

Naslov članka/Article:

Izzivi pri vrednotenju preizkusov nacionalnega preverjanja znanja matematike

Challenges in Assessing the National Assessment of Knowledge Tests in Mathematics

Avtor/Author:

David Janet, Jerneja Bone

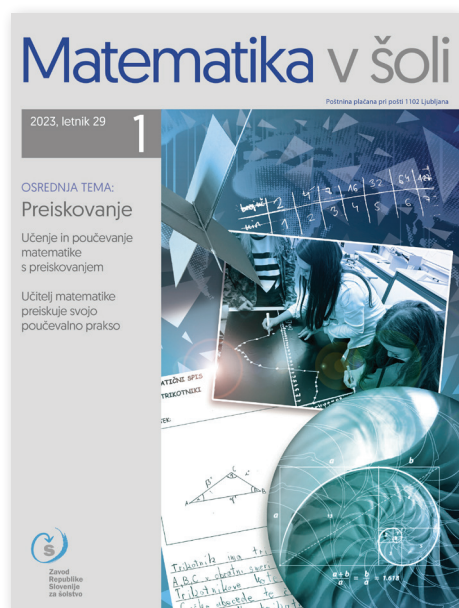
DOI:

<https://doi.org/10.59132/mvs/2023/1/2-12>

CC licenca



Priznanje avtorstva-Nekomercialno-Brez predelav



Matematika v šoli št. 1/2023, letnik 29

ISSN 1318-010X

Izdal in založil: Zavod Republike Slovenije za šolstvo

Kraj in leto izdaje: Ljubljana, 2023

Spletna stran revije:

<https://www.zrss.si/strokovne-revije/matematika-v-soli/>

Izzivi pri vrednotenju preizkusov nacionalnega preverjanja znanja matematike

Challenges in Assessing the National Assessment of Knowledge Tests in Mathematics

David Janet, Državni izpitni center
Jerneja Bone, Zavod RS za šolstvo

Izvelek

Raziskava izpostavlja različna vprašanja pri vrednotenju preizkusov nacionalnega preverjanja znanja (NPZ) matematike. Preverili smo veljavnost in učinkovitost preizkusov za standardizacijo in kontrolnih preizkusov pri ločevanju med natančnimi in manj natančnimi ocenjevalci. Prav tako smo se osredotočili na lastnosti nalog oziroma postavk, ki se povezujejo z natančnostjo pri vrednotenju preizkusov. Rezultati kažejo, da oba kontrolna mehanizma (standardizacija in vrednotenje kontrolnih preizkusov) razmeroma učinkovito ločuje ta natančne ocenjevalce od manj natančnih. Za naloge, ki jih ocenjevalci vrednotijo natančneje, velja, da so povečini zaprtega tipa. Več težav pri vrednotenju ocenjevalcem povzročajo naloge aritmetike in algebre, ki spadajo v rdeče in modro območje ter v območje nad modrim. Tudi naloge, ki vsebujejo dodatna moderirana navodila, ocenjevalci vrednotijo slabše. Ovrednotili smo pomen raziskave in izpostavili nekatere smernice pri oblikovanju postopka vrednotenja preizkusov NPZ v prihodnje.

Ključne besede: nacionalno preverjanje znanja, matematika, vrednotenje, preizkusi, veljavnost kontrolnih mehanizmov vrednotenja

Abstract

The following study addresses various issues in assessing the National Assessment of Knowledge (NAK) tests in Mathematics. We examined the validity and efficiency of standardisation and control tests in discriminating between accurate and less accurate assessors. We also focused on the properties of test items associated with assessment accuracy. The results show that both control mechanisms (standardisation and control tests assessment) are relatively effective in discriminating between accurate and less accurate assessors. Closed items are typically more accurately assessed, whereas arithmetic and algebra items, which fall in the red and blue zones and above the blue zone, are more challenging. Items with additional moderated instructions also tend to be more problematic. The article discusses the significance of the study and issues some guidelines for the future design of the assessment process of the NAK tests.

Keywords: National Assessment of Knowledge, mathematics, assessment, tests, validity of control assessment mechanisms

1 Uvod

Nacionalno preverjanje znanja (NPZ) matematike je znatnega pomena za ugotavljanje ravni matematičnega znanja slovenskih učencev v osnovnih šolah. Gre namreč za zunanje preverjanje znanja, ki nudi pomembne informacije o znanju na sistemski ravni. Še pomembnejše pa je za učence, ker jim omogoča primerjavo z vrstniki tudi na nacionalnem nivoju. Tako je nujno, da so preizkusi NPZ objektivni, zanesljivi in veljavni pokazatelji znanja, pomembno pa je tudi, da je postopek vrednotenja objektivni in za vse učence enak. Čeprav je vrednotenje preizkusov pri predmetu matematika samo po sebi razmeroma objektivno (za-

radi prevladujočega števila zaprtih nalog in enoznačnih rešitev), praksa kaže, da ocenjevalci pri vrednotenju niso zmerom poenoteni. Složnost ocenjevalcev pri vrednotenju preizkusov NPZ dosegamo z različnimi kontrolnimi mehanizmi. Ocenjevalci morajo za te namene ovrednotiti preizkuse za vajo in standardizacijo ter kontrolne preizkuse (Ric, 2022a).

Hellrung in Hartig (2013) sta v raziskavi, v katero sta vključila pregled člankov od leta 1991 do 2011, ugotavljala, kako učitelji razumejo in uporabijo povratne informacije, ki jih prejmejo od zunanjih ocenjevanj in kako to vpliva na učne dosežke učencev. Ugotovila sta, da imajo učitelji veliko težav z razumevanjem in uporabo povratnih informacij pri razvijanju in načrtovanju

nadaljnega poučevanja. Tudi Panhoon in Wongwanich (2014) ugotavljata, da učitelji niso uporabili ugotovitev iz analiz ocenjevanja za izboljšanje uspešnosti svojih učencev. Sporočata, da bo poučevanje učencev uspešnejše, če bodo učitelji prejeli učinkovite povratne informacije o ugotovitvah analiz preizkusov znanj in če bodo te ugotovitve prenesli v svoje poučevanje, kar se bo posledično odražalo na dosežkih učencev.

Dokazi iz obsežne kvalitativne študije (Smith, 1991) o vlogi zunanjih preverjanj in ocenjevanj znanja v osnovnih šolah so privedli do domnev o učinkih takih preizkusov znanja na učitelje. Učitelji so v razgovorih povedali, da doživljajo negativna čustva zaradi objave rezultatov zunanjih ocenjevanj in se odločijo, da bodo storili vse, kar je potrebno, da se izognejo nizkim rezultatom. Hellrung in Hartig (2013) omenjata, da so raziskave, ki se osredotočajo na razumevanje in uporabo povratnih informacij z zunanjih preverjanj znanj z namenom izboljšanja dosežkov učencev še kako potrebne.

Vpliv nacionalnega preverjanja znanja matematike na slovenske učitelje matematike v osnovni šoli ni raziskan, kljub temu da Državni izpitni center v Letnih poročilih o izvedbi NPZ v posameznem šolskem letu pojasni namen in cilj kontrolnega vrednotenja ter način analize kakovosti vrednotenja. V anketnih vprašalnikih, ki jih izpolnjujejo ravnatelji ob izvedbi NPZ, se vprašanje nanaša tudi na to, ali so ravnatelji z učitelji opravili analizo kakovosti vrednotenja. Ugotavljamo pa, da podrobnejša raziskava vpliva analize kakovosti vrednotenja na učitelje in njihovo poučevanje (matematike) v Sloveniji še ni bila opravljena.

Zgornjo trditev, da vpliv nacionalnega preverjanja znanja oz. uporaba povratnih informacij na poučevanje učiteljev matematike v slovenskem šolskem prostoru ni raziskana, potrjuje pregled strokovnih in znanstvenih člankov. Z iskanjem in pregledom literature smo ugotovili, da članki opisujejo rezultate raziskovanja vpliva psiholoških značilnosti mladostnika in izobrazbe staršev na učno uspešnost (Marjanovič Umek idr., 2007), druga raziskava pa preučuje kognitivne in družbenoekonomske dejavnike šolskega uspeha v Sloveniji (Klanjšek idr., 2007), stališča staršev do Nacionalnih preverjanj znanj (Škalič in Ivanuš-Grmek, 2017), dejavnike učne uspešnosti ob zaključku šolanja (Marjanovič Umek idr., 2006). O prepričanjih učiteljev o notranjih dejavnikih učne uspešnosti sta Smrtnik Vitulić in Lesar (2014) ugotovili, da so učitelji kot najpomembnejši dejavnik označili sprejemljivost in čustveno stabilnost, Javornik Krečič (2005) opisuje, kako se z leti šolanja spreminjajo značilnosti ocenjevanja znanja in zahteve učitelja pri ocenjevanju.

Še najbolj uporabo povratnih informacij pri poučevanju učiteljev izpostavi raziskava, ki preučuje, ali lahko različni načini spodbujanja samoevalvacije na šoli vplivajo na spreminjanje prepričanj učiteljev o samoevalvaciji, pa tudi na njihovo vedenje v razredu ter na dosežke učencev. Rezultati so pokazali, da so prepričanja precej stabilna, saj je za spremembo potrebno veliko časa. V raziskavi niso uspeli dokazati vpliva spodbujanja procesa samoevalvacije na spreminjanje prepričanj, kar je zelo verjetno posledica omejitve raziskave. Sprememba prepričanj je osnova za spreminjanje vedenja in dosežkov, vendar pa se to zgodi v daljšem časovnem okviru (Hauptman, 2012, stran 19).

1.1 Zagotavljanje kakovosti pri vrednotenju preizkusov NPZ

Vrednotenje preizkusov NPZ poteka elektronsko, s pomočjo programa RM Assessor³ (RM Assessor, 2019). Ocenjevalcem pred vrednotenjem nalog v programu svetujemo, da pred začetkom elektronskega vrednotenja sami rešijo preizkus ter tako dobijo vpogled v naloge, načine reševanja in problematičnost samih nalog ter da se seznani z moderiranimi navodili za vrednotenje. Vsak ocenjevalec mora pred pričetkom rednega vrednotenja ovrednotiti en preizkus za vajo. Namen te je seznanitev ocenjevalcev s preizkusom ter nalogami, predvsem pa z navodili za vrednotenje. Ko zaključijo z vrednotenjem vaje, lahko preverijo, kako natančno so ovrednotili preizkus. Vsaka naloga se namreč preverja z referenčnimi točkami, ki jih nalogam predhodno dodeli predmetna komisija in glavni ocenjevalec predmeta. Ko ocenjevalci zaključijo s pregledom, lahko preidejo na vrednotenje dveh preizkusov za standardizacijo. Namen teh je dodatna umeritev ocenjevalca z navodili za vrednotenje in (posledično) tudi z ostalimi ocenjevalci. Tudi naloge preizkusov za standardizacijo se primerjajo z vnaprej dodeljenimi referenčnimi točkami. Če se dodeljene točke ocenjevalcev bistveno ne razlikujejo od referenčnih točk, program ocenjevalca avtomatsko potrdi za vrednotenje ostalih preizkusov. Če pa ocenjevalec pri vrednotenju preveč odstopa, mora počakati, da njegova preizkusa pregleda pomočnik glavnega ocenjevalca. Ta mu poda povratno informacijo o njegovem vrednotenju in ga nato potrdi za nadaljnje vrednotenje. Naloga pomočnikov glavnega ocenjevalca je, da podrobneje spremljajo ocenjevalce, ki imajo med vrednotenjem več težav, in jim pomagajo z napotki in usmeritvami. Ocenjevalec lahko zatem nadaljuje z vrednotenjem dejanskih preizkusov. Med njimi so naključno razporejeni – ponavadi trije – kontrolni preizkusi. Ti skrbijo za kakovost, tudi ko ocenjevalci vrednotijo dejanske preizkuse NPZ. Če namreč pri posameznem kontrolnem preizkusu preveč odstopajo, se njihovo vrednotenje začasno ustavi. Pred nadaljevanjem morajo pregledati odstopanja pri posameznih nalogah in sprejeti povratno informacijo (RM Assessor, 2019).

1.2 Specifike preverjanja matematike na NPZ

Vrednotenje preizkusov matematike je v primerjavi s preostalimi predmeti NPZ dokaj specifično. V pretežni meri namreč prevladujejo naloge oziroma postavke kratkih zaprtih odgovorov. Med letoma 2018 in 2022 so v preizkusih NPZ matematike v 9. razredu glavnino (79,5 %) predstavljale postavke kratkih zaprtih odgovorov, sledijo postavke kratkih polodprtih odgovorov (20 %). Ostalih tipov postavk je bilo manj kot odstotek. Podobna je situacija pri predmetu matematika v 6. razredu, kjer je nalog s kratkimi zaprtimi odgovori sicer malenkost manj (73 %), sledijo naloge s kratkimi polodprtimi odgovori (21,5 %), ostalih nalog – predvsem izbirnih tipov – pa je nekoliko več kot v 9. razredu (5,5 %). Prevlada nalog zaprtega tipa je načeloma dobrodošla, vsaj kar se tiče vrednotenja preizkusov. Te naloge so objektivnejše, kar pomeni, da je pri njihovem vrednotenju manj dilem in nedoslednosti. To v praksi pomeni tudi večjo poenotenost med ocenjevalci, posledično pa tudi večjo zanesljivost in objektivnost vrednotenja.

Reši neenačbo, če je $\mathbb{U} = \mathbb{N}$.

$$7x - 2(x + 2) - 4x \leq 0$$

$$\mathfrak{R} = \{ \quad \quad \quad \}$$

Slika 1: Primer naloge tipa kratkih zaprtih odgovorov (Ric, 2021d).



Slika 2: Primer naloge tipa kratkih polodprtih odgovorov (Ric, 2021d).

1.3 Opisi dosežkov na nacionalnem preverjanju znanja in druge karakteristike preizkusa iz matematike

Dosežki na preverjanju NPZ (matematike) so ovrednoteni tako kvantitativno (s točkami in odstotki) kot kvalitativno (z umestitvijo v izbrana območja dosežkov). Slednje pripravijo člani predmetnih komisij in sodelavci Rica (Ric, 2022a). Pri tem dosežke vseh učencev razvrstijo od najnižjega do najvišjega in jih umestijo v štiri območja dosežkov: zeleno, rumeno, rdeče in modro.

- Zeleno območje označuje učence, katerih skupni dosežki določajo mejo spodnje četrtine dosežkov (glede na preostale dosežke). V tem območju je 10 % učencev; njihovi dosežki so višji od spodnjih 20 % in nižji od 70 % preostalih dosežkov.
- Rumeno območje označuje učence, katerih skupni dosežki določajo mejo med polovicama dosežkov. V tem območju je 10 % učencev; njihovi dosežki so višji od spodnjih 45 % in hkrati nižji od 45 % preostalih dosežkov.
- Rdeče območje označuje učence, katerih skupni dosežki določajo mejo zgornje četrtine dosežkov. V tem območju je 10 % učencev; njihovi dosežki so višji od spodnjih 70 % in nižji od 20 % preostalih dosežkov.
- Modro območje označuje učence, katerih skupni dosežki so v zgornji desetini dosežkov. V tem območju je 10 % učencev z najvišjimi dosežki; njihovi dosežki so višji od 90 % dosežkov.
- Nalog, ki so v območju nad modrim, v 65 % primerov ne rešijo niti učenci z najvišjimi skupnimi dosežki. (Ric, 2022a, str. 57)

Za vsako od navedenih območij predmetna komisija določi tudi naloge, ki so jih učenci reševali uspešno – uspešnost reševanja pomeni, da je dano nalogo pravilno rešilo vsaj 65 % učencev z dosežki v danem območju (tako določajo uspešnost tudi v mednarodnih raziskavah znanja) (Ric, 2022a, str. 56). Posledično

lahko naloge razvrstimo tudi kvalitativno. V praksi to npr. pomeni, da naloge v modrem območju uspešno rešujejo najboljši učenci. Podrobnejše pojasnilo območij dosežkov najdemo v Letnem poročilu o izvedbi nacionalnega preverjanja znanja (npr. Ric, 2022a).

Posamezna naloga v nacionalnem preizkusu znanja matematike je ovrednotena z največ šestimi točkami. Naloga je razdeljena na več postavk, vsaka postavka pa je ovrednotena z eno točko. V preizkusu znanja so posamezne vsebine zastopane v različnih odstotkih. V 6. razredu je 55 % preizkusa iz aritmetike in algebre, 25 % iz geometrije in merjenja in 20 % iz drugih vsebin (Ric, 2022b). V 9. razredu je aritmetike in algebre 50 %, geometrije in merjenja 35 % in drugih vsebin 15 % (Ric, 2022c).

1.4 Cilji raziskave

Z našo raziskavo smo želeli poiskati odgovore na nekatere dileme in odprta vprašanja, ki se pojavljajo pri vrednotenju NPZ. Prvi cilj raziskave zajema podajo generalne ocene kakovosti vrednotenja pri predmetu matematika. Z raziskavo smo nameravali potrditi veljavnost kontrolnih mehanizmov standardizacije in vrednotenja kontrolnih preizkusov pri ločevanju bolj in manj natančnih ocenjevalcev preizkusov NPZ matematike. Naslednji cilj zajema prepoznavo nalog, ki jih ocenjevalci vrednotijo manj natančno. Prav tako smo želeli identificirati skupne lastnosti teh nalog (npr. tip odgovorov). Odgovore na zastavljena vprašanja smo poiskali s pomočjo ponovnega ovrednotenja preizkusov izbranih ocenjevalcev, kar nam je omogočilo izračun natančnosti vrednotenja le-teh ter primerjavo s standardizacijo in vrednotenjem kontrolnih preizkusov.

1.5 Raziskovalna vprašanja

V skladu s cilji raziskave smo želeli odgovoriti na več raziskovalnih vprašanj. Eno bolj generalnih se dotika splošne kakovosti vrednotenja preizkusov. Kako natančno v povprečju preizkuse ovrednoti posamezen ocenjevalec? Naslednja dilema se povezuje z vrednotenjem preizkusov za standardizacijo in kontrolnih preizkusov. Vprašanje, ki se pojavlja, je veljavnost omenjenih mehanizmov za ločevanje med natančnimi in manj natančnimi ocenjevalci. Se ocenjevalci, ki naredijo več napak pri vrednotenju preizkusov standardizacije oz. kontrolnih preizkusov, slabše odrežejo tudi pri vrednotenju preostalih preizkusov? Ali torej s kontrolnimi mehanizmi zares detektiramo ocenjevalce, ki so pri vrednotenju NPZ manj natančni? Je pri tem učinkovitejši postopek standardizacije ali vrednotenja kontrolnih preizkusov? Zanimajo nas tudi specifične naloge, ki ocenjevalcem pri vrednotenju povzročajo več preglavic. So to predvsem naloge polodprtega tipa, katerih vrednotenje je načeloma nekoliko subjektivnejše? Gre morebiti za zahtevnejše naloge? Lahko pri nalogah, ki jih ocenjevalci ovrednotijo manj natančno, zapazimo katere druge značilnosti?

2 Metodologija

2.1 Raziskovalni pristop in postopek

V raziskavi smo uporabili prečni pristop zbiranja podatkov in kvantitativne analize njihove obdelave. Sprva smo izračunali

povprečna odstopanja vrednotenj ocenjevalcev od referenčnih vrednotenj pri standardizaciji in vrednotenju kontrolnih preizkusov. To je bilo izhodišče za razvrstitev ocenjevalcev v skupine glede na njihovo natančnost. Pri obeh kontrolnih mehanizmi smo ocenjevalce razdelili v pet skupin. V nadaljnje analize smo zajeli ocenjevalce iz skupin najnižjih, srednjih in najvišjih odstopanj. Ocenjevalce smo torej razdelili v 3 skupine standardizacije in 3 skupine kontrolnih preizkusov. Po načrtu 3×3 smo oblikovali 9 kombinacij skupin. V vsaki kombinaciji skupin smo izbrali 9 do 10 naključnih ocenjevalcev. Za vsakega od izbranih ocenjevalcev pa smo določili 9 naključnih ovrednotenih preizkusov NPZ iz leta 2021. Te so ponovno ovrednotili člani Predmetne komisije za matematiko (postopek je bil sicer del obsežnejšega projekta – vključujočega tudi druge predmete NPZ in njih komisije – podrobneje opisanega v študiji Janet in Cankar (2023)). Točkam ponovno ovrednotenih nalog pravimo referenčne. Na podlagi teh smo lahko izračunali odstopanja med vrednotenji ocenjevalcev in referenčnimi vrednotenji. Za oblikovane skupine ocenjevalcev smo izračunali dva indeksa odstopanj: povprečno kumulativno odstopanje ocenjevalcev in povprečno skupno odstopanje ocenjevalcev. Prvo zajema povprečni seštevek vseh absolutnih odstopanj po nalogah, povprečno skupno odstopanje pa zajema zgolj povprečna absolutna odstopanja na koncu preizkusov. Dva indeksa odstopanj smo uporabili za namene temeljiteše detekcije morebitnih razlik med skupinami.

2.2 Sodelujoči v raziskavi

Pri ponovnem vrednotenju preizkusov NPZ matematike v letu 2021 v 9. razredu je sodelovalo 7 ocenjevalcev, skupaj pa so ponovno ovrednotili 729 preizkusov. Pri ponovnem vrednotenju preizkusov NPZ v 6. razredu za leto 2021 pa je sodelovalo 6 ocenjevalcev, ki so skupaj ovrednotili 810 preizkusov. Ocenjevalci so bili člani predmetne komisije za matematiko.

2.3 Obdelava podatkov

Statistične analize smo izvedli s programom R-studio (verzija 1.3.1093). Sprva smo izračunali opisne statistike odstopanj posameznih skupin. Statistično značilnost razlik v odstopanjih med skupinami standardizacije oziroma kontrolnih preizkusov smo preverili s pomočjo enosmerne ANOVE. Razlike med posameznimi pari skupin pa smo preverili s Tukeyevimi post hoc testi. Za vsako nalogo smo izračunali delež, v katerem je bila ta pravilno ovrednotena. Prav tako smo izračunali povprečne deleže natančnosti vrednotenja nalog za posamezne tipe.

3 Rezultati

3.1 Razlike med kombinacijami skupin standardizacije in kontrolnih preizkusov

Sprva si oglejmo razlike v indeksih odstopanj med kombinacijami skupin standardizacije in kontrolnih preizkusov. Preglednica 1 prikazuje povprečna kumulativna in skupna odstopanja ocenjevalcev v 9. razredu. Pri tem se prva številka oznake skupine naša na skupino standardizacije, druga pa na skupino kontrolnih

preizkusov. Skupina 1 zajema ocenjevalce z najnižjimi odstopanji, skupina 2 ocenjevalce z zmernimi odstopanji, skupina 3 pa ocenjevalce z najvišjimi odstopanji pri obeh kontrolnih mehanizmi. V preglednici pa so seveda prikazana odstopanja skupin pri vrednotenju dejanskih preizkusov NPZ (tistih, ki smo jih zajeli v vzorec). Iz preglednice gre razbrati trend, da ocenjevalci v skupinah, ki so bolj odstopale pri vrednotenju preizkusov za standardizacijo in kontrolnih preizkusov, bolj odstopajo tudi pri vrednotenju dejanskih preizkusov. Rezultati enosmerne ANOVE sicer kažejo, da so razlike med skupinami statistično značilne zgolj v povprečnih kumulativnih odstopanjih ($F(8, 72) = 2,45$, $p = 0,021$), ne pa v povprečnih skupnih odstopanjih ($F(8, 72) = 1,27$, $p = 0,272$).

Preglednica 1: Indeksi povprečnih kumulativnih in skupnih odstopanj kombinacij skupin standardizacije in kontrolnih preizkusov (9. razred).

Skupina	Povprečno kumulativno odstopanje	Povprečno skupno odstopanje
1.1	1,22	0,85
1.2	1,37	0,95
1.3	1,53	0,86
2.1	1,53	1,01
2.2	1,69	1,00
2.3	1,83	1,16
3.1	1,60	0,96
3.2	2,04	1,20
3.3	2,04	1,17

Za preverbo razlik med posameznimi pari skupin smo izvedli Tukeyeve post-hoc teste. Ti kažejo, da značilne razlike beležimo zgolj v povprečnih kumulativnih odstopanjih med skupinama 1.1 in 3.2 ($p = 0,046$) ter 1.1 in 3.3 ($p = 0,046$). Ostale razlike niso statistično značilne.

Postopek smo ponovili pri preverbi razlik v odstopanjih kombinacij skupin standardizacije in kontrolnih preizkusov v 6. razredu (preglednica 2). Spet višja odstopanja najdemo pri kombinacijah skupin, katerih ocenjevalci so pri vrednotenju preizkusov obeh kontrolnih mehanizmov bolj odstopali. Značilne razlike med skupinami smo zabeležili zgolj v povprečnih kumulativnih odstopanjih ($F(8, 81) = 3,614$, $p = 0,001$), ne pa v povprečnih skupnih odstopanjih ($F(8, 81) = 1,873$, $p = 0,076$) – slednjo razliko lahko sicer opredelimo kot mejno statistično značilno. Tukeyevi post hoc testi razkrivajo značilne razlike v povprečnih kumulativnih odstopanjih med skupinama 2.1 in 1.3 ($p = 0,004$). Kot mejno značilne lahko morebiti opredelimo še razlike med pari skupin 1.1 in 1.3 ($p = 0,055$), 1.2 in 1.3 ($p = 0,055$), 2.1 in 2.3 ($p = 0,080$), 2.1 in 3.2 ($p = 0,074$) ter 2.1 in 3.3 ($p = 0,051$). Statistično značilno razliko beležimo tudi v povprečnih skupnih odstopanjih skupin 2.1 in 3.3 ($p = 0,043$).

Preglednica 2: Indeksi povprečnih kumulativnih in skupnih odstopanj kombinacij skupin standardizacije in kontrolnih preizkusov (6. razred).

Skupina	Povprečno kumulativno odstopanje	Povprečno skupno odstopanje
1.1	1,57	1,14
1.2	1,57	1,01
1.3	2,80	1,24
2.1	1,23	0,81
2.2	2,34	1,30
2.3	2,41	1,39
3.1	2,18	1,18
3.2	2,42	1,24
3.3	2,48	1,48

3.2 Razlike med skupinami standardizacije

V nadaljevanju smo združili skupine standardizacije in kontrolnih preizkusov ter po enakem postopku preverili značilnost razlik v povprečnih kumulativnih oziroma povprečnih skupnih odstopanjih. V preglednici 3 so prikazane statistike odstopanj pri standardizaciji za ocenjevalce matematike v 9. razredu. Najvišja odstopanja beležimo v skupini 3 (pri ocenjevalcih, ki so pri standardizaciji najbolj odstopali), sledi skupina 2, najnižja odstopanja pa beležimo v skupini 1 (pri ocenjevalcih, ki so pri standardizaciji odstopali najmanj). To velja za povprečna kumulativna in skupna odstopanja. Preverili smo, ali so razlike med skupinami statistično značilne. To smo ugotovili za povprečna kumulativna odstopanja ($F(2, 78) = 7,514, p = 0,003$), razlike v povprečnih skupnih odstopanjih pa mejijo na statistično značilne ($F(2, 78) = 3,098, p = 0,051$).

Izvedli smo še Tukeyeve post hoc teste. Ti kažejo, da značilne razlike v povprečnih kumulativnih odstopanjih najdemo med 1. in 3. skupino ($p = 0,002$), razlika med 1. in 2. skupino pa meji na statistično značilno ($p = 0,095$). Tudi za razliko med 1. in 3. skupino v povprečnih skupnih odstopanjih lahko trdimo, da meji na statistično značilno ($p = 0,051$).

Preglednica 3: Indeksi povprečnih kumulativnih in skupnih odstopanj skupin standardizacije (9. razred).

Skupina	Povprečno kumulativno odstopanje	Povprečno skupno odstopanje
1	1,37	0,89
2	1,68	1,06
3	1,89	1,11

Preglednica 4 prikazuje statistike odstopanj pri standardizaciji za ocenjevalce matematike v 6. razredu. Spet beležimo zgoraj opisa-

ni trend: najvišja odstopanja najdemo v skupini 3, sledi skupina 2, najnižja odstopanja pa so v skupini 1. Z enosmerno ANOVO smo preverili statistično značilnost razlik med skupinami, pri čemer značilnih razlik v povprečnih kumulativnih odstopanjih nismo ugotovili ($F(2, 87) = 1,481, p = 0,233$), podobno velja za povprečna skupna odstopanja ($F(2, 87) = 1,002, p = 0,368$).

Preglednica 4: Indeksi povprečnih kumulativnih in skupnih odstopanj skupin standardizacije (6. razred).

Skupina	Povprečno kumulativno odstopanje	Povprečno skupno odstopanje
1	1,98	1,13
2	2,00	1,17
3	2,36	1,30

3.3 Razlike med skupinami kontrolnih preizkusov

Enak postopek smo ponovili pri skupinah kontrolnih preizkusov. Preglednica 5 tako prikazuje odstopanja skupin kontrolnih preizkusov za ocenjevalce matematike v 9. razredu. Iz preglednice lahko razberemo, da so odstopanja najvišja v skupini 3 (pri ocenjevalcih, ki so pri vrednotenju kontrolnih preizkusov najbolj odstopali), sledi skupina 2, medtem ko so odstopanja najnižja v skupini 1 (pri ocenjevalcih, ki so pri vrednotenju kontrolnih preizkusov odstopali najmanj). Rezultati enosmerne ANOVE razkrivajo, da so razlike v povprečnih kumulativnih odstopanjih mejno statistično značilne ($F(2, 78) = 2,724, p = 0,072$), razlike v povprečnih skupnih odstopanjih pa ne dosegajo meje statistične značilnosti ($F(2, 78) = 0,983, p = 0,379$).

Tukeyeve post hoc teste smo izvedli zgolj za preverbo razlik v povprečnih kumulativnih odstopanjih. Med 1. in 3. skupino smo zabeležili mejno značilno razliko ($p = 0,067$), razlike med ostalimi skupinami pa ne dosegajo statistične značilnosti.

Preglednica 5: Indeksi povprečnih kumulativnih in skupnih odstopanj skupin kontrolnih preizkusov (9. razred).

Skupina	Povprečno kumulativno odstopanje	Povprečno skupno odstopanje
1	1,45	0,94
2	1,70	1,05
3	1,80	1,07

Postopek smo ponovili še pri skupinah kontrolnih preizkusov ocenjevalcev matematike v 6. razredu. Tudi pri teh zapažamo trend, da najvišja odstopanja beležimo v skupini 3, sledi skupina 2, najnižja odstopanja pa beležimo v skupini 1. Rezultati enosmerne ANOVE indicirajo, da so razlike med skupinami v povprečnih kumulativnih odstopanjih statistično značilne

($F(2, 87) = 7,394$, $p = 0,001$). Podobno lahko zatrdimo za povprečna skupna odstopanja ($F(2, 87) = 3,725$, $p = 0,028$).

Izvedli smo še serijo Tukeyevih post hoc testov. Rezultati teh kažejo, da značilne razlike v povprečnih kumulativnih odstopanjih beležimo le med skupinama 1 in 3 ($p < 0,001$). Tudi razlike v povprečnih skupnih odstopanjih so statistično značilne le med 1. in 3. skupino ($p = 0,021$).

Preglednica 6: Indeksi povprečnih kumulativnih in skupnih odstopanj skupin kontrolnih preizkusov (6. razred).

Skupina	Povprečno kumulativno odstopanje	Povprečno skupno odstopanje
1	1,66	1,04
2	2,11	1,19
3	2,56	1,37

3.4 Analiza po postavkah

Izvedli smo tudi analizo po postavkah. Zanimalo nas je, katere naloge oz. postavke v 6. in 9. razredu so ocenjevalci ovrednotili najbolj skladno z referenčnim vrednotenjem in pri katerih so najbolj odstopali. Za vsako postavko smo izračunali seštevek odstopanj oz. razlik glede na referenčno vrednotenje (maksimalni možen seštevek odstopanj pri vrednotenju NPZ matematike v 9. razredu je tako 729, v 6. razredu pa 810). Števila odstopanj pri posameznih postavkah smo pretvorili v odstotek odstopanj.

Oblikovali smo kriterij za kategorizacijo natančnosti vrednotenja postavk. Med zelo skladne smo uvrstili postavke, pri katerih je odstotek odstopanj nižji od 0,50 %.

Med skladne postavke smo uvrstili tiste, pri katerih je odstopanje med 0,50 % in 3,00 %. Med srednje skladne postavke smo uvrstili postavke z odstotki odstopanj med 3,01 % in 7,00 %. Med neskladne postavke smo uvrstili postavke z odstopanji med 7,01 % in 9,99 %, med zelo neskladne postavke pa tiste z odstopanji nad 10 %.

Preglednica 7: Število postavk glede na odstopanje od referenčnega vrednotenja na preizkusih NPZ matematike v 6. in 9. razredu.

	Št. postavk 6. razred	Št. postavk 9. razred
Zelo skladno vrednotenje	8	5
Skladno vrednotenje	20	28
Srednje skladno vrednotenje	11	11
Neskladno vrednotenje	4	4
Zelo neskladno vrednotenje	7	2

3.4.1 Vrednotenje nalog v 6. razredu

Iz preglednice 8 razberemo, da so razlike v skladnosti vrednotenja posameznih postavk v 6. razredu velike in so v razponu od 0,12 % do 19,88 %.

V letnem poročilu o izvedbi nacionalnega preverjanja znanja v šolskem letu 2020/2021 smo iz specifikacijske tabele razbrali nekatere podatke (področje, taksonomsko stopnjo, indeks težavnosti in območje), ki smo jih predstavili v preglednici 8. Natančnejši pregled petih postavk v 6. razredu, ki so vrednotene zelo

Preglednica 8: Najbolj skladno in neskladno ovrednotene postavke (NPZ matematika, 6. razred).

	Postavka	Odstotek odstopanj	Področje	Taksonomska stopnja	Območje	Indeks težavnosti (IT)
Zelo skladno vrednotenje	5.a	0,12	aritmetika in algebra	I	zeleno	0,93
	1.d	0,37	aritmetika in algebra	III	modro	0,31
	9.c	0,37	druge vsebine	IV	modro	0,26
	4.b.2	0,49	aritmetika in algebra	II	zeleno	0,94
	9.a.1	0,49	aritmetika in algebra	II	zeleno	0,87
Zelo neskladno vrednotenje	6.a.2	10,00	druge vsebine	II	nad modrim	0,15
	7.b.1	10,37	druge vsebine	III	rdeče	0,49
	7.c.1	10,49	aritmetika in algebra	III	modro	0,28
	8.c	10,99	druge vsebine	IV	nad modrim	0,16
	8.a.2	11,11	druge vsebine	III	rdeče	0,48
	5.b.1	13,83	aritmetika in algebra	I	rumeno	0,69
	9.a.2	19,88	aritmetika in algebra	II	zeleno	0,63

skladno z referenčnim vrednotenjem, nam pove, da postavke pokrivajo vse taksonomske stopnje, prevladujejo pa postavke (tri postavke) I. in II. taksonomske stopnje (poznavanje in razumevanje pojmov in dejstev ter izvajanje rutinskih postopkov). Glede na področja učnega načrta ugotavljamo, da je večina postavk (4) s področja aritmetike in algebre, ena pa iz drugih vsebin, pri čemer področja v preizkusih niso zastopana v enakih odstotkih. Postavke so iz dveh območij, največ (3) iz zelenega in dve iz modrega območja. Povprečni dosežek postavk (indeks težavnosti) pa je od 0,26 do 0,94.

V 6. razredu so bile najbolj ovrednotene postavke 5.a, 1.d in 9.c. Zahtevale so odgovor, ki je bil zapisan v moderiranih navodilih za vrednotenje (Slika 3 in Slika 4). Ocenjevalci so pri vrednote-

5. a) Kolikšen del lika je osenčen? Obkroži ustrezni ulomek.



$$\frac{4}{5}$$

$$\frac{5}{8}$$

$$\frac{5}{9}$$

$$\frac{4}{9}$$

$$\frac{6}{10}$$

5.a

1

• obkrožen samo ulomek $\frac{5}{8}$

Slika 4: Naloga 5.a in moderirana navodila za vrednotenje, NPZ 6. razred, 2021.

nju preverili, ali je učenec zapisal dani odgovor ali ne, drugih možnosti ni bilo.

Podroben pregled sedmih postavk, kjer je vrednotenje najbolj odstopalo, nam pove, da postavke uvrščamo v vse štiri taksonomske stopnje, največ (3) pa jih je s III. taksonomske stopnje (uporaba kompleksnih postopkov). Večina postavk (4) je s področja druge vsebine, ostale (3) so s področja števil. Postavke so iz vseh območij, sta pa dve iz območja nad modrim, ena modro

1. d) Dana so števila:

0,9 0,10 0,011 0,12

Zapiši jih po velikosti od najmanjšega do največjega.

_____ < _____ < _____ < _____

1.d

1

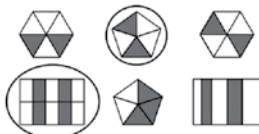
• $0,011 < 0,10 < 0,12 < 0,9$

Slika 3: Naloga 1.d in moderirana navodila za vrednotenje, NPZ 6. razred, 2021.

5. b) Obkroži vsak lik, katerega $\frac{2}{5}$ sta osenčeni.



5.b



5.b.1

1

• obkrožen eden izmed pravih odgovorov in noben napačen

5.b.2

1

• obkrožen še drugi pravih odgovor in noben napačen

Učenec dobi le točko 5.b.1, če obkroži oba pravilna odgovora in še enega napačnega. Točk ne dobi:
– če obkroži vsaj dva napačna odgovora,
– če sta obkrožena en pravih in en napačen odgovor.

Slika 5: Naloga 5.b in moderirana navodila za vrednotenje, NPZ 6. razred, 2021.

in dve rdeče, po ena iz rumenega in zelenega. Povprečni dosežek (IT) je od 0,15 do 0,69.

Največje odstopanje pri vrednotenju je bilo opaženo pri postavkah 9.a.2 in 5.b.1. Po moderiranih navodilih so morali ocenjevalci točko 5.b.1 dodeliti, če je učenec obkrožil enega izmed pravih odgovorov in nobenega napačnega, hkrati pa je ocenjevalec moral upoštevati še dodatna navodila (Slika 5). Ocenjevalci morajo točke niso dodelili, če je bil obkrožen le lik v drugi vrstici ali pa so spregledali dodatna navodila.

Točko pri postavki 9.a.2 so ocenjevalci dodelili le, če je bil upoštevan vrstni red računskih operacij (Slika 6). Lahko so spregledali, da je učenec nakazal deljenje kot stranski račun ali pa so

9. a) Izračuna vrednosti številskih izrazov.
 $0,16 : 0,2 + 2 =$

9.a.2	1	* upoštevan vrstni red računskih operacij (iz postopka je razvidno, da je učenec najprej delil $0,16$ z $0,2$)	
9.a.3	1	* 2,8	V izrazu morajo veljati vse enakosti.

Slika 6: Naloga 9.a in moderirana navodila za vrednotenje, NPZ 6. razred, 2021.

menili, da točko dobi le, če je tudi rezultat deljenja pravilen. Točko učenec dobi, če nakaže, da ima deljenje prednost, količnik pa je lahko napačen.

3.4.2 Vrednotenje nalog v 9. razredu

Iz preglednice 9 razberemo, da so razlike v skladnosti vrednotenja posameznih postavk v 9. razredu velike, od 0,14 % do 17,56 %.

Natančnejši pregled petih postavk v 9. razredu, ki so vrednotene zelo skladno z referenčnim vrednotenjem, nam pove, da postavke pokrivajo vse taksonomske stopnje, prevladujejo pa postavke

(2) IV. taksonomske stopnje (reševanje in raziskovanje problemov). Glede na področja učnega načrta ugotavljamo, da sta dve postavki s področja aritmetike in algebre, dve iz drugih vsebin, ena iz geometrije, pri čemer področja v preizkusih niso zastopana v enakih odstotkih. Zelo skladno vrednotene postavke so iz treh območij, največ (3) iz modrega območja, ena iz rdečega in ena iz območja nad modrim. Povprečni dosežek postavk (indeks težavnosti) pa je od 0,06 do 0,72.

2. Eva je stara pet let in živi skupaj z očetom in babico. Oče je star 26 let, babica pa 57 let.

2. d) Čez koliko let bo vsota njihovih starosti enaka 100 let?

Odgovor: _____

2.d	1	* Čez štiri leta bo vsota njihovih starosti enaka 100.
-----	---	--

Slika 7: Naloga 2.d in moderirana navodila za vrednotenje, NPZ 9. razred, 2021.

V 9. razredu so bile najbolj ovrednotene postavke 2.d, 5.b in 8.a.1. Zahtevale so odgovor, ki je bil zapisan v moderiranih navodilih za vrednotenje (Slika 7 in Slika 8). Ocenjevalci so pri vrednotenju preverili, ali je učenec zapisal tak odgovor ali ne.

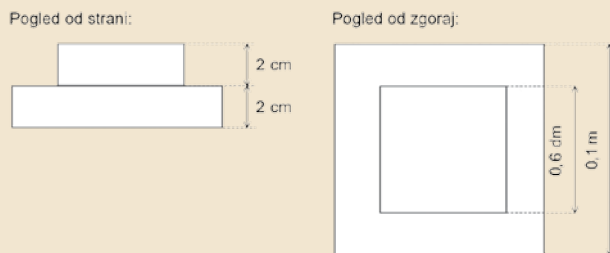
Podroben pregled dveh postavk v 9. razredu, kjer je vrednotenje najbolj odstopalo, nam pove, da postavki uvrščamo na I. in IV. taksonomsko stopnjo. Večina postavk (4) je s področja druge vsebine, ostale (3) so s področja števil. Postavki sta iz rdečega in modrega območja. Povprečni dosežek (IT) je 0,31 in 0,51.

Največje odstopanje pri vrednotenju je bilo opaženo pri postavkah 4.a.1 in 7.c. Po moderiranih navodilih so morali ocenjevalci točko 4.a.1 dodeliti, če je učenec vstavil vrednosti spremenljivk v dani oz. ekvivalentni izraz, pri čemer so morali upoštevati tudi dodatna navodila, ki so narekovala, da učenec dobi točko tudi,

Preglednica 9: Najbolj skladno in neskladno ovrednotene postavke (NPZ matematika, 9. razred).

	Postavka	Odstotek odstopanj	Področje	Taksonomska stopnja	Območje	(IT) indeks težavnosti
Zelo skladno vrednotenje	2.d	0,14	Druge vsebine	IV	Modro	0,43
	5.b	0,14	Aritmetika in algebra	III	Modro	0,24
	8.a.1	0,14	Geometrija in merjenje	I	Nad modrim	0,06
	4.c.1	0,41	Aritmetika in algebra	II	Rumeno	0,72
	9.e	0,41	Druge vsebine	IV	modro	0,26
Zelo neskladno vrednotenje	4.a.1	15,50	Aritmetika in algebra	I	rdeče	0,51
	7.c	17,56	Druge vsebine	IV	modro	0,31

5. Marjan je iz lesa izdelal dve enako visoki pravilni štiristrani prizmi, ki ju je postavil eno na drugo. Narisal je, kako se postavitve vidi od strani in kako od zgoraj, ter zapisal podatke.



5. b) Koliko odstotkov zgornje ploskve večje prizme ne pokrije spodnja ploskev manjše prizme?

Spodnja ploskev manjše prizme ne pokrije _____ %
zgornje ploskve večje prizme.

5.b	1	* 64 (%)
-----	---	----------

Slika 8: Naloga 5.b in moderirana navodila za vrednotenje, NPZ 9. razred, 2021.

4. a) Izračunaj vrednost izraza $x + 2 - (5x + 4y) \cdot xy$, če sta vrednosti spremenljivk $x = 0$ in $y = -10$.
Reševanje:

4.a	4.a.1	1	* vstavljene vrednosti spremenljivk x in y v dani oziroma ekvivalentni izraz	Učenec dobi točko 4.a.1, če se iz njegovega zapisa vidi, da sta vrednosti za x in y pravilno vstavljeni, npr.: $0 + 2 - (5 \cdot 0 + 4 \cdot (-10)) \cdot 0 \cdot (-10)$ oz. $0 + 2 - (0 - 40) \cdot 0$
-----	-------	---	--	---

Slika 9: Naloga 4.a in moderirana navodila za vrednotenje, NPZ 9. razred, 2021.

če se iz učenčevega zapisa vidi, da so vrednosti spremenljivk pravilno vstavljene (Slika 9). Ocenjevalci so lahko pri vrednotenju spregledali dodatno navodilo.

V postavki 7.c ocenjevalec ovrednoti pravilno oz. smiselno utemeljitev (Slika 10). Vseh utemeljitev v moderiranih navodilih ni zapisanih, kar je predstavljalo dodatno težavo pri vrednotenju, saj se je moral ocenjevalec odločiti, ali je utemeljitev matematično sprejemljiva ali ne. Hkrati je moral biti ocenjevalec pozoren na dodatna navodila. Različne možnosti ustreznih in neustreznih utemeljitev so predstavili Bone idr. (2021). Pri tej nalogi utemeljevanja ugotavljamo, da so jo učenci slabše reševali, hkrati pa tudi učitelji slabše vrednotili. Večjo skrb in pozornost je treba nameniti postopkom utemeljevanja, z ustreznimi didaktičnimi pristopi in strategijami (Doz, 2023; Bone idr., 2021).

7. Na kmetiji so nabrali 0,75 tone jabolk.

7. c) Ali bi lahko z vsemi nabranimi jabolki napolnili zaboje, da bi bilo v vsakem po 18 kg jabolk? Utemelji.

Utemeljitev:

7.c	1	* Smiselna utemeljitev. Npr.: – ne, saj 750 ni deljivo z 18, – ne, utemeljeno z računom $750 : 18 = 41,6$ in zapisom, da količnik ni naravno število oz., da se deljenje ne izide.	Oziroma glede na 7.a.1. Učenec točke 7.c ne dobi, če je v odgovoru zapisano število jabolk (npr.: ostane 12 jabolk).
-----	---	--	---

Slika 10: Naloga 7.c in moderirana navodila za vrednotenje, NPZ 9. razred, 2021.

Zaključki

V raziskavi ugotavljamo razmeroma dobro generalno kakovost vrednotenja preizkusov matematike. Povprečna odstopanja različnih skupin ocenjevalcev NPZ matematike v 6. in 9. razredu so namreč razmeroma nizka. Eno temeljnih vprašanj naše raziskave se je dotikalo veljavnosti kontrolnih mehanizmov standardizacije in vrednotenja kontrolnih preizkusov pri vrednotenju NPZ matematike. Rezultati kažejo, da sta mehanizma uporabna pri ločevanju natančnih in manj natančnih ocenjevalcev. Rezultati odstopanj kombinacij skupin standardizacije in kontrolnega vrednotenja kažejo trend, da ocenjevalci, ki pri obeh kontrolnih mehanizmih bolj odstopajo, bolj odstopajo tudi pri vrednotenju dejanskih preizkusov NPZ matematike. Mehanizma sta torej koristna in veljavna za detektiranje manj natančnih ocenjevalcev.

Zanimiv je trend, da četudi smo razlike med kombinacijami skupin zabeležili pri obeh indeksih odstopanj, so statistično značilne predvsem razlike v povprečnih kumulativnih odstopanjih, ne pa v povprečnih skupnih odstopanjih. Razlog gre morebiti iskati v dejstvu, da so naloge matematike predvsem objektivnega tipa, zato se razlike nemara izkažejo v bolj natančni, podrobni meri odstopanj (kumulativnih odstopanjih), ne pa v bolj celostni meri (skupnih odstopanjih). Postavke matematike so pogostoma tudi soodvisne – vrednotijo se

namreč ločeno glede na postopek, izračun, odgovor ipd. Tudi to se lahko odraža v zabeleženih kumulativnih odstopanjih.

Rezultati analiz, ločeni glede na standardizacijo in vrednotenje kontrolnih preizkusov, ne dajo jasnega odgovora, kateri mehanizem je učinkovitejši pri detekciji boljših oziroma slabših ocenjevalcev. Pri vrednotenju NPZ matematike v 9. razredu se namreč za koristnejše izkazuje preizkusi za standardizacijo, pri vrednotenju v 6. razredu pa kontrolni preizkusi. Pri obeh mehanizmih se sicer izkazuje jasen trend, da ocenjevalci, ki bolj odstopajo pri vrednotenju standardizacijskih oz. kontrolnih preizkusov, bolj odstopajo tudi pri vrednotenju dejanskih preizkusov. Rezultati torej implicirajo, da sta oba mehanizma veljavna in uporabna ter ju velja obdržati v postopku vrednotenja preizkusov NPZ tudi v prihodnje. Mehanizma imata tudi nekoliko različno funkcijo, standardizacija skrbi predvsem za umeritev ocenjevalcev z navodili za vrednotenje, kontrolni preizkusi pa so namenjeni doseganju zadovoljive zanesljivosti med samim vrednotenjem.

Naloge oz. postavke, ki so vrednotene skladno z referenčnim vrednotenjem in naloge, ki zelo odstopajo od referenčnega vrednotenja, so različnih taksonomskih stopenj, prevladujejo naloge iz aritmetike in algebre ter drugih vsebin, po večini pa spadajo v rdeče in modro območje ter v območje nad modrim. Indeks težavnosti teh nalog je raznolik. Nismo zaznali bistvenih odstopanj pri enih ali drugih nalogah. Naloge, ki so bolj vrednotene, so naloge zaprtega tipa, kjer ocenjevalci upoštevajo le odgovor oz. rešitev, ki je napisana tudi v moderiranih navodilih. Pri nalogah, kjer je treba upoštevati še dodatna navodila, vrednotenje bolj odstopa, a to še ne pomeni, da takih nalog v naslednjih nacionalnih preizkusih znanja matematike ne bi bilo več smiselno dajati. Pozornost je treba usmeriti v sestavo kakovostnih nalog in zapis navodil za vrednotenje. Ozaveščanje ocenjevalcev o doslednem upoštevanju moderiranih navodil za vrednotenje je ključno pri izboljševanju vrednotenja.

Analiza vrednotenj po nalogah oz. postavkah nam nakazuje, da je v 6. razredu več nalog, kjer je zaznati odstopanja, višja od 10 %. Razlog gre morda iskati v tem, da preizkuse za 6. razred vrednotijo tudi učitelji in učiteljice razrednega pouka. Za namen izboljšanja vrednotenja Predmetna komisija za matematiko pri NPZ izvede izobraževanja, kjer se ocenjevalci seznanijo z načinom, pomenom in postopki vrednotenja ... Iz rezultatov te raziskave lahko sklepamo, da je izobraževanje ocenjevalcev potrebno, saj bo pripomoglo k izboljšanju vrednotenja.

Pričujoča raziskava ponuja obilo relevantnih in uporabnih ugotovitev. V pretežni meri smo namreč potrdili veljavnost kontrolnih mehanizmov pri vrednotenju preizkusov NPZ matematike. Opravili smo analizo nalog/postavk in izpostavili značilnosti teh, ki so bolj oz. manj problematične za vrednotenje. Ugotovitve prav tako ponujajo aplikativna vodila za nadaljnje oblikovanje postopka vrednotenja preizkusov NPZ matematike. Analiza po nalogah daje uporabne usmeritve za oblikovanje nalog, ki so manj težavne pri samem vrednotenju. Naloge, ki so se izkazale težje za vrednotenje (npr. naloge utemeljevanja), naj učitelji uvajajo v proces poučevanja (od preverjanja do ocenjevanja). Učenci bodo pridobili znanja in veščine, kako take naloge reševati, učitelji pa, kako podati učencu povratno informacijo in jih ovrednotiti.

Raziskava ima tudi nekaj omejitev. Je prečnega raziskovalnega pristopa in zajema zgolj podatke enega leta. Vzorec vključuje zgolj del ocenjevalcev NPZ matematike in le del njihovih ovrednotenih preizkusov. Raziskava tudi ni bila kvalitativna. Vse naštet omejuje reprezentativnost študije. Raziskavo bi bilo smiselno ponoviti, tako bi lahko namreč opazovali trende pri vrednotenju. Prihodnje študije bi lahko natančneje – z rabo eksperimentalne metode – preučile učinkovitost kontrolnih mehanizmov (npr. v kolikšni meri se natančnost vrednotenja zaradi njih izboljša). Podobno naj prihodnje študije z drugimi raziskovalnimi pristopi naslovijo tudi dognanja, ki se tičejo karakteristik nalog in navodil za vrednotenje, ki se povezujejo z natančnostjo ocenjevalcev pri njihovem vrednotenju.

Viri in literatura

- Bone, J., Cotič, M. in Felda, D. (2021). Utemeljevanje pri pouku matematike. *Pedagoška Obzorja*, 36(1), 33–52.
- Doz, D. (2023). Mathematical reasoning: what are the issues? V A. Žakelj, M. Cotič, D. Kadijević in A. Lipovec (Ur.), *Selected topics in the didactics of mathematics* (str. 107–125). Založba Univerze na Primorskem.
- Hauptman, A. (2012). Spreminjanje prepričanj učiteljev o samoevalvaciji šole ter vpliv na vedenje učiteljev in dosežke učencev. *Psihološka obzorja*, 21(2), 19–28.
- Hellrung, K. in Hartig, J. (2013). Understanding and using feedback – A review of empirical studies concerning feedback from external evaluations to teachers. *Educational Research Review*, 9, 174–190.
- Janet, D. in Cankar, G. (2023). *Zagotavljanje kakovosti pri vrednotenju preizkusov nacionalnega preverjanja znanja* [Rokopis oddan za objavo]. Državni izpitni center.

- Javornik Krečič, M. (2005). Ocenjevanje znanja z vidika izsledkov empirične raziskave: kako se z leti šolanja spreminjajo značilnosti ocenjevanja znanja in zahteve učitelja pri ocenjevanju. *Vzgoja in izobraževanje*, 36(4/5), 73–81.
- Klanjšek, R., Flere, S. in Lavrič, M. (2007). Kognitivni in družbenoekonomski dejavniki šolskega uspeha v Sloveniji. *Družboslovne razprave*, 23(55), 49–69.
- Marjanovič Umek, L., Sočan, G. in Grgič, K. (2006). Šolska ocena: koliko jo lahko pojasnimo z individualnimi značilnostmi mladostnika in koliko z dejavniki družinskega okolja. *Psihološka obzorja*, 4(15), 25–52.
- Ljubica Marjanovič, U., Sočan, G. in Bajc, K. (2007). Vpliv psiholoških dejavnikov in izobrazbe staršev na učno uspešnost mladostnikov. *Psihološka obzorja*, 16(3), 27–48.
- Panhoon, S. in Wongwanich, S. (2014). An analysis of teacher feedback for improving teaching quality in primary schools. *Procedia-Social and Behavioral Sciences*, 116, 4124–4130.
- Ric (2021a). *Matematika. Navodila za vrednotenje 6. razred*. <https://www.ric.si/nacionalno-preverjanje-znanja/predmeti-npz/predmeti-v-6%20-razredu/matematika/>
- Ric (2021b). *Matematika. Navodila za vrednotenje 9. razred*. <https://www.ric.si/nacionalno-preverjanje-znanja/predmeti-npz/predmeti-v-9%20-razredu/matematika/>
- Ric (2021c). *Matematika. Preizkus znanja 6. razred*. <https://www.ric.si/nacionalno-preverjanje-znanja/predmeti-npz/predmeti-v-6%20-razredu/matematika/>
- Ric (2021d). *Matematika. Preizkus znanja 9. razred*. <https://www.ric.si/nacionalno-preverjanje-znanja/predmeti-npz/predmeti-v-9%20-razredu/matematika/>
- Ric (2021e). *Nacionalno preverjanje znanja. Letno poročilo o izvedbi v šolskem letu 2020/2021*. Ljubljana: Državni izpitni center. <https://www.ric.si/nacionalno-preverjanje-znanja/porocila--analize--raziskave/>
- Ric (2022a). *Nacionalno preverjanje znanja. Letno poročilo o izvedbi v šolskem letu 2021/2022*. Ljubljana: Državni izpitni center. <https://www.ric.si/nacionalno-preverjanje-znanja/porocila--analize--raziskave/>
- Ric (2022b). Struktura in opis preizkusa znanja za NPZ v 6. razredu. Ljubljana: Državni izpitni center. <https://www.ric.si/nacionalno-preverjanje-znanja/predmeti-npz/predmeti-v-6%20-razredu/matematika/>
- Ric (2022c). Struktura in opis preizkusa znanja za NPZ v 9. razredu. Ljubljana: Državni izpitni center. <https://www.ric.si/nacionalno-preverjanje-znanja/predmeti-npz/predmeti-v-9%20-razredu/matematika/>
- RM Assessor (2019). *Kratek vodnik po programu za e-ocenjevanje pri NPZ*. RM Education Ltd.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational researcher*, 20(5), 8–11.
- Škalič, M. in Grmek, M. I. (2017). Stališča staršev do nacionalnega preverjanja znanja ob koncu drugega vzgojno-izobraževalnega obdobja. *Journal of Elementary Education*, 10(1), 57–72.
- Vitulić, H. S. in Lesar, I. (2014). Prepričanja učencev, staršev in učiteljev o notranjih dejavnikih učne uspešnosti. *Journal of Elementary Education*, 7(1), 19–32.

REFERENČNI OKVIRI KOMPETENC

Digitalna bralnica ZRSS



Zavod
Republike
Slovenije
za šolstvo

Vse štiri publikacije lahko brezplačno preberate v digitalni bralnici ZRSS.

<https://www.zrss.si/digitalna-bralnica/referencni-okviri-kompetenc/>