

**Dr. Gašper Cankar**

## Standardiziranje ocen pri splošni maturi

**Povzetek:** Avtor izhaja iz članka Kališnika in Bucika (2001), ki sta predlagala nov model ocenjevanja pri splošni maturi. Predstavljene so raziskave in simulacije, ki so sledile pobudi iz članka in predstavljajo poskuse izboljšane modela ocenjevanja. Težave, pasti in zapleti posameznih različic novih načinov ocenjevanja so opisani in razlogi, zaradi katerih novi model ocenjevanja ni bil vpeljan v praksi, so povzeti v sklepu članka.

**Ključne besede:** splošna matura, ocenjevanje.

UDK: 371.26

Strokovni prispevek

*Dr. Gašper Cankar, Državni izpitni center, Ljubljana; e-naslov: gasper.cankar@guest.arnes.si*

## Uvod

Vsake toliko časa se v krogih strokovnjakov, ki se ukvarjajo z maturo, pojavijo zamisli in težnje po standardiziranju ocen pri maturi. Pobude o standardiziranju ocenjevanja smo lahko spremljali predvsem pri splošni maturi, kjer so bile tudi intenzivno proučevane ter lahko na podlagi simulacij in preverjanj najdemo veliko poučnih primerov o pozitivnih in negativnih učinkih standardiziranja. Članek je izid serije izračunov in simulacij v okviru poskusa vpeljave novega modela ocenjevanja, ki je temeljil na članku Kališnika in Bucika (2001). Zaradi večje berljivosti so predstavljene le ugotovitve opravljenih raziskav, in ne obširne analize, na katerih temeljijo. Čeprav izhaja opravljeno delo iz nekaj let starega članka, pa so predstavljeni problemi še kako aktualni, saj se splošna matura v zadnjem času ni bistveno spreminjala.

## Standardizacija ocen

Standardizacija se v širšem pomenu nanaša na poenotenje postopkov, procedur, načinov izvedbe. V ožjem pomenu predstavlja pretvorbo rezultatov ocenjevanja na enotno aritmetično sredino in standardni odklon, kar omogoča primerjavo med dosežki na različnih lestvicah (Bucik 1997; Nunnally in Bernstein 1994). Večina pobud za standardizacijo ocenjevanja omenja večjo pravičnost takih rezultatov za kandidate, prednost pa je tudi v neoporečnosti izračunavanja skupnih rezultatov kot vsote ali povprečja posameznih dosežkov. V omenjenem predlogu so omenjene vrednosti  $T$ , ki imajo aritmetično sredino 50 in standardni odklon 10 točk.

## Sedanji način ocenjevanja pri splošni maturi

Maturitetni izpiti pri splošni maturi so sestavljeni iz enega ali več delov,

katerih točkovanje je med sabo popolnoma neodvisno. Določena so razmerja med posameznimi deli izpita in kandidatov končni rezultat se izračuna tako, da se točke v posameznih delih izpita skladno z obtežitvijo istega dela preračunajo in seštejejo. Po preračunu ima vsak izpit tako lahko največ 100 točk, ki jim pravimo tudi odstotne točke. Po sedanji praksi predmetne komisije za posamezni predmet nato strokovno določijo meje med ocenami 1–5 oziroma 1–8 pri materinščini in predmetih na višji ravni zahtevnosti. Tako dobi kandidat ocene oziroma točkovne ocene, ki so skupaj s seštevkom točkovnih ocen zapisane v spričevalu in predstavljajo njegov rezultat pri splošni maturi.

### **Predlog Bucika in Kališnika (2001)**

Na tretjem strokovnem posvetu o maturi in v objavljenih člankih (Bucik, Kališnik 2001) se je pojavila pobuda o standardiziranju ocen pri splošni maturi. Kališnik in Bucik (2001) sta predstavila model, ki:

»/.../ naj bi odpravil nekatere dileme sedanje zasnove mature, denimo neprimerno upoštevanje enostranske nadarjenosti ob enakih pogojih za vse kandidate, neuravnoteženost predmetov glede zahtevnosti, samovoljno določeno razmerje med osnovno in višjo ravnijo, samovoljno določeno pogojno pozitivnost in strokovno sporno seštevjanje točk.«

Avtorja v članku predvidevata obvezno izbiro bodisi matematike bodisi angleščine na višji ravni zahtevnosti ter opravljanje vseh drugih predmetov na osnovni ravni zahtevnosti. Prav tako predvidevata normalizacijo in standardizacijo ocen s pretvorbo v lestvico T, ki naj bi omogočala tudi enostavno pretvorbo v lestvico petih šolskih ocen. Prag za pozitivno naj bi še vedno določila predmetna komisija, saj avtorja zagovarjata, da je pri zelo solidnem znanju kandidatov lahko delež negativnih kandidatov poljubno majhen. Za izračun skupnega uspeha pri maturi lahko vrednosti T za vseh pet predmetov povprečimo in po potrebi pretvorimo v lestvico šolskih ocen.

Po predlaganem modelu naj bi se odstotne točke kandidatov pri posameznem predmetu najprej normalizirale in nato standardizirale. Predmetna komisija bi še vedno opredeljevala prag pozitivnosti pri predmetu, vendar naj bi to postalo bolj izjema kot pravilo in meje med drugimi ocenami naj bi se določile samodejno po znanih statističnih kriterijih za vse predmete enako. Pri predmetih z zelo majhnim številom kandidatov bi lahko pretvorba v lestvico petih šolskih ocen potekala neposredno. Podrobnejše opise predlaganih postopkov lahko bralec prebere v izvirnem članku.

Avtorja v članku navajata številne prednosti novih postopkov, vendar ob poskusu implementacije nov model ni zaživel v praksi. Zakaj ne?

## Prostovoljnost izbire višje ravni zahtevnosti

Ena od novosti, ki sta jo predlagala, pomeni zelo veliko spremembo v izvedbi splošne mature za kandidate. V sedanji izvedbi splošne mature lahko kandidat poleg materinščine, ki obstaja le na enotni (višji) ravni zahtevnosti, izbere še nič, enega ali največ dva predmeta na višji ravni zahtevnosti. Če je kandidat pri vseh predmetih kar najbolj uspešen, lahko sega njegov rezultat pri maturi glede na izbrano kombinacijo od 28 do 34. Avtorja ugotavljata, da imajo zaradi vpliva ravni prednost pogumnejši kandidati, ki se odločijo za višjo raven in lahko tako iztržijo več točk, to pa ni nujno posledica le večjega znanja. Prednost naj bi imeli kandidati, nadarjeni za tuje jezike, saj imajo le tuji jeziki med izbirnimi predmeti pri maturi tudi izpite na višji ravni zahtevnosti, ki kandidatu omogočajo, da dobi večje število točk. Zaradi omenjenih težav predlagata kompromisni model, v katerem kandidat izbere bodisi matematiko bodisi angleščino na višjem nivoju. Tako bi se za vse kandidate izenačilo največje število mogočih točk pri maturi.

Njun predlog ni zajemal najpreprostejše rešitve, namreč odprave dveh nivojev zahtevnosti izpitov. Razen zagotovil komisije, ki sestavlja izpite, namreč ni nobenega metodološkega postopka, s katerim bi zagotavljali določeno razmerje med osnovno in višjo ravnjo, zato je razlika med osnovno in višjo ravnjo med predmeti lahko zelo različna. Seveda bi to pomenilo zelo radikalen poseg v splošno maturo, ki v svojih posledicah presega uvajanje novosti v model ocenjevanja.

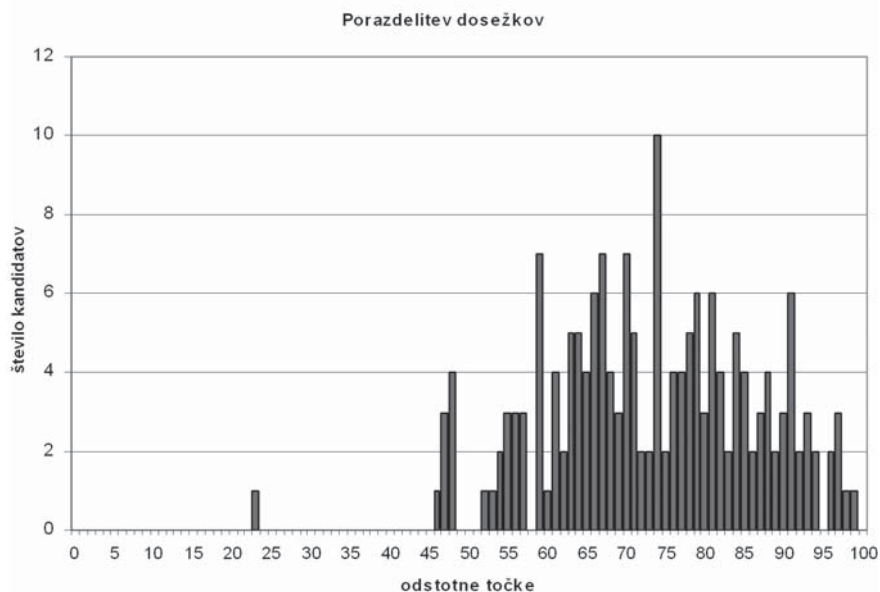
Čeprav nista zagovarjala tako radikalne rešitve, pa za obvezno izbiro zgolj enega predmeta na višji ravni zahtevnosti ni bilo posluha in v vseh simulacijah in poskusih vpeljave novega modela ocenjevanja pri maturi je bilo treba upoštevati prostovoljno izbiro dveh, enega ali nobenega predmeta na višji ravni zahtevnosti.

## Normalizacija: da ali ne?

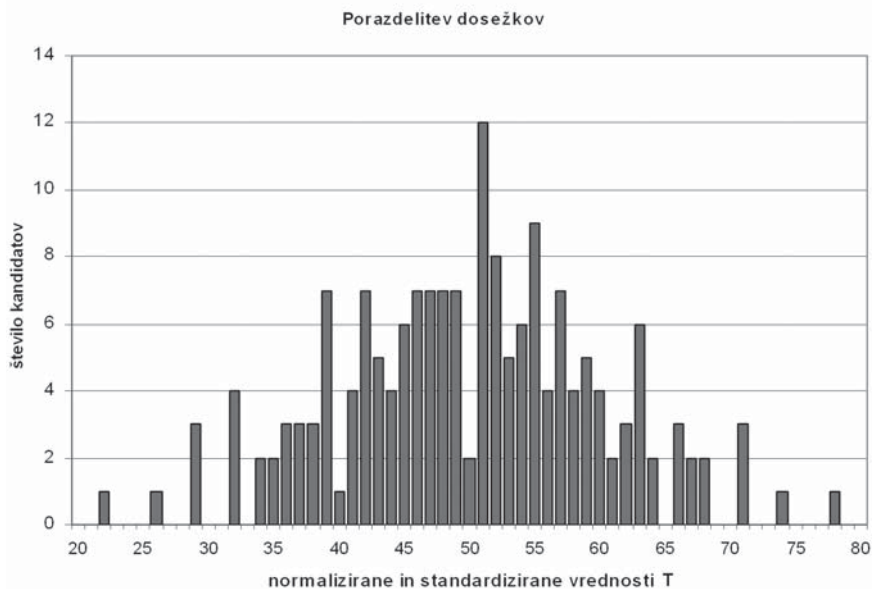
Prvi korak novega modela ocenjevanja naj bi bila normalizacija dobljenih rezultatov pri posameznem predmetu. V tem procesu se prek izenačevanja ploščin empirično dobljeni rezultati kandidatov pretvorijo v ustrezne vrednosti v normalni porazdelitvi. Rezultat je normalna porazdelitev rezultatov, ki je med predmeti tako rekoč enaka.

Poleg paradoksnega dejstva, da smo do normalizacije upravičeni le, kadar ta ne predstavlja bistvene spremembe porazdelitve – kadar je torej empirična porazdelitev zelo blizu normalne in normalizacija sploh ni potrebna, ima normalizacija dosežkov za rezultat neželjeno posledico rušenja razmerij med kandidati. Vzemimo za primer podatke na sliki 1, ki predstavlja dosežke 175 kandidatov pri enem od predmetov splošne mature (vir: RIC 2005). Najslabše se je odrezal kandidat, ki je dosegel 23 odstotnih točk, naslednji pa jih je dosegel že 46. Po normalizaciji (slika 2) razlika med njima skoraj izgine in je npr. enaka kakor

med najboljšim kandidatom in tistim takoj za njim, ki se v empirični porazdelitvi razlikujeta zgolj za eno odstotno točko (99 oziroma 98 odstotnih točk).



Slika 1: Empirična porazdelitev odstotnih točk 175 kandidatov pri enem od izpitov splošne mature (Vir: RIC 2005)



Slika 2: Porazdelitev normaliziranih in standardiziranih vrednosti istih dosežkov 175 kandidatov (Vir: RIC 2005)

Ker se pri normalizaciji izgubijo prava razmerja med znanjem kandidatov in ker pri večini predmetov oblike porazdelitev niti ne upravičujejo uporabe postopkov normalizacije, so normalizirani rezultati neprimerni za postavljanje mej, saj se razlike v znanju ne izražajo v pretvorjenih rezultatih. Enostavna in logična pretvorba v lestvico šolskih ocen iz omenjenih rezultatov tako niti ni mogoča niti pri večini predmetov ni priporočljiva zaradi relativno majhnega števila kandidatov. Od 47 različnih maturitetnih izpitov pri splošni maturi 2007 je bilo le pri 14 predmetih število kandidatov v vseh rokih in terminih skupaj večje od 500 (Letno poročilo za splošno maturo 2007), kar kaže na veliko predmetov z »majhnim« število kandidatov.

### **Spremembe novega modela**

Ohranjanje prostovoljne izbire dveh, enega ali nobenega predmeta na višji ravni zahtevnosti je pomenilo prvo spremembo v predlaganem modelu, ki je onemogočala enostavno uvedbo. Pri poskusih upoštevanja ravni so zato naslednji predlogi predvidevali, da se dobljene vrednosti T pri predmetih na višji ravni zahtevnosti pomnožijo s faktorjem 1,5 in tako ustrezno nagradijo. Faktor je bil določen na podlagi primerjave obeh ravni zahtevnosti pri enem samem predmetu v eni sami generaciji kandidatov in se ne razlikuje bistveno od sedanje rešitve, ki najuspešnejše kandidate nagradi v razmerju 1,6 (8 : 5). Ker je faktor enak za vse predmete v vseh generacijah, ne pomeni pomembnejšega premika od »samovoljno določenega razmerja med osnovno in višjo ravni« (Kališnik in Bucik 2001), saj lahko predvidevamo, da razmerje med obema ravnima ni enako pri vseh predmetih, niti ni konstantno skozi več let.

Določene pozitivne učinke bi lahko dosegli s sidranjem nalog in vsakoletnim določanjem razmerja med osnovno in višjo ravni zahtevnosti. Tako bi se izpita na različnih ravneh zahtevnosti med sabo izenačila, omogočena bi bila primerjava znanja kandidatov na različnih ravneh in primerljivost njihovih ocen. Tudi pri postavljanju mej med ocenami bi v tem primeru lahko predmetna komisija rezultate obravnavala enotno. Edini argument ohranjanju obeh ravni pa bi v tem primeru ostal krajši čas pisanja in posledično manjša obremenitev za obe skupini kandidatov.

### **Težave z neprimerljivimi skupinami pri predmetih**

Ne glede na to, ali so bili v simulacijah dosežki kandidatov normalizirani ali ne, so nastajale pri pretvorbah dosežkov nepravilnosti, ki so bile vezane na neprimerljivost populacij pri posameznih predmetih. Kandidati se za izbirne predmete odločajo na zelo različne načine. Pri nekaterih izbirnih predmetih pri splošni maturi dobimo tako skupine kandidatov, ki izkazujejo odlično znanje, pri drugih pa skupine, ki se po svojem znanju veliko bolj razlikujejo in v katerih so tudi kandidati z relativno slabimi dosežki. Transformacija v vrednosti

T s poprejšnjo normalizacijo ali brez nje izrazi kandidatov dosežek v odnosu do dosežka skupine. Če skupine med predmeti niso primerljive, dosežki kljub pretvorjenim vrednostim niso primerljivi in enotno postavljanje mej v lestvici T vrednosti pri višjih ocenah ni upravičeno. Zaradi velikih razlik med izbirnimi predmeti v številu kandidatov jih med sabo težko primerjamo.

## Vpliv na vpis kandidatov

Pri preizkušanju novih modelov ocenjevanja smo preverjali tudi vpliv morebitnih sprememb na vpis kandidatov. Čeprav je upoštevanje rezultatov mature za vpis na univerzo pristojnost slednjih, pa vpogled v problematiko ni odveč. V zadnjem predlogu modela naj bi predmetne komisije določile mejo za pozitivno in odlično oceno, ki se pretvorita v vrednosti 20 in 50, druge vrednosti pa linearno glede na omenjeni meji. Pri predmetih na višji ravni se te vrednosti še dodatno pomnoži s faktorjem 1,5. Zaradi praktičnih razlogov je lestvico smiselno omejiti navzgor in navzdol, saj sicer najvišje in najnižje vrednosti povprečnih rezultatov niso znane. Če lestvico pri predmetu na osnovni ravni omejimo na 10 oziroma 60 na spodnji in zgornji strani, imamo razpon možnih povprečnih vrednosti kandidatov, ki so pozitivno opravili maturo od 19 (upoštevajoč pogojno pozitivnost) do 78 (zaradi treh predmetov na višji ravni zahtevnosti). Glede na večji možni razpon lestvice veliko manj kandidatov doseže maksimalne ali minimalne vrednosti kot doslej. Porazdelitev je tako bolj zgoščena in standardni odklon glede na razpon med najvišjo in najnižjo vrednostjo manjši.

Najpogosteje se za vpis na študij upoštevata dosežek pri maturi in uspeh zadnjih dveh letnikov srednje šole. Če bi bila za vpis na izmišljeni študij oba kriterija enako pomembna, je bilo leta 2003 razmerje vplivov (preračunanih standardnih odklonov) obeh kriterijev 49 : 51 (matura : šolski uspeh), po novem pa bi bilo razmerje vplivov zaradi zmanjšanega standardnega odklona pri maturi kar 36 : 64 v prid šolskega uspeha. Ves trud za natančnejše in pravičnejše ocenjevanje bi ob nespremenjenih vpisnih pogojih vodil v zmanjšanje vpliva mature na kandidatov vpis.

Kakršne koli spremembe ocenjevalnega modela bi bilo zato vsekakor treba izvajati hkrati s spremembami v vpisnih pogojih ali pa vsaj ob skrbni kontroli le-teh.

## Sklep

Prav gotovo je eden najpomembnejših razlogov za neuspeh vpeljave novega modela odklon od prvotnih predlogov avtorjev. Vpeljava obvezne izbire enega (in samo enega) predmeta na višji ravni zahtevnosti ali pa celo morebitna odprava dveh ravni zahtevnosti bi zelo spodbudila in omogočila vpeljavo predlaganega modela. V procesih simulacije se je predvsem normalizacija izkazala kot neprimerna zaradi velikega števila predmetov z relativno majhnim številom kandidatov. V po-

skusu vpeljave novega modela se je pokazala velika raznolikost in kompleksnost predmetov pri maturi, ki se med sabo precej razlikujejo in jim je bilo zato težko prirediti rešitve, ki bi se zmeraj obnesle zadovoljivo. Vsako prihodnje razmišljanje o novih modelih ocenjevanja mora bodisi vzeti omenjeno stanje za izhodišče bodisi opredeliti spremembe, potrebne za uspešnost vpeljave novega modela ocenjevanja. Veliko predmetov, ki jih vsako leto opravlja le peščica kandidatov, je neprimernih za predlagane spremembe in s svojimi posebnostmi in izjemami onemogočajo uniformen pristop k predmetom in primerjavo med njimi.

Eden od razlogov za predstavitev prednosti in slabosti standardizacije ocen je prav gotovo uporaba pridobljenih izkušenj pri vseh drugih oblikah zunanje-ga preverjanja in ocenjevanja. Izkušnje, razmišljanja in veliko simulacij nam omogočajo nabor spoznanj, ki so lahko v pomoč pri vzpostavitvi optimalnega načina ocenjevanja, ki je kar najpravičnejše za kandidate (in obenem še izvedljivo). Drugi razlog je diseminacija spoznanj in vpliv povratnih informacij. Pri reševanju praktičnih problemov smo povsem lahko spregledali kakšen enostaven postopek, s katerim bi lahko upravičili trud standardizacije ocen in odpravili negativne posledice, ki jih morebiti prinese. S širjenjem spoznanj se širi možnost povratnih informacij, ki lahko prinesejo v ocenjevanje pri maturi pozitivne premike in omogočijo cilj, ki je bil obenem prvotni namen predlagateljev standardizacije ocen – pravičnejše ocenjevanje kandidatov pri maturi.

## Literatura

- Bucik, V. (1997). Osnove psihološkega testiranja. Ljubljana: Oddelek za psihologijo Filozofske fakultete Univerze v Ljubljani.
- Kališnik, M. in Bucik, V. (2001). Izbirnost zahtevnostne ravni izpita in standardiziranje ocen v predlaganem novem modelu mature. *Sodobna Pedagogika*, št. 52, str. 110–122.
- Nunnally, J. C. in Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Šimenc, M. (ur.) (2007). *Letno poročilo – Splošna matura 2007*. Ljubljana: Državni izpitni center.

CANKAR, Gašper, Ph.D.

## STANDARDIZATION OF SCHOOL GRADES IN MATURA

**Abstract:** the author refers to the article written by Kališnik and Bucik (2001), who suggested a new model for assessing in matura examination. There are different simulations and research, following the initiative from the article an representing the attempts of the improved model of assessing. Problems, traps and complications of particular variants of new ways of assessing are described, whereas the reasons why the new model of assessing has not been introduced into practice are summarized in the conclusion of the article.

**Keywords:** matura examination, assessing.