predalček.

Ta predalček je poimenoval konstruktna (construct) veljavnost in ga napolnil z vrsto izredno relevantnih, zanimivih in inovativnih, a nikakor dorečenih in sistematiziranih spoznanj o odnosih med teorijo, podatki in mersko metodo. S tem je odprl pot produktivnemu toku idej, ki je danes še ravno tako živ kot tedaj. Kajti navkljub uspešnim prizadevanjem, da bi koncept konstruktne veljavnosti natančneje opredelili, mu določili meje in notranjo strukturo ter zadovoljive operacionalizacije in kriterije, ta v končni instanci še vedno referira na nikoli razrešena temeljna epistemološka vprašanja in dopušča razlike v interpretaciji. Kar pa seveda ne pomeni, da se ni kljub temu razvil v ključno orodje za raziskovalca, ki hoče svojo teorijo adekvatno in produktivno soočiti z empiričnimi dejstvi.

V pionirskem članku Cronbacha in Meehla (1955), ki delno predstavlja dosežke omenjenega odbora in delno lastne ideje avtorjev, je konstruktna veljavnost ravno tako opredeljena kot ena od treh vrst veljavnosti. Po njunem mnenju pride konstruktna validacija v poštev, če se test interpretira kot mero nekega atributa ali kvalitete, ki jo ni mogoče operacionalno definirati. Raziskovalec odgovarja na vprašanje: kateri konstrukti pojasnjujejo varianco testnih skorov. Čeprav je ta koncept relevanten za vsa področja psihološkega merjenja, je po njunem mnenju še posebej pomemben na področju razvoja osebnostnih testov.

Cronbach in Meehl (1955) menita, da uporaba vsebinske, s-kriterijem-povezane ali konstruktne validacije ni samo stvar razlik v postopku proučevanja, ampak stvar

* Oddelek za psihologijo, Filozofska fakulteta, Aškerčeva 2, Ljubljana

raziskovalčeve orientacije. S-kriterijem-povezana veljavnost zahteva sprejemanje skupine operacij kot aдекватne definicije tistega, kar je merjeno. Če pa raziskovalec meni, da noben njemu dostopen kriterij ni popolnoma veljaven, se nujno obrne h konstruktni validaciji, da bi se izognil "neskončni frustraciji" povezovanja vsakega kriterija z nekim še bolj veljavnim standardom. Pri vsebinski veljavnosti gre za sprejemanje podmene, da vsebinsko področje definira merjeno spremenljivko. Če noben univerz vsebine ni popolnoma aдекватna definicija merjenega (kar je še posebej očitno pri merjenju osebnostnih potez), se je spet treba zateči h konstruktni validaciji.

Najpogostejša praksa sestavljalcev testov tistega časa je bila, da so poleg zanesljivosti testa navajali korelacijo z neko povezano, v večini primerov ne-testno spremenljivko, kot "koeficient veljavnosti". Čeprav je jasno, da je tako pojmovanje prav smešno nezadostno, ga tudi danes pogosto srečamo v navajanju "metričnih karakteristik" testa. Jasen unitaren vedenjski kriterij merjene spremenljivke je izredno težko zagotoviti, če pa je spremenljivka latenten konstrukt, pa si ga je v večini primerov celo težko zamisliti.

Cronbach in Meehl (1955) navajata, da se pogosto zgodi, da validnost testa sčasoma preseže validnost kriterija. Binetov test je bil sprva sprejet, ker so učitelji menili, da se njegovi rezultati v zadostni meri skladajo z njihovimi sodbami o inteligentnosti otrok. V nasprotnem primeru bi ga, podobno kot reakcijske čase in druge preizkušane mere inteligentnosti, zavrgli. Če pa danes s testom otroku izmerimo IQ 135, učitelj pa se pritožuje, kako zabit da je, naša prva reakcija prav gotovo ne bo sklep, da je test zanič. Že Thurstone (1952, cf. Cronbach in Meehl 1955) je ugotavljal, da za večino sodobnih testov mentalne sposobnosti ni mogoče najti kriterijev, ki bi bili bolj veljavni kot ti testi sami. Zato se validacijske študije čedalje bolj zanašajo na kriterije interne konsistentnosti.

Podobno je v kliničnem kontekstu. Psiholog z svojim instrumentom registrira vzorec določenih značilnih vedenj, testni rezultat pa se "validira" s pomočjo psihiatrične diagnoze. Vendar je tudi psihiater svojo diagnozo postavil na osnovi registracije določenih vzorcev vedenja. Nobenega intrinzičnega zagotovila torej ni, da je kriterij bolj veljaven od testnega rezultata. V mnogih primerih je celo mogoče jasno dokazati, da je manj veljaven.

Cronbach in Meehl (1955) svoje uvodno razmišljanje o konstruktni veljavnosti ilustrirata z naslednjim primerom procesa validacije:

Denimo, da smo ugotovili, da mera X korelira .50 z potenjem dlani v situaciji, ko je bilo poskusni osebi sporočeno, da je padla na izpitu. Prediktivno veljavnost X-a aдекватno opišemo s koeficientom in opisom eksperimentalnih pogojev in vzorčenja. Če bi kdo vprašal, "Ali ne obstaja kakšna druga interpretacija za to korelacijo?" ali "Kakšne dokaze še lahko navedete za vašo interpretacijo?", bi ga komaj razumeli, saj sploh nismo podali nobene interpretacije. Ta vprašanja pa postanejo relevantna, če korelacijo navedemo kot dokaz, da test meri anksioznost. Možne so alternativne interpretacije: npr. da test meri "akademske aspiracije". V tem primeru bi pričakovali drugačno korelacijo, če bi potenje dlani inducirali npr. z ekonomsko grožnjo. V takih primerih pa se je torej popolnoma smiselno vprašati po drugih vrstah dokazov.

Dokazno gradivo o konstruktni veljavnosti predstavljajo razne vrste podatkov. Cronbach in Meehl navajata naslednje vire:

- Predvidene razlike med skupinami.
- Interna struktura testa: konsistentnost v skladu s predvidevanji (ne nujno visoka).

– Pričakovane spremembe med situacijami, pod različnimi pogoji, kamor spada tudi test-retest stabilnost (ne nujno visoka) ter eksperimentalni učinki.

– Študije procesa: katero dejansko vedenje res vpliva na testne rezultate.

Navedeni postopki zbiranja dokaznega gradiva o konstruktni veljavnosti ne vsebujejo vnaprej postavljenih kriterijev. Pomembno je, da so rezultati v skladu s predvidevanji. Teorija torej vodi proces konstruktno validacije. Teorija predpostavi mrežo odnosov med obravnavanim konstruktom in drugimi konstrukti, pogoji in procesi. To mrežo Cronbach in Meehl imenujeta nomološka mreža. Nomološka mreža je povzetek njune filozofije znanosti, ki jo skušata čim bolj zvesto izpeljati iz dejanskega procesa znanstvenega odkritja in ima naslednja osnovna načela:

1. V znanosti "pojasniti neko stvar" pomeni formulirati zakone, v katerih se pojavlja. Ta prepletajoči se sistem zakonov je nomološka mreža.

2. Zakoni nomološke mreže povezujejo (a) opazljive (observable) lastnosti ali kvantitete med seboj, (b) teoretične konstrukte z opažanji in (c) teoretične konstrukte med seboj. Ti zakoni so lahko deterministični ali statistični.

3. Nujen pogoj za znanstveno legitimnost konstrukta je, da se pojavlja v nomološki mreži, v kateri se vsaj nekateri zakoni nanašajo na opažanja. Ta zveza z opazljivimi dejstvi je lahko oddaljena in posredna. Konstrukt se ne da reducirati na opažanja, pač pa v kombinaciji z drugimi konstrukti iz mreže implicira opažanja.

4. "Izvedeti več o konstruktu" pomeni elaborirati nomološko mrežo, v kateri se pojavlja, ali povečati razločnost in določnost njenih komponent (nomologemov). V začetku zgodovine nekega konstrukta bo nomološka mreža omejena in povezav bo malo.

5. Obogatitev mreže, npr. dodajanje konstrukta ali relacije, je upravičena, če usvarja nomologeme, ki jih potrjujejo opažanja, ali če reducira število nomologemov, ki so potrebni za predikcijo istih opažanj. Če se opažanja ne skladajo z mrežo, ima ponavadi znanstvenik na voljo več alternativnih poti za modifikacijo mreže, ki se tisti trenutek zdijo vse enako upravičene.

6. Za kvalitativno zelo različne "operacije" lahko rečemo, da se "prekrivajo" oz. "merijo isto stvar", če jih njihov položaj v nomološki mreži veže na isti konstrukt. Zato ni nujno, da to prekrivanje potrjujemo z neposrednim primerjanjem opažanj, ampak se lahko zadovoljimo s posrednim potrjevanjem odnosov v mreži.

Ta načela naj bi tudi pomirila "trdo" usmerjene metodologe, ki se bojijo, da konstruktna veljavnost odpira vrata za navajanje lastnosti in kvalitet testa, ki jih ni mogoče potrditi. Če se nomološka mreža nikjer ne stika z opažanji ali če koraki v posredni inferenci niso eksplicitno in jasno navedeni, o konstruktni validaciji ni mogoče govoriti.

Če raziskovalci ne sprejmejo in uporabljajo vsaj v bistvenih potezah enake nomološke mreže, konstruktna validacija ni mogoča. Uporabnik testa, ki zavrača avtorjevo teorijo, ne more sprejeti njegovih dokazov o konstruktni veljavnosti, ampak se mora validacije lotiti sam. Cronbach in Meehl (1955, s.300) celo zaključujeta, da se proučevanje konstruktno veljavnosti v bistvu ne razlikuje od splošnih znanstvenih postopkov razvijanja in testiranja teorij. Kar seveda raziskovalca z naivnimi epistemološkimi predstavami spravlja v nemajhno zadrego: "Kako se rešiti iz cirkularnosti, ko po eni strani teorije ni mogoče dokazati brez veljavnih mer, po drugi strani pa veljavnosti ni mogoče dokazati brez sprejete teorije?"

Temu enostavnih receptov vaje nemu raziskovalcu Cronbach in Meehl na koncu namenita še milostni strel z ugotovitvijo, da konstruktno veljavnosti ni mogoče povzeti z enostavno številčno oceno, pa če avtorji še tako radi govorijo o koeficientu veljavnosti testa. Ob vsej kompleksnosti problematike ni direktne formule, ki bi

izrazila delež prave variance konstrukta v testnem rezultatu. S pomočjo raznega dokaznega gradiva je mogoče samo grobo oceniti spodnjo in zgornjo mejo tega deleža.

Očitno je, da Cronbach in Meehl v svojih napotkih za konstruktno validacijo nista formulirala definitivnih rešitev. S tem, ko sta sledila uradni trinitarni APA klasifikaciji in ohranila konstruktno veljavnost kot zgolj eno od treh vrst veljavnosti, sta sicer naredila prvi korak, nista pa uspela narediti še drugih potrebnih. Pomembno pa je, da sta uspela mnoge metodološko osveščene raziskovalce prepričati, da je konstruktna veljavnost nujen pogoj za razvoj in preverjanje teorije, kar pomeni, da konstruktna validacija leži v samem srcu napredka družboslovnih znanosti. (Steenkamp in van Trijp 1991)

Čeprav tudi za delo avtorjev, ki so se kasneje trudili zakoličiti dokončno opredelitev, strukturo in operacionalizacijo koncepta konstruktne veljavnosti, težko rečemo, da so se svojemu cilju zelo približali, pa je očitno, da jih prizadevanja vodijo v smeri vsaj dveh ključnih preobratov v pojmovanju:

1. Konstruktna veljavnost ni ena od veljavnosti, ampak je to veljavnost nasploh, ki drugi dve kategoriji (s-kriterijem-povezano in vsebinsko) vključuje kot svoji komponenti. Kritike "trinitarne" APA klasifikacije veljavnosti so postajale čedalje pogostejše. Avtorji so se zatekali k rešitvi, da gre zgolj za tri različne pristope oz. strategije v procesu validacije, ne pa za tri različne vrste veljavnosti (npr. Rosenbaum 1989, Lawshe 1985, cf. Cohen, Montague in Nathanson 1988). Vse to pa je bilo premalo. Kajti, če je veljavnost opredeljena kot "sodba o tem, koliko test dejansko meri tisto, kar naj bi meril" oz. "na dokazih utemeljena sodba o tem, kako ustrezni so zaključki (inferences), ki jih potegnemo iz testnih skorov" (Cohen, Montague in Nathanson 1988, s.123), potem je jasno, da tako sodbo lahko podpre samo celovito zasnovana konstruktna validacija. Ali če, kot Nunnally (1978), definiramo veljavnost kot "znanstveno uporabnost merskega instrumenta", potem ne moremo imeti za zadosten dokaz veljavnosti korelacijski koeficient z nekim izoliranim kriterijem. V skladu s tem Cohen, Montague in Nathanson (1988, s.141) ugotavljajo, da konstruktna veljavnost vedno bolj postaja povezujoč koncept za vse dokaze o veljavnosti. Tudi kriterijsko in vsebinsko veljavnost lahko imamo za komponente konstruktne veljavnosti. Rosenbaum (1988) pa govori kar o s-kriterijem-povezani konstruktni veljavnosti.

2. Čedalje bolj očitno je, da ne moremo govoriti samo o konstruktni veljavnosti testa (ali kakršnekoli mere). Ravno tako, kot je pomembno vprašanje, ali je test veljaven pokazatelj konstrukta, je pomembno tudi vprašanje, ali sploh gre za legitimen, epistemološko sprejemljiv konstrukt. Konstruktna veljavnost se torej ne nanaša zgolj na mero, ampak na kombinacijo mere in konstrukta, pa tudi merske teorije, ki ju povezuje. Tako Steenkamp in van Trijp (1991) ne govorita več o konstruktni veljavnosti testa, temveč o veljavnosti konstrukta. Če gre pri konstruktni veljavnosti za vprašanje, ali se pojmi, o katerih govori teorija, res reflektirajo tudi v empiričnih podatkih, in sicer na tak način, kot predpostavlja teorija, iz tega sledi, da teoretičnega ali empiričnega pola res ne moremo obravnavati izolirano.

Prepletenost teorije in empirije veje iz vsakega resnega pisanja o konstruktni veljavnosti. Za Bentlerja (1978, s.288) je glavni prispevek Cronbacha in Meehla to, da sta poudarila nujno, da procesa validacije ne moremo obravnavati samo kot prakticistično-mersko obrt, ampak da je predpogoj za te postopke določen skup dokazov o obstoju oz. veljavnosti obravnavane spremenljivke kot znanstvenega konstrukta. Poudarek Cronbacha in Meehla na konstruktih pomeni poskus dati vsebinski teoriji relevantno vlogo v tipičnem procesu konstrukcije in validacije testa. Bentler ek-

splicitno enači konstruktno veljavnost s testiranjem ustreznosti kavzalnega modela, ki reprezentira teorijo. To razmišljanje pa nas potegne že globoko v epistemološke vode.

Cirkularnosti predpostavk, da konstrukta ne moremo uporabiti za preverjanje teorije, dokler ni dokazana njegova konstruktna veljavnost, hkrati pa je samo ugotavljanje konstruktne veljavnosti že tudi preverjanje teorije v celoti, ne more razrešiti naivna epistemologija pozitivizma ali iz njega izpeljanih doktrin logičnega empirizma. Epistemologija holističnega konstruktivizma (glej Bagozzi 1984, Vodopivec 1988) lahko mnogo ustrežneje razloži, kako se v dejanskem procesu znanstvenega odkrivanja teorija in empirija medsebojno določata in oplajata. Stopnja razločljive strukture v teoriji v vsakem trenutku nekoliko presega ali nekoliko zaostaja za stopnjo razločljive strukture v podatkih. Prava stopnja artikulacije teorije omogoča smiselno empirično preverjanje in prava stopnja informacije v podatkih omogoča njihovo smiselno teoretično interpretacijo. (Lea, Tarpy in Webley 1987).

Konstruktna veljavnost se dejansko tiče vseh elementov kanonične forme znanstvenih teorij in razmerij med njimi, še posebej korespondenčnih pravil. (Bagozzi 1984, Vodopivec 1988) Specifikacija in analiza korespondenčnih pravil oz. odnosov med teoretičnimi in empiričnimi koncepti je predpogoj za določanje pomena teoretičnim terminom. Konstruktna veljavnost je odvisna od definicije in teoretičnega pomena koncepta, vedenja njegovih mer, ter semantičnih in sintaktičnih odnosov med obema poloma (procesna specifikacija mehanizma reflektiranja). Naša sposobnost, da odkrijemo resnične empirične generalizacije in jih interpretiramo kot dokaz za temeljne teoretične (neopazljive) predpostavke, je fundamentalno odvisna od konstruktne veljavnosti oz. razmerja med teoretičnimi in empiričnimi koncepti. Če niso specificirana korespondenčna pravila in veljavne mere, ustvarimo brezno med gradnjo in razvojem teorije na eni strani in testiranjem teorije na drugi strani, hkrati pa si onemogočimo celo to, da bi lahko artikulirali pogoje in stopnjo falsifikabilnosti teorije in ji tako zagotovili znanstveni status. (Bagozzi 1984)

2. OPREDELITVE KONSTRUKTNE VELJAVNOSTI

Namen doslej napisanega je vzbuditi pri bralcu vtis, ki ga deli tudi avtor: da je vprašanje konstruktne veljavnosti tako kompleksna in živa materija, da mu, kot pri drugih podobnih pojmi, vsega pomena ne more izčrpati nobena definicija, ampak se ta lahko oblikuje samo v produktivni interakciji teoretičnih razmišljanj in empiričnih operacionalizacij. Upam, da je taka interakcija tudi pričujoče delo. Zato ne bom poskusil formulirati lastne definicije konstruktne veljavnosti. Prav pa je, da navedem nekatere definicije drugih avtorjev, z opozorilom, da so te definicije del toka razmišljanj posameznih tekstov in da v niti enem primeru ne gre za eksplicitno formalno definicijo. Najprej je navedenih nekaj definicij konstruktov.

Konstrukt je nek postuliran atribut ljudi, za katerega menimo, da se reflektira v testnem rezultatu. ... Konstrukti so po naravi različni, od tistih, ki so zelo blizu "čistim opisom" (in vključujejo samo ekstrapolacijo odnosov med opažanji), do visoko teoretičnih konstruktov, ki vključujejo hipotetične entitete in procese ali se identificirajo s konstrukti iz drugih znanosti, npr. nevrologije, genetike, fiziologije ... (Cronbach in Meehl 1955)

Konstrukt je postuliran atribut merjenega objekta. (Bentler 1978)

Konstrukti so abstraktni povzetki nekaterih zakonitosti v naravi. Povezani so z konkretnimi, opazljivimi entitetami oz. dogodki. Nekatere konstrukte (npr. agresivnost, storilnostna motivacija) pojmujeemo kot vzroke vedenja. Drugi (npr. ugodje, potrnost) se nanašajo na občutke, ki jih ljudje doživljajo. Spet tretji (npr. verbalna inteligentnost, glasbeni talente) se nanašajo na uspešnost izvajanja določenih povezanih vedenj. (Murphy in Davidshofer 1988)

Konstruktno veljavnost pa ugotovimo tako, da ugotovimo, ali je razmerje med testnim skorom in povezanimi vedenji tako, kot je pričakovano razmerje med konstruktom in temi vedenji. (Murphy in Davidshofer 1988)

Konstruktna veljavnost se nanaša na primernost sklepov, dobljenih na osnovi testnih skorov, glede posameznikove pozicije na določeni vrsti spremenljivke, imenovane konstrukt. Konstrukti niso direktno opazljivi. (Cohen, Montague in Nathanson 1988)

Konstruktno veljavnost ovrednotimo s proučevanjem vprašanja, katere psihološke kvalitete test meri; torej z ugotavljanjem, v kakšni meri razlagalni koncepti ali konstrukti lahko pojasnijo rezultate na testih (Cronbach 1971, cf. Rosenbaum 1989)

Konstruktna veljavnost pove, ali je test dobra mera konstrukta. (Murphy in Davidshofer 1988)

Konstruktna veljavnost nam pove, do kakšne mere konstrukt doseže teoretični in empirični pomen. (Steenkamp in van Trijp 1991)

Konstruktna veljavnost pove, do kakšne mere operacionalizacija meri koncept, ki naj bi ga merila. (Bagozzi 1984)

Konstruktna validacija je proces obvladovanja dokaznega gradiva v podporo sklepu, da ima opažena konsistentnost testnih odgovorov določen pomen, predvsem z ugotavljanjem, do katere mere so empirični odnosi z drugimi merami oziroma njihova odsotnost, konsistentni z tem pomenom... V tem procesu skušamo povezati zanesljivo konsistentnost odgovorov, kot jo povzema testni skor, z ne-testnimi konsistentnostmi v vedenju, ki v skladu s predpostavkami reflektirajo isti temeljni konstrukt... Konstruktna validacija implicira tako... konvergentne dokaze o tem, da je obravnavana mera koherentno povezana z drugimi merami tega konstrukta, ... kot tudi divergentne dokaze o tem, da ta mera ni premočno povezana z merami drugih, distinktnih konstruktov. (Messick 1980, cf. Rosenbaum 1989)

Navedene definicije ne glede na različno stopnjo elaboracije v glavnem konvergirajo. Izjema je definicija Steenkampa in van Trijpa, ki že govori o veljavnosti konstruktov in ne konstruktni veljavnosti testov. Kot pri večini vprašanj je tudi v tem primeru nekoliko samosvoje Cattellovo mnenje:

Vprašanje konstruktno veljavnosti je predvsem vprašanje jasnosti factorske strukture nekega vsebinskega področja oz. vprašanje, ali je za faktorje dokazano, da obstajajo kot neodvisne funkcionalne entitete (s.33) ... In drugič, s kakšno natančnostjo testna baterija meri te faktorje. Potem, ko temeljno raziskovanje nekega področja (npr. motivacije) zadovoljivo odgovori na prvo vprašanje, lahko konstruktno veljavnost ovrednotimo s koreliranjem testnih skal s skori čistih faktorjev, dobljenimi v temeljnem raziskovanju. (Cattell in drugi 1975)

Druga bazična komponenta veljavnosti pa je za Cattella kontekstualna veljavnost (circumstantial validity), ki se nanaša na vprašanje, koliko je korelacija testa x z drugimi faktorji podobna korelaciji teh faktorjev s konstruktom X, katerega mera je test x. Končno pa je pomembna tudi konkretna veljavnost, torej korelacija s čisto vsakdanjimi konkretnimi vedenji. Ker pa ta vedenja sama po sebi redkokdaj v kolikor toliko čisti obliki izražajo konstrukt, bi bilo ta vidik bolje imenovati konkretna relevantnost kot konkretna veljavnost.

3. KOMPONENTE KONSTRUKTNE VELJAVNOSTI

Že Cronbach in Meehl (1955) sta v svojem pionirskem članku o konstruktni veljavnosti vseskozi poudarjala, da konstruktne veljavnosti ne moremo izraziti z eno formulo in enim koeficientom, ampak je treba zbrati več vrst dokaznega gradiva iz različnih virov. Glede tega, kakšne dokaze in na kakšen način, pa so njuna navodila vse prej kot izdelan recept, ki bi povedal: to, vse to in nič več kot to. Podobno lahko trdimo o navodilih za postopek konstruktne validacije in potrebne komponente tega postopka, ki so jih podali kasnejši avtorji. Poglejmo nekaj primerov:

Murphy in Davidshofer (1988) imenujeta svoj postopek konstruktne eksplikacija. Treba je preveriti tri skupine na široko opredeljenih predpostavk:

1. Identificiraj vedenja, ki so povezana z merjenim konstruktom
2. Identificiraj druge konstrukte in na podlagi teorije odloči, ali so ali niso povezani z merjenim konstruktom.

3. Identificiraj vedenja, ki so povezana s temi dodatnimi konstrukti in na osnovi odnosov med konstrukti določi, če so ta vedenja povezana z merjenim konstruktom. Ta sistem razmerij med konstrukti in vedenji je nomološka mreža.

Cohen, Montague in Nathanson (1988) opozarjajo, da samo vrsta različnih dokazov lahko zgradi končno sliko o konstruktni veljavnosti. Naštevajo naslednje kriterije: (s.141)

1. Test je homogen, meri en sam konstrukt
2. Testni skori korelirajo z drugimi merami na način, kot to predvideva teorija, ki pokriva manifestacije dotičnega konstrukta
3. Testni skori se višajo ali nižajo s starostjo, kot je to teoretično predvideno.
4. Testni skori, dobljeni po določenem dogodku ali zgolj po preteku določenega časa (post-test), se razlikujejo od pre-test skorov, kot je to teoretično predvideno.
5. Testni skori različnih skupin se razlikujejo, kot je to teoretično predpostavljeno.

6. Konvergentnost različnih mer istega konstrukta

7. Divergentnost mer tega konstrukta in mer drugih konstruktov

Steenkamp in van Trijp (1991) uporabljata naslednje kriterije:

1. Enodimenzionalnost

2. Navzkrižna veljavnost

3. Konvergentna veljavnost znotraj metode

4. Zanesljivost

5. Stabilnost

6. Konvergentna veljavnost preko metod in diskriminativna veljavnost

7. Nomološka veljavnost

Po mnenju Steenkampa in van Trijpa predstavlja ta lista določeno hierarhijo zahtev – vsak naslednji pogoj je kompleksnejši, zahtevnejši, vsak prejšnji pa je predpogoj za vse kasnejše. Njun postopek je zelo dobro sistematično razdelan, vseeno pa se zdi, da nekateri kriteriji na njuni listi manjkajo in da bi morali biti vključeni kot posebne zahteve.

Še najpopolnejša in najbolj utemeljena se zdi lista zahtev za konstruktno veljavnost, ki jo predlaga Bagozzi (1988) in ki jo bomo uporabljali tudi v tem delu:

- vsebinska veljavnost

- enodimenzionalnost in zanesljivost

- konvergentna veljavnost

- diskriminativna veljavnost

- andecedenčna in konsekventna prediktivna veljavnost

– nomološka veljavnost

V nadaljevanju so bolj ali manj podrobno predstavljeni nekateri vidiki posameznih kriterijev, ki se jih pri površnih predstavitvah psihometrične metodologije ponavadi spregleda in ki nikakor niso sami po sebi umevni.

3.1. ENODIMENZIONALNOST MER

Zadostna enodimenzionalnost mer konstrukta je ena kritičnih prepostavk merske teorije in s tem tudi konstruktne veljavnosti. (Hattie 1985, cf. Steenkamp in van Trijp 1991) Enodimenzionalnost pomeni, da se za vsako skupino indikatorjev (mer) skriva samo en konstrukt in da kovariranje teh indikatorjev ni posledica vpliva večih konstruktov, če jo je raziskovalec pripisal enemu. Ponavadi se enodimenzionalnosti ocenjuje s pomočjo eksploratorne faktorske analize. Ta pristop pa ni zadosten, saj niso znani kriteriji, pri katerih je enodimenzionalnost še sprejemljiva. Bolj natančna sta dva analitična kriterija: interna konsistentnost in eksterna konsistentnost.

Temeljna enačba interne konsistentnosti je

$$r_{ac}/r_{ad} = r_{bc}/r_{bd}$$

kjer so a, b, c in d mere istega konstrukta, n. Ta enakost mora držati v okviru napake vzorčenja. Za preverjanje so potrebne vsaj štiri mere za vsak konstrukt. Enačba eksterne konsistentnosti je

$$r_{ae} = r_{ag}r_{ne}r_{ge}$$

kjer je a mera konstrukta n in e mera konstrukta n'. Ker imamo ponavadi manj kot štiri mere vsakega konstrukta, nam za ugotavljanje enodimenzionalnosti konstruktov ostane preverjanje interkorelacij mer različnih konstruktov, za katere princip eksterne konsistentnosti implicira točno določen vzorec. Je pa ta postopek zamuden, po svojem bistvu eksploratoren in zato ne pripelje direktno do dokončne sodbe o enodimenzionalnosti obravnavane skupine mer. Ta sodba pa je pomembna, kajti enodimenzionalnost je nujni pogoj, na katerem počivajo najpogosteje uporabljeni postopki za ugotavljanje zanesljivosti, predvsem Cronbachov a.

3.2. ZANESLJIVOST

Za zanesljivost se zdi, da je tista merska karakteristika, s katero res ne bi smelo biti težav. Je konceptualno in matematično obdelana, možno jo je povzeti z enim samim koeficientom. Pa vendar je težav z interpretacijo ničkoliko. Zanesljivost namreč ni natančnost merjenja oz. ujemanje izmerjenega rezultata z dejansko izraženo merjenega konstrukta. Zanesljivost pomeni zgolj verjetnost, da bomo pri istem preizkušanju v istih pogojih z istim instrumentom dobili isti rezultat.

Klasično testologijo zanima replikabilnost rezultatov: preko različnih form, na krajši čas in na daljši čas. Zanesljivost pomeni konsistentnost merjenja. Če hočemo razumeti faktorje, ki vplivajo na konsistentnost testnih skorov, se je dobro vprašati, kaj sploh vpliva na variabilnost skorov: (Murphy in Davidshofer 1988)

1. Trajne in splošne karakteristike posameznika: splošni nivo poteze, nivo sposobnosti razumevanja navodil, odgovorni slogi in seti.
2. Trajne, a specifične lastnosti posameznika: poznavanje konkretne vsebine testa, trajen poseben odnos do konkretne vsebine posameznih itemov.
3. Začasne splošne lastnosti posameznika: zdravje, motivacija, utrujenost, pozornost.

4. Specifični začasni dejavniki: trenuten poseben odnos do konkretne vsebine posameznih itemov.

Trajni in trenutni viri variabilnosti lahko prispevajo h konsistentnosti skorov, odvisno od situacije in uporabljene metode ocenjevanja zanesljivosti. Ni pa nujno, da faktorji, ki prispevajo h konsistentnosti, prispevajo tudi k natančnosti (pravilnosti, ustreznosti) merjenja, ki jo opisuje definicija, da "... bo popolna mera konsistentno pripisala števila atributom oseb v skladu z določenimi pravili." (Murphy in Davidshofer 1988)

Splošni model zanesljivosti

Klasična testologija konceptualizira zanesljivost v terminih pravega skora in naključne napake merjenja. Napaka merjenja je torej definirana kot naključna spremenljivka. Ne vemo, kolikšna je napaka v posameznem primeru, vemo pa, kolikšna je varianca napake oz. njen delež v testnem skoru.

$$x = T + e$$

x = skor na testu

T = pravi skor

e = napaka merjenja

Predpostavke teorije pravega skora so:

1. Povprečna napaka = 0

2. Pravi skor in napaka sta nekorelirana

3. Napake različnih skorov so nekorelirane

Varianca testnega skora je enaka vsoti variance pravega skora in variance napake.

$$\sigma_x^2 = \sigma_T^2 + \sigma_e^2$$

Zanesljivost je definirana kot delež pravega skora v testnem skoru.

$$r_{xx} = \sigma_T^2 / \sigma_x^2 = \sigma_T^2 / (\sigma_T^2 + \sigma_e^2)$$

Prav izraz "pravi skor" ima velike zasluge za zmedo pri interpretaciji zanesljivosti. Pravi skor pomeni v skladu z teorijo zanesljivosti vse, kar doprinaša h konsistentnosti merjenja. To pa seveda še zdaleč ni indikator "resničnega" nivoja atributa. Glede tega vprašanja zagrešita napako v interpretaciji pri svoji predstavitvi tudi Murphy in Davidshofer (1988), ki nas s pomočjo tabele 2.1.2.2 skušata naučiti, da je vprašanje, kaj je pravi skor, odvisno od namena merjenja.

Tabela 1: Različna interpretacija virov variabilnosti

faktorji variabilnosti	predmet merjenja	
	mehanične sposobnosti	mehanične sposobnosti pod stresnimi pogoji
razlike v mehanskih sposobnostih	pravi skor	pravi skor
stresni pogoji (razlike v reakciji nanje)	napaka	pravi skor
drugi dejavniki	napaka	napaka

Zakaj je ta interpretacija zmotna? Napaka merjenja je v teorji pravega skora definirana kot naključna spremenljivka. Razlike v reakciji na stresne pogoje v zgornjem primeru pa so sistematičen dejavnik, ki vpliva na konsistentnost rezultatov posameznika. Zato po teorji pravega skora ne morejo biti enkrat pravi skor, enkrat pa napaka, odvisno od želja raziskovalca. Povedano pa seveda ne pomeni, da taka interpretacija sploh ni ustrezna. Le teoretične predpostavke klasične teorije zanesljivosti moramo zamenjati z drugimi.

Ocenjevanje zanesljivosti

Psihometrični učbeniki ponavadi navajajo naslednje metode ocenjevanja zanesljivosti:

1. Test – retest
2. Alternativne (paralelne) oblike
3. Split – half
4. Metode interne konsistentnosti

Precej samosvoja je spet Cattellova sistematizacija, kjer je konsistentnost definirana v generični obliki kot nadredni pojem in vsebuje: (Cattell in drugi 1975)

1. zanesljivost (v času)
2. homogenost (preko delov)
3. transferabilnost (preko različnih vzorcev in kulturnih skupin)

Test – retest metoda ugotavljanja zanesljivosti je že na prvi pogled dokaj sporna, saj je v dobljenem koeficientu korelacije nerazločljivo pomešana stopnja zanesljivosti testa in stopnja stabilnosti dejavnikov, ki vplivajo na konsistentnost tega testa. Majhna test-retest korelacija (stabilnost) ne pomeni nujno slabe zanesljivosti (konsistentnosti) testa. Za nekatere poteze je dokajšnja časovna stabilnost pričakovana, za druge pa sploh ne. Glede koeficientov stabilnosti pri dinamičnih potezah ne smemo pričakovati previsokih vrednosti, saj ima v skladu s teorijo situacijska komponenta pomembno težo v trenutni motivacijski moči nekega faktorja. Visoka stabilnost bi izdala slab in neobčutljiv motivacijski test, ki ne more registrirati sprememb v nivoju ergične tenzije. Nestabilnost v tem primeru torej ne predstavlja lastnosti testa, ampak je odraz fluktuacij v samem nivoju merjene poteze. (Cattell in drugi 1975)

Po drugi strani pa je visoka test-retest korelacija možna celo pri majhni konsistentnosti testa. Test je sestavljen iz določenega števila itemov. Odgovor na vsakem itemu je posledica vplivov mnogih dejavnikov (stabilnih ali nestabilnih). Večino od teh dejavnikov lahko v kontekstu določenega testa imenujemo napaka, ker vplivajo samo na posamezne iteme in ne prispevajo h konsistentnosti testa. Še vedno pa je lahko toliko nesistematičnih dejavnikov stabilnih preko časa, da napihnejo test-retest korelacijo.

Uporabo test-retest zanesljivosti nekateri avtorji opravičujejo na različne načine, npr.: "Nekateri psihometre obtožujejo, da predpostavljajo, da psihološke poteze ostajajo več ali manj stabilne skozi življenje in da se torej ljudi da le malo spremeniti oz. izboljšati. Taka filozofija sploh ni nujna za teorijo napake merjenja. Teorija bi ravno tako držala, če bi se ljudje iz dneva v dan opazno spreminjali; vendar, če bi to držalo, bi se naši poskusi, da o ljudeh sprejemamo praktične odločitve in formuliramo splošne principe človeškega vedenja, spremenili v kaos. Ljudje se spreminjajo, vendar se v večini potez spreminjajo dovolj počasi, da dopuščajo veljavno uporabo psiholoških merskih instrumentov v vsakdanji praksi in v znanstvenih raziskavah." (Nunnally 1978, cf. Cohen, Montague in Nathanson 1988 s.106) Ta prakticistični pogled, ki je sicer v mnogih primerih pravičen, pa nam ne sme zamegliti teoretičnih

predpostavk merske teorije in koncepta zanesljivosti. Test-retest zanesljivost psiholoških instrumentov je bila prav zaradi prakticističnih vzrokov intenzivno raziskovana. Odvisna je od statusa (pacienti ali normalni) in starosti poskusnih oseb, števila itemov, inter-item korelacije in test-retest intervala. Nižja stabilnost je značilna za mere patologije kot za mere normalnih potez. (Schuerger, Zarrella in Hotz 1989)

Interna konsistentnost se nanaša na ujemanje različnih mer istega konstrukta. V tem smislu sta tudi metoda split-half in metoda paralelnih oblik varianti interne konsistentnosti, čeprav precej odstopata.

Metode konsistentnosti med itemi temeljijo na povprečni interkorelaciji itemov neke lestvice. Včasih se imenujejo tudi homogenost testa, vendar ta izraz že implicira predpostavko o enodimenzionalnosti. (Cohen, Montague in Nathanson 1988) Enodimenzionalnost pomeni, da so vse ugotovljene korelacije med itemi posledica tega in zgolj tega, da nanje vpliva isti (merjeni) konstrukt. Metode interne konsistentnosti te predpostavke ne testirajo, zato pa nekatere na njej temeljijo. (Stenkamp in VanTrijp 1991)

Daleč najpogosteje uporabljana mera zanesljivosti je Cronbachov a koeficient, ki je generalizacija Kuder-Richardsonove formule 20 (KR20) za intervalne spremenljivke. Dokazano je, da je a enaka povprečju vseh možnih split-half korelacijskih koeficientov in da statistično predstavlja spodnjo mejo ocene zanesljivosti testa, čeprav še ni jasno, ali gre za oceno zanesljivosti v vzorcu ali v populaciji. (Bentler 1986)

$$KR20 = (k/(k-1))(1 - \sum p_i q_i / \text{var}(\sum x_i))$$

$$\alpha = (k/(k-1))(1 - \sum \text{var}(x_i) / \text{var}(\sum x_i))$$

kjer je k število mer, p_i je proporc tistega odgovora na itemu i, ki šteje v skor, q_i je $1 - p_i$, x_i je rezultat na posameznem itemu in var je varianca.

Cronbachov a koeficient ima vrsto prednosti pred npr. metodo delitve v dva dela (split-half), saj kompenzira napako, ki nastaja zaradi vpliva naključnih dejavnikov pri delitvi skupine mer v dva ali več delov. Na žalost pa raziskovalci, ki ga uporabljajo, pogosto spregledajo, da počiva na vrsti prepostavk, za katere je treba potrebno preveriti, ali držijo. Predvsem α sam po sebi ne preverja predpostavke o enodimenzionalnosti mer, na kateri počiva. Če je skupna varianca v neki skupini mer posledica vpliva več skupnih latentnih dimenzij, bomo dobili visoko oceno zanesljivosti, ki ne ustreza dejanskemu stanju in dejansko predstavlja popolnoma nesmiselno število. Drugi pogoj za uporabo koeficienta a je enaka velikost variance pravega rezultata pri vseh merah v skupini.

Visoka konsistentnost je sicer zaželjena lastnost testa, vendar je visoka konsistentnost pogosto artefakt, ki so ga sestavljenci ustvarili na račun drugih lastnosti testa. Še posebej pri osebnostnih testih je malo verjetno, da so odgovori na posameznih testnih itemih formata papir-svinčnik v veliki meri determinirani prav s proučevanim konstruktom.

Na izraženost latentne poteza lahko sklepamo samo z opazovanjem vedenj, v katerih se odraža (narava, trajanje in raznovrstnost odgovorov in situacije, v katerih se pojavljajo). Bolj kot je spekter situacij širok, z večjo veretnostjo lahko sklepamo, da gre res za neko stabilno karakteristiko posameznikovega vedenja. (Furnham in Henderson 1982)

Podobno mnenje pogosto poudarja Cattell. Visoka homogenost ni nujno zaželjena lastnost faktorske lestvice. Zaradi drugih metričnih kriterijev ponavadi ni zaželjena maksimalna, ampak optimalna homogenost mer (merjena npr. s koeficientom a). Prevelika homogenost ogroža transferabilnost testa in njegovo robustnost. Itemi tes-

ta morajo vzorčiti celotno vsebinsko področje manifestacij merjenega faktorja. Po Cattellovem mnenju je zaželen samo poseben primer homogenosti, namreč homogenost ekvivalentnih oblik, od katerih naj bi vsaka pokrivala celotno področje manifestacij faktorja. (Cattell in drugi 1975) Za to zadnje stališče pa lahko trdimo, da je v luči teorije pravega skora zmotno, saj faktorjem konsistentnosti prišteje tudi napihovanje korelacije med paralelnima formama zaradi nesistematičnih vplivov, ki se podvajajo v obeh paralelnih oblikah.

3.3. KONVERGENTNA IN DISKRIMINATIVNA VELJAVNOST

Kriterij konvergentne veljavnosti zahteva, da merjenje istega konstrukta, kadar uporabimo karseda različne merske metode, da enake ali vsaj podobne rezultate. Diskriminativna veljavnost pa zahteva, da so rezultati merjenja obravnavanega konstrukta različni od drugih sorodnih, vendar različnih konstruktov, tudi če je pri merjenju uporabljena ista metoda. Oba kriterija torej predstavljata zelo različna vidika konstruktne veljavnosti. V istem razdelku ju predstavljamo zato, ker se ju da najbolje preveriti z isto vrsto podatkov in z isto metodo: analizo matrike multiplih potez – multiplih metod (*multitrait-multimethod MTMM*).

MTMM matrike za analizo konvergentne in diskriminativne veljavnosti sta prva uporabila Campbell in Fiske (1959) in se pri tem eksplicitno sklicevala na preverjanje konstruktne veljavnosti. Konvergentna in diskriminativna veljavnost sta zanj dokaz, da je poteza reakcijska tendenca, ki jo je mogoče opazovati v različnih (eksperimentalnih) pogojih in ki jo lahko smiselno razločimo od drugih potez.

Njena izhodišča so bila sledeča:

1. Pri tipični validacijski študiji je v ospredju konvergenca, torej potrditev z neodvisnimi merskimi postopki. Neodvisnost od metode je skupni imenovalec mnogih glavnih tipov veljavnosti.
2. Pri upravičevanju operacionalizacije novega konstrukta (npr. osebnostne poteze) pa se pokaže, da je nujna komponenta konstruktne veljavnosti tudi diskriminativna validacija. Novi konstrukt ali test, ki ga meri, lahko razveljavi premajhna različnost od že uveljavljenih konstruktov in testov.
3. Vsak test oz. mero nasploh lahko imamo za kombinacijo metoda-poteza, za celoto, ki jo posamezna poteza tvori z uporabljenjo metodo. Sistematična varianca testnih rezultatov je lahko tudi posledica lastnosti merskega postopka, ne samo posledica razlik v izraženosti poteze.
4. Če hočemo proučiti diskriminativno veljavnost, in če hočemo oceniti relativni vpliv poteze in metode na varianco mere, ni dovolj, da uporabimo več metod, ampak moramo s temi metodami tudi meriti več (po možnosti sorodnih) potez. Matrika interkorelacij več potez, merjenih z več metodami, ki pa morajo biti enake za vse poteze, je MTMM matrika.

MTMM matrika je sestavljena iz več delov, ki jih pri analizi konvergentne in divergentne veljavnosti primerjamo med sabo. (tabela 2). Glavno diagonalo matrike sestavljajo tri *diagonale zanesljivosti* (a), kjer so navedene zanesljivosti posameznih mer, po možnosti ocenjene z metodo interne konsistentnosti. Zanesljivosti lahko interpretiramo kot *ena poteza – ena metoda* vrednosti MTMM matrike. Pod vsako diagonalo zanesljivosti je trikotnik *različne poteze – ena metoda* (b). Diagonala zanesljivosti in trikotnik *različne poteze – ena metoda* tvorijo blok *ena metoda*. Blok *različne metode* tvorijo *diagonala veljavnosti* (c), v kateri se nahajajo vrednosti *ena*

poteze – različne metode, in dva trikotnika različne poteze – različne metode pod in nad diagonalo veljavnosti (d). Trikotnika pod in nad diagonalo veljavnosti nista simetrična.

Tabela 2: Primer MTMM matrice

	METODA 1			METODA 2			METODA 3		
	JS	MS	SS	JS	MS	SS	JS	MS	SS
METODA 1									
JS	a .98								
MS	b .50	.94							
SS	.42	.59	.96						
METODA 2									
JS	c .65	d .10	.15	a .92					
MS	d .02	.61	.18	b .40	.89				
SS	.04	.10	.66	.60	.52	.86			
METODA 3									
JS	c .59	d .08	.11	c .69	d .00	.04	a .80		
MS	d .01	.55	.09	d .00	.72	.02	b .10	.85	
SS	.06	.11	.58	.02	.05	.68	.50	.46	.88

S stališča tega diagrama lahko vprašanje konvergentne in diskriminativne veljavnosti obravnavamo s štirih vidikov:

1. Koeficienti v diagonalah veljavnosti se morajo statistično pomembno razlikovati od nič in biti dovolj veliki, da je nadaljnja analiza sploh smiselna. Ta kriterij potrjuje konvergentno veljavnost.

2. Koeficienti v diagonalni veljavnosti morajo biti višji kot koeficienti, ki ležijo v isti vrsti in stolpcu pripadajočih trikotnikov različne poteze – različne metode. Vrednost veljavnosti mere mora biti višja kot korelacija te mere z neko drugo, ki nima niti iste metode niti ne meri iste poteze. Ta zahteva se zdi tako minimalna in tako očitna, da se je skoraj ne zdi vredno omenjati, vendar Campbell in Fiske navajata, da so v literaturi pogosti primeri, ko niti ta zahteva ni izpolnjena.

3. Tretja dovolj očitna zahteva je, da dve meri iste poteze, dobljeni z različnima metodama, korelirata višje kot mere različnih potez, merjene z isto metodo. Za vsako mero moramo torej primerjati njene vrednosti v diagonalni veljavnosti z njenimi vrednostmi v trikotniku različne poteze – ena metoda. Campbell in Fiske navajata, da je kršenje tega pravila pri raziskovanju individualnih razlik prej tipično kot redkost.

4. Približno enak vzorec korelacij med potezami se mora pojavljati v vseh trikotnikih več potez – ena metoda in trikotnikih več potez – več metod. Če torej poteza MS korelira s potezo SS višje kot s potezo JS, se mora ta odnos pokazati v vseh trikotnikih. Odstopanja od tega pravila morajo biti v okviru napake vzorčenja in napake nezanesljivosti.

3.4. FAKTORJI METODE PRI PSIHOLOŠKEM MERJENJU

Vsaka merjena spremenljivka je kombinacija poteza-metoda. V vsakem psihološkem merskem instrumentu so določene lastnosti stimulusov namerno zasnovane tako, da zastopajo potezo, ki naj bi jo test meril. Druge lastnosti stimulusov pa zastopajo uporabljeno metodo in so lahko neodvisno prisotne v instrumentih za merjenje popolnoma različnih potez. (npr. petstopenjska Likertova lestvica, "res – ni res" format odgovorov itd.) Test, ocenjevalna lestvica ali druge mere skoraj brez izjeme povzročajo sistematično varianco v odgovorih zaradi obeh skupin lastnosti. Do tiste mere, do katere irelevantna varianca metode prispeva k dobljenim skorom, so ti skorji neveljavni. Ta vir neveljavnosti je prvi odkril Thorndike (1920) v obliki halo efekta pri ocenjevanju. Raziskave na laboratorijskih živalih so odriale močne vplive aparata, ki so bili ponavadi močnejši kot faktorji psiholoških procesov. (Tyron 1942). Pri instrumentih tipa 'papir in svinčnik' se je varianca metode pojavljala v obliki faktorjev testne forme (Vernon 1957) ali odgovornih setov in slogov.

S problemom variance metode se je prvi sistematično začel ukvarjati Cronbach (1941, cf. Rorer 1965), ko je proučeval zakonitosti ugibanja na testih znanja. Ugotovil je, da nekateri subjekti pri nalogah, pri katerih ne poznajo pravega odgovora, konsistentno izbirajo pozitivne odgovore (da, res je, se strinjam). Za to tendenco, ki jo je imenoval popustljivost (acquiescence), je menil, da je osebna lastnost prav tako kot npr. prijaznost ali veselost. V svojih kasnejših člankih je Cronbach spoznal, da popustljivost predstavlja samo enega od množice možnih odgovornih setov. Odgovorni set je definiriral kot vsako tendenco, ki povzroči, da oseba na testne iteme konsistentno odgovarja drugače, kot bi, če bi ji bila testna vsebina predstavljena v drugačni obliki. Glede tega je zapisal (Cronbach 1946, cf. Campbell in Fiske 1959): "Ponavadi predpostavljamo, ...da test meri tisto, kar določa vsebina itemov. Vendar je končni rezultat...kompozit učinkov, rezulturajoč iz forme prav tako kot iz vsebine itemov."

Kasneje so drugi avtorji odkrili izredno pomembne vplive faktorjev metode na rezultate osebnostnih vprašalnikov. Ugotovili so npr., da je treba glavne skupne faktorje osebnostnih inventorijskih, kot so MMPI ali California Personality Inventory, ki uporabljajo format "je res / ni res" ("da/ne") ali "se strinjam / se ne strinjam", interpretirati v smislu odgovornega seta in sloga, ne pa v smislu vsebine itemov. (Rorer 1965) Nasploh se je pokazalo, da je abstraktne konstrukte mnogo težje meriti kot konkretne in da so abstraktni konstrukti še bolj podvrženi vplivu napake merjenja. Pri nekaterih merjenjih stališč je npr. sistematična napaka metode prispevala 70% variance (Cote and Buckley 1987, cf. Bagozzi in Yi 1991).

Rorer (1965) je predlagal, da naj se namesto izraza odgovorni set uporablja izraz odgovorni slog. Izraz set naj bi tudi v psihometričnem smislu uporabljali konsistentno z njegovim pomenom na drugih psiholoških področjih. Tam set pomeni, da je respondent na nek način, zavedno ali nezavedno, motiviran za to, da bi ustvaril določeno podobo o sebi, torej za določeno vrsto odgovora, npr., da bo zaradi kakršnegakoli razloga izgledal boljši, bolj bolan, bolj agresiven itd. Za odgovorne tendence, za katere ni mogoče najti neke eksplicitne ali implicitne motivacije, zato Rorer predlaga izraz odgovorni slogi. Za slog je značilno, da ni odvisen od vsebine itema, ampak od formalnih značilnosti itema in odgovornih kategorij. Zato tudi velja, da bolj kot je eksplicitna in transparentna vsebina itema, večjo vlogo pri izkrivljanju odgovorov bodo odigrali seti. Po drugi strani pa imajo pri bolj implicitni, skriti vsebini itemov večji pomen odgovorni slogi. Za vpliv setov je potrebno vsaj nekaj

vsebine itema, za vpliv slogov pa vsaj nekaj nejasnosti oz. nedoločnosti, večsmiselnosti (ambiguity) itema. (Rorer 1965)

Napako merjenja lahko razložimo v naljučno in sistematično komponento (varianca metode). Termin metoda se nanaša na obliko merjenja na različnih nivojih abstraktnosti, npr. vsebina specifičnih itemov, vrsta lestvice, sintaktične in pragmatične lastnosti stavčnih konstrukcij, tip lestvice, format odgovora in splošni kontekst. Na bolj abstraktnem nivoju predstavljajo faktorji metode izkrivljenosti (biases) odgovorov, kot so halo efekt, socialna zaželenost, nagnjenost k da– oz. ne– odgovorom, popustljivost in podredljivost.

Vsaka od dveh komponent napake ima lahko resne posledice, ki zamegljijo empirične rezultate in privedejo do napačnih zaključkov. Naključna napaka oslabi opažene povezave med spremenljivkami. Zato naključna napaka vodi v napake inference tipa 2 oz. b. V klasični teoriji zanesljivost tudi varianca metode prispeva h konsistentnosti mer in se zato obravnava kot del pravega skora. Zato so vplivi sistematične napake metode pogosto spregledani. Sistematična napaka metode lahko napihne ali zmanjša opažene korelacije med konstrukti. Bolj natančno, vplivi metode napihnejo korelacije med dvema konstruktoma, kadar je dejanska korelacija med uporabljenima metodama večja od dejanske korelacije med proučevanima konstruktoma. To pomeni, da uporaba podobnih metod skoraj vedno napihne korelacijski koeficient. Uporaba enake metode (korelacija = 1) pa korelacijo napihne v vsakem primeru. Po drugi strani pa bodo faktorji metode korelacijo med konstrukti zmanjšali, kadar je korelacija med metodama nižja kot dejanska korelacija med konstruktoma.

Ker kontaminacija zaradi naključne napake in faktorjev metode ogroža veljavnost izsledkov raziskave, je pomembno, da razvozljamo popačujoče učinke teh dejavnikov, predno testiramo teorijo. To lahko dosežemo z uporabo MTMM matrik. Taka triangulacija omogoča dekompozicijo informacije v tri vire: prispevek naključne napake, prispevek sistematičnih dejavnikov metode in čisti prispevek proučevane poteze (konstrukta).

V MTMM matriki se varianca metode kaže v razliki med korelacijami v trikotniku različne poteze – ena metoda in paralelnimi korelacijami v trikotnikih različne poteze – različne metode. Ob odsotnosti faktorjev metode in ob približno enakih zanesljivostih testov bi te vrednosti morale biti približno enake.

Diskriminativna validacija. Običajni razlog za sodbo o neveljavnosti testa so nizke korelacije v diagonalni veljavnosti, torej slaba konvergentna veljavnost. Sodba o neveljavnosti pa je lahko tudi posledica previsoke korelacije z merami, ki naj bi merile druge poteze. Simptoma nezadostne diskriminativne veljavnosti sta naslednja:

- Vrednosti v trikotnikih različne poteze – različne metode so tako visoke kot ustrezne vrednosti v diagonalni veljavnosti.
- Vrednosti v trikotniku različne poteze – ena metoda so tako visoke kot zanesljivosti.

MTMM matrike v literaturi

MTMM matrike bolj poredko srečamo celo v metodološki oz. psihometrični literaturi. Še posebej redke pa so v literaturi, ki se ukvarja z vsebinskimi vprašanji. Če se avtorji že potrudijo za poglobljenejšo validacijo obravnavanih konstruktov, uporabijo fragmente MTMM matrike: dve ali tri metode in eno potezo (diagonala veljavnosti, dopolnjena z nekaj zanesljivostmi) ali nekaj potez, merjenih z isto ali vsako s po eno različno metodo. Ne en ne drug način ne omogočata razločevanja

med vplivi metod in vplivi potez, zato tudi ne moreta nuditi zadostnega dokaza za konvergentno in diskriminativno veljavnost. Analiza tistih dotlej publiciranih podatkov, iz katerih je bilo mogoče konstruirati MTMM matrike, pa je pokazala, da je uporaba te metodologije še kako potrebna, saj se je v večini študij pokazal zelo močan vpliv metode, medtem ko v nekaterih študijah sploh ni bilo mogoče identificirati vpliva poteze.

Eden najpomembnejših vzrokov za to, da se MTMM matrike tako redko uporabljajo, so pomanjkljivosti pristopa Campbella in Fiskea, ki jih navajata Bagozzi in Yi (1990):

1. Odsotnost kvantifikabilnih kriterijev.
2. Ne moremo upoštevati vpliva različnih zanesljivosti.
3. Implicitne predpostavke analize so take, da so v praksi dokaj poredko dosežene, še posebej predpostavka o medsebojni neodvisnosti metod (odsotnost korelacij med metodami), v določenih primerih pa tudi predpostavka o medsebojni neodvisnosti metod in potez (nekatero poteze vplivajo na določene odgovorne sete).
4. Analiza Campbella in Fiskea se konča pri validaciji oz. invalidaciji, ne daje pa opore za nadaljnjo analizo teoretičnega modela odnosov med konstrukti. Če z analizo korelacij v MTMM matriki ugotovimo, da je vpliv variance napake večji od vpliva prave variance poteze, se tu zgodba konča: nismo več upravičeni testirati odnosov med konstrukti. Pa tudi, če je varianca napake manjša in torej smemo naprej, v nadaljnji analizi ne moremo izolirati čistih vplivov med konstrukti, ampak nam jih varianca metode in naključna napaka še vedno zamegljujeta.

Treba je poudariti, da se te kritike nanašajo zgolj na analitični postopek Campbella in Fiskea (interpretacija korelacij), ne pa na njune osnovne zamisli (MTMM matrika, konvergentna in diskriminativna veljavnost), ki so absolutno upravičene in fundamentalne.

3.5. VSEBINSKA VELJAVNOST

Vsebinska (content) veljavnost zahteva, da so merjena vedenja reprezentativen vzorec manifestacij merjenega atributa (konstrukta). Pri tem je pomembna pokritost vsebinskega področja, pa tudi oblika merjenja manifestacij atributa – test papir-svinčnik lahko npr. dobro kaže delavčevo znanje o delu, ne pa njegovo sposobnost za opravljanje dela.

Vsebinsko področje (domain) predstavlja celotno množico vedenj, v katerih se manifestira nek konstrukt. Ima svoje meje (boundaries) in svojo strukturo. Detajlna primerjava med strukturo vsebinskega področja in strukturo testa nam da osnovo za ugotavljanje vsebinske veljavnosti. (Murphy in Davidshofer 1988) Vsebinsko veljavnost ugotovimo tako, da analiziramo načrt vzorčenja, ki je vodil sestavo testa. Vsebinsko področje, njegovo strukturo in meje je mogoče dokaj dobro opredeliti pri testih znanja ali usposobljenosti za delo. Dosti težje ga je opredeliti pri testih sposobnosti. Vsebinsko področje za osebnostne teste je skoraj nemogoče definirati, kaj šele strukturirati. (Cohen, Montague in Nathanson 1988) To pa so ravno tisti merjeni atributi, kjer se najbolj resno postavlja vprašanje konstruktne veljavnosti. Zato nekateri avtorji še vedno zmotno obravnavajo vsebinsko in konstruktno validacijo kot alternativna pristopa. Vendar je tudi pri merjenju latentnih potez kot del postopka konstruktne validacije nujno vsaj oceniti, koliko uporabljene mere ustrezajo splošno sprejetim definicijam merjene poteze in njenih vedenjskih manifestacij.

Čeprav so nekateri avtorji predlagali izračun koeficienta vsebinske veljavnosti, ki kaže, v kolikšni meri instrument pokriva relevantne vsebine, se ta v praksi skoraj ne uporablja. Še posebej za instrumente, ki niso testi znanja, je zahtevane vhodne parametre tega koeficienta nemogoče oceniti. Vsebinska veljavnost je po svojem bistvu veljavnost merjenja, ne pa veljavnost odločitev, sprejetih na osnovi rezultatov merjenja. (Murphy in Davidshofer 1988)

3.6 PREDIKTIVNA VELJAVNOST

Prediktivna veljavnost je tisti pojem, preko katerega začetni kurzi psihološke metodologije ponavadi uvedejo problematiko veljavnosti in mnogi raziskovalci kar ostanejo pri enačenju obeh pojmov. Vendar to ne pomeni, da je vprašanje prediktivne veljavnosti nedvosmiselno rešeno. Pri prediktivni veljavnosti moramo ločiti dve, temeljno različni pojmovanji, ki pa jih avtorji na žalost običajno zamenjujejo. Prvo pojmovanje je implicitno v priporočilih glede validacije, ki jih je izdalo Ameriško psihološko združenje. To obravnava prediktivno veljavnost kot eno od dveh podkategorij s kriterijem povezane veljavnosti (criterion-related). Čeprav tudi ta pristop skriva zelo kompleksno problematiko in mnogo subtilnih detajlov (Rosenbaum 1989), lahko njegovo logiko na kratko predstavimo takole:

Veljavnost testa dokažemo tako, da njegove rezultate koreliramo z neko drugo mero. Da bi ta korelacija res odražala veljavnost proučevanega test, mora biti uporabljena mera (kriterij) boljša mera merjenega konstrukta kot obravnavani test. Po možnosti mora ta mera čisto, brez napake odslikavati merjeni konstrukt. Cronbach (1971) to definira takole: S-kriterijem-povezana veljavnost primerja testne skore ali predikcije, narejene na njihovi podlagi, z zunanjo spremenljivko, za katero menimo, da predstavlja direktno mero obravnavanega vedenja ali karakteristike. (cf. Rosenbaum 1989)

Glede na to, ali je kriterij merjen istočasno s testom ali kasneje, govorimo o sočasni ali prediktivni veljavnosti. Cohen, Montague in Nathanson (1988) po drugi strani trdijo, da časovna sukcesija sploh ni bistvena in da je glavna razlika v tem, da imamo pri sočasni veljavnosti že vnaprej sistematično selekcioniran vzorec, z vsemi posledicami, ki jih to ima za korelacijske koeficiente. Prediktivna validacija pa naj bi bila opravljena na reprezentativnem, neselekcioniranem vzorcu. Že iz samih teh težav je jasno, da je s to, največkrat uporabljeno konceptualizacijo prediktivne veljavnosti nekaj narobe. Če se naš test imenuje "test potencialne šolske uspešnosti", res lahko trdimo, da korelacija z dejanskim šolskim uspehom pomeni validacijo tega testa. Če pa se naš test imenuje "test splošne inteligentnosti", je taka interpretacija popolnoma nesmiselna. Kdor namreč trdi, da je korelacija tega testnega rezultata s šolskim uspehom koeficient veljavnosti, s tem implicitno trdi, da je šolski uspeh boljša mera inteligentnosti kot pa uporabljeni test. Očitno je torej, da je ta, običajna konceptualizacija prediktivne veljavnosti uporabna zgolj za mere strogo operacionalno definiranih spremenljivk, nikakor pa ne za mere hipotetičnih konstruktov.

Druga konceptualizacija prediktivne veljavnosti, ki se je držim tudi v tem delu, pojmuje prediktivno veljavnost v bistvu kot del nomološke veljavnosti. Prediktivna validacija se nanaša na vprašanje, koliko mere obravnavanega konstrukta korelirajo z merami konstruktov, na katere ta konstrukt v skladu z našo teorijo vzročno deluje, in z merami tistih konstruktov, ki vzročno delujejo nanj. V prvem primeru govorimo o konsekvenci prediktivni veljavnosti in v drugem primeru o antecedenci prediktivni veljavnosti.

Tu torej ne trdimo, da druge mere, s katerimi naš test koreliramo, merijo isti konstrukt. Tudi ni nujno, da visoke korelacije pomenijo večji dokaz veljavnosti. Korelacije morajo biti čim bližje tistemu velikostnemu redu vplivov, ki ga predpostavlja naša teorija. Pri interpretaciji teh korelacijskih koeficientov, kakor tudi pri interpretaciji koeficientov nomološke veljavnosti, pa moramo seveda upoštevati atenuacijo korelacij, ki jo povzroča nepopolna zanesljivost merjenja.

3.7. NOMOLOŠKA VELJAVNOST

Nomološka veljavnost se nanaša na potrditev predpostavljenih odnosov obravnavanega konstrukta z drugimi konstrukti v nomološki mreži. Pomemben del nomološke veljavnosti predstavljata antecedenčna in konsekvenčna prediktivna veljavnost. Vendar se nomološka veljavnost ne nanaša samo na tiste konstrukte, ki so z direktnimi kavzalnimi vplivi povezani z obravnavanim konstruktom, ampak tudi na tiste, kjer so ti kavzalni vplivi indirektni, na tiste, ki skupaj z obravnavanim konstruktom vplivajo na tretje konstrukte, na tiste, s katerimi ima obravnavani konstrukt skupne antecedente, in na tiste, kjer pričakovana asociacija s stališča kavzalnosti ostaja nepojasnjena. Nomološka veljavnost se običajno preverja s korelacijskimi in regresijskimi koeficienti. Tudi tu moramo biti pozorni na prilagoditev ugotovljenih korelacij zaradi nepopolne zanesljivosti merjenja.

4. UPORABA KONSTRUKTNE VALIDACIJE

Sam pojem konstruktne veljavnosti je bil doslej samo delno elaboriran, tako konceptualno kot matematično (Rosenbaum 1989). Tako navedene opredelitve, kot tudi klasifikacije komponent konstruktne veljavnosti se precej razlikujejo. Lahko bi rekli, da avtorji nizajo potrebne pogoje konstruktne veljavnosti, nikjer pa niso dokončno definirani zadostni pogoji. Dejstvo je, da je bila teorija konstruktne veljavnosti, kot sta jo zasnovala Cronbach in Meehl, sicer revolucionarna in daljnosežna, a vse preveč nedoločna glede operacionalizacije, da bi si sestavljanci testov z njo dosti pomagali. (Bentler 1978) Predvsem je manjkala primerna metoda za sistematično in sintetično implementacijo teorije konstruktne veljavnosti. Ta situacija se je kasneje spremenila. Bentler je že leta 1978 spoznal, da metodologija modelov strukturnih enačb predstavlja za družboslovne znanstvenike orodje, s katerim težave pri operacionalizaciji koncepta konstruktne veljavnosti lahko premagajo.

Pomanjkljive smernice za operacionalizacijo koncepta konstruktne validacije so privedle do tega, da so uporabljeni postopki dokazovanja konstruktne veljavnosti v veliki večini primerov delni, pomanjkljivi, enostranski, v nekaterih primerih pa celo popolnoma zgrešeni. Pregled člankov z raznih področij psihološke znanstvene literature zadnjih let, v katerih avtorji trdijo, da so ugotavljali ali celo dokazali konstruktno veljavnost določenih spremenljivk, nudi dovolj opore za gornjo trditev.

Tako npr. Watson, Tilleskjer in Jacobs (1990) kot dokaz konstruktne veljavnosti preprosto navajajo korelacije z eksternimi kriteriji, brez specifikacije, ali gre za antecedenčne ali konsekvenčne spremenljivke, ali gre za kakšen drug odnos med teoretičnimi spremenljivkami. Poleg tega ne vključujejo nobenega dokaza za diskriminativno veljavnost svojega konstrukta, čeprav se zdi dokaj podoben nekaterim drugim, ki jih uporabljajo kot kriterije. Leroux in drugi (1990) podobnemu postopku računanja korelacij z eksternimi kriteriji dodajajo še podatke o zanesljivosti, vendar

se ne potrudijo niti toliko, da bi korelacijske koeficiente korigirali zaradi nepopolne zanesljivosti merjenja. Cianni in Horan (1990) sprejemata kot kriterij konstruktne veljavnosti že zgolj dejstvo, da eksperimentalni tretma povzroči predvidene učinke v obravnavani spremenljivki. Podobno lahko trdimo za študijo Smitha, Sandersa in Alexandra (1990), ki sicer zaupanje v svoje mere potrdijo s podatki o zanesljivosti, nerazčističene pa ostanejo vse dileme okrog ugotavljanja zanesljivosti, ki smo jih navedli v prejšnjih delih tega poglavja. Meier (1991) je z uporabo postopnega vključevanja neodvisnih spremenljivk v multiplo regresijsko enačbo in s pomočjo kriterija dodatne pojasnjene variance v odvisni spremenljivki ugotavljal diskriminativno veljavnost obravnavanega konstrukta. Čeprav je metoda delno adekvatna, pa seveda ne upošteva nikakršnega teoretskega modela odnosov med konstrukti, ne upošteva odnosov med konstruktom in njegovimi merami, pa še cele vrste drugih kriterijev konstruktne veljavnosti. Kremer (1990) je uporabil za ugotavljanje konvergentne in diskriminativne veljavnosti MTMM matriko, vendar se je je lotil v skladu s klasičnim postopkom Campbella in Fiskea (1959), zato je bil v naslednji fazi prisiljen iskati latentno strukturo, ki se skriva za podatki, s pomočjo eksploratorne faktorске analize, in jasno obtičal brez zaključkov. Nasploh dosti raziskovalcev enači konstruktno veljavnost svojih spremenljivk s skladnostjo med predvideno in z eksploratorno faktorško analizo dobljeno faktorško strukturo svojih podatkov (npr. Klemchuk, Hutchinson in Frank 1990, Cranton in Smith 1990). Še najbolj ugoden vtis od vseh pregledanih člankov zadnjih dveh let, ki jim je konstruktna veljavnost ena od ključnih besed, z metodološkega vidika pušča študija Kinga in Kinga (1991), čeprav zgolj kvalitativno kompilira rezultate drugih študij o sindromu vietnamskih veteranov. Celovit in sistematičen pregled vseh vidikov konstruktne veljavnosti več kot odtehta odsotnost kvantitativne metodologije sintetiziranja teh rezultatov. Po drugi strani pa O'Brienova (1990) študija dokazuje, da tudi aplikacija LISREL sama po sebi, če upošteva samo nekatere vidike konstruktne veljavnosti in predvsem ne modelira vpliva metode, lahko da samo delno, in še to hudo napačno oceno.

5. VIRI

Bagozzi, R.P. (1984). A prospectus for theory construction in marketing. *Journal of marketing*, 48, 11-29.

Bagozzi, R.P. in Y. Yi (1990). Assessing method variance in multitrait-multimethod matrices: The case of self-reported affect and perceptions at work. *Journal of applied psychology*, 75/5, 547-560.

Bagozzi, R.P. in Y. Yi (1991). Multitrait-multimethod matrices in consumer research. *Journal of consumer research*, 17, 426-439.

Bentler, P.M. (1978). The interdependence of theory, methodology, and empirical data: Causal modeling as an approach to construct validation. V D.B. Kandel (ur.), *Longitudinal research on drug use: Empirical findings and methodological issues*. Washington: Hemisphere publishing corporation.

Bentler, P.M. (1986). Structural modeling and psychometrika: An historical perspective on growth and achievements. *Psychometrika*, 51/1, 35-51.

Block, J. (1990). More remarks on social desirability. *American psychologist*. 45, 1076-107.

Campbell, D.T. in D.W. Fiske (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56/2, 81-105.

Cattell, R.B., J.L. Horn, A.B. Sweney in J.A. Radcliffe (1975). Motivation analysis test. manual, Champaign, IL, IPAT.

Cianni, M. in J.J. Horan (1990). An attempt to establish the experimental construct validity of cognitive and behavioral approaches to assertiveness training. *Journal of counseling psychology*, 37/3, 243-247.

Cohen, R.J., P. Montague, L.S. Nathanson in M.E. Swerdlik (1988). *Psychological testing*. Mountain view, CA: Mayfield publishing company.

Cranton, P. in R.A. Smith (1990). Reconsidering the unit of analysis: A model of student ratings of instruction. *Journal of educational psychology*, 82/2, 207-212.

Cronbach, L.J. in P.E. Meehl (1955). Construct validity in psychological tests. *Psychological bulletin*, 52/4, 281-302.

Furnham, A. in M. Henderson (1982). A content analysis of four personality inventories. *Journal of clinical psychology*, 38/4, 818-825.

King D.W. in L.A. King (1991). Validity issues in research on Vietnam veteran adjustment. *Psychological bulletin*, 109/1, 107-124.

Klemchuk, H.P., C.B. Hutchinson in R.I. Frank (1990). Body dissatisfaction and eating-related problems on the college campus: Usefulness of the eating disorder inventory with a nonclinical population. *Journal of counseling psychology*, 37/3, 297-305.

Kremer, J.F. (1990). Construct validity of multiple measures in teaching, research, and service and reliability of peer ratings. *Journal of educational psychology*, 82/2, 213-218.

Lea, S.E.G., Tarpy R.M. in P. Webley (1987). *The individual in the economy*. Cambridge university press.

Leroux, M.D., K.R. Vincent in R.H. McPherson (1990). Construct validity of the diagnostic inventory of personality and symptoms: External correlates. *Journal of clinical psychology*, 46/3, 285-291.

Meier, S.T. (1991). Tests of the construct validity of occupational stress measures with college students: Failure to support discriminant validity. *Journal of counseling psychology*, 38/1, 91-97.

Murphy, K.R. in C.O. Davidshofer (1988). *Psychological testing*. Prentice-Hall international.

O'Brien, T.P. (1990). Construct validation of the Gregorc Style Delineator: An application of lisrel 7. *Educational and psychological measurement*, 50, 631-636.

Rorer, L.G. (1965). The great response-style myth. *Psychological bulletin*, 63/3, 129-156.

Rosenbaum, P.R. (1989). Criterion-related construct validity. *Psychometrika*, 54/4, 625-633.

Schuerger, J.M., K.L. Zarrella in A.S. Hotz (1989). Factors that influence the temporal stability of personality by questionnaire. *Journal of personality and social psychology*, 56/5, 777-783.

Smith, T.W., J.D. Sanders in J.F. Alexander (1990). What does the Cook and Medley hostility scale measure? Affect, Behavior, and attributions in the marital context. *Journal of personality and social psychology*, 58/4, 699-708.

Steenkamp, J.-B.E.M. in H.C.M. van Trijp (1991). The use of LISREL in validating marketing constructs. *International journal of research in marketing*, 8/4, 283-299.

Vodopivec, B. (1988). Epistemološki vidiki analize kovariančnih struktur. *Anthropos*, 4-6, 46-57.

Watson, C.G., C. Tilleskior in L. Jacobs (1990). The construct validity of an after-effect-based subtyping system for alcoholics. *Journal of clinical psychology*, 46/4, 507-517.