

Študentske evalvacije Univerze v Ljubljani

Vasja Vehovar



Univerza v Ljubljani
Fakulteta *za družbene vede*

Struktura

1. Splošen uvod v evalvacijo pedagogov
2. Splošen uvod v študentske ankete
3. Metodologija študentskih anket UL
4. Bipolarna lestvica 1-5
5. Pregled izpisov
6. Povratna zanka

1. Splošen uvod v evalvacijo pedagogov

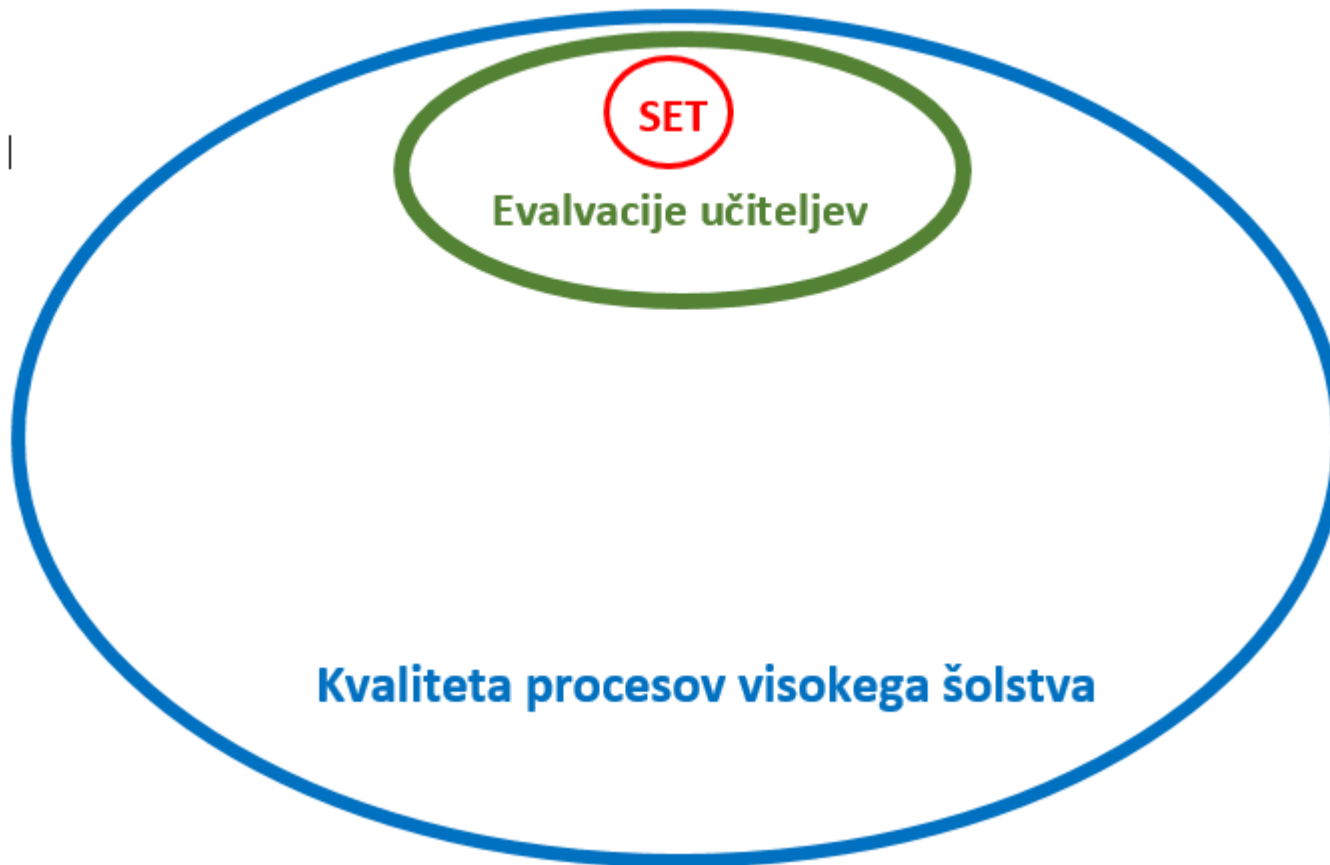
Pojav in razvoj študentskih evalvacij

- Študentske evalvacije oziroma ankete so poznane kot **SET** (student evaluation of teaching), včasih tudi **SETE** (student evaluation of teaching **effectiveness**), saj gre v osnovi za ocenjevanje učinkovitosti poučevanja;
- Pojavile so se v ZDA 1923, U of Pennsylvania (Max Freid), institucionalizirati so se pričele 1927 na Purdue University (Herman Remmers).
- Posebej so se razrasle s splošnimi koncepti kvalitete procesov (npr. TQM), kjer velja načelo, da se vsi koraki procesov evalvirajo;
- Konec 1970-tih let je SET uporabljala že tretjina univerz v ZDA, danes je to standard na praktično vseh univerzah v ZDA;
- SET so eno najbolj raziskovanih področij visokega šolstva; obstaja tisoče knjig, deset tisoče člankov, razprave, konference;
- SET vprašanja imajo ordinalne lestvice (5 ali 7 kategorij), kjer se ocenjuje splošno zadovoljstvo s predmetom oziroma z pedagogom (CORE) ter različne komponente (od pravičnosti ocenjevanja do urejenosti gradiv na spletu).

SET kot povratna zanka

- Če razumemo poučevanje kot poslovni proces, v katerem porabi študent pri predmetu 100-150 ur dela, je za izboljšavo procesa primerno, da porabi še 1-2 minuti (0.01%) časa za evalvacijo.
- Podobno je za optimizacijo procesov primerno, da tudi pedagog za npr. 200-300 ur dela pri 60 urnem predmetu (tretjina plače), porabi nekaj časa za refleksijo, evalvacijo in izboljšanje procesov.
- Ena skrajnost: nobene formalne povratne zanke, niti SET.
- Druga skrajnost: pretiran obseg evalviranja in pripisovanje prevelike ali napačne teže rezultatom SET.

Umestitev SET



Nameni evalvacije pedagogov

Osnovni namen evalvacije pedagogov je merjenje kvalitete in učinkovitosti (effectiveness) poučevanja oziroma prenosa znanja od pedagoga na študente.

Za katere namene se evalvacije pedagogov uporabljajo?

1. **Formativni:** izboljšanje pedagoškega dela;
2. **Sumativni:** letna evalvacija, zaposlovanja, napredovanja (kar je seveda zelo kritično in občutljivo);
3. **Programski:** akreditacije, monitoring programov.

Za vsakega od treh namenov se priporoča različna kombinacija evalvacijskih pristopov poučevanja; nikakor ne samo SET.

Pristopi k evalvaciji pedagogov

(* zvezdica označuje najbolj pogoste pristope)

1. *Študentske ankete (SET)
2. *Evalvacije predavanj s strani kolegov (peer review)
3. *Pregled gradiv s strani kolegov
4. Nasveti mentorjev
5. *Evalvacije s strani zunanjih ekspertov
6. Samo-evalvacije pedagogov
7. Video (samo-oceni ga pedagog in/ali kolega)
8. *Študentski intervjuji
9. Evalvacija diplomantov (nekaj let po diplomi)
10. Evalvacije zaposlovalcev (organizacije, ki zaposlujejo diplomante)
11. Administrativne evalvacije (analiza aktivnosti, letni razgovori)
12. Razvoj, raziskovanje, prezentacije, aktivnosti na področju poučevanja
13. Nagrade in priznanja za pedagoško delo
14. *Analiza rezultatov: ocene, znanje, učni izidi, kompetence, dosežki
15. *Portfolio/dosje pedagoga in dokumentacije

Standardi in priporočila

Za področje evalvacije pedagogov v terciarnem sektorju izobraževanja so razviti številni standardi in priporočila, posebej v Severni Ameriki, npr. SEPT (APA), CAUT (CA), EEOC (US).

V Evropi je veliko spodbudo pomenilo sprejetje priporočil na področju kvalitete procesov v visokem šolstvu ESG 2005 (ENQA).

Evalvacije pedagogov v Evropi

Raziskava (Trend 2015, EUA) med evropskimi institucijami terciarnega izobraževanja kaže naslednje:

- 98% zbira neko povratno informacijo študentov;
- 93% uporablja SET;
- 89% ima vzpostavljen regularen sistem evalvacije poučevanja;
- 72% ima sistem za obravnavo kritičnega segmenta pedagogov;
- 42% uporablja analizo port folia (plus 20% pripravlja);
- 37% uporablja evalvacije kolegov (plus 20% pripravlja).

Najdlje so šli v UK, kjer je Vlada letos (2016) sprejela sklep, da SET postane standardna (izvaja se centralno za vse univerze v UK) in obvezna komponenta za Teacher evaluation framework (TEF).

2. Splošen uvod v SET

Prednosti SET

SET ima v primerjavi z drugimi pristopi veliko prednosti:

- Rezultati so kvantitativni, številke so načeloma enostavne za obdelavo, analizo, interpretacijo in za povratne zanke.
- Mogoče so enostavne analize trendov, agregirane analize na različnih nivojih in primerjave po različnih skupinah.
- Izvedba preko spleta je sicer tehnično kompleksna, ko pa se enkrat vzpostavi, je enostavna in avtomatizirana.
- V primerjavi z drugimi pristopi se izvede izjemno hitro.
- Je neprimerljivo cenejši od vseh alternativ.
- Zagotavlja široko in podrobno povratno informacijo, saj se informacija pridobi prav od vsakega študenta za vsak predmet.

Problemi SET

Ker je SET pogosto osnova za sumativne namene (nagrajevanje, napredovanje in zaposlovanje), je razumljivo, da je izpostavljen obsežnemu preverjanju in številnim kritikam. Glavni problemi oziroma očitki so naslednji:

1. SET določajo zunanji dejavniki, na katere pedagog nima vpliva.
2. SET sploh ne meri učinkovitosti pedagoškega dela.
3. SET meri predvsem pedagogovo všečnost.
4. SET ima vrsto metodoloških težav.

Problemi SET #1

Vpliv zunanjih dejavnikov

SET je opazovalna študija

Ker študent ne ocenjuje v laboratorijski pogojih – niti niso mogoči eksperimenti (tako kot sicer ne v družboslovju) – je dokazovanje vzročnosti težavno. Glavni faktorji, ki lahko (povsem mimo pedagoga) vplivajo na SET, so naslednji:

1. Lastnostni pedagoga: spol, starost, seniornost, značaj, izgled;
2. Lastnosti študentov: spol, starost, letnik, srednja šola, točke na maturi, pričakovana višina ocen, interesi, program;
3. Vsebina predmeta: nepriljubljene (npr. Statistika 😞) - bolj priljubljene (npr. Zgodovina popularne glasbe);
4. Posebnosti generacije (npr. vpisni pogoji, številčnost);
5. Način (pisemsko/spletno) in čas ocenjevanja (pred/po izpitu);
6. Ostale okoliščine: govorice, incidenti, posebnosti, dogodki.

Izpostavljeni zunanji dejavniki

Posebej pogosto so v raziskavah izpostavljeni naslednji dejavniki, ki lahko vplivajo na SET ocene določenega pedagoga (ne glede na njegovo kvaliteto, učinkovitost in prizadevanje):

- spol študenta (ženske naj bi ocenjevale bolj ugodno);
- spol pedagoga (moški naj bi prejeli boljše ocene);
- velikost razreda (boljše ocene so v skupinah z manj študenti);
- nivo študija (podiplomski ima višje ocene kot dodiplomski);
- ocene (bolje ocenjeni pedagogi dajejo študentom višje ocene).

Različne raziskave empirično dokazujejo, da so vplivi navedenih dejavnikov na ocene SET lahko zelo veliki.

Posledice

Če naj bi zunanji dejavniki – na katere nima pedagog nobenega vpliva – odločilno vplivali na SET ocene določenega pedagoga, potem je SET seveda neprimeren instrument za tako pomembne odločitve.

Na tej osnovi so nasprotniki pristopu SET odrekli kvaliteto in s tem tudi legitimnost.

Zunanji dejavniki: Realna vloga

- Raziskave so pokazale, da je učinek zunanjih dejavnikov v splošnem razmeroma majhen in nekonsistenten, čeprav obstajajo tudi primeri, kjer je ta učinek velik.
- Kadar obstajajo večji učinki, je pri primerjavah in interpretacijah potrebna previdnost. Če so na neki fakulteti npr. pri velikih skupinah ocene bistveno nižje kot pri majhnih, potem neposredne medsebojne primerjave pedagogov, ki učijo zgolj velike oziroma zgolj male skupine, niso korektne.
- Zaradi vpliva zunanjih dejavnikov je varno primerjati predvsem medletne izvedbe istega predmeta pri istem pedagogu in istih pogojih (enak program) ali pa različna predmeta istega pedagoga v istem letu. Pri primerjavah na agregiranih nivojih (programi, oddelki, fakultete, pedagogi) lahko prihaja do več nekontroliranega vpliva in interakcije zunanjih dejavnikov.

Miti in raziskovalna dejstva o zunanjih dejavnikih

Pri izpostavljanju zunanjih dejavnikov, ki naj bi odločilno vplivali na ocene SET, gre torej večinoma za pretiravanja in preveliko posploševanje posamičnih raziskav izvedenih v specifičnem okolju. Hkrati gre tudi za ignoriranje širšega nabora raziskav, ki kažejo drugače.

Velika vloga zunanjih dejavnikov je bila namreč – na osnovi izčrpnih meta-analiz – ovržena že v prejšnjem tisočletju:

LAWRENCE M. ALEAMONI. 1999. Student Rating Myths Versus Research Facts 1924-1998

Avtor v članku obravnava in zavrne 16 najpogostejših mitov na področju SET, ki se nanašajo na vlogo zunanjih dejavnikov, od povezave z ocenami in spolom do velikosti razreda.

Zunanji dejavniki: Primer UL in FDV

Učinki zunanjih dejavnikov – ki so neodvisni od pedagogovega prizadevanja – so torej praviloma majhni. To lahko pokažemo tudi na primeru UL oziroma FDV:

- Variiranje nivoja SET ocen med članicami UL (2016)
- Variiranje nivoja SET ocen na FDV 2015 zaradi:
 - Različne zahtevnosti predmeta (glede na število KT);
 - Različnega števila študentov pri predmetu;
 - Nivoja študija (dodiplomski-podiplomski).

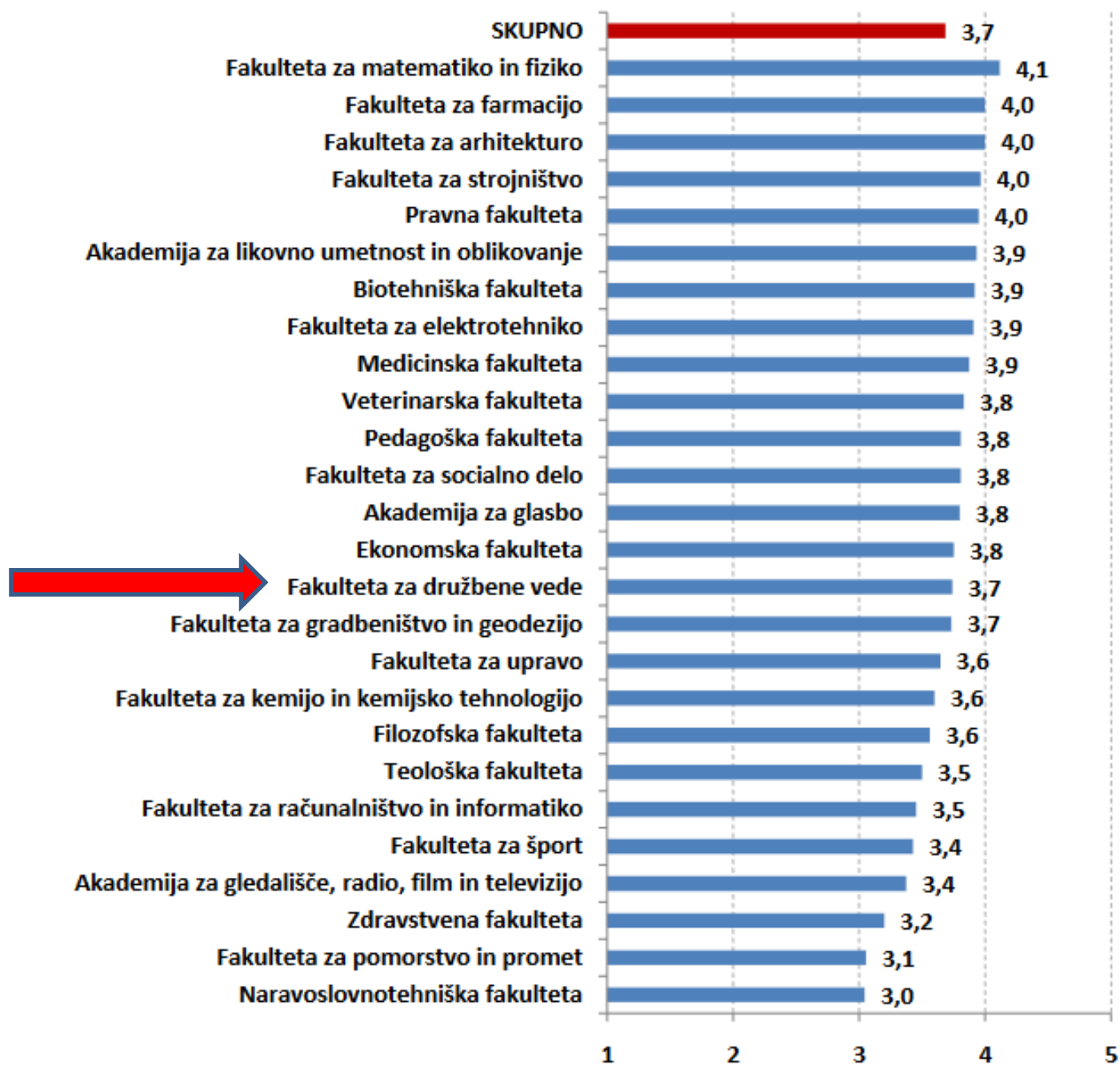
Primer UL: Splošni nivo zadovoljstva

Splošen nivo zadovoljstva študentov je vsekakor pomemben zunanji dejavnik, ki bi lahko vplival tudi na SET.

Na naslednji strani je ilustracija splošnega nivoja zadovoljstva študentov na članicah UL – kjer poleg prizadevanj, kvalitete in učinkovitosti pedagogov – seveda nastopajo tudi zunanje okoliščine, ki lahko vplivajo na SET (mimo pedagoga). Navedimo nekatere:

- prostori (npr. prostorska stiska);
- vpisni pogoji (npr. le kritični študenti z visokim številom točk);
- specifičnost profila (npr. zdravnik, matematik, umetnik);
- velikost skupin;
- in še cela vrsta drugih (domnevno) znanih in neznanih dejavnikov.

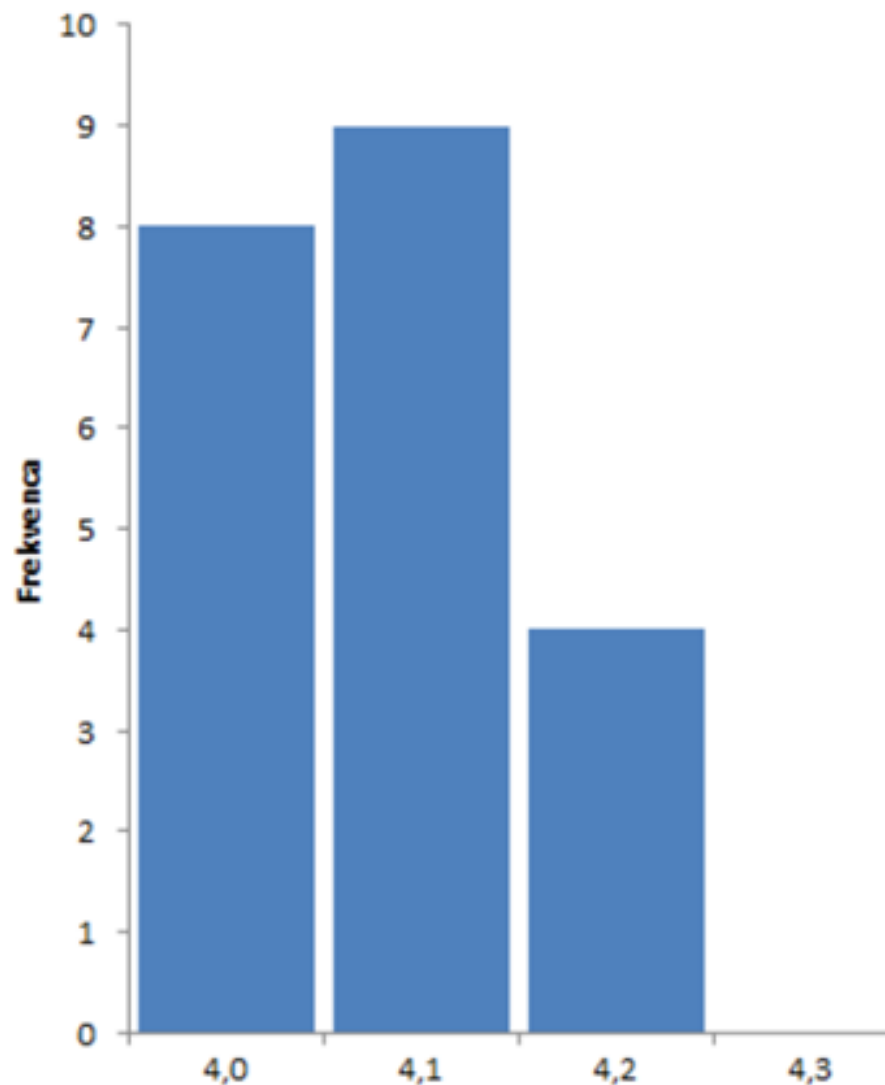
Primer splošnega zadovoljstva s študijem na UL (Raziskava ŠOU 2010)



Primer UL: Splošni nivo SET

- V primerjavi z velikim razponom splošnega zadovoljstva študentov (slika zgoraj) nivo ocen SET na članicah UL variira bistveno manj (slika spodaj).
- Razlog za to so radikalno ožja in natančna SET vprašanja, ki se nanašajo na specifične komponente učnega procesa, na kar zunanji dejavniki ne morejo veliko vplivati.
- Variiranje povprečnih ocen med članicami UL za komponento SET “splošno zadovoljstvo s predmetom” (slika spodaj) je zato v ozkem intervalu 4.0-4.2, medtem ko je bilo splošno zadovoljstvo (slika zgoraj) v širokem intervalu 3.0-4.1.
- Majhno variiranje SET med članicami ilustrira in potrjuje, da je vpliv zunanjih dejavnikov na SET v bistvu zelo šibek.

Članice UL glede na povprečno zadovoljstvo s predmeti (SET 2016, prvo polletje, preliminarne ocene)



Primer FDV: Zunanji dejavniki in SET

Tudi podatki FDV potrjujejo odsotnost vpliva zunanjih dejavnikov oziroma zgolj šibek vpliv na nivo SET:

- **Zahtevnost predmeta**, ki jo študenti ocenijo glede na število porabljenih ur, ni povezana s SET - korelacijski koeficient (označen z # v tabeli spodaj) znaša namreč le **0.01**. pedagog, ki bolj obremenjuje študente, torej nima nižjih SET ocen. Tudi sicer lahko v spodnji tabeli razberemo, da zahtevnost predmeta ni povezana s skoraj nobeno SET komponento razen (zelo šibko) s percepcijo pridobljenih kompetenc (0.23) in pridobljene kritičnosti (0.14). Kljub temu lahko pri določenih predmetih zahtevnost vpliva na SET. Seveda zahtevnost še ne pomeni tudi ocen; vpliv ocen bi bilo potrebno preveriti posebej.
- **Število vpisanih študentov** je šibko povezano s SET „splošno zadovoljstva s predmetom“, saj je korelacijski koeficiente samo **-0.13**, kar pomeni, da večje število študentov pri predmetu le minimalno vpliva na SET ocene (znižuje).
- **Stopnja študija**: SET ocene na prvi stopnji se – zaradi vrste zunanjih okoliščin – sicer nekoliko razlikujejo od nivoja SET ocen na drugi stopnji, vendar so razlike majhne, večinoma ena ali dve desetinki (npr. **4.1 - 4.3**).

Korelacija SET komponent na FDV

(* zvezdica pomeni ocene podane PO izpitu)

Komponente SET	1. Zadovolj.	12. KT	8. Kompeten.
1. Splošno zadovoljstvo	1.00	# 0.01	0.69
2. Deli predmeta usklajeni	0.83	- 0.01	0.61
3. Spodbuda kritičnemu razm.	0.85	0.14	0.65
4. Študijska literatura	0.76	0.05	0.59
5. Dobra obveščенost	0.55	- 0.07	0.41
6. Gradivo na spletu	0.54	0.03	0.40
7. Ustrezno sprotно preverjanje	0.77	- 0.10	0.51
8* Pridobljene kompetence	0.69	0.23	1.00
9* Izpit - vsebinsko ustrezen	0.66	0.01	0.78
10* Izpit - jasne naloge	0.59	- 0.09	0.62
*11 Izpit - kriteriji ocenjevanja	0.57	- 0.06	0.69
*12 Poraba časa gled na KT	# 0.01	1.00	0.23

Problemi SET #2

Všečnost poučevanja

Ali SET meri všečnost?

Nasprotniki pristopa SET izpostavljajo, da SET meri predvsem všečnost, ki nato določa komponente SET, posebej oceno splošnega zadovoljstva. Številne študije dejansko kažejo na povezanost med všečnostjo ter ocenami SET.

- Pogosto se navaja *Half a minute* (Ambady & Rosenthal, 1993), kjer je ocena 30 sekundnega videa uvodnega nastopa pedagoga brez zvoka (le izgled in govorica telesa) visoko korelirala (0.76) s SET oceno zadovoljstva ob zaključku predmeta.
- Zelo znan je „dr. Fox“ efekt (*A Paradigm of Educational Seduction*, Naftulin et al, 1976) izobraževalnega zapeljevanja, kjer je karizmatični in ekspresivni - vendar prazen in vsebinsko nekvaliteten - predavatelj prejel boljše SET evalvacije kot vrhunski strokovnjak v korektnem, a monotonem predavanju.
- Posebej intrigantne so številne raziskave, ki (razmeroma konsistentno) dokazujejo povezavo med SET in (neodvisno in objektivno izmerjeno) lepoto pedagoga (Npr. *Beauty in the classroom*, Hammers&Parker, 2005).

Večina teh študij je vezanih zgolj na specifično okolje, mnoge pa so tudi metodološko sporne. Kljub temu ne gre zanikati, da ima všečnost lahko nek učinek na SET, podobno kot ga imajo zunanje okoliščine.

Štirje viri všečnosti pedagoga

Komponente všečnosti pedagoga lahko strukturiramo v štiri skupine, glede na to, kako nastanejo oziroma kako so pridobljene:

1. Psihofizičen profil: čednost, lepota, postavnost, privlačnost, simpatičnost, empatičnost, čustvena inteligenca, seksapil, karizmatičnost, ekstrovertiranost, šarm, priljudnost itd.
2. Nadgrajene fizične lastnosti: lepotni posegi in dodatki, urejenost, oblačenje, ličila, uporaba osebnega stilista.
3. Nadgrajene splošne vedenjske lastnosti: šolan nastop, izvežban glas, naštudirana govorica telesa, igralski prijemi.
4. Specifični predavateljski prijemi za večanja všečnosti (ki jih ločujemo od siceršnje **prizadevnosti** pri poučevanju); nekateri od teh so sprejemljivi, nekateri pa morda tudi niso (podrobnosti na naslednji strani).

Sprejemljivo večanje všečnosti

Nekaj klasičnih predavateljskih prijemov večanja všečnosti, ki se npr. v ZDA rutinsko uporabljajo:

- Dodatna in posebna prilagoditev izgleda in osebne urejenosti za predavanja (npr. kravata v primerih, ko to pomaga);
- Sistematični napor za memoriziranje študentov po imenu, kar se nato tudi uporablja pri njihovem naslavljanju;
- Vztrajanje, da študent pedagoga kliče po imenu (tikanje?);
- Deljenje privatne telefonske številke za pomoč;
- Izrecno in sistematično ponavljajoče spodbujanje študentov, da se lahko obrnejo na pedagoga za kontakt in pomoč;
- Izrecna zahvala za vsako postavljeno vprašanje v smislu, da je bilo to zelo dobro vprašanje ter preverjanje, če je odgovor v redu;
- Izčrpna in izjemno zavzeta podpora ter odprtost v primeru, ko ima študent kakšne predloge, ideje, projekte ali probleme itd.;
- Sistematično vnaprej naštudirane in dozirane šale (npr. na 20 minut).

Sporni prijemi večanja všečnosti

V okoljih, kjer je SET odločilnega pomena za zaposlitev ali napredovanja, so pedagogi v svoji stiski razvili prijeme, ki so lahko sporni:

- Najbolj učinkovita za višanje SET je hrana, posebej piškoti, pa tudi „drugi snacki, ali pijača v bifeju, predvsem na zaključnem srečanju oz. ob evalvaciji.
- Prilagoditev zadnje naloge (assignment), tako da je lahka in zanimiva, morda celo zabavna, hkrati pa se pred evalvacijo ne posreduje njene ocene.
- V okoljih, kjer imajo študenti srednješolsko mentaliteto, lahko pomaga tudi krajšanje predavanj, kasnejši začetek, odpoved predavanj ipd.
- Popuščanje pri izpitnem režimu: toleriranje zamud pri oddaji nalog, možnost ponovnega neformalnega opravljanja izpita, naknadne možnosti večanja ocen, dodatna predavanja ali seanse za pripravo na izpit ipd.
- Višanje ocen, nižanje izpitnih kriterijev in poenostavljanje snovi, vključno s strategijo „teaching-to-the-test“ namesto „real teaching“.
- Nepotrebni in pretirani prijemi za poživitev (npr. šale, video, gosti, izleti), kjer je glavni ali izključni cilj večanje všečnosti, ne pa pedagoška odličnost.
- V primeru pisemskega izpolnjevanja v učilnici učinkuje tudi ustvarjanje vtisa naglice (ocene so potem višje), kot tudi dejstvo, da pedagog ostane v učilnici.

Preprečevanje spornih prijemov

Sporne prijeme je mogoče omejiti ali regulirati:

- Administrativna kontrola (npr. prisotnost v predavalnici, na GU).
- Reguliranje obsega, timinga in nivoja osebnega druženja s študenti s kodeksi, priporočili, pravilniki ipd.
- Reskaliranje ocen: mnoge univerze zahtevajo, da je povprečje ocen pri vseh predmetih izravnano, npr. enako 8.0.
- Izenačevanje zahtevnosti predmetov v smislu porabe ur glede na KT – predmeti, kjer so velika odstopanja navzgor ali navzdol, morajo prilagoditi svojo zahtevnost.
- Paralelni pristopi k evalvaciji (npr. samoevalvacija, portfolio analiza, evalvacija kolegov, eksterna evalvacija, učni izidi, pogovori s študenti) so daleč najučinkovitejši način za odkrivanje spornih praks.
- Spodbujanje pedagoške prizadevnosti pa je edini pravi način za prekanaliziranje naporov od vsečnosti k prizadevnosti (glej spodaj).

Pedagoška prizadevnost

Pedagoška prizadevnost se nanaša na napore, ki so – v skladu s pedagoško stroko – potrebni za doseganje pedagoške odličnosti, kvalitete in učinkovitosti. Dosega na različne načine:

- pedagog uporabi klasična spoznanja in priporočila pedagoške stroke; v tem okviru je na voljo izredno bogata strokovna literatura, spletna gradiva, multimedijски pripomočki ter standardna izobraževanja.
- pedagog se poleg tega lahko tudi kontinuirano (samo)izobražuje v pogledu pedagoškega dela, zanima za dobre prakse, odlične zgledе in preizkuša oziroma uvaja sodobne pristope in tehnike.
- pedagog pozna komponente konkretnega SET vprašalnika, in poskrbi zanje, npr. če ima SET vprašalnik komponento „urejenosti gradiv na spletu“, učitelj za to dobro poskrbi.
- **Predvsem in v prvi vrsti pa pedagoška prizadevnost pomeni, da pedagog za pripravo poučevanja porabi dovolj časa. Enako velja za čas, potreben za skrbnost in promptnost pri popravilnih nalogah, seminarjev in izpitov, ter za čas potreben za formalno in neformalno kontaktiranje s študenti.**

Všečnost in prizadevnost

V splošnem ima lahko všečnost nek vpliv na SET, vendar – podobno kot zunanji dejavniki (večinoma do desetinko ali dve) - bistveno manjši kot pa je vpliv pedagoške prizadevnosti pedagoga.

- **Dokaz 1:** Mnogi pedagogi lahko opazijo pri svojih različnih predmetih (v istem letu) velike razlike v SET ocenah, kar je praviloma posledica zaradi razlik v prizadevnosti, kompetentnosti ter motiviranosti oziroma entuziazma za različne vsebine.
- **Dokaz 2:** Pri istem predmetu lahko pedagog opazi premembe SET ocen, vsakič ko pomembno spremeni nivo svoje pedagoške prizadevnosti. Praviloma imajo vse večje medletne spremembe razloge, ki jih je mogoče identificirati s spremembo prizadevnost (včasih pa seveda tudi z zunanjimi okoliščinami).
- **Dokaz 3:** Na fakultetah, kjer imajo mehanizem za obravnavo nizkih ocen, pedagogi z nizkimi SET ocenami po dodatnem izobraževanju in prizadevanju skoraj vedno bistveno popravijo svoje SET ocene (ne glede na všečnost).

Problem SET #3

Učinkovitost poučevanja

Ali SET meri učinkovitost?

- Obstajata dve večletni eksperimentalni (in NE opazovalni) študiji, kjer so ugotovili, da SET ocene niso povezane z učinkovitostjo: Carrell and West (2008) ter Braga, Paccagnella and Pellizzari (2011).
- Nasprotniki SET zato trdijo, da SET ne meri učinkovitosti poučevanja.
- Vendar dve študiji, ki to problematizirata, nista dovolj, še posebej, ker sta učinkovitost definirali na specifičen način: z ocenami, ki so jih študenti (izmerjeni s SET) nato dosegali pri nadaljevalnih predmetih.
- Če pa SET dejansko ne meri učinkovitosti poučevanja, potem je to siceršnji problem vseh evalvacijskih metod. SET ocene so namreč dobro korelirane z drugimi evalvacijskimi pristopi (samo-evalvacija, kolegialna evalvacija, eksterna evalvacija, učni izidi, intervjuji).
- V takem primeru je učinkovitost pač treba meriti drugače. Kako? Nasprotniki SET ne navajajo rešitev, ki bi merile učinkovitost in bi lahko na letni osnovi služile za formativne in sumativne namene.

Problem SET #4

Metodološke težave

Metodološki problemi SET

Obstajajo resni metodološki problemi SET. Vendar je mogoče z ustreznim pristopom njihov negativen vpliv preprečiti:

- Nizke stopnje odgovorov lahko pomenijo pristranske rezultate (posebej, če bolj/manj kritični študenti v SET ne sodelujejo).
- Če študenti ne zaupajo tajnosti odgovorov, potem v evalvacijah olepšujejo.
- Pojavijo se lahko neprimerni, žaljivi komentarji za pedagoge.
- Ordinalna lestvica – posebej v primeru asimetrije, bipolarnosti, sploščenost – je lahko neprimerna za izračun povprečij.
- Problematična je lahko veljavnost in zanesljivost SET meritev.
- Pojavijo se lahko učinki načina anketnega ocenjevanja (splet, papir...).
- Ocenjevanje pred izpitom ali po izpitu (ali celo oboje, kot na UL).
- Kako doseči sodelovanje študentov? Prisiljevanje lahko generira zadostovanje (satisficing), kjer študenti odgovarjajo minimalistično.
- Kdaj vabiti v anketo: ob prijavi vsakega predmeta posamično ali ob semestru za vse predmete hkrati? Kdaj in kako pogosto opominjati?
- Obstaja tudi vrsta vprašanj pri statističnih analizah in interpretacijah.

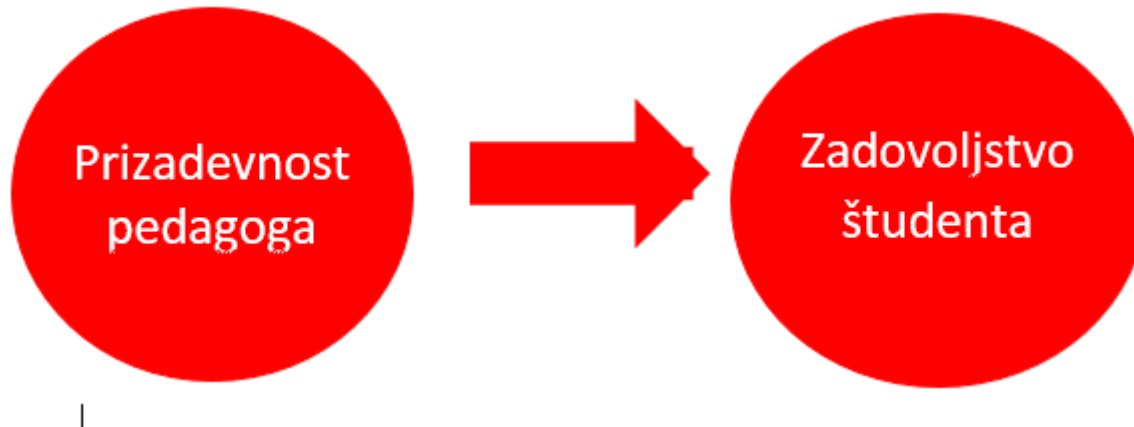
Problemi SET

Povzetek konceptualnih težav

Kaj merijo SET vprašanja

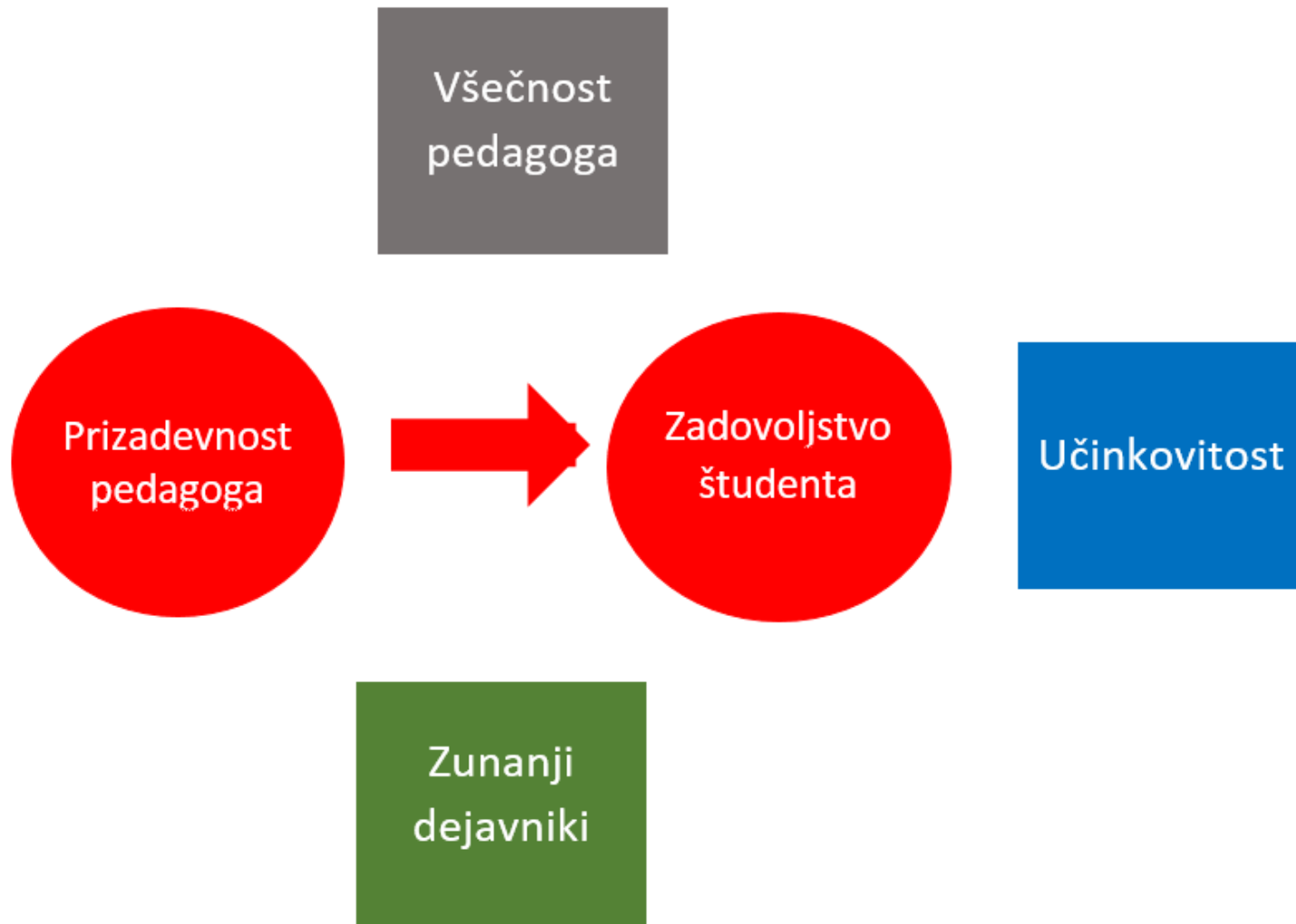
SET meritve se nanašajo na dva povezana sklopa vprašanj:

- Specifične komponente SET: gre za različne vidiki prizadevnosti pedagoga (npr. korektnost odnosov, urejenost gradiv na splelu, pravičnost ocenjevanja itd.) kot jih percipirajo študent;
- Splošna komponenta SET (CORE): zadovoljstvo študenta s pedagogom oziroma s predmetom



Komponente prizadevnosti pedagoga vplivajo na SET zadovoljstvo študenta.

Kakšno je razmerje SET do zunanjih dejavnikov, vsečnosti in učinkovitosti?

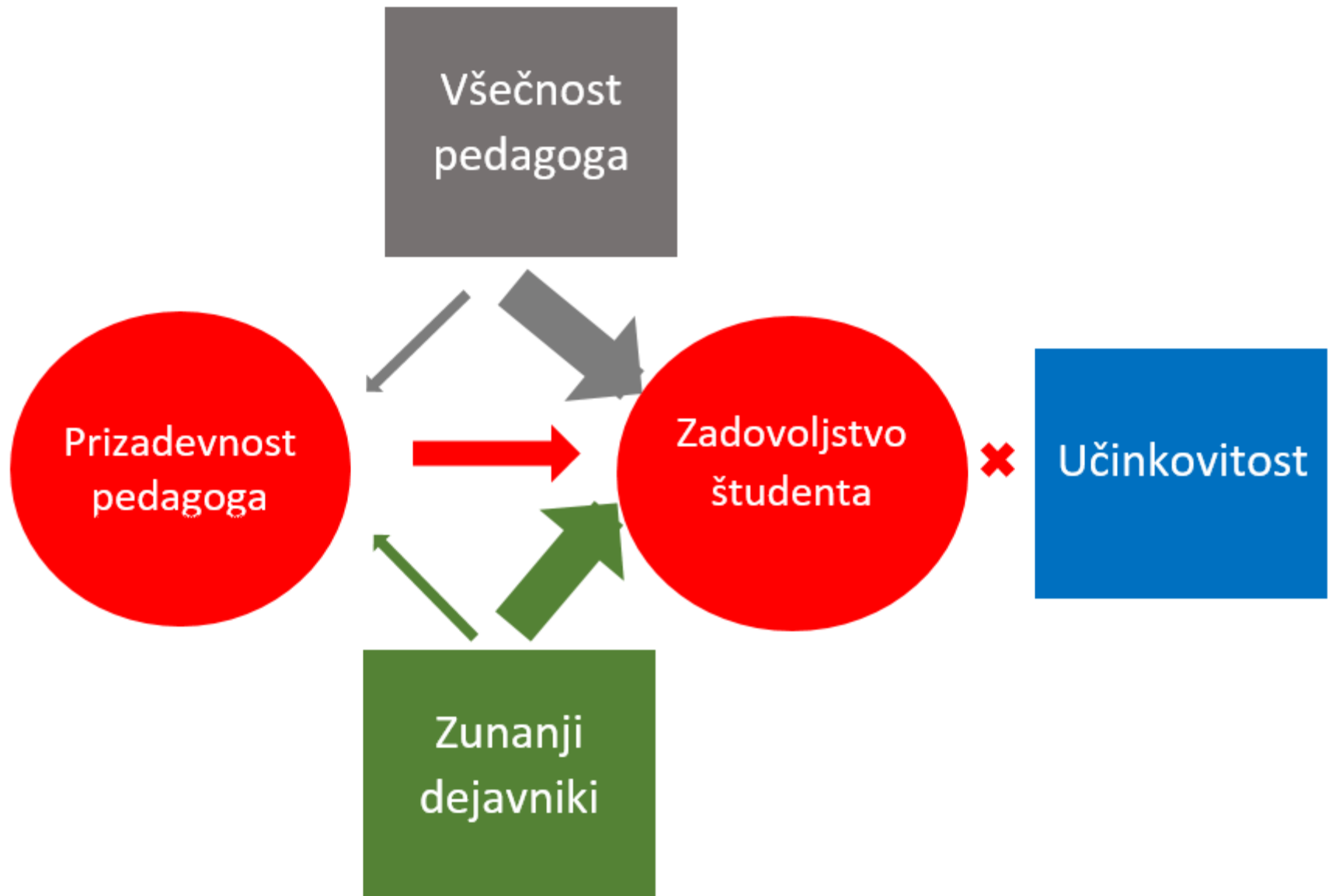


Kritiki SET

Če povzamemo kritike SET, so glavne pripombe naslednje (shematsko so prikazane tudi na sliki spodaj):

- Na SET pomembno vplivajo zunanji dejavniki, na katere pedagogi nimajo vpliva;
- Na SET pomembno vpliva tudi všečnost pedagoga, na kar pa pedagog ima določen vpliv;
- Percepcija prizadevnosti pedagoga in SET zadovoljstvo so visoko korelirani, oboje pa je predvsem rezultat zunanjih dejavnikov in všečnosti
- SET ne meri pedagoške učinkovitosti.

SET zato usmerja pedagoge v napore za povečanje všečnosti, ne pa v napore za večanje pedagoške prizadevnosti.



Zagovor SET

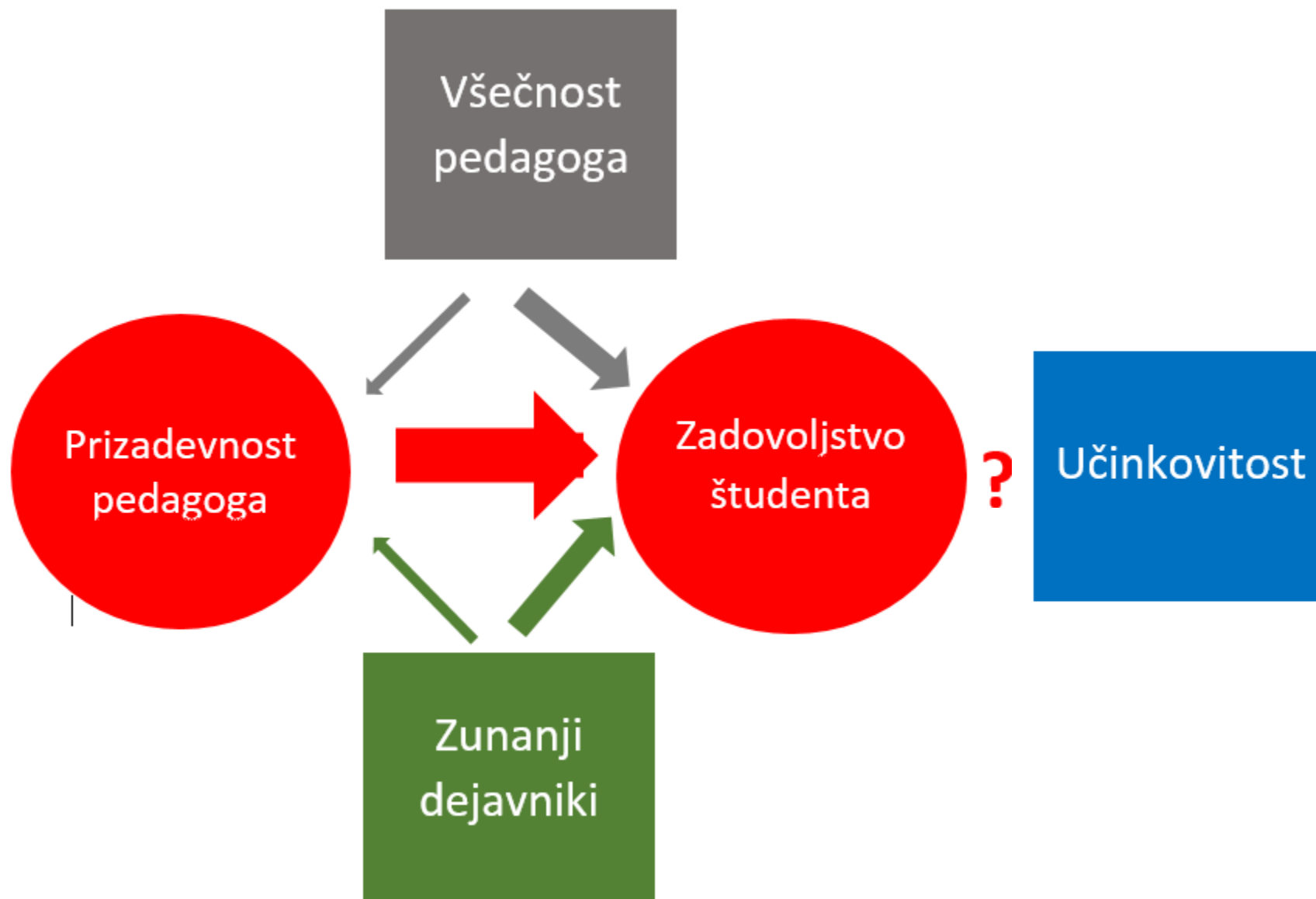
Učinkovitost:

- Morda res še ni raziskano, ali SET meri učinkovitost ali ne.
- **Vendar: tudi če ocene SET niso povezane z učinkovitostjo poučevanja, to še ne pomeni, da prizadevnost in zadovoljstvo nista pomembni.**
- **Študentu je namreč treba ponuditi oboje, UČINKOVITO in tudi PRIJAZNO poučevanje.**

Zunanji dejavniki imajo majhen vpliv (kar smo pokazali tudi na primeru UL in FDV). Če pa ta vpliv obstajaj, je pri interpretacijah in primerjavah na agregirani nivojih potrebna dodatna previdnost (lahko pa se izvede statistične korekcije).

Všečnost lahko ima nek vpliv na SET, vendar bistveno manjši kot je vpliv pedagoške prizadevnosti pedagoga.

Na tej osnovi je na sliki spodaj popravljena shema odnosov med koncepti.



Problemi SET

Ilustracije anomalij v ZDA

Skrajnosti v ZDA I

The mere mention of faculty evaluation to many college professors conjures up mental images of the “shower scene” from Psycho. They’re thinking: “Why not just whack me now, rather than wait to see those student ratings again.”

Berk, 2013

Skrajnosti v ZDA II

Among faculty, student evaluations of teaching (SET) are a source of pride and satisfaction—and frustration and anxiety.

High-stakes decisions including tenure and promotions rely on SET.

Yet it is widely believed that they are primarily a popularity contest; that it's easy to “game” ratings; that good teachers get bad ratings and vice versa; and that rating anxiety stifles pedagogical innovation and encourages faculty to water down course content.

Stark & Freishtat, 2014

Skrajnosti v ZDA III

... She was an assistant professor and she has a number burned into her mind: 4.7. That's the average student-evaluation score that she has to reach to get the tenure.

Stacey Patton, Morgan State University

Skrajnosti v ZDA IV

A Republican state senator in Iowa, Mark Chelgren, proposed a bill that would have fired professors, even tenured ones, who scored low on their student evaluations.

Mr. Chelgren cited high student-loan debt to argue that students should be able to hold their professors accountable through metrics. “Professors need to understand that their customers are those students”

www.npr.org

Dozorevanje vloge SET

- V začetnih letih je SET ocenjeval izgled, čustveno stabilnost, zagnanost pedagoga ... predvsem v formativne namene;
- Kasneje se je fokus postopno selil na zadovoljstvo študentov, pridobivanje kompetenc ... in tudi v sumativne namene.
- Do okoli leta 2000 se je SET predvsem razvijal, standardiziral in utrjeval, ter zagovarjal in odvrčal dvome oz. napade.
- Danes prevladuje v stroki precej enotno mnenje:
 - SET je **nepogrešljiva** komponenta evalvacije pedagogov;
 - SET je posebej pomemben za **odkrivanje izrazito slabih** pedagoških praks;
 - SET **ne sme biti edina** evalvacijska komponenta, posebej ne za sumativne odločitve – potrebni so še drugi, paralelni, evalvacijski pristopi;
 - SET se sicer sooča z vrsto metodoloških izzivov, ki pa jih je mogoče z dovolj skrbnim pristopom premagati.

2. Metodologija UL

Pravilnik in specifikacije UL

Pravilnik in specifikacije:

- Metodologija je v osnovi določena s Pravilnikom UL in je neodvisna od projekta tehnične implementacije v 1KA;
- Pravilnik UL podrobno opredeljuje tudi vse ankete:
 - anketa predmeta in izvajalca pred izpitom in nato še anketa po izpitu (kjer je vključen tudi poseben vprašalnik o praksi),
 - doktorska anketa za 1 in 2 letnik,
 - splošna anketa.
- V Tehničnih specifikacijah UL so definirani vsi izpisi in metodološke zahteve;
- Vse navedeno je dostopno na <http://ul.1ka.si>.

Kako se zagotavlja sodelovanje?

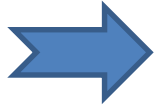
- Študenti se morajo odzvati (odgovoriti ali zavrni) na ankete vseh predmetov pred prvo prijavo v semestru (in ne pri vsaki prijavi na izpit posamično);
- Ni prisile: študent lahko sodelovanje vsake ankete zavrne, lahko tudi preskoči katerokoli vprašanje;
- Izvedba v sistemu 1KA – podrobnosti so v „Izvedbenih priporočilih“, <http://ul.1ka.si>;
- Vabila, opomniki in nadzor nad neopravljenimi anketami v celoti upravlja in oblikuje visokošolski sistem članice;
- Vprašanje promocije (motiviranja študentov) je zelo pomembno in je problem UL, vodstev fakultet, ŠS in tudi vseh pedagogov.
- Zavrnitve sodelovanja v SET s strani študentov so se na članicah UL ustalile pod 40%, kar je sprejemljivo.

Priprava vprašalnika

Visokošolska informatika članice
priprava podatkov o predmetih



vodstvo
članice
posebnosti



DINAMIČNO GENERIRAN VPRAŠALNIK 1KA



podatki o študijski programih
informatični sistem UL

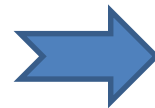
Potek anketiranja

Informacijski sistem UL

Podatek o letniku in študijskem programu študenta



Usklajena anketa: Anketna vprašanja



ZAPIS ODGOVOROV



Zaznamek v profilu študenta
(opravi anketo DA/NE)

Visokošolska informatika članic

Kaj točno je v zapisu?

Zapis se kreira v realnem času iz opisanih treh virov.

V zapisu študentove evalvacije je samo naslednje:

- Odgovori na anketna vprašanja;
- Podatki članice o ocenjevanem predmetu;
- Podatki UL o programu in letniku študenta.

Česa ni v zapisu študentove evalvacije:

- Nobenih parapodatkov (čas, IP številka ipd.);
- Nobenih identifikatorjev, šifer in osebnih podatkov.

Kako se podatki procesirajo?

strežnik z 1KA anketnimi podatki



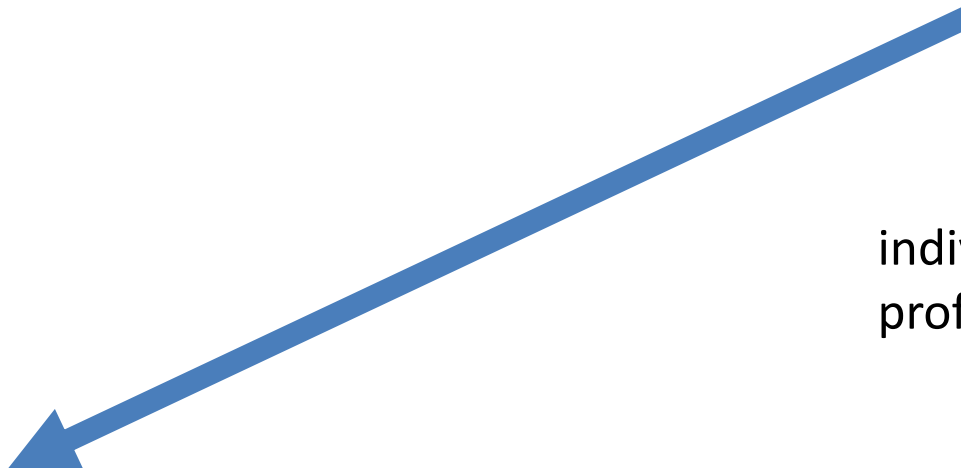
procesiranje v programu R na strežniku



izvoz PDF analiz v dokumentni sistem UL (Government connect – GC)



pedagogi prejmejo individualne analize v svoj profil v spletnem referatu



ostale analize gredo v GC profil pooblaščne osebe na članici, ki jih nato nadalje distribuira

3. Bipolarna lestvica 1-5

Lestvica

Znanost o merskih lestvicah je zelo široka in mnoge razprave o številu kategorij, orientaciji, postavitvi, označevanju, formatiranju itd. so še vedno odprte. V primeru implementacije SET za UL so odločitve naslednje:

- Opcije odgovorov so brez števil, zgolj opisne kategorije;
- Šele v procesiranju se lestvico oštevilči (1-5) in računa povprečje.
- Bipolarna (torej so 1 in 2 negativni vrednosti, in 2 ne pomeni zadostno) lestvica namesto unipolarne;
- Pet kategorij namesto sedem;
- Možnost nevtralnih kategorij je omogočena (vrednost 3);
- Možnost ne-vsebinskih kategorij (ne vem, ne želim) je omogočena;

Lestvica SET za UL

V celoti gledano sem s predmetom zadovoljen....

**sploh se ne
strinjam**

**se bolj ne
strinjam kot
strinjam**

**niti se ne
strinjam, niti
se strinjam**

**se bolj
strinjam, kot
ne strinjam**

**popolnoma
se strinjam**

*nimam
dovolj
informacij*

*ne želim
odgovoriti*



Interpretacije lestvice 1-5

Glavni mejniki za **povprečja** – in s tem povezan **delež pozitivnih ocen (vrednosti 4 in 5)**, ki se vse bolj uveljavlja za interpretacijo ordinalnih spremenljivk – so naslednji:

- **< 3.0**: več negativnih (1, 2) kot pozitivnih (4, 5) ocen; delež pozitivnih ocen (4, 5) je v grobem **pod 30%** – gre za kritično nizke ocene;
- **3.0 ≤ 3.5**: večina ocen ni pozitivnih (1, 2, 3); delež pozitivnih ocen (4, 5) je običajno **pod 50%** – potrebna je obravnava in izboljšanje;
- **3.5 ≤ 4.0**: glavnina ocen ima vrednost 4, vendar je ocen, ki niso pozitivne (1, 2, 3), več kot odličnih ocen (5); povezani delež pozitivnih ocen (4, 5) je običajno **pod 80%** – gre za standard na številnih univerzah in pomeni, da se pedagoga ne zaposli, ne vabi na poletne šole, onemogoči napredovanje ipd.; v formativnem pogledu ocena vsekakor sporoča, da je obstaja jasen prostor za izboljšave;
- **≥ 4.5**: večina enot (več kot 50%) je podala oceno odlično (5), delež pozitivnih ocen (4,5) je običajno **nad 90%** – gre za indic pedagoške odličnosti in pravilnik UL daje pri habilitaciji dodatne točke za pedagoško komponento.
- **≥ 4.8**: več kot tri četrtine enot je dalo oceno odlično (5) – gre za pedagoško izvedbo, ki v pogledu SET evalvacije skoraj nima napak.

Statistična značilnost

SET ocene imajo določeno vrednost oziroma odstopanja tudi zaradi slučajnih variacij, ki nastajajo zaradi vzorčenja in merskih napak:

- Vendar: pri SET imamo celotno populacijo, zato ne potrebujemo statističnega sklepanja in intervalov zaupanja.
- Upoštevati pa je treba upoštevati zanesljivost (reliability) odgovorov, ker bi študent pri ponovljenem merjenju morda odgovoril drugače.
- Variiranje zanesljivosti zavzame pri SET okoli 10-20% variance, ki je giblje v intervalu od 0.3 (ocene nad 4.0) do 0.8 (nižje ocene).
- Običajni (5%) „Interval zaupanja“, ki nastaja zaradi (ne)zanesljivosti odgovorov je pri $n=10$ med ± 0.1 in ± 0.2 , npr. 4.0 ± 0.2 . Pri $n=40$ oziroma $n=90$ pa je še 2 oziroma 3 krat ožji, pač za faktor $\sqrt{n/10}$.
- Pri primerjavah dveh predmetov, vsak z $n=10$, je lahko statistično značilna že razlika 0.2-0.3 (npr. $4.0 \rightarrow 4.2$). Če je n večji, se širina ustrezno zoži za odgovarjajoči faktor $\sqrt{n/10}$.
- Če primerjamo isto populacijo, npr. komponente določenega predmeta, ali pa predmete z istimi študenti, imamo odvisne vzorce in intervali zaupanja se običajno zožijo najmanj za 3-4 krat; v takem primeru je zagotovo statistično značilna že desetinka (npr. 4.0 - 4.1)

Pomembnost razlik

Razlike, ki niso statistično značilne, niso pomembne in ne morejo postati predmet vsebinskih interpretacij o vzrokih.

Na drugi strani statistična značilnost še ne pomeni pomembnost. Res pa je, da so zaradi statističnih okoliščin večinoma značilne še razlike za desetinko.

Kot kriterij pomembnosti pogosto uporabimo 5% ali 10% relativno razliko. V primeru SET ocen na skali 1-5 velja:

- Povečanje za desetinko (npr. 4.0 → 4.1) je v specifičnem okviru lestvice 1-5, že razmeroma pomembno, saj 0.1 pomeni 3% vrednosti ocene, povečanje 4.0 → 4.2 pomeni 5%.; povečanje 4.0 → 4.4 pa že 10%.;
- Dejstvo tudi je, da se SET ocene gibljejo praktično zgolj v intervalu 3.5-5.0, zato sprememba za 0.1 dejansko pomeni kar 7% celotnega razpona; sprememba 4.0 → 4.4 pa skoraj tretjino (27%) celotnega razpona.

Vsebinske interpretacije

- Ker je večina SET razlik statistično značilnih, imajo zato nek svoj vzrok: v prvi vrsti je to pedagoška prizadevnost, v določenih primerih pa vstopajo tudi zunanji dejavniki ali vsečnost pedagogov.
- Večinoma je glavni in najmočnejših vzrok, ki pojasnjuje razlike, torej prizadevnost oziroma napake v izvedbi. Identifikacija konkretnih vzrokov je stvar vsebinske razprave in poznavalci (predvsem pa vpleteni pedagog) jih večinoma lahko izluščijo, posebej, če gre za značilne in relativno velike razlike.
- Na ocene SET načeloma lahko – praviloma sicer minimalno - vplivajo tudi zunanji dejavniki (npr. velikost razreda) in vsečnost pedagoga (npr. izgled). Več kot zunanjih dejavnikov kontroliramo pri primerjavah, bolj verjetno pa lahko razlike pojasnimo zgolj s prizadevnostjo pedagoga.
- Tudi ko vseh okoliščin ne kontroliramo, pa zunanji dejavniki in vsečnosti praviloma vplivajo največ za nekaj desetink (obstajajo pa tudi izjeme).
- Če dejavnikov ne kontroliramo, so primerjave med pedagogi, programi, letniki, predmeti lahko motene in s tem tudi nekorektne do pedagogov.

Primerjave pri istih študentih

- **Primerjave SET komponent določenega predmeta.** Komponente, ki so ocenjene nižje jasno kažejo na možnosti izboljšanja, kar usmerja pedagoga, kje in kako izboljšati pedagoški proces. Pri predmetih je pogosto najbolj kritična komponenta „nivo pridobljenih kompetenc“, pri pedagogih pa komponenta „spodbujanje kritičnega razmišljanja“. Motečih vplivov zunanjih dejavnikov in vsečnosti v tem primeru ni, ravno tako ni slučajni vplivov.
- **Primerjave SET komponent za različne predmete.** Slučajni vpliv zaradi različnih populacij to ne moti in enako velja za zunanje dejavnike, ki so vezani na generacijo študentov. Načeloma gre tudi tokrat večino razlik pripisati kvaliteti oziroma prizadevnosti pedagoga, vendar lahko v tem primeru primerjave moti različna narava predmeta (oziroma specifična privlačnosti vsebin) ter različna vsečnost pedagoga.

Medletne primerjave predmetov

Medletne primerjave SET komponent istega (nespremenjenega) predmeta pri istem izvajalcu so zelo pomembne in pogosto v ospredju analiz.

Kot smo pokazali, so že pri 10 študentih statistično značilne razlike za 0.2-0.3 (npr. 4.1—4.3), pri predmetih z več študenti pa še bistveno manjše razlike.

Glavni dejavnik razlik je pedagoška prizadevnost oziroma specifične pomanjkljivosti izvedbe. Poleg tega seveda lahko (praviloma v razmeroma omejenem obsegu nekaj desetink, razen če ne gre za zelo nenavadne in izredne enkratne spremembe oziroma dogodke) vplivajo tudi zunanje okoliščine:

- Posebnost generacije (npr. bolj ali manj kritični, omejitev vpisa, prepisi, tujci, število študentov).
- Specifične zunanje okoliščine na fakulteti (npr. novi prostori) ali izven fakultete (npr. študentske demonstracije, nemiri)
- V primeru, da se predmet oziroma njegova izvedba spremenila (npr. urnik, vsebina ali celo nov program), pa je to seveda lahko potencialni razlog.

Primerjava programov

- Ko imamo opravka z agregati SET ocen, je slučajno variiranje še bistveno manjše in enako velja tudi za razpon. Oboje se zmanjšuje če primerjamo pedagoge (vsak ima namreč v povprečju 2.5 predmeta), programe (imajo lahko tudi več deset predmetov) in fakultete, kjer je pogosto je več sto predmetov, zato so statistično značilne praviloma že desetinke (npr. 4.0→4.1)
- Primerjava programov je posebej pomembna, saj je razpon njihovih povprečij na določeni fakulteti pri npr. skupnem povprečju 4.2, pogosto zgolj v intervalu 3.9-4.4. Razlika za eno odstotno točko zato pomeni kar petino celotnega razpona.
- Če primerjamo **isti program med dvema letoma**, ima prizadevnost pedagogov posebej velik vpliv, saj so študenti podobni (v večletnih programih pa v veliki meri celo isti, npr. triletni program 67%, dvoletni 50%), pedagogi pa skoraj isti. Vse to zožuje moč slučajnih vpliv in vpliv zunanjih dejavnikov zgolj na zelo posebne vplive (npr. novi prostori, študenti na izmenjavi, politične razmere).
- Če primerjamo **različne programe med sabo v istem letu**, pa imajo poleg pedagoške prizadevnosti lahko določen vpliv predvsem spremenjeni profil študentov, narava programa oziroma predmetov ter izvedbene posebnosti.

Napaka zaradi neodgovorov

- Vse zgoraj navedene interpretacije veljajo ob predpostavki, da študenti niso izraziteje zavračali sodelovanja.
- Izkušnje na FDV sicer kažejo, da padec sodelovanja od 65% na 45% ni spremenil nivoja ocen. Ko je bilo v SET vključeno še vprašanje o pogostosti obiskovanj predavanj, pa so analiza kazale, da tudi pogostost obiskovanj predavanj (ki je tesno povezano tudi z zavračanjem sodelovanja), ne vpliva na nivo SET.
- Dejstvo tudi je, da ni nekega posebnega teoretskega razloga (kot npr. obstaja v primeru, ko alkoholik težje odgovarja o rabi alkohola), da bi na tej osnovi nerespondenti imeli drugačen nivo SET ocen.
- Obstajajo seveda študije, ki dokazujejo, da so ne-sodelujoči študenti v pogledu SET ocen drugačni (npr. „slabi“ študenti) z domnevno nižjimi SET ocenami.
- Upoštevati tudi velja, da so ocene pristranske za produkt deleža manjkajočih ter razlike med respondenti in nerespondenti. Celo če bi respondenti dali oceno 4.0, nerespondenti pa bi dali 3.0 (kar je absurdno in nerealno velika razlika), imamo pri npr. deležu 30% nerespondentov le napako 0.3 (4.0 namesto 3.7).

4. Pregled izpisov

Pregled SET izpisov na članicah UL

- a) Anketa predmetov in pedagogov PRED in PO izpitu:
 - 1. *osebni izpis za pedagoga;
 - 2. *izpis trendov za pedagoga in njegove predmete;
 - 3. izpis za predmet;
 - 4. izpisi za skrbnika programa;
 - 5. hierarhični izpis za predstojnika katedre, oddelka, področja ipd.;
 - 6. izpis za vodstvo (in managerski sistem);
 - 7. izpis za koordinatorja prakse;
 - 8. izpis za kadrovske službo in ŠS (habilitacija);
 - 9. izpis za javnost in študente #.
- b) Splošna anketa
- c) Doktorska anketa za 1. in 2. letnik

** izpis se avtomatsko prenese v profil pedagoga v spletnem referatu.*

v pripravi

5. Povratna zanka

Rezultati SET za tipično članico UL

Rezultati za SET „zadovoljstvo s predmeti“ na tipični članici UL:

- Povprečje zadovoljstva s komponentami predmetov je 4.2.
- Porazdelitev ocen je naslednja:
 - Predmetov z oceno 4.5 in več je okoli tretjina,
 - Predmetov z oceno med 4.0 in pod 4.5 je okoli polovica;
 - Predmetov z oceno **pod 4.0 do vključno s 3.5 je okoli šestina**
 - Predmetov z oceno **pod 3.5 je nekaj odstotkov.**
- Oceno 4.8 in več dosega 5%, **pod 3.0 pa je tudi nekaj predmetov;**
- Podobno se porazdeljujejo tudi ocene za pedagoge.
- Smo z deleži **pod 4.0, pod 3.5** in pod **3.0** zadovoljni?

Izzivi evalvacije pedagogov

- Kako pomembna je splošna kvaliteta procesov in evalvacija?
- Kako pomembna je kvaliteta pedagoških procesov?
- Kako se v to vključuje evalvacija pedagogov?
- Se bodo upoštevala priporočila ESG (ENQA)?
- Ali in kako se bodo procesi evalvacije formalizirali?
- Koliko resursov, če sploh, smo pripravljeni dodatno nameniti:
 - za izboljšano evalvacijo kvalitete procesov,
 - za evalvacijo pedagoških procesov in
 - za dodatno procesiranje ter promocijo SET
- Se bo poleg SET sistemsko (obvezno) vzpostavilo še kak drug pristop za evalvacijo pedagogov?

SET vprašanja za članice UL

1. Ali se bo na nivoju članice (ali UL) formaliziralo interpretacijo in procesiranje SET rezultatov v **formativne** namene - ali pa se bo to prepuščalo katedram?
2. Kateri konkretni SET mejniki (npr. 4.5, 4.0, 3.5, 3.0) bodo vstopali v **sumativne** namene (npr. napredovanja, nagrajevanja, zaposlovanja), če sploh?
3. Ali se bo za **sumativne** namene evalvacije pedagogov sistemsko (formalno in obvezno) vzpostavilo, poleg SET, še kak drugi način evalviranja pedagogov?
4. Se bo postavilo petletne SET cilje za povprečja ali deleže pod 3.0 , 3.5, 4.0?
5. Se bo upoštevalo navodilo iz Pravilnika UL, da mora dekan poročati rektorju o pogovoru z 10% najnižje ocenjenimi pedagogi?
6. Se bo vzpostavil sistematičen mehanizem obravnave (npr. izobraževanja) povsod tam, kjer so SET ocene kritično nizke (npr. pod 3.0 ali pod 3.5)?
7. Bodo agregirane SET ocene programov objavljene javno, interno ali sploh ne?
8. Bodo rezultati pedagoga javni (kot npr. v ZDA ali v MB)? Z njegovim soglasjem?
9. Ali in kako bo formaliziran in reguliran dostop do podatkov ter arhiviranje?
10. Ali se bo dodatno vlagalo v večanje zaupanja študentov glede tajnosti SET odgovorov ter spodbujanje sodelovanja (tako s strani vodstva kot pedagogov)?

Nekaj ključnih virov

- Aleamoni, L. M. 1999. Student Rating Myths Versus Research Facts 1924-1998.
- Berk, R. A. 2005. Survey of 12 Strategies of Measure Teaching Effectivness.
- Pounder, J.S. 2007. Is Student Evaluation of Teaching Worthwhile?
- Spooren, P., Brockx, B., Mortelmans, D. 2013. On the Validity of Student Evaluation of Teaching: The state of Art.
- Stark, B., Freishtat, R. 2014. An Evaluation of Course Evaluations.

Komentarji, vprašanja, pripombe

<http://1ka.si/set>