

PRIMERJAVA BIPOLARNIH IN UNIPOLARNIH OCENJEVALNIH LESTVIC V ŠTUDENTSKIH EVALVACIJAH POUČEVANJA

NIKA DOBROVOLJC, JANJA JEREBIC,
ALENKA BREZAVŠČEK, ŠPELA KAJZER

Univerza v Mariboru, Fakulteta za organizacijske vede, Kranj, Slovenija
nika.dobrovoljc@um.si, janja.jerebic@um.si, alenka.brezavscek@um.si,
spela.kajzer1@um.si

Prispevek empirično proučuje vpliv unipolarnih in bipolarnih ocenjevalnih lestvic na rezultate študentskih evalvacij poučevanja. V okviru pilotne raziskave sta dve primerljivi skupini študentov z uporabo različnih ocenjevalnih lestvic ocenjevali enake trditve o pedagoškem procesu. Analize porazdelitev ocen, povprečnih vrednosti in deležev odgovorov v najvišjih kategorijah kažejo, da uporaba bipolarne lestvice pri večini vprašanj vodi do statistično značilno večjega zgoščanja odgovorov v zgornjem delu lestvice ter izrazitejšega učinka stropa. Ugotovitve potrjujejo teoretična predvidevanja o vplivu polarnosti lestvic na odzivno vedenje anketirancev ter kažejo, da izbira ocenjevalne lestvice vpliva na variabilnost in interpretabilnost rezultatov evalvacij. Prispevek osvetljuje metodološke posledice izbire ocenjevalne lestvice in odpira razpravo o njeni ustreznosti v kontekstu visokošolskih evalvacijskih praks.

DOI
[https://doi.org/
10.18690/um.fov.3.2026.14](https://doi.org/10.18690/um.fov.3.2026.14)

ISBN
978-961-299-124-1

Ključne besede:
ocenjevalne lestvice,
bipolarna lestvica,
unipolarna lestvica,
variabilnost ocen,
porazdelitev ocen,
pristranskost,
učinek stropa

DOI
[https://doi.org/
10.18690/um.fov.3.2026.14](https://doi.org/10.18690/um.fov.3.2026.14)

ISBN
978-961-299-124-1

Keywords:
rating scales,
bipolar scale,
unipolar scales,
score variability,
score distribution,
response bias,
ceiling effect

A COMPARISON OF BIPOLAR AND UNIPOLAR RATING SCALES IN STUDENT EVALUATIONS OF TEACHING

NIKA DOBROVOLJC, JANJA JEREBIC,

ALENKA BREZAVŠČEK, ŠPELA KAJZER

University of Maribor, Faculty of Organizational Sciences, Kranj, Slovenia
nika.dobrovoljc@um.si, janja.jerebic@um.si, alenka.brezavscek@um.si,
spela.kajzer1@um.si

This paper empirically examines the effects of unipolar and bipolar rating scales on the outcomes of student evaluations of teaching. In a pilot study, two comparable groups of students evaluated identical items using different rating scales. Analyses of score distributions, mean values, and top-category response shares indicate that the use of a bipolar scale leads to a statistically significantly higher concentration of responses at the upper end of the scale for most items, reflecting a more pronounced ceiling effect. The findings support theoretical expectations regarding the influence of scale polarity on response behaviour and demonstrate that rating format selection affects score variability and result interpretability. The paper discusses the methodological implications of rating scale choice and contributes to the ongoing debate on appropriate evaluation practices in higher education.



University of Maribor Press

1 Uvod

Študentske evalvacije poučevanja (angl. *Student Evaluations of Teaching* – SET) predstavljajo osrednji element zagotavljanja kakovosti pedagoškega dela v visokem šolstvu in enega ključnih virov povratnih informacij o učinkovitosti poučevanja. Kljub njihovi razširjeni uporabi se pogosto domneva, da je izbira numerične ocenjevalne lestvice zgolj tehnična odločitev brez bistvenih posledic za rezultate. Vendar psihometrijske raziskave in študije pristranskosti v odzivanju kažejo, da se anketiranci na bipolarne (npr. od -2 do $+2$) in unipolarne (npr. od 1 do 5) lestvice odzivajo različno. Predhodne analize poročajo o sistematičnih vzorcih odzivanja, zlasti o izogibanju negativnim kategorijam pri bipolarnih lestvicah, kar vodi v zgoščanje odgovorov v zgornjem delu skale, zmanjšano variabilnost ocen ter pojav izrazitega učinka zgornje meje oziroma t. i. učinka stropa (angl. *ceiling effect*). Slednje se pokaže še posebej izrazito pri ocenjevanju oseb (Höhne et al., 2020; Höhne et al., 2021; Menold, 2023; Steinberg in Rogers, 2020; Nemani et al., 2025).

Namen prispevka je empirično proučiti, ali in v kolikšni meri tip ocenjevalne lestvice vpliva na ocene pedagoškega procesa in pedagoškega osebja ter osvetliti metodološke in praktične posledice izbire ocenjevalnega formata za visokošolske institucije.

2 Teoretična izhodišča in pregled sorodnih raziskav

Iz vidika polarnosti delimo ocenjevalne lestvice na *unipolarne* in *bipolarne*. Pri unipolarnih lestvicah (npr. od 1 do 5) vse vrednosti izražajo različno stopnjo iste lastnosti, pri čemer nevtralna referenčna točka ni nujno eksplicitno opredeljena, medtem ko bipolarne lestvice (npr. od -2 do $+2$) vključujejo jasno določeno sredinsko nevtralno vrednost (0) ter omogočajo razlikovanje med negativnim in pozitivnim polom ocenjevanja.

Čeprav se ocenjevalne lestvice v praksi pogosto obravnavajo kot nevtralen tehnični element pri izvedbi mnenjske raziskave, pregled literature pokaže, da značilnosti lestvic neposredno vplivajo na kakovost in verodostojnost zajetih podatkov ter veljavnost merjenja (DeCastellarnau, 2018). Ključno teoretično izhodišče predstavlja ugotovitev, da anketiranci numeričnih vrednosti ne uporabljajo le za beleženje odgovora, temveč kot vir informacij za interpretacijo pomena vprašanja (Schwarz,

1999; Schwarz et al., 1991). Ilustrativen primer tega mehanizma predstavlja študija (Schwarz et al., 1991), v kateri so avtorji primerjali vpliv lestvice na samooceno uspešnosti v življenju. Rezultati so pokazali izrazit učinek numeričnih vrednosti: medtem ko je na bipolarni lestvici (–5 do +5) le 13 % anketirancev izbralo vrednosti v spodnji polovici razpona, je na formalno ekvivalentni unipolarni lestvici (0 do 10) vrednosti v spodnji polovici izbralo kar 34 % vprašanih. Analize so potrdile, da razlika izhaja iz spremenjene semantične interpretacije pola »sploh ne uspešen«. V kombinaciji s številko 0 so anketiranci ta opis razumeli kot »odsotnost izjemnih dosežkov«, medtem ko so ga v kombinaciji s številko –5 interpretirali kot »prisotnost izrazitih neuspehov«.

Ta pojav Cabooter et al. (2016) podrobneje pojasnjujejo z učinkom simetrije (angl. *symmetry effect*). Sklicevaje se na Gannon in Ostroma (1996) utemeljujejo, da format lestvice sproži proces aktivacije kategorij. Pri unipolarnih lestvicah anketiranci aktivirajo le eno kategorijo, kar pušča interpretacijo spodnjega dela lestvice dvoumno. V takšnih primerih se anketiranci pogosto oprejo na t. i. hevrstiko »sredina pomeni tipično« (angl. *Middle Means Typical Heuristic*; Tourangeau et al., 2000) in se zatekajo k srednjim vrednostim, ki jih dojemajo kot varno, normativno izbiro. Nasprotno pa bipolarne lestvice (z dvema eksplicitnima poloma) aktivirajo obe nasprotni kategoriji (npr. »slabo« in »dobro«). Lestvica se percipira kot simetrična, kjer leva stran jasno označuje negativno, desna pa pozitivno valenco. Ta jasna simetrična delitev sproži dva komplementarna psihološka mehanizma, ki skupaj vodita do izrazite pozitivne pristranskosti:

1. Odpor do izgube (angl. *Loss Aversion*): V skladu s teorijo vedenjske ekonomije (Tversky in Kahneman, 1991) anketiranci dojemajo negativne numerične vrednosti na levi strani simetrične lestvice (npr. –2) kot obliko kaznovanja ali izgube. Ker imajo izgube na posameznika večji psihološki učinek kot enakovredni dobički, se anketiranci podzavestno izogibajo izbiri odgovorov v negativnem spektru, da ne bi povzročili »škode« ocenjevani osebi.
2. Težnja k strinjanju (angl. *Agreeableness*): Hkratna jasna ločitev »dobrega« od »slabega« aktivira socialno normo vljudnosti. Ker imajo ljudje naravno težnjo k strinjanju in vzdrževanju pozitivnih odnosov, simetrična valenca odgovorov še dodatno okrepi pritisk k izbiranju pozitivnih ocen (Cabooter et al., 2016).

Sinergija teh dveh učinkov pojasnjuje, zakaj empirične raziskave pri bipolarnih lestvicah dosledno zaznavajo višjo stopnjo pozitivne asimetrije v primerjavi z unipolarnimi lestvicami, kjer je spodnji pol semantično bolj dvoumen (Cabooter et al., 2016).

Empirično relevantnost teh mehanizmov v specifičnem okolju študentskih evalvacij potrjujejo rezultati Sedlmeierja (2006), ki je v svoji študiji ugotovil, da so ocene pedagogov podvržene močnemu učinku stropa, kjer se velika večina odgovorov zgošča na pozitivnem delu lestvice. Rezultati te študije kažejo, da manipulacija formata ocenjevalne lestvice statistično značilno vpliva na premik povprečne ocene (angl. *mean shift*). Izkaže se, da uporaba ocenjevalnih lestvic, ki spodbujajo pozitivno pristranskost (kot so bipolarne lestvice s simetričnim pozitivnim polom), umetno zviša povprečne ocene vseh ocenjevanih pedagogov. Avtor ob tem ugotavlja, da takšen premik ne poveča nujno variance med dobrimi in slabimi pedagogi, temveč le zamakne celotno porazdelitev navzgor, kar institucijam otežuje identifikacijo dejanske kakovosti poučevanja.

Problem merske neinvariantnosti pa ni omejen le na izobraževanje. Sorodne ugotovitve so bile potrjene tudi na drugih področjih, od merjenja dohodkovne neenakosti (Höhne et al., 2021) do dinamičnih analiz psiholoških stanj in obnašanja (Nemani et al., 2023), kar potrjuje robustnost zaznanih metodoloških učinkov.

3 Empirična raziskava

3.1 Zasnova raziskave

Učinek ocenjevalne lestvice smo želeli empirično preveriti. Osnovna ideja je bila ta, da se dvema primerljivima skupinama anketirancev ponudi enak anketni vprašalnik, pri čemer ena skupina na vprašanja odgovarja na unipolarni, druga skupina pa na analogni bipolarni lestvici. Na podlagi izsledkov v literaturi in izkušenj smo postavili naslednje raziskovalne hipoteze:

H1: Pri več kot polovici vprašanj iz anketnega vprašalnika bo porazdelitev ocen med obema skupinama statistično značilno različna.

H2: Pri več kot polovici vprašanj iz anketnega vprašalnika bo povprečna ocena v skupini, ki uporablja bipolarno lestvico, statistično značilno višja kot v skupini, ki uporablja unipolarno lestvico.

H3: Pri več kot polovici vprašanj iz anketnega vprašalnika bo razlika v povprečnih ocenah med obema skupinama srednja ali visoka (Cohen's $d \geq 0,5$).

H4: Pri več kot polovici vprašanj iz anketnega vprašalnika bo delež odgovorov v najvišji kategoriji statistično značilno višji pri uporabi bipolarne ocenjevalne lestvice kot pri uporabi unipolarne lestvice. Učinek stropa bo torej bolj izrazit v skupini, ki uporablja bipolarno ocenjevalno lestvico.

3.2 Izvedba raziskave

Pilotna študija je bila izvedena med študenti Fakultete za organizacijske vede Univerze v Mariboru (FOV UM). Kot ciljno populacijo smo izbrali študente, ki so v študijskem letu 2025/26 vpisani v 1. letnik rednega študija na enega od VS študijskih programov 1. stopnje (OMIS, OMKIS ali IPS) in imajo v predmetniku predmet Osnove matematike v organizacijskih vedah. Glede na podatke v akademskem informacijskem podsistemu (AIPS) obsega izbrana populacija 208 študentov. V raziskavi, ki je bila izvedena v sklopu zaključnega kolokvija pri izbranem predmetu v decembru 2025, je sodelovalo vseh 80 študentov, ki so se kolokvija udeležili.

Vprašanja za anketni vprašalnik smo povzeli iz ankete o pedagoškem delu in obremenitvi študentov, ki jo UM vsako študijsko leto izvaja na vseh članicah z namenom pridobitve povratne informacije o kakovosti izvedbe študijskega procesa. V sklopu te ankete so študenti pozvani, da na lestvici od -2 do 2 izrazijo strinjanje z vnaprej podanimi trditvami. Trditve, ki smo jih vključili v našo raziskavo, so naslednje:

T1: Visokošolski učitelj/sodelavec predava zanimivo, razumljivo, vključuje sodobne metode poučevanja, uporablja ustrezno literaturo, spodbuja razmišljanje in intelektualno radovednost.

T2: Visokošolski učitelj/sodelavec spodbuja in upošteva pobude študentov, svetuje in usmerja pri opravljanju študijskih obveznosti.

T3: Visokošolski učitelj/sodelavec je ob dogovorjenem času (vaje, govorilne ure, e-pošta idr.) dostopen in odziven.

T4: Študent pridobi ustrezna pričakovana znanja, spretnosti in veščine, ki naj bi jih poznal, obvladal in razumel.

Sodelujoči v raziskavi so bili razdeljeni v 2 skupini, A in B. **Skupini A** je bila za ocenjevanje strinjanja s podanimi trditvami T1 – T4 predložena unipolarna lestvica (1 - zelo slabo, 2 - slabo, 3 - povprečno, 4 - dobro, 5 - zelo dobro). **Skupina B** pa je uporabljala analogno bipolarno lestvico (–2 - zelo slabo, –1 - slabo, 0 - povprečno, 1 - dobro, 2 - zelo dobro), ki je tudi uradna ocenjevalna lestvica v študentskih anketah za ocenjevanje kakovosti pedagoškega dela na UM. Raziskava je za obe skupini potekala sočasno, pri čemer sodelujoči niso vedeli, da se na anketnih vprašalnikih uporabljajo različne ocenjevalne lestvice, niti kateri skupini pripadajo.

Da bi lahko preverili in utemeljili primerljivost obeh skupin, smo anketni vprašalnik dopolnili še s kontrolnim vprašanjem:

V5: Kako v splošnem ocenjujete celotno izvedbo predmeta?

Pri tem vprašanju je bila v obeh skupinah uporabljena unipolarna lestvica (1 - zelo slabo, 2 - slabo, 3 - povprečno, 4 - dobro, 5 - zelo dobro).

4 Rezultati

4.1 Vzorčni rezultati

Struktura vzorca sodelujočih v raziskavi glede na pripadnost posamezni skupini je prikazana v tabeli 1.

Tabela 1: Struktura vzorca glede na pripadnost posamezni skupini

Skupina	Frekvenca	Odstotek
A – unipolarna	39	49 %
B – bipolarna	41	51 %

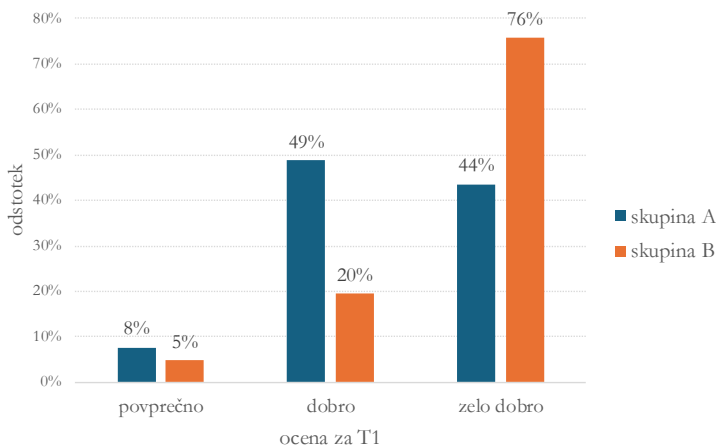
V nadaljevanju podajamo odgovore obeh skupin anketiranih (A – unipolarna in B – bipolarna) v smislu strinjanja s posameznimi trditvami (T1 – T4).

Porazdelitev ocen strinjanja s trditvijo T1 je za obe skupini predstavljena v tabeli 2, grafično pa je prikazana na sliki 1.

Tabela 2: Porazdelitev ocen za strinjanje s trditvijo T1

Skupina A			Skupina B		
Ocena	Frekvenca	Odstotek	Ocena	Frekvenca	Odstotek
1 - zelo slabo	0	0 %	-2 - zelo slabo	0	0 %
2 - slabo	0	0 %	-1 - slabo	0	0 %
3 - povprečno	3	8 %	0 - povprečno	2	5 %
4 - dobro	19	49 %	1 - dobro	8	20 %
5 - zelo dobro	17	44 %	2 - zelo dobro	31	76 %
Skupaj	39	100 %	Skupaj	41	100 %

Iz tabele 2 je razvidno, da se v obeh skupinah čez 90% odgovorov nahaja v zgornjih dveh kategorijah (4 in 5 oz. 1 in 2), medtem ko najnižji dve kategoriji (1 in 2 oz. -2 in -1) sploh nista zastopani. Slika 2 nadalje pokaže, da je delež ocen v najvišji kategoriji pri skupini B (bipolarna) občutno višji kot pri skupini A (unipolarna).

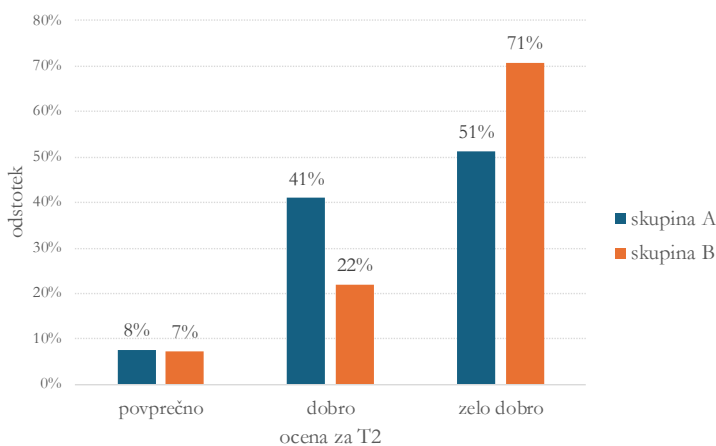


Slika 1: Grafični prikaz porazdelitve ocen za strinjanje s trditvijo T1 za obe skupini

Porazdelitev ocen strinjanja s trditvijo T2 je za obe skupini podana v tabeli 3, grafično pa je prikazana na sliki 2.

Tabela 3: Porazdelitev ocen za strinjanje s trditvijo T2

Skupina A			Skupina B		
Ocena	Frekvenca	Odstotek	Ocena	Frekvenca	Odstotek
1 - zelo slabo	0	0 %	-2 - zelo slabo	0	0 %
2 - slabo	0	0 %	-1 - slabo	0	0 %
3 - povprečno	3	8 %	0 - povprečno	3	7 %
4 - dobro	16	41 %	1 - dobro	9	22 %
5 - zelo dobro	20	51 %	2 - zelo dobro	29	71 %
Skupaj	39	100 %	Skupaj	41	100 %



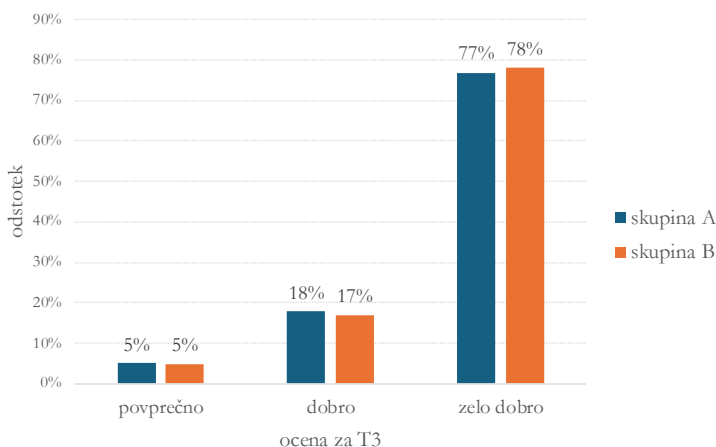
Slika 2: Grafični prikaz porazdelitve ocen za strinjanje s trditvijo T2 za obe skupini

Podobno kot pri rezultatih strinjanja s trditvijo T1 se tudi tu pokaže, da se v obeh skupinah čez 90% odgovorov nahaja v zgornjih dveh kategorijah, delež ocen v najvišji kategoriji pa je tudi tokrat pri skupini B višji kot pri skupini A.

Tabela 4: Porazdelitev ocen za strinjanje s trditvijo T3

Skupina A			Skupina B		
Ocena	Frekvenca	Odstotek	Ocena	Frekvenca	Odstotek
1 - zelo slabo	0	0 %	-2 - zelo slabo	0	0 %
2 - slabo	0	0 %	-1 - slabo	0	0 %
3 - povprečno	2	5 %	0 - povprečno	2	5 %
4 - dobro	7	18 %	1 - dobro	7	17 %
5 - zelo dobro	30	77 %	2 - zelo dobro	32	78 %
Skupaj	39	100 %	Skupaj	41	100 %

Porazdelitev ocen strinjanja s trditvijo T3 je razvidna iz tabele 4, njen grafični prikaz pa je na sliki 3.



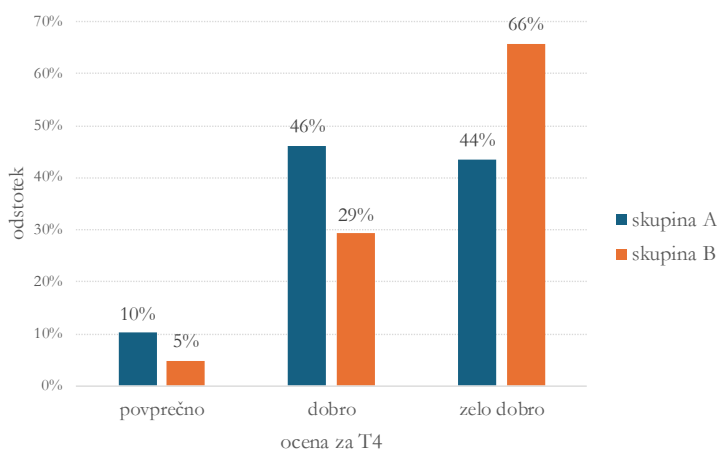
Slika 3: Grafični prikaz porazdelitve ocen za strinjanje s trditvijo T3 za obe skupini

Iz tabele 4 in slike 3 lahko razberemo, da so v primeru ocenjevanja strinjanja s trditvijo T3 rezultati v obeh skupinah zelo podobni in primerljivi.

Tabela 5 in slika 4 prikazujeta porazdelitev ocen strinjanja z zadnjo trditvijo, T4.

Tabela 5: Porazdelitev ocen za strinjanje s trditvijo T4

Skupina A			Skupina B		
Ocena	Frekvenca	Odstotek	Ocena	Frekvenca	Odstotek
1 - zelo slabo	0	0%	-2 - zelo slabo	0	0%
2 - slabo	0	0%	-1 - slabo	0	0%
3 - povprečno	4	10%	0 - povprečno	2	5%
4 - dobro	18	46%	1 - dobro	12	29%
5 - zelo dobro	17	44%	2 - zelo dobro	27	66%
Skupaj	39	100%	Skupaj	41	100%



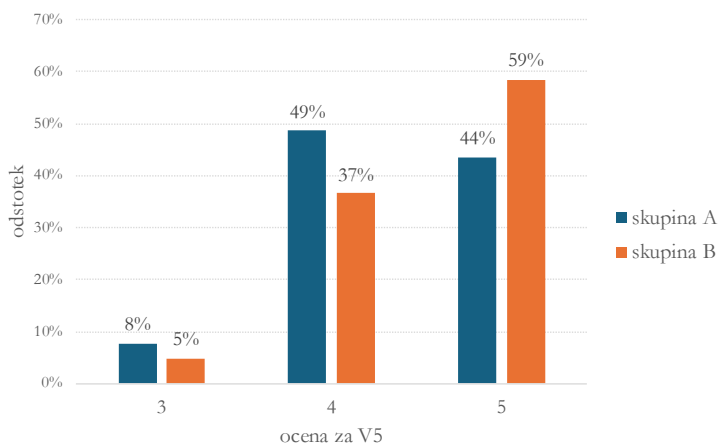
Slika 4: Grafični prikaz porazdelitve ocen za strinjanje s trditvijo T4 za obe skupini

Podobno kot pri prejšnjih trditvah, se tudi v tem primeru izkaže, da je v obeh skupinah čez 90 % v zgornjih dveh kategorijah. Delež ocen v najvišji kategoriji je tudi tokrat pri skupini B znatno višji kot pri skupini A.

Zadnje vprašanje (V5) je bilo kontrolno, zato je bila za obe skupini uporabljena enaka ocenjevalna lestvica. Porazdelitev odgovorov obeh skupin podaja tabela 6, grafično pa so prikazani na sliki 5. Vzorčni rezultati nakazujejo, da se odgovori na kontrolno vprašanje med obema skupinama nekoliko razlikujejo. Ali so te razlike statistično značilne, pa bomo preverili z ustreznimi statističnimi testi v naslednjem poglavju.

Tabela 6: Porazdelitev odgovorov na kontrolno vprašanje V5

Ocena	Skupina A		Skupina B	
	Frekvenca	Odstotek	Frekvenca	Odstotek
1 - zelo slabo	0	0 %	0	0 %
2 - slabo	0	0 %	0	0 %
3 - povprečno	3	8 %	2	5 %
4 - dobro	19	49 %	15	37 %
5 - zelo dobro	17	44 %	24	59 %
Skupaj	39	100 %	41	100 %



Slika 5: Grafični prikaz porazdelitve odgovorov na kontrolno vprašanje V5 za obe skupini

4.2 Analiza raziskovalnih hipotez

Vse statistične analize smo izvedli s programom IBM SPSS Statistics 31.

Za namen preverjanja primerljivosti skupin A in B, smo za kontrolno vprašanje V5 izvedli test razlike populacijskih povprečij pri 5 % stopnji značilnosti testa. Pri tem smo najprej interpretirali rezultat Levenovega testa, na podlagi katerega smo lahko predpostavili enakost varianc v obeh skupinah ($p = 0,772 > 0,05$). Skladno s to ugotovitvijo smo nadalje odčitali dvostransko p -vrednost t -testa za razliko populacijskih povprečij in dobili 0,198. Ker je dvostranski $p = 0,198 > 0,05$, statistično značilnih razlik v povprečni oceni odgovorov na kontrolno vprašanje V5 med skupinama ne moremo potrditi. S tem smo dokazali, da sta skupini primerljivi.

Da smo lahko za potrebe analize raziskovalnih hipotez H1 - H4 rezultate obeh skupin med seboj primerjali, smo morali najprej ocenjevalni lestvici uskladiti. To smo naredili tako, da smo vse odgovore skupine B (za vse 4 trditve T1 – T4), ki so bili podani na bipolarni lestvici, linearno pretvorili na unipolarno lestvico ($-2 \rightarrow 1$, $-1 \rightarrow 2$, ipd.).

Nadalje smo izračunali opisne statistike za vsa vprašanja anketnega vprašalnika (T1 – T4 in V5) za obe skupini. Rezultati za skupino A (unipolarna lestvica) so podani v tabeli 7, za skupino B (bipolarna lestvica po linearni pretvorbi) pa v tabeli 8. Razvidno je, da so vzorčna povprečja v vseh kategorijah vprašalnika za skupino B nekoliko višja. Največjo razliko med povprečnima ocenama vidimo pri trditvi T1, kjer je absolutna razlika med povprečnima ocenama 0,35, najmanjšo razliko pa pri trditvi T3, kjer je absolutna razlika enaka 0,01.

Tabela 7: Opisne statistike za vprašanja anketnega vprašalnika – skupina A (unipolarna lestvica)

	T1	T2	T3	T4	V5
N	39	39	39	39	39
Povprečje	4,36	4,44	4,72	4,33	4,36
Mediana	4	5	5	4	4
Modus	4	5	5	4	4
St. odklon	0,628	0,641	0,560	0,662	0,628
Min.	3	3	3	3	3
Maks.	5	5	5	5	5

Tabela 8: Opisne statistike za vprašanja anketnega vprašalnika – skupina B (bipolarna lestvica po linearni pretvorbi)

	T1	T2	T3	T4	V5
N	41	41	41	41	41
Povprečje	4,71	4,63	4,73	4,61	4,54
Mediana	5	5	5	5	5
Modus	5	5	5	5	5
St. odklon	0,559	0,623	0,549	0,586	0,596
Min.	3	3	3	3	3
Maks.	5	5	5	5	5

Za namen preverjanja razlike v porazdelitvi ocen na populaciji (H1) smo izvedli neparametrični Mann-Whitney test pri stopnji značilnosti 0,05. Rezultati testa so pokazali, da statistično značilne razlike v porazdelitvi ocen med obema lestvicama lahko potrdimo pri trditvah T1 ($p = 0,006 < 0,05$) in T4 ($p = 0,045 < 0,05$), medtem ko razlike v porazdelitvi ocen pri T2 ($p = 0,105 > 0,05$) in T3 ($p = 0,905 > 0,05$) niso statistično značilne. Hipotezo H1 lahko torej potrdimo le delno.

Za preverjanje hipoteze H2 smo izvedli enostranski test razlike populacijskih povprečij, s katerim smo testirali, če je populacijsko povprečje ocen pri uporabi bipolarne lestvice (skupina B, po ustrezni linearni pretvorbi) višje kot pri uporabi unipolarne lestvice (skupina A). Rezultati so pokazali, da lahko pri 5 % stopnji značilnosti testa statistično značilno višje povprečne ocene pri uporabi bipolarne lestvice potrdimo pri trditvah T1 ($p=0,005<0,05$) in T4 ($p=0,026<0,05$), medtem ko pri trditvah T2 ($p=0,082>0,05$) in T3 ($p=0,456>0,05$) tega ne moremo trditi. Če izvedemo test pri 10 % stopnji značilnosti, pa lahko statistično značilno višje povprečne ocene pri uporabi bipolarne lestvice potrdimo tudi pri trditvi T2. Rezultati testov torej nakazujejo, da lahko hipotezo H2 potrdimo.

V sklopu preverjanja hipoteze H3 smo za vse štiri trditve izračunali vrednosti Cohenovega koeficienta d. Velikost učinka, izražena s Cohenovim koeficientom d, je bila za vse štiri trditve srednja (med 0,55 in 0,63). Izračunane vrednosti kažejo, da med obema skupinama obstajajo zmerne, a praktično pomembne razlike v povprečnih ocenah trditvev. Ugotavljamo torej, da izbira lestvice vpliva na izide merjenja. Hipotezo H3 potrdimo.

Za potrebe analize hipoteze H4 smo pri posameznih trditvah primerjali delež odgovorov v najvišji kategoriji (delež vrednosti »5« za skupino A in delež vrednosti »2« za skupino B). Vzorčni deleži odgovorov v najvišji kategoriji za posamezne trditve za obe skupini so prikazani v tabeli 9.

Tabela 9: Vzorčni deleži odgovorov v najvišji kategoriji

	Skupina A (delež vrednosti »5«)	Skupina B (delež vrednosti »2«)
T1	0,44	0,76
T2	0,51	0,71
T3	0,77	0,78
T4	0,44	0,66

Razvidno je, da so vzorčni deleži pri vseh trditvah višji pri skupini B. Z izjemo trditve T3 so razlike občutne. Pri stopnji značilnosti testa 5 % smo izvedli test deleža, ki je potrdil statistično značilne razlike pri trditvah T1 ($p=0,002<0,05$), T2 ($p=0,038<0,05$) in T4 ($p=0,023<0,05$), pri trditvi T3 pa rezultat testa ni bil statistično značilen ($p=0,453>0,05$). Efekt stropa je torej izražen pri 3 od 4 testiranih trditvev, kar pomeni, da hipotezo H4 lahko potrdimo.

5 Zaključek

Pričujoča raziskava potrjuje teoretična predvidevanja iz psihometrične in metodološke literature, da tip ocenjevalne lestvice pomembno vpliva na odzivno vedenje anketirancev v študentskih evalvacijah poučevanja. Empirični rezultati kažejo, da uporaba bipolarne ocenjevalne lestvice sistematično povečuje koncentracijo odgovorov v najvišjih kategorijah in krepi učinek stropa, s čimer vpliva na porazdelitev ocen in zmanjšuje njihovo diskriminativnost. Rezultati analize nadalje kažejo, da je vpliv izbire ocenjevalne lestvice izrazitejši pri postavkah, ki temeljijo na subjektivni in vrednostni presoji pedagoškega dela (T1, T2), medtem ko je pri trditvi T3, ki se nanaša na relativno objektivno in vedenjsko opazno značilnost (dostopnost in odzivnost), ta vpliv bistveno manj izrazit. To nakazuje, da učinek polarnosti ocenjevalne lestvice ni univerzalen, temveč je pogojen tudi z naravo merjenega konstrukta, pri čemer so subjektivne evalvacije bolj dovzetne za odzivne pristranskosti, povezane s formatom ocenjevalne lestvice.

Z vidika visokošolskih institucij to pomeni, da izbira ocenjevalne lestvice ne sme biti zgolj tehnična odločitev, temveč metodološka izbira, ki neposredno vpliva na veljavnost, primerljivost in interpretabilnost evalvacijskih rezultatov. Ugotovitve zato odpirajo prostor za razmislek o morebitni prilagoditvi ali ponovni presoji uporabljene ocenjevalne lestvice, zlasti v kontekstih, kjer se rezultati evalvacij uporabljajo za spremljanje kakovosti poučevanja in podpora institucionalnemu odločanju.

Rezultate naše raziskave je seveda treba interpretirati ob upoštevanju njenih omejitev, saj temelji na pilotni študiji z relativno majhnim vzorcem, omejenim na en predmet in eno visokošolsko institucijo. Kljub temu ugotovitve ponujajo pomembne indikacije metodoloških učinkov izbire ocenjevalne lestvice, ki jih je smiselno preveriti v obsežnejših in longitudinalnih študijah.

Viri in literatura

- Cabooter, E., Weijters, B., Geuens, M., & Vermeir, I. (2016). Scale format effects on response option interpretation and use. *Journal of Business Research*, 69(7), 2564–2571. doi:10.1016/j.jbusres.2015.10.138
- DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: A literature review. *Quality & Quantity*, 52(4), 1523–1559. doi:10.1007/s11135-017-0533-4

- Gannon, K. M., & Ostrom, T. M. (1996). How meaning is given to rating scales: The effects of response language on category activation. *Social Cognition*, 14(3), 155–179. doi:10.1521/soco.1996.14.3.155
- Höhne, J., Krebs, D., & Kühnel, S. (2020). Measuring Income (In)equality: Comparing Survey Questions With Unipolar and Bipolar Scales in a Probability-Based Online Panel. *Social Science Computer Review*, 40, 108 - 123. <https://doi.org/10.1177/0894439320902461>.
- Höhne, J., Krebs, D., & Kühnel, S. (2021). Measurement properties of completely and end labeled unipolar and bipolar scales in Likert-type questions on income (in)equality.. *Social science research*, 97, 102544 . <https://doi.org/10.1016/j.ssresearch.2021.102544>.
- Menold, N. (2023). Verbalization of Rating Scales Taking Account of Their Polarity. *Field Methods*, 35, 378 - 391. <https://doi.org/10.1177/1525822x231151314>.
- Nemani, A., Hufschmidt, B., Kohl, V., Sendig, L., Ebert, M., Bonarius, D., Hofmann, S., & Stangier, U. (2025). Bipolar vs. unipolar scaling in dynamic network analyses of Ecological Momentary Assessment data. *PLOS One*, 20. <https://doi.org/10.1371/journal.pone.0314102>.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), 93-105. doi:10.1037/0003-066X.54.2.93
- Schwarz, N., Knäuper, B., Hippler, H. J., Noelle-Neumann, E., & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, 55(4), 570-582. doi:10.1086/269282
- Sedlmeier, P. (2006). The role of scales in student ratings. *Teaching of Psychology*, 33(4), 262-267. doi:10.1207/s15328023top3304_7
- Steinberg, L., & Rogers, A. (2020). Changing the Scale: The Effect of Modifying Response Scale Labels on the Measurement of Personality and Affect. *Multivariate Behavioral Research*, 57, 79 - 93. <https://doi.org/10.1080/00273171.2020.1807305>.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039-1061. doi:10.2307/2937956