Dr. Janez Sagadin

Veljavnost kvantitativnih empiričnih raziskav na vzgojno-izobraževalnem področju

Povzetek: Za veljavno šteje avtor raziskavo, ki nas pripelje do veljavnih izsledkov (ki držijo, so verodostojni). Opredeljuje tele pogoje za veljavnost kvantitativnih empiričnih raziskav vzgoje in izobraževanja: veljavnost postopkov zbiranja podatkov/merjenja kot pogoj za veljavnost zbranih podatkov, veljavnost postopkov analize podatkov kot pogoj za veljavnost rezultatov analize, veljavnost posplošitev in veljavnost vzročnih razlag ali sklepov. Nakazuje, kako tem pogojem zadostiti in v čem so vzroki motenj veljavnosti v okviru posameznih pogojev. Ugotavlja, da se lahko popolni izpolnitvi pogojev in s tem tudi popolni veljavnosti raziskav le bolj ali manj približamo. (Ne omejuje pa navedenih pogojev in tudi sicer ne vsega metodološkega v svoji razpravi samo na raziskave vzgoje in izobraževanja.) V zvezi s prvim pogojem posebno opozarja na prednost merskega modela konfirmatorne faktorske analize v primerjavi z merskim modelom klasične testne teorije. Pri obravnavi drugega pogoja je poudarek na veljavnosti statističnih postopkov analize podatkov (statistična veljavnost raziskave). V okviru obravnave tretjega pogoja so v ospredju pogoji za veljavnost posplošitev z verjetnostnih vzorcev na ciljne populacije (statistična indukcija), poudarjen pa je tudi vsebinski vidik posploševanja. Za neverjetnostne vzorce je nakazana pot do notranje veljavnosti ugotovitev in arbitrarnost sklepanja na ciljne populacije (ker uporaba statistične indukcije pri teh vzorcih ni možna). Razprava o veljavnosti posplošitev pri eksperimentalnih raziskavah pa je organsko vpeta v obravnavo veljavnosti vzročnih razlag. Avtor navaja pogoje za obstoj vzročne zveze med dvema spremenljivkama oziroma kriterije za presojo, ali gre za tako zvezo. Pove, da je za kvantitativne empirične raziskave značilen induktivno-deduktivni pristop ali postopek: najprej deduktivno pridobljena in teoretsko preverjena (verificirana) kavzalna hipoteza (hipoteza o vzročni zvezi) in nato empirično preizkušanje te hipoteze, namreč preizkušanje, ali ima hipoteza (in vzročna razlaga v njej) podporo v empiričnih podatkih, ustrezno zbranih in analiziranih. Če jo ima, štejemo v tem smislu vzročno razlago za veljavno. Empirično preizkušanje kavzalnih hipotez je v prispevku podrobneje prikazano za eksperimentalne raziskave (za slučajnostne in za neslučajnostne eksperimente, zlasti za tiste, ki so bolj uporabni na vzgojno-izobraževalnem področju). Poudarjena je vloga notranje veljavnosti eksperimenta pri tem preizkušanju. Za neeksperimentalne raziskave je omenjeno preizkušanje nakazano v krajši obliki, tudi postopki, ki jih opredeljuje SEM (S*tructural* E*quation* M*odel(l)ing*, modeliranje strukturnih enačb).

Ključne besede: veljavnost (raziskave, postopkov zbiranja podatkov/merjenja, podatkov, postopkov analize podatkov, empiričnih ugotovitev analize, posplošitev, notranja, zunanja, vzročnih razlag), statistična veljavnost raziskave, statistična indukcija, kavzalna hipoteza, empirično preizkušanje kavzalnih hipotez.

UDK: 37.012

Izvirni znanstveni prispevek

Dr. Janez Sagadin, redni profesor, v pokoju, Univerza v Ljubljani, Filozofska fakulteta, Oddelek za pedagogiko in andragogiko, Slovenija

SODOBNA PEDAGOGIKA 3/2009, 140–157

Uvod[1]

Raziskava, ki je veljavna, nas pripelje do *veljavnih izsledkov* (ugotovitev, rezultatov), to je do izsledkov, ki držijo, so verodostojni. Tako je z veljavnostjo tudi pri empiričnih raziskavah na vzgojno-izobraževalnem področju oziroma pri empiričnih raziskavah vzgoje in izobraževanja. Prizadevanje za čim višjo veljavnost raziskave se začne že, ko se odločamo, opredeljujemo in načrtujemo, *kaj in kako bomo raziskovali*, in pripravljamo vse potrebno za začetek raziskave, ter se nadaljuje v izvajanju raziskave. Kvantitativno in kvalitativno empirično raziskovanje imata glede tega sicer vsako svoje posebnosti, najdemo pa med njima tudi podobnosti/analogije. Pri obeh so pogoji za njuno veljavnost: *veljavnost postopkov zbiranja podatkov kot pogoj za veljavnost zbranih podatkov, veljavnost postopkov analize podatkov kot pogoj za veljavnost rezultatov analize, veljavnost posplošitev* (kadar posplošujemo) *in veljavnost vzročnih razlag ali sklepov* (kadar gre za vzročne razlage ali sklepe). Glede odgovorov na vprašanje, kako tem pogojem zadostiti, pa se omenjeni vrsti raziskovanja bistveno razlikujeta med seboj. V tem prispevku se bomo posvetili odgovorom na to vprašanje in problemom v zvezi z veljavnostjo le za kvantitativno raziskovanje in še za to samo toliko, kolikor dopušča obseg enega samega prispevka.

Veljavnost podatkov. Veljavnost postopkov zbiranja podatkov

Temelj veljavnosti vsake empirične raziskave, tako kvantitativne kot kvalitativne ali v tem pogledu mešane raziskave, je *veljavnost empiričnih podatkov*, na katerih gradimo raziskavo. Tako je, ne glede na to, kakšne ontološke in epistemološke nazore imamo in katero raziskovalno metodo ali metode uporabimo pri raziskavi. To pa ne pomeni, da cilji raziskav ter potrebni podatki in postopki njihove validacije niso prav nič odvisni od omenjenih nazorov.

Do veljavnih podatkov nas lahko pripeljejo le veljavni postopki zbiranja podatkov ali merjenja. Za področje vzgoje in izobraževanja so značilni pojavi, ki niso neposredno merljivi in o katerih govorimo kot o *latentnih (konstruktnih) spremenljivkah*. Latentne spremenljivke (znanje, sposobnosti, spretnosti, stališča, interese, motivacijo, učne navade itn.) merimo posredno z indikatorji, imenovanimi *manifestne (indikatorske, merske) spremenljivke*, kot so rezultati (za naloge) pri različnih testih in (za postavke) na lestvičnih instrumentih, odgovori na različnih vprašalnikih itn. Tako so latentne spremenljivke, operacionalizirane z manifestnimi spremenljivkami. Preveč bi bilo pričakovati, da bo operacionalizacija lahko tako popolna, da bodo merske spremenljivke v celoti izražale latentne spremenljivke. Menim, da je v tem primarni (seveda pa ne edini) izvor omejitev veljavnosti podatkov. Prizadevati pa si moramo, da bi opravili operacionalizacijo čim bolje. Osrednje mesto imajo pri operacionalizaciji *instrumenti za zbiranje po-*

[1] V uvodu in nadaljevanju prispevka bom uporabil tudi dele povzetka svojega predavanja na Pedagoško-andragoških dnevih 2009 (Sagadin 2009). Avtor.

datkov/merjenje, s svojimi *merskimi značilnostmi.* Pri preverjanju in zagotavljanju veljavnosti instrumentov se ne moremo in ne smemo zanašati in opirati samo na njihovo zanesljivost. Zanesljivost je sicer pogoj za veljavnost, a ne zadostni pogoj. Če instrumenta ne sestavimo in validiramo dovolj pretehtano in temeljito, da bi mogel dovolj prinašati podatke, ki jih od njega želimo, nas niti njegova popolna zanesljivost, ki je sicer neuresničljiv ideal, ne more ali ne bi mogla obvarovati pred njegovo pomanjkljivo veljavnostjo. Naj posebej opozorimo na *merski model konfirmatorne faktorske analize* v okviru SEM (S*tructural* E*quation* M*odel(l)ing,* modeliranje strukturnih enačb), ki prinaša napredek v primerjavi s klasičnimi metodami ocenjevanja zanesljivosti in veljavnosti merskih instrumentov,[2] to je z metodami klasične testne teorije.[3] Izhodišče napredka je v tem, da so sestavni del merskega modela konfirmatorne faktorske analize tudi latentne spremenljivke (ne le manifestne, kot je pri klasičnih metodah); to omogoča, da se pri presojanju in določanju zanesljivosti in veljavnosti instrumentov opremo na regresijo med merskimi in latentnimi spremenljivkami (več glej v Blunch 2008, str. 127–156). Popolne zanesljivosti in veljavnosti instrumentov pa seveda tudi po tem modelu ne moremo doseči.

Veljavnost postopkov analize podatkov. Statistična veljavnost raziskave

Zbrane podatke je treba obdelati/analizirati. Postopki analize morajo biti veljavni, da nas pripeljejo do veljavnih *empiričnih* ugotovitev. Za kvantitativno raziskovanje so značilne *statistične metode* analize podatkov. Če so v instrumentih za zbiranje podatkov kakšne postavke, pri katerih odgovarjajo respondenti s svojimi besedami, je treba pred statistično analizo opraviti dovolj veljavno kategorizacijo odgovorov. Na statistične metode se ne spomnimo šele takrat, ko je na vrsti analiza podatkov, ampak je načrtovanje statistične analize podatkov integralni del načrtovanja raziskave, ko namenu/ciljem raziskave ustrezno načrtujemo potrebne podatke in postopke njihovega zbiranja, spremenljivke (ustrezne podatkom) ter postopke analize podatkov (za različne vrste spremenljivk); pri vzorčnih raziskavah vse to načrtujemo in usklajujemo tudi z načini izbire vzorcev.[4] Med različnimi vrstami raziskav so sicer glede tega tudi razlike, ni pa med njimi razlik glede zahteve, *da se morajo statistične metode analize podatkov prilegati podatkom ali vrstam spremenljivk,*[5] *pri vzorčnih raziskavah pa tudi načinom izbire vzorcev, na katerih zberemo podatke*. Veljavnost statističnih postopkov analize podatkov, rekli bi lahko *statistična veljavnost raziskave*, je med bistvenimi pogoji za veljavnost empiričnih ugotovitev raziskave. V korist te veljavnosti je treba gledati na izbiro

[2] Glej te metode v Sagadin 1993, Bucik 1997.

[3] Ta je tradicionalno deležna tudi kritik. Več o tem glej v Bucik 1997, str. 124–128, Blunch 2008, str. 27–46.

[4] To ne pomeni, da lahko glede analize podatkov prav vse vnaprej podrobno predvidimo in načrtujemo in da se pozneje načrta le slepo držimo.

[5] Po merski ravni ter (pri intervalnih in razmernostnih) tudi z vidika zveznosti. Seveda tudi na vidik porazdelitve ne smemo pozabiti.

vzorcev in na analizo na vzorcih zbranih podatkov kot na nerazdružljivo celoto. Tako se (lažje) izognemo tudi nevarnosti, da bi izbrali vzorec, ki se mu nobena razpoložljiva metoda statistične analize podatkov ne bi dovolj prilegala.

Popolna izpolnitev omenjene zahteve glede prileganja statističnih metod spremenljivkam in vzorcem je v raziskovalni praksi žal težko dosegljiv ideal, zato se moramo zadovoljiti tudi z manj popolnim izpolnjevanjem te zahteve. Okoliščin, zaradi katerih je tako, je več. Zaplete se lahko že pri *okviru vzorčenja* [6] (vzorčnem okviru), če ni popoln; nato lahko nastane med zbiranjem podatkov še osip oseb v vzorcu; oboje prizadene reprezentativnost vzorca glede na ciljno populacijo.[7] Če povrhu uporabimo, kot se rado zgodi, na drugačnih verjetnostnih vzorcih metode obdelave podatkov, ki veljajo samo za enostavne slučajnostne vzorce, pa se prileganje metod vzorcem sploh bistveno poruši.[8]

Naše stvarne (merske) spremenljivke, kakor jih določajo zbrani podatki, pa lahko imajo (ne glede na zaplete z okvirom vzorčenja in z osipom oseb v vzorcu med zbiranjem podatkov) težave pri raznih *predpostavkah* za uporabo določenih statističnih metod. Tu mislim na predpostavke glede merske ravni, zveznosti in porazdelitve spremenljivk. Spremenljivke predpostavk ne izpolnjujejo popolnoma (absolutno), ampak se lahko taki popolnosti približujejo le v zadovoljivi meri. Denimo, pri nekaterih značilnih spremenljivkah so zadrege glede njihove merske ravni in zveznosti ter zato tudi glede tega, katere statistične metode se jim prilegajo. Naj posebno spomnimo na rezultate na raznih *lestvičnih instrumentih*, ko je instrument sestavljen iz postavk, na katere odgovarjajo respondenti tako, da izbirajo med odgovori, ki sestavljajo *ordinalno lestvico*, nato odgovore ponderiramo s celoštevilskimi ponderji ter na podlagi ponderjev določimo vsakemu respondentu njegovim odgovorom ustrezen številski rezultat. Sem sodi *lestvica stališč Likertovega tipa*[9] *in podobni instrumenti*.[10] Pri analizi rezultatov na takih instrumentih pa vendarle uporabljamo tudi statistične metode, ki se prilegajo popolnim intervalnim, zveznim in normalno porazdeljenim spremenljivkam. Tudi sklicevanje na centralni limitni izrek[11] dvoma o popolni korektnosti takega ravnanja ne odpravi. To ni mišljeno kot očitek, zaradi katerega bi se morali taki praksi (ki pa tudi ni samo naša, slovenska), odpovedati, le na pridržke ne smemo povsem pozabiti. Po drugi strani je treba poudariti, da pa so lestvični instrumenti za merjenje na področju vzgoje in izobraževanja zelo uporabni; možnosti, ki jih

---

[6] To je pri seznamu (oseb ali skupin oseb ali institucij …), po katerem izbiramo vzorec.

[7] Pozneje bomo o teh zapletih spregovorili še podrobneje.

[8] Take rabe metod obdelave podatkov se bomo dotaknili še v zvezi s posploševanjem.

[9] Z običajno lestvico odgovorov: Se popolnoma strinjam (ali: Se močno strinjam), Se strinjam (ali: Se v glavnem strinjam), Mi je vseeno (ali: Sem nevtralen, ali: Se ne morem odločiti), Se ne strinjam (ali: Se v glavnem ne strinjam), Se nikakor ne strinjam (ali: Se sploh ne strinjam, ali: Se močno ne strinjam). Pri pozitivnih postavkah ponderiramo odgovor Se popolnoma strinjam (ali Se močno strinjam) s 5, sledijo ponderji 4, 3, 2 in 1 (1 za odgovor Se sploh ne strinjam/Se močno ne strinjam), pri negativnih postavkah pa je obratno. Več o lestvici stališč Likertovega tipa glej npr. v Sagadin 1993, str. 176–183.

[10] Ordinalni sta npr. tudi lestvici odgovorov: Vedno, Pogosto, Redko, Nikoli, Zelo, Precej, Nekoliko, Nič. In podobne lestvice.

[11] Za kakšno sklicevanje gre, glej v Blunch 2008, str. 33.

ponujajo, bi morali izrabiti še bolj kot doslej.

Pomislekov o raznih predpostavkah, na katere se opiramo pri statistični analizi podatkov, o njihovi težki izpolnljivosti in s tem povezanimi motnjami veljavnosti postopkov analize bi lahko izrazili še več; tudi pomislekov o predpostavkah, ki se ne nanašajo ravno neposredno na merske spremenljivke. Več kot v preprostejši bivariantni analizi je takih predpostavk v bolj sofisticirani multivariantni analizi. Navedli bi lahko tudi kakšne napake in spornosti v postopkih analize podatkov, ki so vgrajene celo v računalniške programske pakete.[12] Popolna veljavnost postopkov analize podatkov je torej težko uresničljiv ideal.

Veljavnost posplošitev z vzorcev na ciljne populacije

Pri preudarjanju o veljavnosti posplošitev z vzorcev na ciljne populacije moramo natančno razlikovati posploševanje z verjetnostnih vzorcev od posploševanja z neverjetnostnih vzorcev.

Posploševanje z verjetnostnih vzorcev. Statistična indukcija

Verjetnostne vzorce izbiramo iz osnovnih populacij z verjetnostnim vzorčenjem. Pri verjetnostnem vzorčenju izbiramo enote (elemente) slučajnostno, verjetnost izbire je znana vnaprej in je različna od nič. Poznamo več načinov verjetnostnega vzorčenja: enostavno slučajnostno, sistematično,[13] skupinsko, večstopenjsko, stratificirano, poststratificirano in večfazno vzorčenje ter še druge načine.[14] Podatke, zbrane na verjetnostnih vzorcih, analiziramo po statističnih metodah, ki se prilegajo podatkom oziroma spremenljivkam in načinu izbire vzorcev ter hkrati tudi vrsti/tipu vsakokratne raziskave in njenemu predmetu oziroma namenu/ciljem. V analizo je vključeno tudi posploševanje z vzorcev na osnovne/ciljne populacije. Gre za statistično indukcijo, to je za verjetnostno sklepanje z vzorcev na ustrezne populacije z uporabo metod inferenčne statistike. Pri tem sta temeljni obliki takega sklepanja ocenjevanje (točkovno in intervalno) parametrov populacij (tudi delnih populacij in razlik med njimi) ter preizkušanje statističnih hipotez za populacije. Statistična indukcija je poglavitna metoda empiričnega posploševanja pri kvantitativnem raziskovanju. Z verjetnostnih

---

[12] Na primer Osborne, Costello in Kellow (2008) upravičeno opozarjajo na sporno rabo metode glavnih komponent za ekstrakcijo faktorjev v okviru eksploratorne faktorske analize, na nenatančnost Kaiserjevega oziroma Guttman-Kaiserjevega kriterija za retencijo (obdržanje) faktorjev in na v splošnem nerealno predpostavko ali »zahtevo« po ortogonalnosti (ali popolni medsebojni nekoreliranosti) faktorjev pri rotaciji varimaks. Te pomanjkljivosti so vgrajene v razne popularne računalniške programske pakete, tudi v SPSS in SAS. Navajajo tudi ustreznejše rešitve teh problemov.

[13] Če v seznamu enot, po katerem sistematično izberemo enote v vzorec, ni kakšnega trenda ali ponavljanja enot z določenimi lastnostmi, ki bi bile v zvezi s proučevanim pojavom, lahko tak sistematični vzorec obravnavamo kot enostavni slučajnostni vzorec.

[14] Podrobnejšo obravnavo enostavnih slučajnostnih vzorcev glej v Sagadin 2003. O drugih verjetnostnih vzorcih glej več v Kalton in Vehovar 2001; Thompson 2002.

vzorcev lahko posplošujemo seveda samo na populacije, ki jih vzorci reprezentirajo. Posplošitve z vzorcev na ciljne populacije so veljavne pod temile pogoji: (1.) da je okvir vzorčenja popoln, da namreč obsega vse enote/elemente (npr. učence[15]) ciljne populacije in nobenega tujega elementa, ki ne sodi v ciljno populacijo; (2.) da iz tega okvira izberemo reprezentativni verjetnostni vzorec (večji vzorci imajo ob enakih drugih pogojih prednost pred manjšimi) ciljne populacije; (3.) da nam uspe zbrati potrebne in veljavne podatke od vseh enot v vzorcu; (4.) da podatke analiziramo po ustreznih statističnih metodah, prilegajočih se spremenljivkam in vzorcem. Čim večji so odkloni od teh pogojev, tem bolj je prizadeta veljavnost posplošitev. V težavah smo, če je bil vzorčni okvir pomanjkljiv in smo zato v vzorec zajeli tudi osebe, ki niso (več) enote te ciljne populacije, in/ali če pri izboru vzorca nismo upoštevali nekaterih oseb, ki pa so enote te populacije, ali če smo koga izbrali dva- ali večkrat. Zgodi se tudi lahko, da nekatere osebe iz vzorca, ki sicer so enote dane populacije, pri zbiranju podatkov zavrnejo sodelovanje, ali da ob zbiranju podatkov niso dosegljive in nam ne uspe navezati stika z njimi, ali da pri anketiranju po pošti neomajno vztrajajo pri neodzivu, ali da niso zmožne sodelovati in podobno. Zaradi naštetih pojavov se nam vzorec lahko tako skrči (osuje), da trpita njegova reprezentativnost in nepristranskost ter je zato prizadeta tudi veljavnost posplošitev na ciljno populacijo. Temu so izpostavljene še zlasti anketne raziskave z velikimi heterogenimi ciljnimi populacijami, ko potrebujemo za dovolj natančno ocenjevanje parametrov dokaj velike in reprezentativne vzorce, ki jih oblikujemo s smotrnim kombiniranjem različnih načinov verjetnostnega vzorčenja, upoštevaje pri tem za vsako populacijo tudi in še zlasti, kako homogena oziroma heterogena je po značilnostih, povezanih s predmetom raziskave; hkrati pa tudi mimo racionalnosti glede stroškov zbiranja podatkov in mimo izvedljivosti tega zbiranja po organizacijski plati ne moremo. Tudi evalvacijske raziskave (sumativna evalvacija) na takih populacijah oziroma vzorcih sodijo sem. Na voljo so sicer metode za zmanjšanje vpliva omenjenih pojavov na veljavnost posplošitev, popolnoma pa neugodnih vplivov ne moremo odpraviti. (Več o vsem tem v Kalton in Vehovar 2001, str. 66–90)

Z motnjami veljavnosti posplošitev, o kakršnih smo govorili, se srečujemo tudi pri drugih vrstah raziskav. Osipu oseb v vzorcih so na splošno najbolj izpostavljene dolgotrajnejše panelne[16] in tudi dolgotrajnejše eksperimentalne raziskave. (K eksperimentalnim raziskavam se bomo še vrnili.) Pri takih raziskavah je pogosto tako, da se sčasoma spreminja tudi ciljna populacija, to pa je še dodatna težava v zvezi s posploševanjem izsledkov. Veljavnost posplošitev na ciljne populacije je torej odvisna od različnih okoliščin, tudi od takih, v katerih se različne vrste raziskav med seboj razlikujejo.

Bistveno lahko prizadenemo veljavnost posplošitev (tudi če omenjenih motenj ni), če uporabimo na drugih verjetnostnih vzorcih za analizo podatkov statistične

[15] Mišljeni so učenci obeh spolov.

[16] Panelna raziskava je vzdolžna (longitudinalna) raziskava na istem verjetnostnem vzorcu oseb iz ciljne populacije; na istih osebah opravimo zbiranje podatkov/merjenje v več zaporednih časovnih trenutkih.

metode, ki veljajo le za enostavne slučajnostne vzorce. Tako je na primer, če izberemo slučajnostni vzorec šolskih oddelkov učencev (oddelek enota izbora), ali če izberemo slučajnostno na prvi stopnji vzorec šol (šola enota izbora) in na drugi stopnji vzorec učencev (učenec enota izbora) na teh šolah, ali če izberemo kak drug kompleksni verjetnostni vzorec, pri obdelavi podatkov pa uporabimo, kakor se pogosto dogaja, računalniške statistične pakete, narejene za enostavne slučajnostne vzorce. (Prim. Kalton in Vehovar 2001, str. 66, 100)

Poudarimo še vsebinski vidik posploševanja. Kvantitativnim kazalcem ne smemo pripisovati širšega vsebinskega ozadja, kot jim gre. Upoštevati moramo, iz kakšnih konceptualnih opredelitev pojmov (konstruktnih, latentnih spremenljivk), kot so znanje, stališča, interesi, motivacija itn. (pač glede na to, kaj raziskujemo), smo izhajali, kako in koliko smo jih z instrumenti za zbiranje podatkov/merjenje operacionalizirali[17] in kako je z veljavnostjo podatkov, ki smo jih zbrali in analizirali. Rezultati merjenja lahko izražajo teoretske zamisli samo v tistem delu, ki je zajet z operacionalizacijo in za katerega zberemo dovolj veljavne podatke. S tem so določene meje vsebinskega ozadja kvantitativnih kazalcev.

Veljavnost posplošitev pri eksperimentalnih raziskavah bomo pa obravnavali skupaj z veljavnostjo vzročnih razlag.

Neverjetnostni vzorci in posploševanje

Neverjetnostne vzorce izbiramo z neverjetnostnim vzorčenjem. Pri neverjetnostnem vzorčenju enot ne izbiramo slučajnostno, ampak na druge načine, zlasti namerno, priložnostno in samoizbirno. Tako izberemo na primer: namerni vzorec oddelkov ali šol za sprotno, predvsem formativno evalvacijo pri vpeljevanju neke novosti; priložnostni vzorec učencev najbližje šole za preliminarni preizkus nekega instrumenta za zbiranje podatkov; samoizbirni vzorec šol, ki same ponudijo svoje sodelovanje pri uvajanju neke programske novosti. Pri samoizbirnem vzorcu ne smemo pozabiti na možnost večjega prizadevanja šol v korist uspeha novosti, kot če ne bi šlo za samoizbirni vzorec, zato je treba ugotovitve o novosti gledati kritično tudi v tej luči. Če opravi izbiro vzorca ekspert ali skupina ekspertov (poznavalcev predmeta nameravane raziskave in značilnosti osnovne populacije), govorimo o ekspertni izbiri. Na neverjetnostnih vzorcih lahko uporabljamo metode deskriptivne (opisne) statistike, pri čemer pa moramo (v korist statistične veljavnosti) uporabljati tiste metode opisne statistike, ki se prilegajo podatkom. Tako pridemo do empiričnih ugotovitev, ki so veljavne za vzorec oziroma za vzorčno populacijo. Tu lahko govorimo o notranji veljavnosti ugotovitev. Presoja o zunanji veljavnosti ugotovitev, namreč o tem, kako veljajo zunaj vzorčne populacije, kako ustrezajo osnovni/ciljni populaciji, pa je, kadar je smiselna, bolj ali manj subjektivna in arbitrarna, pač odvisno od tega, kako je bil vzorec izbran, kaj in kako smo raziskovali, kaj smo ugotovili, kako homogena ali heterogena je ciljna populacija z vidika predmeta raziskave, kako je podobna

[17] Naj spomnim na tisto, kar smo povedali o operacionalizaciji latentnih spremenljivk že prej v prispevku.

vzorčni populaciji itn. Omenjene presoje se ne izogibamo a priori, izogibamo pa se nekritičnega in prenagljenega presojanja in sklepanja. Če smo si prizadevali, da bi vzorec bil velik in čim bolj reprezentativen za ciljno populacijo, bo presoja manj subjektivna in arbitrarna kot sicer. Toda pri sklepanju z neverjetnostnega reprezentativnega vzorca na ciljno populacijo ne moremo uporabiti statistične indukcije. Če jo uporabimo, pa sklepamo dejansko na hipotetično populacijo, iz katere bi bil isti vzorec izbran z verjetnostnim vzorčenjem; če torej uporabimo pri analizi podatkov postopke, ki veljajo za enostavne slučajnostne vzorce, sklepamo dejansko na hipotetično populacijo, iz katere bi bil izbran isti vzorec z enostavnim slučajnostnim vzorčenjem.

Veljavnost vzročnih razlag

Vemo, da so na področju vzgoje in izobraževanja pojavi zelo kompleksno vzročno-posledično med seboj prepleteni in da ima vsak pojav praviloma več vzrokov. V tej luči gledamo na vzročne zveze (vzročno-posledične zveze/odnose) med spremenljivkami tudi v tem prispevku. S pojmom *vzrok* bomo zajeli poleg vzroka v ožjem, ontološkem smislu tudi *razlog* in *motiv*.[18] Če bi poskušali v omenjeni kompleksni vzročno-posledični prepletenosti pojavov najti posameznemu pojavu vse njegove vzroke in povrhu še vsakemu vzroku natančno kvantitativno določiti, kolikšen je njegov vpliv na pojav, bi ciljali praviloma veliko previsoko. To drži tudi ob upoštevanju napredka metodologije družboslovja na področju vzročne (kavzalne) analize pojavov v zadnjih desetletjih.

V kvantitativnem empiričnem raziskovanju iščemo in preizkušamo t. i. *nomotetične vzročne* zveze med pojavi, ki pa so ravno *vzročne zveze med pojavom ustreznimi spremenljivkami.* Za kvantitativne empirične raziskave je značilen induktivno-deduktivni pristop ali postopek: najprej deduktivno pridobljena in teoretsko preverjena (verificirana) kavzalna hipoteza (hipoteza o vzročni zvezi) in nato empirično preizkušanje te hipoteze, namreč preizkušanje, ali ima hipoteza (in vzročna razlaga v njej) podporo v empiričnih podatkih, ustrezno zbranih in analiziranih

Pogoji za obstoj vzročne zveze med dvema spremenljivkama

Ko presojamo, ali je med dvema spremenljivkama vzročna zveza, upoštevamo pogoje za obstoj take zveze oziroma kriterije za presojo, ali gre za tako zvezo. Samo po sebi se razume, *da se mora vzrok zgoditi časovno pred posledico*. Torej se mora

[18] *Filozofijski rječnik* 1965 označuje razlog kot spoznavni (hrv. spoznajni) vzrok (str. 417) in motiv kot spodbudni (hrv. poticajni) vzrok (prav tam); opozarja na razliko med logično zvezo: razlog – naslednik (hrv. posljedak) in ontološko zvezo: vzrok – učinek /posledica (str. 340). Dodajmo, da bi temu lahko pridružili še psihološko zvezo: motiv – posledica.
Po *Slovarju slovenskega knjižnega jezika* 1994 je vzrok tisto, »*kar naredi, da kaj nastane, se zgodi* (str. 1567); razlog tisto, »*kar utemeljuje, podpira kako ravnanje, odločitev*« (str. 1126); motiv tisto, »*kar povzroča kako dejanje, ravnanje; nagib, spodbuda*« (str. 577).

tisto, o čemer zberemo podatke za neodvisno spremenljivko (domnevni vzrok), zgoditi časovno pred tistim, o čemer zberemo podatke za odvisno spremenljivko (domnevno posledico/učinek). To pa ne pomeni, da morata biti zbiranji podatkov za neodvisno in odvisno spremenljivko vedno časovno ločeni; na primer pri istem anketnem zbiranju podatkov lahko zberemo podatke za obe spremenljivki, pri čemer pa moramo upoštevati, kateri podatki so pri anketirancih bolj izpostavljeni pozabljanju in spreminjanju v času (od začetka pojava do trenutka zbiranja podatkov o njem), kar prizadene veljavnost takih podatkov. Potrebni pogoj za obstoj vzročne zveze med dvema spremenljivkama je tudi, da mora biti *med njima empirična zveza, sovariiranje*.[19] Brez tega ni vzročne zveze med spremenljivkama. Omenjena dva pogoja sta potrebna za obstoj vzročne zveze med danima spremenljivkama, nista pa zadostna. Ni vsaka empirična zveza med dvema spremenljivkama vzročna, lahko je tudi samo navidezna, neprava, ne resnična, prava vzročna zveza. Tako je, če je empirična zveza le posledica sovariiranja vsake izmed obeh spremenljivk z neko tretjo spremenljivko ali spremenljivkami. Denimo, med dosežki učencev pri dveh testih, ki merita povsem različno znanje, se lahko kaže navidezna vzročna zveza, ko so učenci, ki so na splošno (povprečno) uspešnejši pri prvem testu, na splošno (povprečno) uspešnejši tudi pri drugem. To navidezno vzročno zvezo bi lahko izrazili tudi v obliki pozitivne korelacije med rezultati učencev pri obeh testih. Podobno je z nevzročnim sovariiranjem učnih uspehov učencev pri dveh ciljno in vsebinsko različnih učnih predmetih. Navidezna vzročna zveza nastaja, ker so pri obeh testih ali pri obeh učnih predmetih uspešnejši učenci, ki so na boljšem glede splošnih dejavnikov učenčeve uspešnosti, kot so sposobnosti, učne zmožnosti, motiviranost, učne navade, prizadevnost, družinsko okolje itn. Ker ni vsaka empirična zveza med dvema spremenljivkama tudi vzročna, je tretji in najpomembnejši pogoj za obstoj vzročne zveze med dvema spremenljivkama, *da vzročna zveza med njima ni le navidezna, ampak pristna, resnična*[20]. Kot smo že omenili, ima v vzgoji in izobraževanju vsak pojav praviloma več vzrokov, zato z vzročno zvezo med dvema spremenljivkama ne mislimo in ne iščemo samo tako tesne zveze, ko bi vzročna spremenljivka bila edini in tako ves vzrok posledične spremenljivke. Raziskovalno najzahtevnejši je tretji pogoj/kriterij. Tem temeljnim pogojem/kriterijem za obstoj vzročne zveze med dvema spremenljivkama se v novejšem času pridružuje kriterij ali zahteva po *ugotovitvi t. i. kavzalnega ali vzročnega mehanizma*, že tradicionalno pa še kriterij ali zahteva, da je treba čim natančneje določiti kontekst – razmere, okoliščine, za katere/kakršne določeni kavzalni sklep velja. Pri vzročnem mehanizmu gre za ugotovitev in uvedbo *posredovalnih spremenljivk* med domnevno vzročno in domnevno posledično

[19] Mišljeno ni samo sovariiranje v obliki korelacije v ožjem pomenu in v obliki kontingence. Za sovariiranje gre tudi, če se na primer pri didaktičnem eksperimentu od skupine do skupine oziroma od tretmaja do tretmaja spreminja tudi povprečni dosežek učencev na testu za kriterij primerjave tretmajev glede učinkovitosti; ali če se denimo pri neeksperimentalni raziskavi od stratuma do stratuma ali od delne populacije do delne populacije spreminja tudi odstotek soglašanja anketirancev z nekim ukrepom. In tako naprej.

[20] Ne smemo v statistični značilnosti/pomembnosti videti zadostnega dokaza, da gre za pristno vzročnost.

spremenljivko, ki posredujejo vzročno zvezo od domnevno vzročne do domnevno posledične spremenljivke, in sicer tako, da domnevno vzročna spremenljivka vpliva vzročno na posredovalne spremenljivke, te pa na posledično spremenljivko. Tako je vzročni mehanizem proces vzročnega vplivanja od neodvisne spremenljivke po posredovalnih spremenljivkah do odvisne spremenljivke. (Gl. in prim. Schutt 2006, str. 175, 189, 192–193; Morgan in Winship 2007, str. 286–287) Odkrivanje omenjenega procesa vzročnega vplivanja pravzaprav tudi ni nekaj povsem novega v primerjavi s tradicijo, je pa res, da nekoč zahteva po »vzročnem mehanizmu« ni bila tako izrecno artikulirana. Dodal bi tudi, da v vzročnem mehanizmu ali procesu med dvema spremenljivkama ne smemo pozabiti na možno neposredno vzročno zvezo med njima, ker pač lahko gre hkrati za neposredno in posredno vzročno zvezo med dvema spremenljivkama.

Poglejmo najprej, kako je z izpolnjevanjem navedenih pogojev pri eksperimentalnih raziskavah, nato pa se bomo na kratko pomudili še ob neeksperimentalnih raziskavah.

Eksperimentalne raziskave: izpolnjevanje navedenih pogojev, preizkušanje kavzalnih hipotez, veljavnost vzročnih razlag, posploševanje

Glede izpolnjevanja navedenih pogojev smo na splošno v najlažjem položaju pri eksperimentalnih raziskavah, predvsem pri slučajnostnih eksperimentih. Vzemimo enega izmed tipičnih modelov eksperimenta,[21] *eksperiment z eno eksperimentalno skupino* (ki je deležna eksperimentalnega tretmaja) *in eno kontrolno skupino* (s kontrolnim tretmajem), s katerim (namreč z eksperimentom) empirično preizkušamo vnaprejšnjo, deduktivno pridobljeno, dobro teoretsko podprto in teoretsko preverjeno (verificirano) hipotezo o učinkovitosti eksperimentalnega tretmaja oziroma o njegovi večji učinkovitosti v primerjavi s kontrolnim tretmajem ter o vzročni razlagi te večje učinkovitosti. Pri slučajnostnem eksperimentu razvrščamo osebe v primerjalne (eksperimentalne in kontrolne)[22] skupine slučajnostno (tj. z randomizacijo), in sicer z namenom, da bi bile primerjalne skupine izenačene do slučajnih razlik glede vplivov/učinkov na kriterijsko (odvisno) spremenljivko (ali spremenljivke), ki jih povzročajo relevantni dejavniki, vezani na osebe kot posameznike in v katerih osebe variirajo, se med seboj razlikujejo. Ob taki izenačenosti skupin omenjeni vplivi/učinki ne motijo

[21] O možnih modelih eksperimenta (slučajnostnega in neslučajnostnega, enofaktorskega in dvofaktorskega) na pedagoškem, predvsem na šolskem področju ter o drugih vprašanjih eksperimentalnih raziskav sem podrobneje pisal že v Sagadin 1977, 1991, str. 68–88; 1994, 2003, str. 346–430.

[22] V metodologiji izrazje ni poenoteno. Sam uporabljam izraz »primerjalne skupine« v pomenu skupin, ki jih primerjamo po dosežkih pri eksperimentu, torej v pomenu eksperimentalnih in kontrolnih skupin skupaj. Izraza kontrolna skupina pri tem ne uporabljam samo v pomenu t. i. ničelne skupine (ki ni deležna nobenega tretmaja – »ničelni« tretma), ampak tudi v pomenu skupine, ki dela »po starem« (ali je deležna nekega alternativnega tretmaja v primerjavi z eksperimentalno skupino). Podrobneje pišem o tem v Sagadin 1991, str. 68–69. V sodobni metodologiji srečujemo rabo izraza »kontrolna skupina« v obeh omenjenih pomenih; srečujemo pa tudi omejitev izraza »kontrolna skupina« samo na ničelno skupino, za skupino z alternativnim tretmajem pa poimenovanje »primerjalna skupina«.

eksperimentalnih izsledkov, ker ne vplivajo na razlike primerjalnih skupin glede rezultatov na instrumentu (ali instrumentih), ki merijo učinke eksperimentalnih in kontrolnih tretmajev in ki prinašajo podatke za kriterijsko spremenljivko (ali spremenljivke).[23] Randomizacija je v izenačevanju skupin glede omenjenih vplivov dovolj učinkovita, če so skupine dovolj velike. Če na primer izberemo iz neke ciljne populacije velik reprezentativni verjetnostni oziroma slučajnostni vzorec oseb in te osebe nato slučajnostno razvrstimo v eksperimentalno in kontrolno skupino, smo že pri dovolj velikih primerjalnih skupinah, izenačenih v dejavnikih, vezanih na osebe kot posameznike, do slučajnih razlik. Ti dejavniki so torej kontrolirani z randomizacijo. Če ni nobenih drugih, nekontroliranih vplivov (ki bi jih povzročali dejavniki, ki jih randomizacija ne zajame)[24] na kriterijsko spremenljivko/spremenljivke ali na rezultate na omenjenih instrumentih in so razlike med eksperimentalno in kontrolno skupino v teh rezultatih posledica razlik v učinkovitosti eksperimentalnega in kontrolnega tretmaja, je *eksperiment notranje veljaven*. Kolikor bolj so omenjene razlike med eksperimentalno in kontrolno skupino res posledica razlik v učinkovitosti eksperimentalnega in kontrolnega tretmaja, ne pa kakšnih nekontroliranih vplivov, toliko bolj drži, da je eksperiment notranje veljaven. Notranja veljavnost se nanaša na veljavnost empiričnih ugotovitev o teh razlikah znotraj eksperimentalne situacije. Če so pri notranje veljavnem eksperimentu te razlike v korist eksperimentalnemu tretmaju, to pomeni, da empirične ugotovitve eksperimenta *podpirajo hipotezo*, ki jo z eksperimentom preizkušamo, *ter vzročno razlago in sklepanje o* prednosti, večji učinkovitosti eksperimentalnega tretmaja v tej hipotezi. V skladu s tem gre za *veljavno vzročno razlago in za veljavni vzročni sklep ali sklepe*. Ker smo (pri naši zamisli) izhajali pri oblikovanju eksperimentalne in kontrolne skupine iz reprezentativnega slučajnostnega vzorca ciljne populacije, pa lahko eksperimentalne izsledke posplošimo na tako ciljno populacijo, v čemer je zunanja veljavnost eksperimenta. Če nastane osip oseb že v omenjenem reprezentativnem vzorcu ali/in pozneje med potekom eksperimenta, je veljavnost takega posploševanja prizadeta. To je pač podobno, kot smo omenili že pri neeksperimentalnih raziskavah. Samo po sebi se razume, da je potrebni pogoj za omenjeno zunanjo veljavnost eksperimenta tudi njegova notranja veljavnost, saj bi drugače šlo za posploševanje neveljavnih izsledkov.

Če ne izhajamo od reprezentativnega slučajnostnega vzorca, marveč od neke nereprezentativne skupine oseb,[25] ki jih slučajnostno razvrstimo v eksperimentalno in kontrolno skupino, je takšen eksperiment še vedno slučajnosten in ob enakih pogojih, kot smo jih omenili že pri eksperimentu, pri katerem izhajamo od repre-

---

[23] Podatki za kriterijske spremenljivke so torej podatki o učinkih eksperimentalnih in kontrolnih tretmajev, ki so jih deležne primerjalne skupine, skupaj z ustreznimi podatki za »ničelni tretma« v morebitni nvelni kontrolni skupini.

[24] Več o dejavnikih te vrste pozneje, ko bomo pri didaktičnih eksperimentih srečali poleg dejavnikov, vezanih na osebe kot posameznike, še druge dejavnike, ki jih je treba v korist notranje veljavnosti didaktičnega eksperimenta kontrolirati.

[25] Na primer od skupine odraslih kandidatov za kakšen izobraževalni tečaj, kot je bilo pri diplomski raziskavi Franje Čeh (1981).

zentativnega slučajnostnega vzorca, tudi notranje veljaven, toda posploševanje, kakršno je mogoče, če izhajamo od reprezentativnega slučajnostnega vzorca, tu ni mogoče. Sklepamo sicer lahko dane eksperimentalne situacije na druge, *podobne situacije*, kadar je to smiselno. Pri takem sklepanju dane situacije na druge situacije se opiramo na *vzročno razlago* razlik v učinkovitosti med eksperimentalnim in kontrolnim tretmajem in *na empirično podobnost* nove situacije z eksperimentalno situacijo. Zavedati pa se vendarle moramo (možne) *subjektivnosti in arbitrarnosti* pri takem sklepanju ter biti zato kritični in ne prenagljeni.

V nakazanih primerih je šlo za eksperiment z osebo kot eksperimentalno enoto. Mogoč pa je tudi slučajnostni eksperiment s skupino oseb, npr. z oddelkom učencev kot eksperimentalno enoto. Tedaj izberemo najprej izhodiščni vzorec oddelkov, ki jih nato slučajnostno razvrstimo v eksperimentalno in kontrolno skupino oddelkov. Če je izhodiščni vzorec slučajnostni reprezentativni vzorec oddelkov iz ciljne populacije oddelkov, lahko eksperimentalne izsledke posplošimo na tako ciljno populacijo. Če oddelkov v izhodiščni vzorec ne izberemo slučajnostno, pa je možnost posploševanja na ciljno populacijo oddelkov odvisna od tega, kako reprezentativen (čeprav ni slučajnosten) je vzorec oddelkov za ciljno populacijo. Reprezentativnost vzorca namreč lahko dosežemo tudi brez uporabe slučajnostnega vzorčenja. Pri analizi podatkov pa moramo (tako tedaj, ko izhajamo od slučajnostnega izhodiščnega vzorca oddelkov, kot tedaj, ko izhodiščni vzorec ni slučajnosten) upoštevati, da gre za oddelek kot eksperimentalno enoto, če hočemo, da bodo rezultati analize veljavni. *Model eksperimenta z oddelkom kot eksperimentalno enoto se šolskemu področju dobro prilega in glede tega presega model z osebo kot eksperimentalno enoto*.

Pri eksperimentih, o katerih smo govorili, smo imeli v mislih veliki primerjalni skupini, pri kakršnih je randomizacija uspešna pri izenačevanju skupin v dejavnikih, vezanih na eksperimentalne enote. Če so primerjalne skupine premajhne, je lahko izenačenost skupin nezadostna. Zato je v takih primerih smotrno relevantne dejavnike, vezane na osebe kot posameznike, kontrolirati dodatno še s *statistično kontrolo z analizo kovariance po klasičnem postopku ali po postopku z multiplo linearno regresijo*[26] in tako zvišati veljavnost eksperimenta. Denimo, pri eksperimentalnih in kontrolnih skupinah velikosti šolskih oddelkov že moramo misliti na to. Takšna kontrola je lahko koristna tudi pri večjih skupinah. Še tem bolj pa je potrebna pri *neslučajnostnih eksperimentih* (imenovanih tudi *kvazieksperimenti*[27]). Če bi na primer v posamezni šoli izvajali eksperiment tako, da bi bil en obstoječi oddelek učencev ali dijakov istega razreda eksperimentalna in en oddelek kontrolna skupina,[28] bi imel izid takšnega

[26] Oba postopka prikazujem v Sagadin 2003, str. 388-430, in sicer za značilni model eksperimenta z eno eksperimentalno in eno kontrolno skupino in s tem z eno umetno spremenljivko. Model z več umetnimi spremenljivkami glej npr. v Tacq 1997.

[27] Po klasifikaciji, ki sta jo vpeljala Campbell in Stanley in po kateri se kvazieksperimenti razlikujejo od »pravih« (slučajnostnih) eksperimentov v tem, da pri kvazieksperimentih ni slučajnostnega razvrščanja eksperimentalnih enot na eksperimentalne in kontrolne skupine, drugače pa so podobni »pravim«. (Gl. Sagadin 1991, str. 70)

[28] Ker obstoječih oddelkov pač ne moremo preprosto razdirati, da bi nato lahko učence slučajnostno razvrščali v eksperimentalne in kontrolne skupine za raziskovalne potrebe. Več možnosti za

neslučajnostnega eksperimenta na splošno zelo nizko *notranjo veljavnost*, zaradi česar neslučajnostnih eksperimentov v taki obliki ne uporabljamo; uporabljamo jih dopolnjene z že omenjeno statistično kontrolo poglavitnih dejavnikov, vezanih na učence/dijake kot posameznike, oziroma ustreznih spremenljivk, imenovanih (kot že vemo) *sospremenljivke*, ki bi lahko sicer povsem nekontrolirano vplivale na kriterijsko spremenljivko/spremenljivke, s tem pa tudi na empirične ugotovitve eksperimenta, na veljavnost teh ugotovitev in na veljavnost naše vzročne razlage razlik v učinkovitosti med tretmaji. Pri eksperimentih na didaktičnih ali specialnodidaktičnih področjih tako kontroliramo prejšnje (to predvsem) in sposobnosti učencev/dijakov; če je le mogoče in smotrno (glede na to, kaj eksperimentalno preizkušamo), pa kontroliramo tako tudi še kaj drugega. Tako poskušamo čim bolj zvišati notranjo veljavnost neslučajnostnih eksperimentov (brez katere tudi sklepanje o možnostih uporabe njihovih izsledkov v drugih podobnih situacijah ne bi bilo smiselno).[29] Statistično kontrolo izvajamo, kot smo že nakazali, z metodo analize kovariance po klasičnem postopku ali po postopku z multiplo linearno regresijo, ki je danes bolj v rabi. Menim, da bi bilo smotrno vzeti za eksperimentalno in kontrolno skupino oddelka posameznega razreda, ki bi bila *čim bolj izenačena med seboj v sospremenljivkah* (po porazdelitvah, aritmetičnih sredinah in standardnih odklonih), ter *na tako izenačenih oziroma podobnih oddelkih uporabiti še omenjeno statistično kontrolo.* Tu mislim na eksperimentalno raziskavo na šoli, ki ima več oddelkov istega razreda. *Kombinacija izenačevanja* (ki zmanjša razlike med eksperimentalno in kontrolno skupino v sospremenljivkah) *in analize kovariance oziroma multiple linearne regresije* bi lahko ugodno vplivala na notranjo veljavnost eksperimenta.[30]

V korist veljavnosti eksperimentov kontroliramo tudi *dejavnike, ki so vezani na oddelke kot celote*, za kar prideta v poštev zlasti tile možnosti: poskrbimo, da so ti dejavniki čim bolj *konstantni* od oddelka do oddelka (tu je v ospredju učitelj), ali da jih *izločimo* iz eksperimenta oziroma jih preprečimo (npr. razne motnje pouka). Teh dejavnikov torej ne kontroliramo statistično, kot kontroliramo dejavnike, ki so vezani na učence/dijake (ali na druge eksperimentalne enote) kot posameznike. Misliti pa moramo tudi na relevantne dejavnike, v katerih se utegneta eksperimentalna in kontrolna skupina razlikovati, pa jih z nobenim od navedenih načinov kontrole ne zajamemo. Sem sodijo na primer šolske ocene znanja učencev iz časa pred eksperimentom (če ocen zaradi pridržkov v zvezi z njihovo mersko ravnijo ne vključimo v analizo kovariance), družinsko okolje

slučajnostni eksperiment kot pri običajnem šolskem pouku se najde pri raznih tečajih in podobnih izobraževalnih oblikah, bodisi za odrasle ali za mladino. Tako je bilo npr. na tečaju za odrasle pri že omenjeni diplomski raziskavi Franje Čeh. Uporaba eksperimentalne raziskovalne metode na področju vzgoje in izobraževanja pa tudi ni omejena samo na didaktične probleme. Glej primer eksperimenta, ki ga omenja Shutt (2006, str. 181, 190, 192), izvedli pa so ga Bushman in sodelavci.

[29] Povedano ne pomeni, da postanejo neslučajnostni eksperimenti tako enako veljavni kot slučajnostni eksperimenti, če tudi pri teh uporabimo enako kontrolo.

[30] O metodah izenačevanja eksperimentalnih in kontrolnih skupin v sospremenljivkah v drugačnih situacijah, pri drugačnih modelih kvazi eksperimentov ter o kombinaciji izenačevanja in regresijske analize glej v Stuart in Rubin 2008.

učencev, relevantna stališča učencev (pred eksperimentom), spol učencev in drugo. Pri *kritični interpretaciji rezultatov analize kovariance oziroma regresijske analize* so nam lahko omenjene in podobne primerjave eksperimentalne in kontrolne skupine v pomoč.

Seveda je pri tem potrebno tudi zadostno poznavanje teoretskega ozadja raziskovane problematike, brez katerega pri vrtanju v vzročno-posledične odnose med pojavi ne gre. Opora na teoretsko ozadje, kot dobro vemo, ni na vrsti šele pri vzročni interpretaciji empiričnih ugotovitev eksperimenta, ampak že pri teoretskem koncipiranju novosti (ki jo bomo eksperimentalno preizkusili), skupaj s hipotetičnim predvidevanjem njenih učinkov in njihove vzročne razlage.[31] Toda teoretsko koncipirano novost empirično preizkusimo z eksperimentom le v taki obliki in toliko, kakor jo pri eksperimentu operacionaliziramo in konkretiziramo, in pod pogojem, da so empirične ugotovitve eksperimenta veljavne. To moramo upoštevati pri vzročni razlagi teh ugotovitev.

Med možnimi motnjami notranje veljavnosti didaktičnega eksperimenta in veljavnosti vzročne razlage tudi nikakor ne smemo pozabiti na možnost, da si učenci[32] v eksperimentalni skupini že ob védenju, da so tudi oni del eksperimenta in raziskovani, bolj prizadevajo za učno uspešnost, kot bi si sicer (Hawthorne Effect). Povsem zanemariti pa ne smemo niti možnosti, da se učenci kontrolne skupine čutijo prizadete, ker niso deležni novosti, in poskušajo z večjo prizadevnostjo za učno uspešnost to kompenzirati in tekmovati v dosežkih z eksperimentalno skupino (John Henry Effect). Motnjo lahko pomeni tudi vnos kakšnih sestavin eksperimentalnega tretmaja v kontrolno skupino. Zanemariti tudi ne smemo možnosti, da se podobna občutja in ravnanja kot pri učencih lahko pojavljajo tudi pri učiteljih. Gotovo pa je nevarnost za motnje vzročnih razlag te vrste manjša, če učitelji niso postavljeni le v vlogo izvajalcev načrtov raziskovalcev, ampak lahko tudi ustvarjalno sodelujejo pri oblikovanju eksperimentalnih tretmajev in pri vrednotenju njihovih učinkov, in če je med raziskovalci in učitelji med eksperimentom dovolj delovnih stikov, v korist ustreznega poteka eksperimenta in nadzora nad eksperimentalno situacijo.

Na opisane načine poskušamo najprej zadostiti omenjenim *prvim trem pogojem/kriterijem za obstoj vzročne zveze* med neodvisno spremenljivko, katere vrednosti sta tretmaja,[33] in odvisno (kriterijsko) spremenljivko (ali spremenljivkami, če jih je več); posegamo pa tudi že v naslednja dva kriterija. Omenjena vzročna zveza je lahko neposredna in/ali posredna, prek posredovalnih spremenljivk. Tu pa smo že pri *vzročnem mehanizmu ali vzročnem procesu*. Vzemimo, da je bil eksperimentalni tretma zasnovan (na podlagi teoretično preverjene raziskovalne hipoteze ali hipotez) ter izvajan tako, da naj bi imel prednost pred kontrolnim (»starim«) tretmajem tako glede neposrednega vzročnega vpliva na učenčevo pridobivanje znanja kakor tudi glede posrednega vpliva, prek (tudi v eksperimentalni tretma vgrajenega) sistematičnega razvijanja učinkovitejših

[31] Glej več o tem v Sagadin 1994.
[32] Ali dijaki, študenti, udeleženci izobraževanja za odrasle.
[33] Ali več tretmajev, če bi uporabili model eksperimenta z več primerjalnimi skupinami.

učnih navad in avtonomne učne motivacije pri učencih, ter da so empirične eksperimentalne ugotovitve (na podlagi zbranih veljavnih podatkov in opravljene ustrezne analize teh podatkov) potrdile, da je eksperimentalni tretma v tem uspel, prekosil kontrolni tretma.[34] Če je tako, je očitno, da je vzročni mehanizem ali proces vzročnega vplivanja takšen, da je med neodvisno spremenljivko (z vrednostma eksperimentalni in kontrolni tretma) in kriterijsko spremenljivko (znanje) neposredna in posredna vzročna zveza, pri čemer sta posredovalni spremenljivki učne navade (pri novem tretmaju učinkovitejši kot pri starem) in učna motivacija (pri novem tretmaju višja, avtonomnejša kot pri starem). Pri tem ne smemo pozabiti, da omenjeni posredovalni spremenljivki v odnosu do neodvisne spremenljivke nista samo posredovalni, ampak tudi njeni neposredni kriterijski spremenljivki. Tudi to sodi k vzročnemu mehanizmu oziroma k vzročni razlagi učinkovitosti eksperimentalnega tretmaja. K vzročni razlagi pritegnemo tudi vse drugo, s čimer jo je mogoče dodatno podkrepiti, toda razvidno mora biti, kaj od tega smo ugotovili empirično in kaj le domnevamo. Tudi za eksperimentalni in kontrolni tretma upoštevamo, kako sta bila v praksi operacionalizirana in izvedena, ter za *konstruktne spremenljivke* (znanje, učne navade, učno motivacijo), kako ali koliko smo jih empirično izrazili z *merskimi spremenljivkami*.

Kako je tu s posploševanjem eksperimentalnih ugotovitev zunaj dane eksperimentalne situacije? Statistična analiza podatkov nam pokaže, kolikšna je verjetnost, da bi se prednost eksperimentalnega tretmaja pred kontrolnim tretmajem (če se je pokazala pri eksperimentu) ponavljala sistematično, če bi eksperiment ponavljali v razmerah, kakršne je predstavljala eksperimentalna situacija. Posploševanje s statistično indukcijo na razmere zunaj dane eksperimentalne situacije pa ni mogoče.[35] Vzročna razlaga učinkovitosti eksperimentalnega tretmaja v kontekstu eksperimentalne situacije na šoli, kjer je eksperiment potekal, pa nam omogoča sklepanje o možnostih uporabe eksperimentalnih izsledkov v podobnih situacijah na drugih šolah. Takšno *sklepanje torej opremo na omenjeno vzročno razlago in na empirično podobnost situacij*. Sklepanje je olajšano, če eksperimenta ne izvedemo le na eni šoli, ampak vsaj na nekaj šolah. Če hočemo pri tem zaznati še morebitni vpliv socialno-ekonomsko-kulturnega *konteksta/okolja* šole na eksperimentalne izsledke ali odvisnost učinkovitosti eksperimentalnega tretmaja od okolja šole, izvedemo eksperiment na šolah v različnih okoljih, da lahko primerjamo med seboj eksperimentalne izsledke iz različnih okolij. Ob vsem tem pa vendarle ne smemo pozabiti, da eksperiment sicer poteka v šolskih razmerah, toda eksperimentalna situacija ima svoje posebnosti v primerjavi z vsakdanjimi šolskimi razmerami, v katerih ni raziskovalca, njegove vloge in opore učitelju, in tudi ni motivacijskih učinkov na učence in učitelja (lahko tudi na starše učencev, če so pritegnjeni k sodelovanju pri eksperimentu), ki jih prinaša (kot smo že nakazali) sodelovanje pri eksperimentu, in podobno. Zato je treba upoštevati možnost, da

---

[34] Tako kot je bilo vse to pri doktorski eksperimentalni raziskavi Branke Čagran (1992). Eksperiment je potekal na šestih osnovnih šolah, na treh v mestnem in na treh v primestnem okolju.

[35] Spredaj smo pokazali, kako priti do eksperimentalne in kontrolne skupine slučajnostnega eksperimenta, pri katerem je mogoče posploševanje na ciljno populacijo s statistično indukcijo.

v novih razmerah, sicer na videz podobnih eksperimentalnim razmeram (le brez omenjenih posebnosti eksperimentalne situacije), novost ne bo delovala enako kot v eksperimentalni situaciji. To se lahko zgodi še zlasti, če je novost vsebinsko in/ali izvedbeno zahtevna, po eksperimentu pa njeno uvajanje v vsakdanjo šolsko prakso ni primerno pripravljeno in se ne izvaja ustrezno.

Raziskovanje vzročnih zvez z neeksperimentalnimi kvantitativnimi empiričnimi raziskavami

Tudi za te kvantitativne raziskave je značilen deduktivno-induktivni postopek: najprej deduktivno pridobljena in teoretsko preverjena hipoteza o vzročni zvezi in nato empirično preizkušanje te hipoteze. Vzročna zveza je morda neposredna, ali posredna, ali hkrati neposredna in posredna. O tem preudarimo že pri postavljanju vnaprejšnje hipoteze. Preprostejši postopek neeksperimentalnega empiričnega preverjanja kavzalnih hipotez vključuje le preprostejše, *bivariantne statistične postopke*, ki zajamejo v analizo hkrati le podatke za po dve merski spremenljivki. Po tej poti iščemo *pristno empirično podporo kavzalni hipotezi* (ali hipotezam, če jih je več). Pri vzorčni raziskavi preizkusimo tudi *statistično značilnost*/*pomembnost* empirične zveze. Toda statistična značilnost sama po sebi še ni zadostni dokaz, da je empirična zveza med spremenljivkama res vzročna,[36] ampak (statistična značilnost) pomeni le, da lahko statistično ničelno hipotezo o obstoju empirične zveze med spremenljivkama v osnovni oziroma ciljni populaciji zavrnemo, ker je tveganje za njeno zavrnitev dovolj majhno, in da ima skladu s tem kavzalna hipoteza podporo v empirični zvezi med spremenljivkama tudi v osnovni populaciji.[37] Če nam glede na namen raziskave preprosta bivariantna analiza podatkov ustreza, jo seveda uporabimo. Poleg opisane, za kvantitativno raziskovanje značilne situacije z vnaprejšnjo kavzalno hipotezo je še druga možna situacija, ko take vnaprejšnje hipoteze nimamo, ampak nas šele empirične ugotovitve *eksplorativne analize podatkov* navajajo na sklepanje o možnih vzročnih zvezah med spremenljivkami. Tako je, denimo, ko odgovore anketirancev (slučajnostnega vzorca) na različna anketna vprašanja soočimo med seboj in opazimo, da so med odgovori na različna vprašanja oziroma med ustreznimi spremenljivkami statistično značilne empirične zveze. S primerjanjem različnih ugotovljenih empiričnih zvez poskušamo ugotoviti, kje bi lahko šlo za prave vzročne zveze, in hkrati *iščemo ustrezne*, *teoretsko utemeljene vzročne razlage*. Gotovost, da so naše ugotovitve o vzročnih zvezah med spremenljivkami veljavne, pa pri tem postopku ne v prvi situaciji (pri vnaprejšnjih kavzalnih hipotezah) in ne v drugi (brez vnaprejšnjih kavzalnih hipotez) ni tolikšna kot pri eksperimentu.

Več gotovosti kot bivariantna analiza lahko omogoča *multivariantna analiza*. S starejšo analizo te vrste zajamemo vzročne odnose med več merskimi spremenljivkami hkrati, v obliki t. i. *modela analize poti (»path analysis«) za*

[36] Kot smo opozorili že v opombi 20.

[37] Tu smo predpostavili, da je tudi smer ugotovljene empirične zveze med spremenljivkama takšna, kot smo jo predvideli v vnaprejšnji hipotezi.

merske spremenljivke. (Prikaz te analize glej npr. v Pedhazur 1997, str. 765–840.) Novejša in še bolj sofisticirana je analiza, ki zajema vzročne odnose med latentnimi spremenljivkami, v obliki *modela analize poti za latentne spremenljivke*. Ta model pomeni kombinacijo modela analize poti za merske spremenljivke in *modela konfirmatorne faktorske analize*, pri katerem gre za vzročne odnose med latentnimi spremenljivkami in njihovimi merskimi indikatorji/latentnimi spremenljivkami. Navedeni modeli sodijo v že omenjeno modeliranje strukturnih enačb, SEM. Pri teh modelih izhajamo od vnaprejšnjih, razčlenjenih/specificiranih, teoretsko dobro podprtih in preverjenih kavzalnih hipotez, SEM pa omogoča s svojimi postopki empirično preizkušanje teh hipotez. (Gl. in prim. Mueller in Hancock 2008, str. 489 idr. Več glej v: Blunch 2008, Byrne 2006.)[38] Ti postopki sicer prinašajo napredek pri neeksperimentalnem raziskovanju vzročnih zvez med spremenljivkami, vendar imajo kljub temu glede veljavnosti vzročnih razlag na splošno še vedno prednost eksperimentalne raziskave, predvsem slučajnostni eksperimenti.

Sklepne misli

V tem prispevku nismo iskali poti do popolne veljavnosti kvantitativnih empiričnih raziskav vzgoje in izobraževanja, ker take poti ni,[39] nakazovali smo le poti za dosego čim višje veljavnosti eksperimentalnih in neeksperimentalnih raziskav. Med raziskavami o vzročno-posledičnih odnosih med spremenljivkami so glede doseganja veljavnosti izsledkov na splošno še vedno na boljšem eksperimentalne raziskave, predvsem slučajnostni eksperimenti, v primerjavi z neeksperimentalnimi raziskavami, tudi če te uporabljajo novejše, bolj sofisticirane postopke, ki jih opredeljuje SEM.

Literatura

Blunch, N. J. (2008). Introduction to Structural Equation Modelling using SPSS and AMOS. Los Angeles: Sage.

Bucik, V. (1997). Osnove psihološkega testiranja. Ljubljana: Filozofska fakulteta, Oddelek za psihologijo.

Byrne, B. M. (2006). Structural Equation Modelling with EQS: Basic concepts, applications, and Programming. Mahwah, New Jersey: Lawrence Erlbaum.

Čagran, B. (1992). Raziskovalno razvijanje sodobnejših oblik načrtovanja, posredovanja, izvrševanja in preverjanja domačih nalog v osnovni šoli. Doktorska disertacija. Ljubljana: Filozofska fakulteta, Oddelek za pedagogiko.

Čeh, F. (1981). Izobraževalna učinkovitost metode predavanja, metode pisanih del – referata z zagovorom in metode demonstracije na ekskurziji. Diplomska naloga. Ljubljana: Filozofska fakulteta, PZE za pedagogiko.

[38] Za SEM so na voljo računalniški programi: AMOS, EQS, LISREL in drugi.

[39] To seveda ne velja le za raziskave vzgoje in izobraževanja, pa tudi sicer ne mislimo, da velja vse, kar smo metodološkega povedali v tem prispevku, samo za te raziskave.

Filozofijski rječnik (1965). Zagreb: Matica hrvatska.

Kalton, G. in Vehovar, V. (2001). Vzorčenje v anketah. Ljubljana: Fakulteta za družbene vede.

Morgan, L. in Winship, C. (2007). Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge etc.: Cambridge University Press.

Mueller, R. O. in Hancock, G. R. (2008). Best Practices in Structural Equation Modelling. V: Osborne, J. W. (ur.). Best Practices in Quantitative Methods. Los Angeles etc. Sage, str. 488–508.

Osborne, J. W., Costello, A. B. in Kellow, J. T. (2008). Best Practices in Exploratory Factor Analysis. V: Osborne, J. W. (ur.). Best Practices in Quantitative Methods. Los Angeles etc.: Sage, str. 86–99.

Pedhazur, E. J. (1997). Multiple Regression in Behavioral Research: Explanation and Prediction. Fort Worth etc.: Harcourt Brace College Publisher.

Sagadin, J. (1997). Poglavje iz metodologije pedagoškega raziskovanja, II. del: Statistično načrtovanje eksperimentov. Ljubljana: Pedagoški inštitut pri Univerzi v Ljubljani.

Sagadin, J. (1991). Razprave iz pedagoške metodologije. Ljubljana: Znanstveni inštitut Filozofske fakultete.

Sagadin, J. (1993). Poglavja iz metodologije pedagoškega raziskovanja (2. izd.). Ljubljana: Zavod Republike Slovenije za šolstvo in šport.

Sagadin, J. (1994). Glavne faze eksperimentalne pedagoške raziskave. Sodobna pedagogika, 45/111, št. 5–6, str. 201–212.

Sagadin, J. (2003). Statistične metode za pedagoge. Maribor: Obzorja.

Sagadin, J. (2009). Veljavnost kvantitativnih empiričnih raziskav vzgoje in izobraževanja. V: Ličen, N (ur.). Pedagoško-andragoški dnevi 2009 – Misliti vzgojo: Problemi oblikovanja vzgojno-izobraževalnega koncepta (zbornik referatov). Ljubljana: Filozofska fakulteta, Oddelek za pedagogogiko in andragogiko, str. 55–59.

Schutt, R. K. (2006). Investigating the Social World: The Process and Practice of Research (5. izd.). Thousand Oaks etc.: Sage.

Slovar slovenskega knjižnega jezika (1994). Ljubljana: DZS.

Stuart, E. A. in Rubin, D. B. (2008). Best Practices in Quasi-Experimental Designs: Matching Methods for Causal Inference. V: Osborne, J. W. (ur.). Best Practices in Quantitative Methods. Los Angeles etc. Sage, str. 155–176.

Tacq, J. (1997). Multivariate Analysis Techniques in Social Science Research: From Problem to Analysis. London etc.: Sage.

Thompson, S. K. (2002). Sampling (2. izd.). New York: Wiley.

Janez SAGADIN, Ph.D. (University of Ljubljana, Slovenia)

THE VALIDITY OF QUANTITATIVE RESEARCHES IN THE EDUCATION FIELD

Abstract: The author considers a study as valid if it can provide valid and credible results. The conditions for the validity of quantitative empirical research in the education field are the following: the validity of data collecting procedures / measurements as a condition for the validity of collected data, the validity of data analysis results, the validity of generalization and the validity of causal explanations or conclusions. He points out how to satisfy these conditions, explaining the causes of validity disturbance, in the framework of the particular condition. He finds out that we can only more or less approach the absolute fulfilment of conditions and consequently also the absolute validity of the research. In connection with the first condition, he particularly emphasizes the advantage of the measurement model of the confirmatory factor analysis over the measurement model of the classical test theory. When treating the second condition, he points out the validity of statistic procedures of data analysis (statistic validity of the research). In the framework of treating the third condition, he considers the conditions for the validity of generalization, from probability samples to target populations (statistic induction), pointing out also the contents view of generalization. Since the employment of statistic induction is not applicable to improbability samples, he suggests, in that case, the direction to the internal validity of findings. In addition, he shows the arbitrariness of concluding about target populations. The discourse on the validity of generalization, regarding experimental studies, is organically included into the treatment of the validity of causal explanations. The author states the conditions for the existence of the causal connection between two variables. He also states the assessment criteria if there exists such a connection. He explains that the inductive – deductive approach or procedure is typical of quantitative empirical research. First, there is a deductively achieved and theoretically checked (verified) causal hypothesis (hypothesis about a causal connection) and then the empirical testing of this hypothesis. It verifies if the hypothesis (including its causal explanation) is supported by empirical data, properly collected and analysed. In that case, the causal explanation is considered as valid. In the article, the empirical testing of causal hypotheses is more specifically presented for experimental studies (for random and non-random experiments, especially for those, which are more applicable to the field of education). Regarding the testing, he emphasizes the role of internal validity of experiments. Concerning non-experimental studies, the above mentioned testing is presented in a shorter form, as well as the procedures regulated by SEM (Structural Equation Modelling).

Key words: validity of (studies, data collecting procedures/measurements, data, data analysis procedures, empirical analysis findings, generalization, causal explanations), internal validity, external validity, statistic validity of studies, statistic induction, causal hypothesis, empirical testing of causal hypotheses.