

Zdenko Lapajne, zasebni raziskovalec v Ljubljani

UDK 372.880:371.274/.276

Urška Zobec, Republiški izpitni center v Ljubljani

## Analiza izbranih maturitetnih postavk

Zunanje preverjanje znanja v slovenskih šolah terja od učiteljev in kritične javnosti poleg znanja o učni snovi in načinih poučevanja tudi več znanja o tem, kako znanje preverjamo in vrednotimo. Še bolj so se povečale zahteve po takem znanju pri tistih izvedencih, ki pri zunanjem preverjanju — hote ali nehote — prevzemajo odgovornejše in kritiki bolj izpostavljene vloge sestavljalcev nalog, ocenjevalcev izdelkov in članov izpitnih komisij na različnih ravneh.

Za kakovostno meritev izobraževalnih učinkov daljšega izobraževalnega procesa je potrebno, da učitelji in načrtovalci meritev kolikor je mogoče soglašajo o *izobraževalnih ciljih* in o *didaktičnem izboru* najvažnejše učne snovi, katere znanje upravičeno pričakujemo od učenca (tudi: dijaka ali študenta) zlasti v primeru, ko je od pouka minilo že razmeroma veliko časa. To je še posebno pomembno pri tistih meritvah znanja in drugih lastnosti, na podlagi katerih se sprejemajo za kandidata in njegov razvoj zelo pomembne odločitve. Od izvedencev, ki sodelujejo pri načrtovanju in izvedbi takšnih meritev, delo zahteva tudi vednost o različnih načinih preverjanja znanja; ta pa je uporabna le, če je organsko prepletena s strokovnim poznavanjem učnega predmeta in razvojne stopnje učencev, ki jim je meritev namenjena.

Kakovostna meritev znanja je praviloma skupinsko delo, pri katerem od izvedencev za posamezne predmete sicer ne moremo pričakovati specialističnega in dokaj tehničnega znanja s področja teorije meritev v pedagogiki in psihologiji, informatike ali matematične statistike. Prav tako od izvedencev za ta vprašanja ne moremo pričakovati poglobljenega študija določenega šolskega predmeta, katerega znanje preverjamo; vendar neposredno in posredno prizadetim koristi, če programi zunanjega preverjanja znanja nastajajo v ozračju tvornega sodelovanja med izvedenci različnih disciplin. Pisca — psiholog in sociologinja — sva prispevek pripravila za strokovni posvet o maturi, ki je bil v Škofji Loki 8. in 9. decembra 1995. Dopolnila sva ga po razpravah v delovnih skupinah za družboslovje in jezikoslovje in temeljitih pripombah tedanjega predsednika republiške predmetne komisije Janeza Dularja. Bralec, ki bolje od naju obvlada jezikoslovje in pouk slovenskega jezika in književnosti, bo prav gotovo lahko obogatil pripombe k posameznim postavkam; najin glavni namen je bil, da takšne bralce seznaniva s količinskimi pristopi k merjenju znanja na kolikor je mogoče nazoren način. Članom omenjenih skupin se zahvaljujema, odgovarjava pa za morebitne preostale napake.

### Analiza testnih postavk

Splošni pojem »kakovostna meritev« v bolj strokovnem izrazju delimo na posamezne lastnosti meritev, kot so *objektivnost*, *zanesljivost* in *veljavnost*, ter lastnosti z njimi podprtih pedagoških odločitev, kot sta *pravičnost* in *sprejemljivost* za vrednostni sistem prizadetih.

*Analiza testnih postavk* je močno orodje, ki lahko postopoma privede do boljnih meritev izobraževalnih dosežkov. Čeprav njene izide dobimo šele *po* opravljeni meritvi, lahko kritičnim analitikom daje številne namige, kako je mogoče izboljšati pouk in meritve v prihodnosti ter izločiti tiste načine preverjanja znanja, ki so se izkazali za manj primerne. Naloge z zadovoljivimi značilnosti v določnem vzorcu lahko shranimo v »banko«; »vloge« se nam lahko v naslednjih letih »obrestujejo« tako, da bo izdelava kakovostnih in v času primerljivih merskih postopkov hitrejša in cenejša. Seveda to velja le za tiste, ki so v banko vložili dovolj veliko »premoženje« in jim ga niso razvrednotile nenehne spremembe učnih ciljev in snovi.

Izvedenci za razvoj meritev v pedagogiki in psihologiji pogosto poročajo o izidih analize testnih postavk, izraženih s števkami. Pri tem pričakujejo od izvedencev za pouk in preverjanje znanja pri posameznih predmetih ne le obvladovanje generičnih pojmov, kot so *težavnost*, *diskriminativnost*, *objektivnost* in *zanesljivost*, temveč tudi precej podrobno znanje o različnih statističnih postopkih, ki jih moramo izbrati v skladu z naravo podatkov.

Tak način sporočanja pogosto povzroči težave pri sporazumevanju in celo spore o tem, »kaj je v čigavi pristojnosti«. Takšni spori največkrat ne vodijo k boljšim meritvam, ker splošni izvedenci za merjenje sami, brez ustvarjalnih poznavalcev določenega učnega predmeta, ne morejo razviti boljnih meritev; predmetni specialisti pa brez uporabnega merskega znanja pogosto porabijo veliko časa za ukrepe, ki žal ne vodijo do boljše meritve.

Strokovna literatura o merjenju znanja in drugih človeških lastnosti postaja čedalje bolj tehnična in matematizirana, tako da zlahka odvrne zlasti izvedence za pouk družbenih ved, jezikov ali umetnosti. Zato sva se odločila, da bova v tem prispevku skušala uvesti bralca v analizo testnih postavk intuitivno in grafično podprto — med statističnimi pojmi se bova zadovoljila že z *aritmetično sredino*.

## Kaj je postavka?

Tako slovenska strokovna literatura kot dosedanja raba v različnih predmetnih komisijah kaže na nedosledno poimenovanje delov testa oz. v našem primeru maturitetnega izpita. Srečamo izraze *naloga*, *vprašanje*, *problem*, *tema* in še bi se kaj našlo. Razmeroma pogosto tisto, kar imenujemo naloga ali vprašanje, vsebuje več *delnih nalog* ali *podvprašanj*.

Za primer si oglejmo začetek analize neumetnostnega besedila v okviru junijskega maturitetnega preizkusa iz slovenskega jezika s književnostjo. Kandidat je najprej prebral daljše besedilo o idrijskih starožitnostih, potem pa naj bi odgovoril na prvo nalogo, sestavljeno iz petih trditev:

(Štev. točk)

1. Obkrožite črke **samo pred tremi** odgovori, ki ustrezajo vsebini priloženega besedila.
  - a) Idrijska kamšt je najstarejša v Evropi.
  - b) Premer njenega kolesa znaša 12,5 metra.
  - c) Ohranjena kamšt je danes v Idriji edina, včasih pa jih je bilo še več.
  - č) Angleški potopisec Walter Pope je bil med občudovalci idrijskih znamenitosti.
  - d) Idrijska kamšt še ni vpisana v register svetovne dediščine pri Unescu.

(6)

Zunanji ocenjevalec je po točkovniku presodil, ali črke, ki jih je obkrožil kandidat, označujejo trditve z oporo v prebranem besedilu. V tem primeru je dobil kandidat dve točki za vsako pravilno označeno trditev. Skupno število točk za vseh pet trditev, ki lahko zavzame vrednosti {0, 2, 4, 6}, je ocenjevalec vpisal v list za odgovore, ki ga je mogoče brati z optičnim čitalcem.

V iskanju primerne izraza za angleški *item* (ki se v strokovni terminologiji širi tudi v druge evropske jezike, npr. v nemščino) bova *postavko* opredelila kot najmanjši del izpitne pole, ki ga samostojno točkujemo. V primeru z opisanim načinom točkovanja imamo torej *eno* samo postavko.

Število točk, ki ga je kandidat dobil z odgovorom na neko postavko, imenujemo njegov *dosežek na postavki* (angl. *item score*).

Poudariti je treba, da na vprašanje *Kaj je postavka?* odgovarjajo sestavljanci izpitne pole in točkovnika, ne pa analitiki. Zgornjih pet trditev bi bilo namreč mogoče točkovati tudi na več drugačnih načinov; eden med njimi je, da bi ocenjevalec pri vsaki trditvi posebej ugotovil, ali kandidatov odgovor ima oporo v prebranem besedilu ali je nima. Pravilno označena trditev bi štel denimo eno točko:

	(Štev. točk)
Obkrožite črke samo pred tistimi odgovori, ki ustrezajo vsebini priloženega besedila.	
1. Idrijska kamšt je najstarejša v Evropi.	(1)
2. Premer njenega kolesa znaša 12,5 metra.	(1)
3. Ohranjena kamšt je danes v Idriji edina, včasih pa jih je bilo še več.	(1)
4. Angleški potopisec Walter Pope je bil med občudovalci idrijskih znamenitosti.	(1)
5. Idrijska kamšt še ni vpisana v register svetovne dediščine pri Unescu.	(1)

Pri takem načinu točkovanja imamo *pet* postavk, točke pa lahko sešteva ali kako drugače preračunava kasneje računalnik. Na odločitve o opredelitvi postavk lahko potemtakem razen vsebinskih razlogov vplivajo tudi administrativni, pa še kakšna omejitev uporabljene tehnologije.

## Empirična karakteristična funkcija postavke

Dosežke kandidata pri vseh postavkah, ki sestavljajo izpit nekega predmeta, seštejemo in dobimo njegov *testni dosežek* (angl. *test score*). Pri tem nas ne zanimajo podrobnosti o tem, kako je izpit sestavljen z vidika odmorov, tehnične ureditve izpitnih pol in delitve dela med zunanji ocenjevalci; zanimajo nas vsi podatki, ki štejejo pri določitvi kandidatove končne ocene.

*Empirična karakteristična funkcija postavke* je graf, pri katerem na abscisni osi prikažemo testni dosežek  $X$ , na ordinatni pa aritmetično sredino dosežkov na neki postavki  $g$ , ki so jih ustvarili tisti kandidati s testnim dosežkom  $X$  točk; označimo jo  $M_g(X)$ . Na sliki 1 je empirična karakteristična funkcija omenjene prve postavke preizkusa slovenskega jezika, točkovane tako, kot je bilo v navadi v junijskem roku. Zaradi primerljivosti grafov za različne predmete in načine točkovanja sva brez škode za splošnost nekoliko pretvorila vrednosti na obeh oseh:

1. Na abscisni osi je testni dosežek  $X$  izražen z *odstotnimi točkami*: testni dosežek delimo z največjim možnim številom točk in pomnožimo s 100, tako da odstotne točke lahko načelno segajo od 0 do 100. (Porazdelitev tako preračunanih odstotnih točk za vse kandidate v junijskem roku je na sliki 2, s katere med drugim zvmemo, da je najboljši kandidat zbral 95 % najvišjega možnega števila točk, najslabši pa 9 %. Uporabljen je bil torej večji del intervala možnih ocen — ne pa ves.)
2. Na ordinatni osi je  $M_g(X)$  izražen z deležem največjega možnega števila točk: dosežek na postavki  $g$  delimo z največjim možnim številom točk pri tej postavki. Najnižja vrednost je 0 (pri prvi postavki 0 točk), najvišja pa 1 (kar pri tej postavki ustreza 6 točkam).

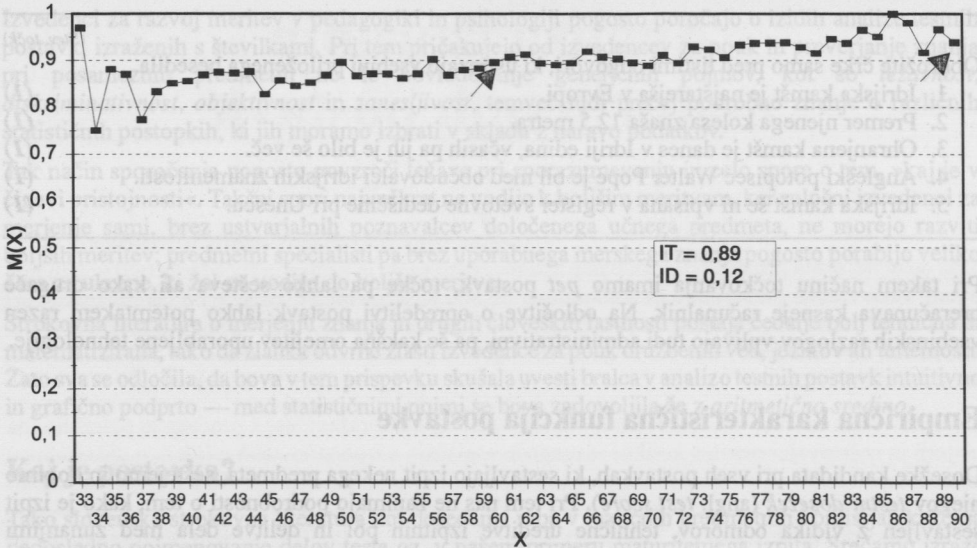
### Primer:

Bralec bo povedano lažje razumel, če se na sliki 1 v mislih posveti skrajni piki desno zgoraj, na katero kaže puščica. Podatki s slike 2 nam povedo, da je 90 odstotnih točk pri slovenščini zbralo 11 kandidatov. Njihov povprečni dosežek na prvi postavki je  $0,939 \times 6 = 5,636$  točke. Največ kandidatov pa je zbralo 60 odstotnih točk: bilo jih je 283, na prvi postavki pa so v povprečju zbrali  $0,896 \times 6 = 5,378$  točke. (Vrednosti 0,939 in 0,896 lahko vsaj približno odčitamo s slike 1.)

Empirične karakteristične funkcije postavk so nazorno pomagalo v analizi postavk zlasti takrat, ko imamo na voljo zelo velike vzorce. Tedaj lahko večina pik temelji na večjem številu kandidatov. V našem primeru imamo podatke o 7.535 kandidatih. Kljub tako velikem vzorcu pa je po sliki 2 izredno malo kandidatov, ki so zbrali zelo veliko ali zelo malo točk. Na vseh grafih sva v želji po preglednosti

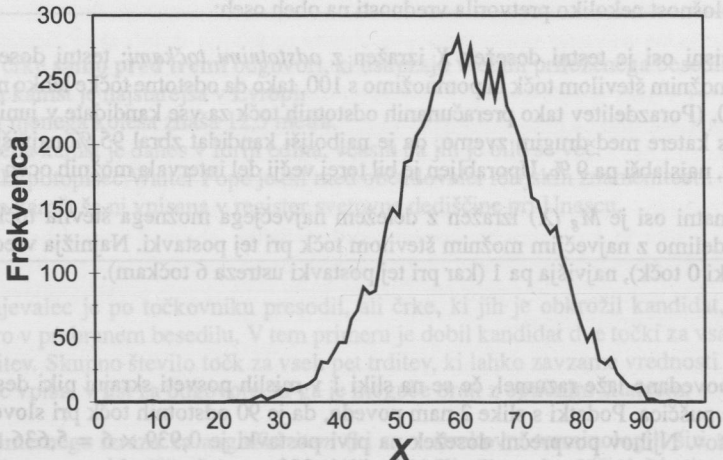
VIR: REPUBLIŠKI IZPITNI CENTER, 1995

SLIKA 1: POSTAVKA 1



VIR: REPUBLIŠKI IZPITNI CENTER, 1995

SLIKA 2: PORAZDELITEV ODSTOTNIH TOČK  
MATURA 1995, JUNIJSKI ROK, N = 7.535



narisala samo tiste pike, ki temeljijo na več kot desetih kandidatih; to je 58 pik v intervalu od 33 do vključno 90 odstotnih točk. Izven okvira grafa pa so namenoma ostali podatki o zelo redkih kandidatih, ki so zbrali od 9 do 32 ali od 91 do 95 odstotnih točk.

Praktične izkušnje izvedencev z empiričnimi karakterističnimi funkcijami postavk v zelo velikih vzorcih so lahko dober intuitiven uvod v sodobno testno teorijo, katere temeljni pojem je karakteristična funkcija postavke v splošeni obliki (Hambleton in Swaminathan, 1985).

## 1. postavka

Kaj nam slika 1 pove o prvi postavki? Primerjajmo jo z rezultati klasične analize testnih postavk (kot jo izračuna podprogram RELIABILITY statističnega paketa SPSS)! Tam preberemo, da je njena težavnost  $IT$  0,89, diskriminativnost  $ID$  pa 0,12. Indeks težavnosti pove, koliko točk lahko pričakuje povprečni kandidat; čim višji je, tem lažja je postavka. Da je prva postavka »lahka«, na grafu vidimo tako, da je velika večina kandidatov zbrala več kot 80 % možnega števila točk. Premislimo najprej kandidatove možnosti, da problem reši brez znanja, samo z ugibanjem: v uvodu zve, da so med petimi trditvami natanko tri z oporo v besedilu. Če svoje tri glasove po naključju razporedi med pet trditev, lahko pričakuje več kot polovico, točneje  $3/5$  ali 0,6 možnega števila točk.

Če odmislimo manjša nihanja zaporednih točk grafa, v splošnem kandidati z boljším uspehom pri celotnem izpitu zberejo tudi pri tej postavki več točk. Pravimo, da postavka *razločuje* ali (pozitivno) *diskriminira* boljše kandidate od slabših; ker pa njena karakteristična funkcija narašča precej počasi, diskriminativnost postavke ni zelo visoka. (To se vidi iz ne pretirano visokega indeksa diskriminativnosti 0,12 — vendar sva bralcu obljubila, da bo shajal samo s povprečji; brez količnika korelacije, ki lahko zavzame vrednosti med -1 in 1!)

Kaj pomeni izjava, da je postavka »lahka« ali da »šibko diskriminira«? Poudariti je treba, da ne velja splošno, temveč samo za tisti vzorec kandidatov, v katerem smo analizirali postavko kot del nekega testa. Odmislimo možnost, da se izpitna pola v praksi hitro razve, in se vprašajmo, kakšni bi bili rezultati analize iste postavke, če bi v septembrskem roku uporabili povsem enako izpitno polo! Odgovor je odvisen od tega, ali v drugem roku kandidirajo podobni kandidati kot v prvem ali pa morda bistveno slabši.

Odgovor na vprašanje o tem, ali velja prvo postavko izločiti iz nastajajoče banke postavk, se razlikuje med sestavljalci testov v okviru klasične, na norme usmerjene testne teorije, in sestavljalci, ki prisegajo na kriterijski pristop. (Podrobneje ga opisujejo Popham (1978), Sagadin (1991) in Hambleton (1994).) Večina klasičnih sestavljalcev bi takó láchko postavko izločila, ker ne pove veliko novega o merjeni lastnosti. Klasični sestavljalec daje prednost postavkam čim višje diskriminativnosti, take pa so najpogosteje v razponu srednje težavnosti.

Sestavljalci *kriterijskega* testa bi se najprej povprašali o tem, v kakšnem razmerju so miselni procesi, ki jih sproža postavka, do zapisanih učnih ciljev. Če postavka smiselno preverja neki pomemben učni cilj, nam odgovori povedo, da ga je večina učencev pač dosegla. Zgodovinsko gledano se je potreba po kriterijskih testih pojavila najprej pri programiranem pouku, kjer programer učno snov razdeli v številne majhne enote, da bi jih učencu lahko ponudil tudi učni stroj ali računalnik. Ker učenec napreduje po programu k naslednji enoti potem, ko prvo obvlada, se pojavi vprašanje, kako to ugotoviti. Zato so bili prvi kriterijski testi kratki in usmerjeni na preverjanje pozameznih podrobnih učnih ciljev. O učencu želimo zvedeti le, ali določeno enoto obvlada ali ne. Klasično konstruirani test učitelju pesimistično sporoča, da povprečni učenec ne zna rešiti približno polovice postavk, vendar omogoča — če je dobro sestavljen — razvrstitev učencev od boljših do slabših glede na številnejše in širše opredeljene izobraževalne cilje. Dosežek posameznega učenca vrednotimo tako, da ga primerjamo s porazdelitvijo dosežkov podobnih učencev, s statistično ugotovljenimi normami. Če se učenci in učitelji bolj potrudijo in bolje naučijo snov, bo treba najti

težje in diskriminativne postavke. Statistični kazalci klasične analize postavk so konstruktorju kriterijskega testa le v pomoč pri izvedenski vsebinski presoji.

Poleg merskega vidika je pri meritvah znanja, kakršna je matura, ki so za učenca pomembne in zato zanj pomenijo stres, pomemben tudi motivacijski vidik. Na začetek zavestno uvrščamo tudi postavke, ki so za veliko večino kandidatov lahke, da dobijo zaupanje vase in da že na začetku izpita ne zgubijo nesorazmerno veliko časa z zahtevnimi problemi.

Ne glede na teoretična izhodišča pa velja pri prvi postavki premisliti, s katerimi ukrepi in dodatnimi sporočili bi lahko dosegli, da bi dosežek na postavki več povedal o merjeni lastnosti (razumevanju prebranega besedila), manj pa o naključju.

## 11. postavka

Po zelo lahki postavki nizke diskriminativnosti si na sliki 3 oglejmo tudi primer postavke, ki zelo dobro diskriminira maturante s slabim jezikovnim znanjem slovenščine.

(Štev. točk)

Napišite knjižni izgovor naslednjih besed (z naglasom):

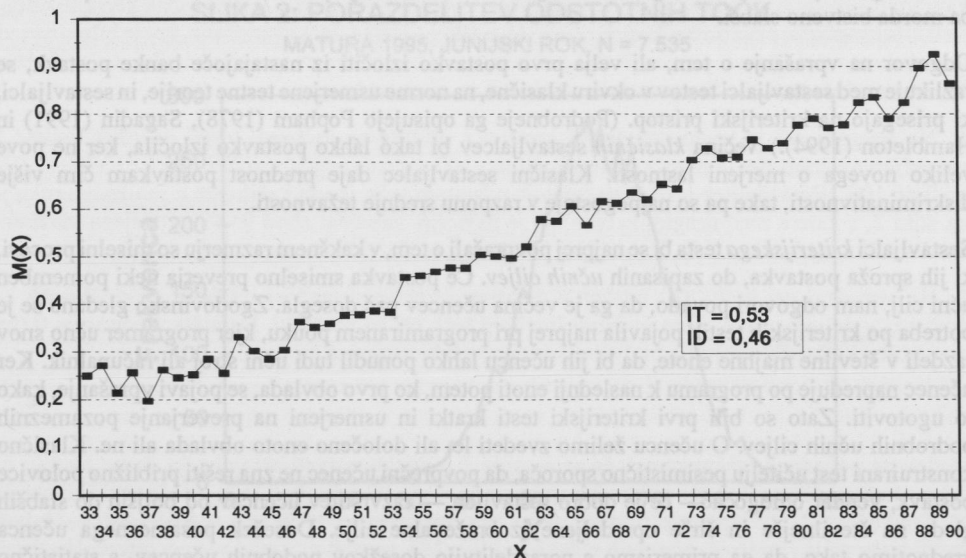
Evropa	[ _____ ]
občudovati	[ _____ ]
prizadevanj	[ _____ ]
jašek	[ _____ ]

(8)

Šolano psihometrično oko v nalogi opazi več možnosti, kako bi jo lahko razdelili na postavke. Posebna postavka bi bila lahko vsaka beseda ali celo vsako težavno mesto. Naslednja možnost je

VIR: REPUBLIŠKI IZPITNI CENTER, 1995

SLIKA 3: POSTAVKA 11



še bolj podrobna: zanimala bi nas lahko pravilnost transliteracije slehernega *grafema* v fonetični zapis. V najini analizi sta nas uporabljeni točkovnik in ocenjevalni list kajpak prisilila, da nalogo obravnavamo celostno, kot eno postavko. Težavnost 0,53 pove, da je povprečni maturant povsem pravilno zapisal izgovor dobre polovice besed, diskriminativnost 0,46 pa je med najvišjimi v izpitni poli. Kot laika domnevava, da postavka preverja pomemben učni cilj, ki ne le dijaku v šoli, temveč tudi intelektualcu kasneje omogoča uporabo slovarjev in pravorečnih vprašanjih in praktično obvladovanje nekaterih občutljivih zadreg knjižnega govora. Zato bo nad postavko najbrž navdušen tudi konstruktor kriterijskega testa. Karakteristična funkcija narašča skoraj linearno, vendar je to morda posledica združevanja podatkov o fonetičnem zapisu štirih morda različno težkih in diskriminativnih besed, od katerih je vsaka pravilno zapisana kandidatu prinesla po dve točki.

## 14. postavka

na sliki 4 je primer razmeroma težke postavke, ki pa odlično diskriminira.

(Štev. točk)

Preberite geselski članek besede srebro iz SSKJ.

(...)

a) Poimenujte naglasni tip, na katerega nas opozarjajo podatki iz zaglavja tega geselskega članka. (2)

b) Kateri izmed uporabljenih kvalifikatorjev v tem geselskem članku se nanaša:

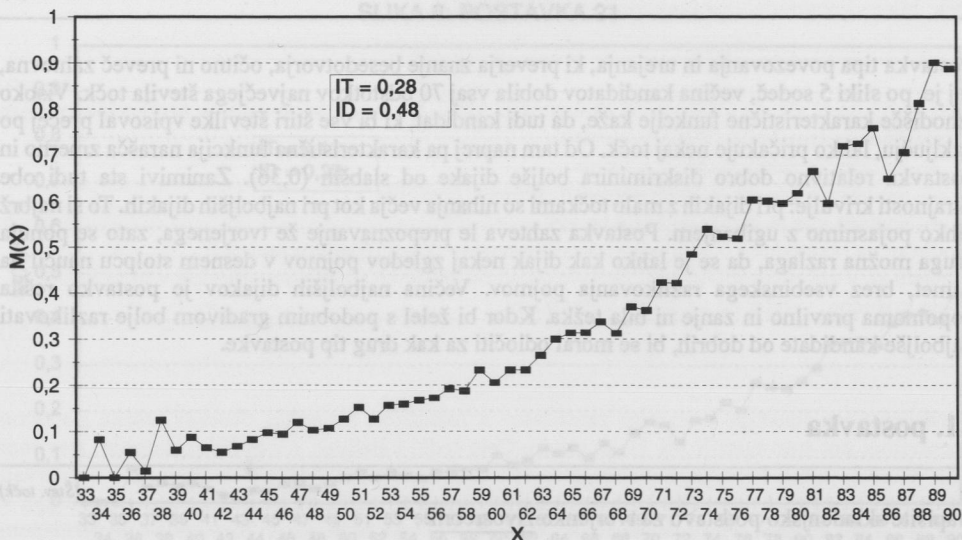
— na socialnozvrstno pripadnost geselske besede \_\_\_\_\_ (1)

— na pripadnost strokovnemu jeziku? \_\_\_\_\_ (1)

\_\_\_\_\_ (4)

VIR: REPUBLIŠKI IZPITNI CENTER, 1995

SLIKA 4: POSTAVKA 14



Pregled vsebine ponovno pokaže problematičnost opredelitve postavke, saj problem vsebuje dve delni nalogi, ki bi ju lahko točkovali kot samostojni postavki; še več: glede na način točkovanja bi jo lahko razdelili na tri samostojne postavke. Pri izbranem načinu točkovanja je postavka razmeroma težka, saj je povprečni maturant dosegel precej manj kot tretjino (28 %) možnega števila točk. Če bi v točkovniku nalogo razdelili v več postavk, bi nam rezultati analize povedali, kaj je povzročalo dijakom več težav; vsekakor pa rezultati kažejo, da je za uporabo in razumevanje slovarskega gesla potrebno kar precej znanja. Vsa vprašanja so odprtega tipa, tako da si slabši kandidati ne morejo pomagati z ugibanjem in se levi del karakteristične funkcije praktično začenja pri ničelnem povprečnem dosežku na postavki. Tudi zato je diskriminativnost izredno visoka (0,48) in kaže, da dijaki z višjim skupnim številom točk iz slovarskega gesla razberejo bistveno bogatejša sporočila. Karakteristična funkcija v srednjem delu narašča počasneje kot v drugem delu, v najboljši tretjini kandidatov. Postavka v uporabljeni formulaciji najbolje razlikuje med boljšimi in najboljšimi kandidati. Če bi jo želeli olajšati za povprečne kandidate, bi to lahko storili tako, da bi odprta vprašanja »zaprli«; lahko pa bi zahtevne pojme (kot je npr. »naglasni tip«) najprej uvedli s primeri. Čeprav so na tem področju napovedi tvegane, pa se zdi, da bi v želji po olajšanju postavke najbrž znižali visoko diskriminativnost.

## 20. postavka

(Štev. točk)

Vsaki besedi iz levega stolpca pripišite ustrezno številko iz desnega.

- |                  |   |                |
|------------------|---|----------------|
| a) cesarski      | — | 1 sklop        |
| b) drogovje      | — | 2 zloženska    |
| c) nadvojvoda    | — | 3 sestavljenka |
| č) sedemdeset    | — | 4 izpeljanka   |
| d) vodoravno     | — |                |
| e) oddati        | — |                |
| f) osemindvajset | — |                |
| g) zdravljenje   | — |                |

(4)

Postavka tipa povezovanja in urejanja, ki preverja znanje besedotvorja, očitno ni preveč zahtevna, saj je, po sliki 5 sodeč, večina kandidatov dobila vsaj 70 odstotkov največjega števila točk. Visoko izhodišče karakteristične funkcije kaže, da tudi kandidat, ki bi vse štiri številke vpisoval precej po naključju, lahko pričakuje nekaj točk. Od tam naprej pa karakteristična funkcija narašča zmerno in postavka relativno dobro diskriminira boljše dijake od slabših (0,36). Zanimivi sta tudi obe skrajnosti krivulje: pri dijakih z malo točkami so nihanja večja kot pri najboljših dijakih. To si najbrž lahko pojasnimo z ugibanjem. Postavka zahteva le prepoznavanje že tvorjenega, zato se ponuja druga možna razlaga, da se je lahko kak dijak nekaj zgledov pojmov v desnem stolpcu naučil na pamet, brez vsebinskega razlikovanja pojmov. Večina najboljših dijakov je postavko rešila popolnoma pravilno in zanje ni bila težka. Kdor bi želel s podobnim gradivom bolje razlikovati najboljše kandidate od dobrih, bi se moral odločiti za kak drug tip postavke.

## 21. postavka

(Štev. točk)

Napišite skladenjsko podstavo za tvorjenko živosrebrn.

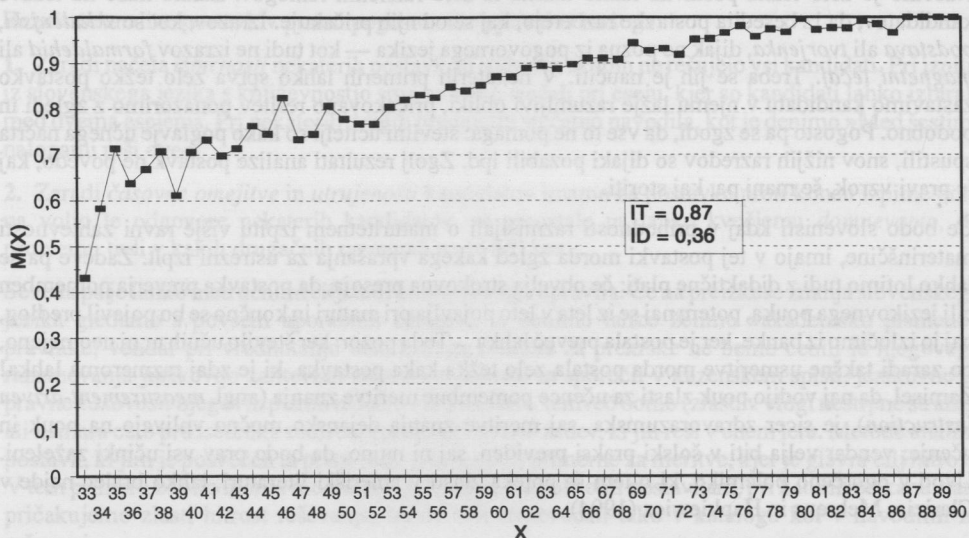
(2)



Tudi ta postavka (slika 6) preverja znanje s področja besedotvorja in je bila za dijake očitno zelo trd oreh: povprečni kandidat je namreč zbral le 11 odstotkov možnega števila točk in najboljši kandidati niso presegli polovice možnega števila točk. Celotni maturant z najvišjim skupnim dosežkom ni dobil obeh točk! Kljub temu da je postavka težka, zelo dobro diskriminira (0,32). V nasprotju z

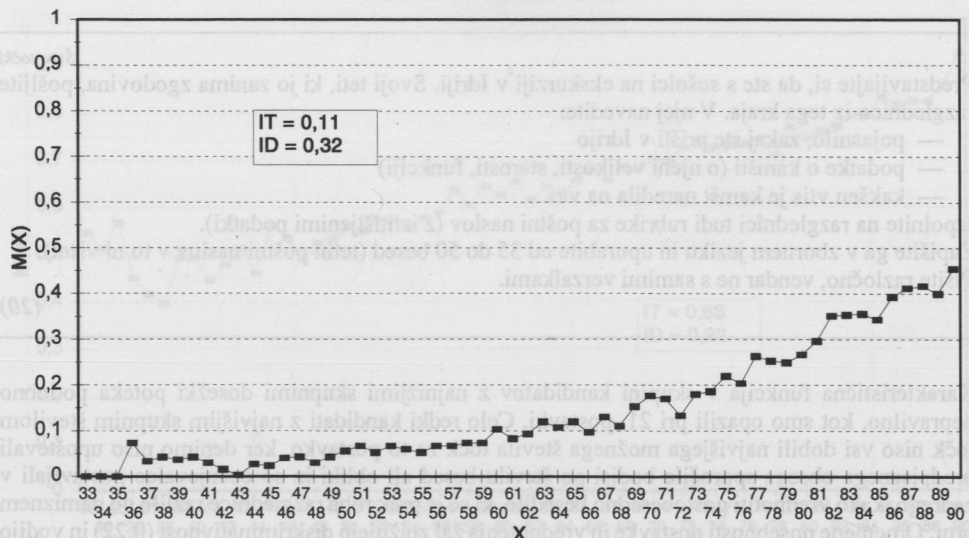
VIR: REPUBLIŠKI IZPITNI CENTER, 1995

**SLIKA 5: POSTAVKA 20**



VIR: REPUBLIŠKI IZPITNI CENTER, 1995

**SLIKA 6: POSTAVKA 21**



dvajseto postavko, ki je zahtevala prepoznavanje že tvorjenega, pa ta zahteva poznavanje besedotvornega algoritma oziroma logično razmišljanje. S tem bi morda lahko pojasnili čuden potek karakteristične funkcije v spodnjem delu. Najslabši kandidati (s 36-40 odstotki točk) so v povprečju nepričakovano dosegli nekaj več točk od tistih z 42-43 odstotki; od tam naprej funkcija sprva zelo počasi in šele v najboljši tretjini kandidatov hitreje narašča. Najslabši kandidati so si morda pri reševanju pomagali z zdravo pametjo, vendar bi bilo za presojo te domneve treba vsebinsko presoditi njihove odgovore.

Načelno je zelo težka postavka lahko takšna iz zelo različnih razlogov. Lahko kaže na težave kandidatov, da iz besedila postavke razberejo, kaj se od njih pričakuje. Izrazov, kot so *skladenjski*, *podstava* ali *tvorjenka*, dijaki ne pozna iz pogovornega jezika — kot tudi ne izrazov *formaldehid* ali *magnetni tečaj*. Treba se jih je naučiti. V nekaterih primerih lahko sprva zelo težko postavko zastavimo kandidatu v njemu lažje razumljivi obliki, pričakovano rešitev ponazorimo z zgledi in podobno. Pogosto pa se zgodi, da vse to ne pomaga: številni učitelji so kako poglavje učnega načrta spustili, snov nižjih razredov so dijaki pozabili ipd. Zgolj rezultati analize postavk ne povedo, kaj je pravi vzrok, še manj pa, kaj storiti.

Če bodo slovenisti kdaj v prihodnosti razmišljali o maturitetnem izpitu višje ravni zahtevnosti materinščine, imajo v tej postavki morda zgled kakega vprašanja za ustrezní izpit. Zadeve pa se lahko lotimo tudi z didaktične plati: če obvelja strokovna presoja, da postavka preverja pomemben cilj jezikovnega pouka, potem naj se iz leta v leto pojavlja pri maturi in končno se bo pojavil predlog, naj jo izločimo iz banke, ker je postala preveč lahka ... Toda pozor: ker število učnih ur ni neomejeno, bo zaradi takšne usmeritve morda postala zelo težka kaka postavka, ki je zdaj razmeroma lahka! Zamisel, da naj vodijo pouk zlasti za učence pomembne meritve znanja (angl. *measurement-driven instruction*), je sicer zdravorazumska, saj meritve znanja dejansko močno vplivajo na pouk in učenje; vendar velja biti v šolski praksi previden, saj ni nujno, da bodo prav vsi učinki zaželeni. Uvod v ogorčeno polemiko, ki o tem še poteka zlasti v ameriški literaturi, lahko bralec najde v Clarizio, Mehrens in Hapkiewicz (1994).

## 32. postavka

Poslednja postavka izpitne pole s področja sporočanja je lahko kandidatu prinesla razmeroma veliko točk. Njena karakteristična funkcija je na sliki 7.

(Štev. točk)

Predstavljajte si, da ste s sošolci na ekskurziji v Idriji. Svoji teti, ki jo zanima zgodovina, pošljite razglednico iz tega kraja. V njej navedite:

- pojasnilo, zakaj ste prišli v Idrijo
- podatke o kamšti (o njeni velikosti, starosti, funkciji)
- kakšen vtis je kamšt naredila na vas.

Izpolnite na razglednici tudi rubrike za poštni naslov (z izmišljenimi podatki).

Zapišite ga v zbornem jeziku in uporabite od 35 do 50 besed (tetin poštni naslov v to ni vštet).

Pišite razločno, vendar ne s samimi verzalkami.

(20)

Karakteristična funkcija v skupini kandidatov z najnižjimi skupnimi dosežki poteka podobno nepravilno, kot smo opazili pri 21. postavki. Celotni redki kandidati z najvišjim skupnim številom točk niso vsi dobili najvišjega možnega števila točk za to postavko, ker denimo niso upoštevali predpisanega obsega sporočila bodisi po številu besed ali obliki in so ocenjevalce postavljali v zadrego, kako vrednotiti gostobesedno besedilo, ki se je prek roba razglednice razlilo po namiznem prtu. Omenjene posebnosti postavke in vrednotenja žal znižujejo diskriminativnost (0,22) in vodijo do vprašanja, če ni bilo morda število točk za to postavko nekoliko precenjeno. Vsiljuje se možnost,

da to postavko (in še nekatere druge) nekateri kandidati lahko rešijo precej uspešno že z osnovnošolskim znanjem jezika in morda nekoliko boljšimi jezikovnimi, morda pa tudi ne zgolj jezikovnimi sposobnostmi (kamor morda sodi tudi kultura reševanja testov, disciplina, urejenost in čitljivost rokopisa, pozorno branje navodil ali ekonomično ravnanje s časom); tudi brez znanja opisne slovnice na srednješolski ravni torej, ki prinaša točke pri večini drugih postavk.

## Vpliv časovne meje ali izbirnosti

Rezultate analize postavk v praksi pogosto omejujeta dva razloga:

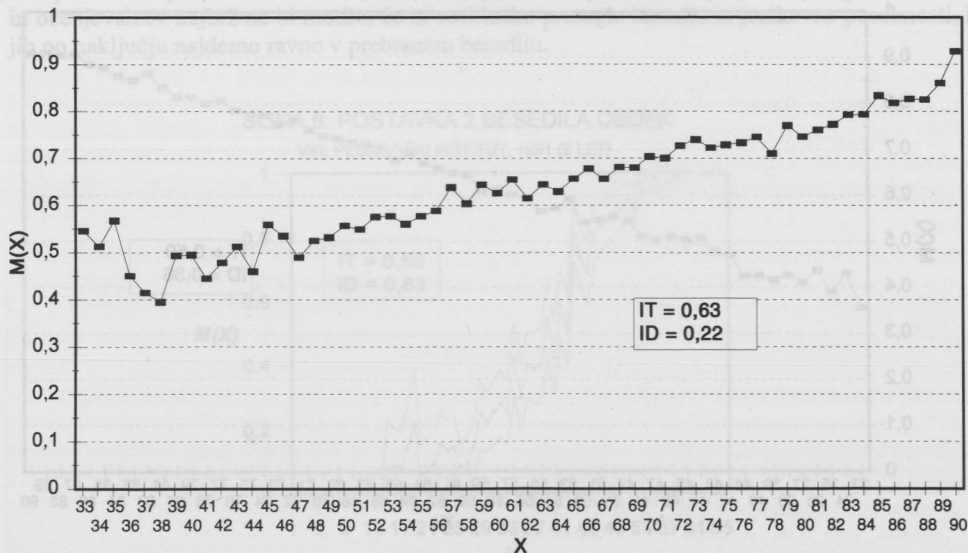
1. Zaradi načela *izbirnosti* nekaterih postavk ali njihovih skupin ne rešujejo vsi kandidati. Pri izpitu iz slovenskega jezika s književnostjo smo se s tem srečali pri eseju, kjer so kandidati lahko izbirali med dvema esejema. Pri nekaterih drugih predmetih srečamo navodila, kot je denimo »Med šestimi nalogami reši dve«.

2. Zaradi *časovne omejitve* in *utrujenosti* kandidatov imamo za postavke proti koncu izpitne pole na voljo le odgovore nekaterih kandidatov, za preostale pa lahko kvečjemu *domnevamo*, da problemu tudi v kakih drugačnih razmerah ne bi bili kos.

Seveda pa je lahko med učnimi cilji tudi *hitrost* nekega opravila. Če na preizkuse znanja slovenskega jezika gledamo s povsem uporabnih izhodišč, si denimo lahko želimo »akademsko pismene« pravnike; vendar pri vrednotenju dela kakega sodnika za prekrške ne bomo cenili le njegovega razumevanja jezikovno zahtevnih zakonskih besedil in sporočil v kazenskem spisu, jezikovne in pravne *kakovosti* njegovih pisnih izdelkov in podobno, temveč bomo (zlasti v vlogi nestrpne stranke ali namara celo predsednika sodišča) cenili tudi *število* zadev, ki jih reši v enem letu. Metode analize postavk, ki jim je posvečen ta prispevek, niso najbolj primerne za meritve, kjer je glavni cilj hitrost. V tem primeru bi bilo morda modro del izpitnega časa posvetiti postavkam, pri katerih od kandidata pričakujemo zlasti hitrost reševanja, ter to tudi napovedati tako v katalogu kot v navodilih za reševanje.

VIR: REPUBLIŠKI IZPITNI CENTER, 1995

SLIKA 7: POSTAVKA 32



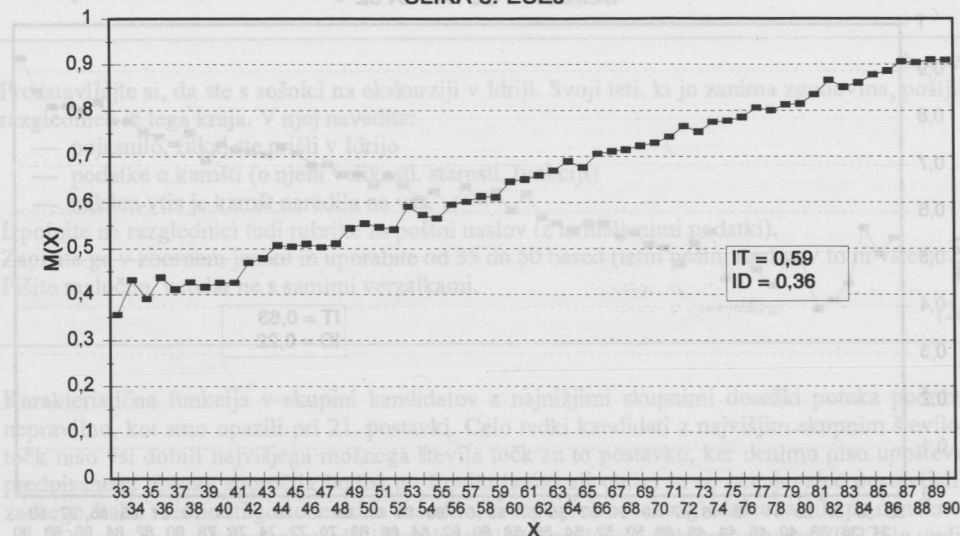
Esej

Priznati moramo, da je nekoliko nasilno, če celotni esej analiziramo kot eno samo postavko. Kandidat je namreč lahko izbiral med dvema vrstama esejev, navodila za točkovanje pa so bila razmeroma podrobna in so poudarjala tako znanje o jeziku kot o književnosti. Kljub temu smo na sliki 8 prikazali tudi empirično karakteristično funkcijo za esej, in to za povprečno število točk pri obeh ocenjevalcih. Karakteristična funkcija zelo pravilno in skoraj premosorazmerno narašča ter zelo dobro diskriminira (0,36). Morda se začne malce previsoko in za koga konča malo prenizko. Povprečni kandidat je za esej dobil 59 odstotkov možnega števila točk, najslabši kandidati pa okrog 40 odstotkov možnega števila točk. V celoti maturitetnega preizkusa slovenskega jezika in književnosti je to morda celo zaželeno, saj z esejem preverjamo ne le zmožnost pisnega sporočanja, temveč tudi zmožnost interpretacije oziroma primerjalne analize literarnih besedil. Ko naj bi kandidat pokazal temeljno znanje sporočanja v učnem jeziku, ki je najpogosteje tudi materni, je najbrž prav, da tudi najslabši doseže vsaj nekaj točk.

Bolj problematičen pa se zdi zgornji del krivulje, saj niti vsi maturanti z najvišjimi skupnimi ocenami niso dobili najvišjega možnega števila točk za esej. Tu se moramo ustaviti ob estetskem vrednotenju maturitetnih esejev, kar je izredno zapleten proces, v katerem poleg čitljivosti rokopisa, preverljivih dejstev, slovničnih napak in drugih zahtev točkovnika igra pomembno vlogo tudi vrednotenje in presojanje prebranega v luči osebnega književnega in jezikovnega okusa ocenjevalca, med avtorji in ocenjevalci pa lahko zija tudi bolj ali manj širok generacijski prepad. Tako le težko najdemo kak maturitetni esej, ki bi popolnoma zadostil *vsem* kriterijem po mnenju *različnih* zunanjih ocenjevalcev. Najbrž bi bilo podobno, če bi ocenjevalcem med dijaške eseje podtaknili tudi eseje nobelovcev za literaturo. Ne smemo pozabiti tudi dejstva, da je bil leta 1995 ne le dijakom in njihovim učiteljem, temveč tudi ocenjevalcem šolski esej precej nov in so se z njegovimi zahtevami pogosto srečali šele ob pripravah na maturo. Nekatere posebnosti vrednotenja eseja, ki jih v luči testne teorije obravnavamo kot »merske napake«, bi lahko ublažila enotnejša didaktika pouka slovenskega jezika in književnosti kot enega šolskega predmeta, več izkušenj z ocenjevanjem pri

VIR: REPUBLIŠKI IZPITNI CENTER, 1995

SLIKA 8: ESEJ



študiju bodočih profesorjev, v času pripravnštva in izmenjav izkušenj na seminarjih za učitelje, ter kolegialno (so)ocenjevanje izbranih pisnih izdelkov dijakov tudi v nižjih razredih srednje šole.

## Omejitve analize postavk

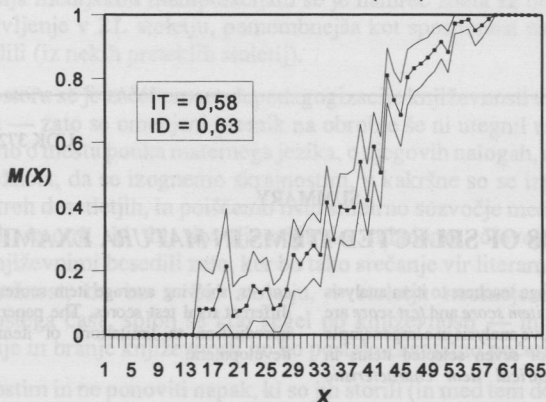
Stabilne empirične karakteristične funkcije postavk lahko narišemo le za zelo velike vzorce. V reprezentativnem vzorcu populacije je mogoče določiti tudi *interval zaupanja*, v katerem z vnaprejšnjo verjetnostjo pričakujemo populacijsko aritmetično sredino. Na sliki 9 je interval zaupanja empirične karakteristične funkcije najbolj diskriminativne postavke slovenske oblike bralnega testa za devetletne učence v mednarodni raziskavi (Elley, Gradišar in Lapajne, 1995). Kljub zelo velikemu vzorcu 3.300 učencev je interval zaupanja precej širok in sili v previdnost pri posploševanju rezultatov analize testnih postavk v manjših slučajnih vzorcih, ko skušamo oceniti parametre karakteristične funkcije po dostopnih podatkih, domnevah o njeni obliki ali smiselnem agregiranju kandidatov v najmanj dve skupini. Zelo pogosto uporabljan postopek približne ocene diskriminativnosti je bil naslednji: točkovane liste za odgovore so najprej razvrstili po številu točk v tri skupine: 27 % najboljših, 27 % najslabših in preostalo večino. Potem so pogostost odgovorov na posamezne postavke primerjali le v skrajnih skupinah najboljših in najslabših (podrobneje Guilford, 1954). Takšne »bližnjice« do analize testnih postavk so bile pomembne pred uporabo optičnega čitalca in računalnika v razvoju testov, saj so olajšale zamudno administrativno delo.

Pri uporabi analize postavk pri maturi se moramo torej zavedati, da ne daje uporabnih rezultatov v zelo majhnih vzorcih (z manj kot približno 30 kandidati) ali v vzorcih z zelo zmanjšano razpršenostjo merjene lastnosti. Ta omejitev nas seveda ne skrbi pri obveznem predmetu, kot je slovenski jezik s književnostjo, pač pa zlasti pri jezikih, ki se jih uči malo srednješolcev.

Zamisel o razvoju banke maturitetnih postavk v primeru preizkusov znanja jezikov na prvi pogled omejuje zahteva, da se posamezne postavke nanašajo na prebrano besedilo. Na podobne primere naletimo tudi pri izpitnih polah za druge predmete, kjer denimo pri zgodovini ali geografiji zastavimo več vprašanj o isti tematski karti. Če podrobneje proučimo razmerje med problemi postavk in prebranim besedilom pri junijski maturi leta 1995 iz slovenskega jezika in književnosti, res opazimo del postavk, ki so z besedilom tesno povezane; pri drugih (npr. pri enajsti) pa kandidatov in ocenjevalcev najbrž ne bi motilo, če bi vsebinsko presegle besedje in jezikovne posebnosti, ki jih po naključju najdemo ravno v prebranem besedilu.

SLIKA 9: POSTAVKA 2 BESEDILA DEDEK

VIR: PEDAGOŠKI INŠTITUT, 1991 (9 LET)



## Uporabljena literatura

- Clarizio, H. F., Mehrens, W. A. in Hapkiewicz, ur. (1994). *Contemporary Issues in Educational Psychology*. Sixth ed. New York: McGraw-Hill.
- Elley, W. B., Gradišar, A. in Lapajne, Z. (1995). *Kako berejo učenci po svetu in pri nas? Mednarodna raziskava o bralni pismenosti*. Nova Gorica: Educa.
- Guilford, J. P. (1954). *Psychometric Methods*. New York: McGraw-Hill.
- Hambleton, R. K. (1994). *Criterion-Referenced Measurement*. V: Husén, T. in Postlethwaite, T. N. (ur.) *The International Encyclopedia of Education*. Second ed. Oxford: Pergamon.
- Hambleton, R. K. in Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Nijhof.
- Nunnally, J. C. in Bernstein, I. H. (1994). *Psychometric Theory*. Third ed. New York: McGraw-Hill.
- Popham, W. J. (1978). *Criterion-Referenced Measurement*. Englewood Cliffs: Prentice Hall.
- Sagadin, J. (1991). *Razprave iz pedagoške metodologije*. Ljubljana: ZI FF.
- Toporišič, J. (1965-1970). *Slovenski knjižni jezik 1-4*. Maribor: Obzorja.
- Toporišič, J. (1992). *Enciklopedija slovenskega jezika*. Ljubljana: Cankarjeva založba.

Zdenko Lapajne  
Urška Zobec

UDK 372.880:371.274/.276

### SUMMARY

## ITEM ANALYSIS OF SELECTED ITEMS IN *MATURA* EXAMINATIONS

The paper introduces language teachers to item analysis in nontechnical terms: *item*, *item score* and *test score* are defined. Results of quantitative analysis in a large sample ( $N = 7,535$ ) are shown for seven selected items in graphical form with *empirical item characteristic*

*curves*, showing average item scores for students with different total test scores. The paper ends with a brief introduction to limitations of item analysis in test development.