



Psihometrijska analiza ispita državne mature

Školska godina 2009./ 2010.

OSOBNNA ISKAZNICA

NACIONALNOG CENTRA ZA VANJSKO VREDNOVANJE OBRAZOVANJA

Puni naziv: *Nacionalni centar za vanjsko vrednovanje obrazovanja*

Skraćeni naziv: Centar ili NCVVO

Adresa: 10000 Zagreb, Petračićeva 4

Pravno ustrojbeni oblik: javna ustanova

Osnivač: Republika Hrvatska

Odgovorna osoba: ravnatelj Goran Sirovatka, dipl.ing.

Telefon: 01/ 4501 800

Telefax: 01/ 4501 801

E-pošta: ravnatelj@ncvvo.hr

Web: www.ncvvo.hr

Matični broj poslovnog subjekta: 1943430

Šifra djelatnosti: 75120

Osnivanje ustanove: 04. ožujka 2005. godine, Trgovački sud u Zagrebu

Autori izvješća:

Josip Šabić, dipl. psiholog

Natalija Ćurković, dipl. psiholog

dr. sc. Jasminka Buljan Culej

U nekim dijelovima ovoga izvještaja upućujemo čitatelja na sadržaje koji se nalaze na mrežnim stranicama Nacionalnoga centra za vanjsko vrednovanje obrazovanja (www.ncvvo.hr).

1. UVOD

Nakon provedbe ispita državne mature 2010. godine, Istraživačko-razvojni odjel Centra proveo je psihometrijske analize svih ispita. Rezultati provedenih analiza predstavljeni su svakoj stručnoj radnoj skupini za onaj predmet za koji je ta skupina izrađivala ispit. Na temelju rezultata dobivenih nakon provedenih psihometrijskih analiza i rasprava provedenih s članovima stručnih radnih skupina koji su izrađivali ispite, proizlaze objašnjenja vezana uz psihometrijska i sadržajna obilježja pojedinih ispita, koja će biti prikazana u daljnjem tekstu.

Cilj psihometrijske analize bio je utvrditi razinu unutarnje valjanosti svakoga pojedinačnoga ispita te obilježja svakoga pojedinoga zadatka. Ovakva analiza potrebna je radi utvrđivanja stupnja sigurnosti u to da su rezultati postignuti na nekome ispitu doista pravi pokazatelji znanja određenih nastavnih sadržaja. Nadalje, povratne informacije o obilježjima pojedinih zadataka i cjelokupnoga ispita važne su članovima stručnih radnih skupina kako bi mogli napredovati u izradbi novih zadataka i ispita.

Ispitivanjem unutarnje valjanosti željelo se dobiti odgovore na ova pitanja:

- Jesu li dobiveni rezultati pouzdani?
- Je li ispit prikladan za učenike na kojima se primjenjuje?
- Jesu li zadatci u ispitu prihvatljive kvalitete?
- Postižu li bolji učenici (učenici boljega školskoga uspjeha) u pojedinim zadacima bolji uspjeh od onih lošijega školskoga uspjeha?
- Jesu li zadatci međusobno povezani?
- Može li se sa sigurnošću tvrditi da ispit mjeri ono što bi trebao mjeriti (znanja pojedinih sadržaja)?

Prilikom izradbe ispita svaka stručna radna skupina nastoji napraviti takav ispit kod kojega će odgovor na svako od gore navedenih pitanja biti potvrđan. Kako bi se odgovorilo na navedena pitanja, nužno je provesti niz analiza te protumačiti dobivene parametre.

Radi utvrđivanja već navedenih obilježja izračunati su sljedeći parametri za svaki ispit: aritmetička sredina, standardna devijacija, standardna pogreška mjerenja, postignuti raspon bodova u odnosu na maksimalni mogući, Cronbachov α koeficijent kao pokazatelj pouzdanosti ispita, težina zadataka i koeficijenti diskriminativnosti zadataka (Formule za izračunavanje svih navedenih parametara nalaze se u Prilogu I.). Kod zadataka višestrukoga izbora izračunati su postotci biranja pojedinoga odgovora kako bismo mogli analizirati kvalitetu i funkcioniranje ometača. Rezultati analize ometača nalaze se na mrežnim stranicama Centra (<http://www.ncvvo.hr/>). Radi dobivanja potpunijih informacija, učenici su prema postignutim rezultatima u ispitu podijeljeni u kvartile te su, s obzirom na kvartile, izračunati postotci biranja pojedinih odgovora na pitanjima višestrukoga izbora te postotci dobivanja određenoga broja bodova na zadacima otvorenoga tipa. Da bi se za svaki pojedini ispit mogli izračunati navedeni parametri, nužno je da za svaki ispit postoji barem 40 rezultata, tj. da je taj ispit pisalo barem 40 učenika. Iako je u brojnim statističkim priručnicima minimalan broj entiteta za izračunavanje ovih pokazatelja 100 (Petz, 2004), mi smo ponešto ublažili ovaj kriterij kako bismo za što veći broj ispita mogli izračunati psihometrijska obilježja. Međutim, valja upozoriti da na ispitima na kojima je sudjelovalo manje od 100 učenika, nije moguće očekivati potpuno stabilne pokazatelje. Zbog premaloga broja učenika koji su ih pisali, psihometrijska obilježja nije bilo moguće izračunati za ispite iz ovih predmeta: Francuski jezik (osnovna razina), Španjolski jezik (osnovna razina), Latinski jezik (viša i osnovna razina), Mađarski jezik te Grčki jezik.

2. OBILJEŽJA ISPITA DRŽAVNE MATURE

2.1. Aritmetička sredina

Najčešća mjera središnje vrijednosti nekoga skupa rezultata jest aritmetička sredina koja predstavlja težište rezultata. Ako je ispit prikladan za određenu skupinu učenika, onda bi se aritmetička sredina trebala nalaziti na polovici mogućega raspona rezultata. Ako je ona pomaknuta prema nižim ili višim vrijednostima, znači da je ispit bio pretežak ili prelagan za određenu skupinu učenika te s takvim ispitom nije moguće postići maksimalno razlikovanje učenika.

Prema dobivenim aritmetičkim sredinama ispita u odnosu na maksimalan mogući broj bodova te uzevši u obzir standardnu devijaciju, ispiti se mogu podijeliti u lagane, prosječne i teške. Tablica 1. prikazuje podjelu ispita državne mature s obzirom na ove tri kategorije.

Tablica 1. Podjela ispita državne mature s obzirom na njihovu težinu

TEŽINA ISPITA DRŽAVNE MATURE	PREDMET IZ KOJEGA SE PISAO ISPIT DRŽAVNE MATURE
Lagan	Hrvatski jezik (viša razina) Hrvatski jezik (osnovna razina) Engleski jezik (viša razina) Engleski jezik (osnovna razina) Njemački jezik (viša razina) Talijanski jezik (viša razina) Talijanski jezik (osnovna razina) Francuski jezik (viša razina) Španjolski jezik (viša razina) Talijanski materinski jezik (viša razina) Talijanski materinski jezik (osnovna razina) Srpski jezik Likovna umjetnost Glazbena umjetnost
Prosječan	Matematika (viša razina) Matematika (osnovna razina) Njemački jezik (osnovna razina) Fizika Geografija Povijest

	Politika i gospodarstvo Psihologija Sociologija Informatika Biologija Logika Vjeronauk Etika
Težak	Kemija Filozofija

2.2. Raspon

Idući pokazatelj koji govori o primjerenosti ispita za određenu skupinu učenika jest raspon. To je razlika između najvišega i najnižega postignutoga rezultata kod primjene ispita, koja pokazuje opseg dobivenih numeričkih vrijednosti u skupu rezultata, a može poslužiti kao približni orijentacijski indeks raspršenja rezultata (Field, 2005). Očekuje se da su ispiti izrađeni tako da je na njima moguće postići maksimalan raspon (od nula do maksimalnoga mogućega rezultata) budući da takav raspon omogućuje najbolje razlikovanje učenika s različitom količinom znanja. Ako na ispitu ni jedan učenik ne postiže maksimalan mogući broj bodova, znači da takav ispit nije prikladan za ciljnu skupinu učenika, odnosno da je pretežak.

Problem može biti i ako je najniža postignuta vrijednost prilikom primjene ispita daleko od nule. To znači da ispit sadrži prevelik broj laganih zadataka koje rješavaju svi učenici zbog čega ponovno nije moguće razlikovanje boljih od lošijih učenika.

Prema navedenim kriterijima, ispiti državne mature koji u potpunosti zadovoljavaju zahtjev za maksimalnim rasponom rezultata jesu: ispiti iz Matematike (viša i osnovna razina), ispiti iz Njemačkoga jezika (viša i osnovna razina), ispit iz Engleskoga jezika (osnovna razina) te ispit iz Fizike.

2.3. Standardna devijacija

Raspon predstavlja grubu mjeru raspršenja rezultata koja ne daje informaciju o obliku distribucije. Standardna devijacija predstavlja mjeru raspršenja rezultata koja pokazuje koliko se „gusto” rezultati nekoga mjerenja grupiraju oko aritmetičke sredine (Petz, 2005). Ova se mjera koristi kao standard za mjerenje varijabiliteta rezultata. To je vrlo praktična mjera raspršenja rezultata oko aritmetičke sredine, jer ako su poznate vrijednosti aritmetičke sredine i standardne devijacije, tada se može odrediti izgled distribucije uz uvjet da je ona normalna.

2.4. Standardna pogreška mjerenja

Standardna pogreška mjerenja jest procjena pogreške rezultata postignutoga na ispitu, a koja se određuje iz stupnja njegove pouzdanosti. Ona je izražena u izvornim jedinicama mjerenja, a omogućuje izračunavanje granica unutar kojih se s određenim stupnjem vjerojatnosti nalazi „pravi” rezultat mjerenja (Petz, 2005). Ako postignuti rezultat učenika označimo s x , a standardnu devijaciju sa sd , tada uz sigurnost od 68% možemo tvrditi da se pravi rezultat učenika nalazi u intervalu $x \pm 1 sd$. Uz sigurnost od 95% možemo tvrditi da se pravi rezultat učenika nalazi u intervalu $x \pm 2 sd$, a uz gotovo stopostotnu sigurnost možemo tvrditi da se pravi rezultat učenika nalazi u intervalu $x \pm 3 sd$. Ovaj pokazatelj vrlo je važan jer on govori kolikoj se pogrešci izlažemo pri zaključivanju o postignutim rezultatima na ispitu. Osobito je važno voditi računa o standardnoj pogrešci mjerenja pri rangiranju učenika na temelju postignutoga rezultata na ispitu. Ako ispit ima veliku standardnu pogrešku mjerenja, onda jednostavno rangiranje učenika prema postignutome rezultatu može biti netočno i na štetu učenika.

2.5. Pouzdanost (Cronbachov α koeficijent)

Obilježje mjernoga postupka (primjene ispita) koje se odnosi na točnost mjerenja naziva se pouzdanost, a obično se iskazuje Cronbachovim α koeficijentom. On govori kolika je prosječna korelacija među svim zadacima u testu. Cronbachov α koeficijent po svojoj je naravi korelacijski koeficijent pa kao takav varira između 0 i 1. Viša vrijednost ukazuje na veću međusobnu povezanost zadataka, odnosno na veću pouzdanost. Za različite vrste mjernih instrumenata prihvatljive su različite razine vrijednosti ovoga koeficijenta. Kod

ispitivanja znanja poželjno je da ovaj koeficijent iznosi barem 0,90 (Kehoe, 1997). Cronbachov α koeficijent je ovisan o broju zadataka i to tako da što je broj zadataka veći, to je ispit pouzdaniji, tj. Cronbachov α koeficijent je veći. Stoga je kod ispita s malim brojem zadataka teško očekivati vrlo visoke vrijednosti ($\geq 0,90$). Prema navedenim kriterijima ispite državne mature iz sljedećih predmeta možemo smatrati pouzdanim instrumentima: Hrvatski jezik (viša razina), Engleski jezik (viša i osnovna razina), Njemački jezik (viša razina), Talijanski jezik (viša razina), Srpski jezik, Matematika (viša razina), Fizika, Biologija, Geografija, Kemija, Povijest, Filozofija, Psihologija, Likovna umjetnost i Glazbena umjetnost.

Najveće odstupanje od zadanoga kriterija prisutno je kod ispita iz Logike (0,82).

2.6. Težina zadataka

Težina zadatka jest proporcija učenika koji su dali točan odgovor u dihotomnim zadacima. Ona nije pokazatelj je li zadatak dobar ili nije, već samo predstavlja težinu toga zadatka za određenu skupinu učenika (Osterlind, 2001). Ako zadatak nije dihotoman, težina se može izračunati tako da se aritmetička sredina zadatka podijeli s brojem bodova koje je u tom zadatku maksimalno moguće postići.

Da bi ispit bio prikladne težine, barem pola zadataka trebalo bi biti prosječne težine, odnosno trebalo bi ih moći uspješno riješiti od 40% do 60% učenika. Drugu polovinu trebali bi činiti teški i lagani zadatci i to tako da su ravnomjerno raspoređeni, odnosno da postoji podjednak broj teških i laganih zadataka. Međutim, poželjno je da ispit ne sadrži prevelik broj preteških i/ili prelaganih zadataka. Preteškim zadacima smatraju se oni koje uspješno rješava 10% ili manje učenika, dok su prelagani oni zadatci koje rješava 90% ili više učenika. Određen broj takvih zadataka potreban je zbog mogućnosti razlikovanja učenika jako dobrih i jako loših postignuća, točnije ispit ne bi smio sadržavati više od 10% zadataka koji spadaju u ove dvije kategorije (Državni izpitni center, 2007). Pritom je važno da broj vrlo teških i vrlo laganih zadataka u ispitu bude ujednačen. Ispiti u kojima postoji ukupno više od 10% vrlo teških i vrlo laganih zadataka su: Hrvatski jezik (viša razina) – 11,0%, Hrvatski jezik (osnovna razina) – 10,1%, Španjolski jezik (viša razina) – 51,6%, Talijanski jezik (viša razina) – 21,9%, Talijanski jezik (osnovna razina) – 34,6%, Francuski jezik (viša razina) –

28,4%, Njemački jezik (osnovna razina) – 13,5%, Engleski jezik (viša razina) – 13,0%, Engleski jezik (osnovna razina) – 19,2%, Talijanski materinski jezik (viša razina) – 32,8%, Talijanski materinski jezik (osnovna razina) – 44,3%, Kemija – 12,0%, Etika – 20,0%, Politika i gospodarstvo – 18,3% i Filozofija – 10,9%.

2.7. Diskriminativnost zadataka

Diskriminativnost ili diskriminativna valjanost jest obilježje zadatka koje opisuje „sposobnost” zadatka da mjeri individualne razlike među učenicima, a koje su odraz njihovih stvarnih razlika u znanju određenih sadržaja (Haladyna, 2004). Kod zadataka koji su visoko diskriminativni možemo s velikim stupnjem sigurnosti tvrditi da oni učenici koji postižu bolji rezultat na tom zadatku, postižu i bolji ukupan rezultat na ispitu. Stoga se može reći da je ovo obilježje zadatka izravni pokazatelj njegove kvalitete (Osterlind, 2001). Diskriminativnost zadataka izražava se preko koeficijenta diskriminativnosti (KD) koji se računa kao korelacija pojedinoga zadatka i ukupnoga rezultata na testu ako se iz ukupnoga rezultata isključi taj zadatak (Norusis, 1998). Prema tome, KD nam govori koliko je pojedini zadatak povezan s rezultatom na cijelome testu. Poželjno je da ta povezanost bude što veća. Osim što viša vrijednost KD-a ukazuje na veću povezanost zadatka s ukupnim rezultatom na testu, ona nam govori i o tome da taj zadatak dobro razlikuje (diskriminira) učenike s obzirom na njihovo znanje. Niski KD (oko nule) govori da je povezanost zadatka i ukupnoga rezultata na razini slučaja pa takve zadatke treba izbjegavati. Zadatak koji je negativno povezan s ukupnim rezultatom ukazuje da učenici s lošijim znanjem bolje rješavaju taj zadatak od učenika s boljim znanjem. Takvi zadatci obično imaju neki problem u samoj izradbi.

Minimalan prihvatljivi iznos KD-a je 0,2 (Tucker, 2007), a u dobro konstruiranom ispitu ne bi smjelo biti više od 20% zadataka koji imaju niži KD od ove minimalne vrijednosti (Državni izpitni center, 2007). Ispiti koji ne udovoljavaju ovom kriteriju, odnosno imaju više od 20% zadataka s manjim KD od 0,2 su: Hrvatski jezik (viša razina) – 20,9%, Hrvatski jezik (osnovna razina) – 39,3%, Španjolski jezik (viša razina) – 31,3%, Francuski jezik (viša razina) – 26,6%, Talijanski materinski jezik (viša razina) – 32,8%, Talijanski materinski jezik (osnovna razina) – 44,3%, Etika – 42,2%, Logika – 34,7% i Politika i gospodarstvo – 31,7%.

2.8. Valjanost

Valjanost je ključni koncept u konstrukciji ispita (Osterlind, 2001). To je karakteristika koja nam pokazuje mjeri li primijenjen ispit i u kojem stupnju upravo ono što smatramo da mjeri (Petz, 2005). Jedna od glavnih metoda ispitivanja valjanosti jest faktorska analiza. To je temeljna multivarijatna metoda, a sastoji se od niza statističko-matematičkih postupaka kojima se veći broj zadataka (manifestnih varijabli) nastoji sažeti u manji skup faktora latentnih varijabli. Osim redukcije broja početnih varijabli, primarni je cilj ove metode utvrditi povezanost zadataka s pojedinim faktorom koji možemo mjeriti pomoću ispita (mjernoga instrumenta). Ako je cilj pojedinoga ispita mjeriti jedan predmet mjerenja (npr. poznavanje gradiva Povijesti za treći razred gimnazije), onda se faktorskom analizom nastoji potvrditi da ispit mjeri upravo samo taj jedan predmet mjerenja. Ako se pokaže da ispit mjeri više predmeta mjerenja, onda više nije opravdano govoriti o jednome ispitu, već o više njih te u skladu s tim nije moguće ni ukupan rezultat takvoga ispita izražavati jednom ocjenom, već svaki utvrđeni predmet mjerenja treba ocjenjivati zasebnom ocjenom.

Cilj stručnih radnih skupina bio je izraditi ispite koji će mjeriti jedan predmet mjerenja. Ova pretpostavka provjerena je tako da su provedene faktorske analize svih ispita koji su zadovoljavali uvjete za provođenje faktorske analize. Glavni je uvjet za provođenje postojanje višestruko većega broja entiteta (učenika) od broja varijabli (ispitnih pitanja) (Field, 2005; Tacq, 1997). Ovaj uvjet nije bio zadovoljen kod ovih ispita: Srpski jezik, Španjolski jezik (viša razina), Talijanski jezik (osnovna razina), Francuski jezik (viša razina), Talijanski materinski jezik (viša razina), Talijanski materinski jezik (osnovna razina), Etika i Glazbena umjetnost. Stoga kod ovih ispita nije bilo moguće provesti faktorsku analizu.

Kod nekih ispita faktorskom su analizom dobivena dva faktora. To su ispiti iz Hrvatskoga jezika (viša razina) i Filozofije. Kod ispita iz Hrvatskoga jezika (osnovna razina) i Geografije dobivena su tri faktora. Iako rezultati faktorske analize nisu jedini pokazatelj valjanosti, dobivanje dvaju ili više faktora umjesto očekivanoga jednoga upućuje na ozbiljno narušene pretpostavke o valjanosti ovih ispita.

3. OPĆA OCJENA KVALITETE PRIMIJENJENIH ISPITA DRŽAVNE MATURE

Kao što je ranije rečeno, cilj psihometrijske analize ispita jest utvrđivanje valjanosti ispita. Tako se može reći da je ocjena kvalitete ispita zapravo ocjena njegove valjanosti. Pritom treba naglasiti da se valjanost ne odnosi na ispit kao takav, već na mogućnost donošenja istinitih zaključaka na temelju rezultata dobivenih pojedinom primjenom toga ispita (AERA/APA/NCME Standards, 1985, prema Osterlind, 2001). Prema tome, valjanost ispita nije moguće izraziti jednim konkretnim koeficijentom valjanosti, već samo složenim sudom utemeljenim na integraciji rezultata niza analiza (Petz, 2005). Tako se svi parametri koji su opisani u prethodnome odjeljku nalaze u tablicama 3. – 6. i služe upravo tomu da se njihovom integracijom donesu zaključci o valjanosti pojedinih ispita. Uvidom u rezultate prikazane u prethodnome tekstu i u tablicama, može se uočiti da pojedini ispiti zadovoljavaju neke kriterije, a druge pak ne. Ovakvi primjeri samo naglašavaju složenost donošenja ocjena o tome je li neki ispit valjan ili nije. Isto tako, ocjena o valjanosti ne donosi se samo na temelju psihometrijskih parametara, već i na temelju kvalitete postupka provedbe ispita te saznanja o djelovanju nekih psiholoških varijabli kao što su motivacija, umor i sl. No, ovdje se ne ćemo baviti tim različitim utjecajima koje često nije moguće ni kontrolirati ni evidentirati.

Prema psihometrijskim pokazateljima koji su nam na raspolaganju, može se zaključiti da se pet ispita prilikom ove primjene pokazalo osobito kvalitetnima, a to su ispiti državne mature iz Matematike (viša i osnovna razina), Fizike, Psihologije i Povijesti. Raspodjela ispita prema razini valjanosti prikazana je u tablici 2.

Tablica 2. Raspodjela ispita državne mature prema razini valjanosti

RAZINA VALJANOSTI	PREDMET IZ KOJEGA SE PISAO ISPIT DRŽAVNE MATURE
Visoka	Matematika (viša razina) Matematika (osnovna razina) Fizika Psihologija Povijest
Zadovoljavajuća	Srpski jezik Talijanski jezik (viša razina)

	<p>Talijanski jezik (osnovna razina) Njemački jezik (viša razina) Njemački jezik (osnovna razina) Engleski jezik (viša razina) Engleski jezik (osnovna razina) Informatika Biologija Kemija Glazbena umjetnost Likovna umjetnost Vjeronauk Sociologija Geografija Filozofija</p>
Niska	<p>Hrvatski jezik (viša razina) Hrvatski jezik (osnovna razina) Španjolski jezik (viša razina) Francuski jezik (viša razina) Talijanski materinski jezik (viša razina) Talijanski materinski jezik (osnovna razina) Etika Logika Politika i gospodarstvo</p>

Detaljni prikazi rezultata psihometrijskih analiza nalaze se u tablicama 3. – 6. te u tekstovima u kojima su opisani svi ispiti zasebno. Ovdje treba još jednom naglasiti da ispite državne mature iz Francuskoga jezika (osnovna razina), Španjolskoga jezika (osnovna razina), Latinskoga jezika (viša i osnovna razina), Mađarskoga jezika te Grčkoga jezika nije bilo moguće psihometrijski analizirati zbog premaloga broja učenika koji su pisali te ispite.

Legenda za tablice 3. – 6.

N - broj učenika koji su pisali pojedini ispit državne mature

Broj zadataka - broj zadataka u ispitu

O - broj zadataka otvorenoga tipa

Z - broj zadataka zatvorenoga tipa

E – broj kriterija za ocjenjivanje zadatka esejskoga tipa (uradak učenika na svakom kriteriju je u analizama tretiran kao zasebna čestica)

Raspon - raspon rezultata koji su učenici postizali

MAX - maksimalan mogući broj bodova koji je na pojedinome ispitu moguće ostvariti

M - aritmetička sredina uratka na ispitu

SD - standardna devijacija uratka na ispitu

SPM - standardna pogreška mjerenja uratka na ispitu

IT<0,1 - broj zadataka na ispitu koji imaju indeks težine manji od 0,1

IT>0,9 - broj zadataka na ispitu koji imaju indeks težine veći od 0,9

IT% - postotak zadataka u ispitu koji imaju indekse težine manje od 0,1 ili veće od 0,9 (ne bi smjelo biti više od 10%)

IP(α) – indeks pouzdanosti (Cronbachov α koeficijent)

KD% - postotak zadataka koji imaju koeficijent diskriminativnosti (KD) manji od 0,2 (ne bi smjelo biti više od 10%)

FA - broj faktora dobiven faktorskom analizom ispita (metoda ekstrakcije: glavne komponente; rotacija: Oblimin)

"-" - analize nisu provedene zbog premaloga broja učenika

Tablica 3. Prikaz psihometrijskih obilježja ispita državne mature iz Hrvatskoga jezika i Matematike

Predmet	Hrvatski jezik (viša razina)	Hrvatski jezik (osnovna razina)	Matematika (viša razina)	Matematika (osnovna razina)
N	21207	12269	9626	23934
Broj zadataka	91	89	45	33
O	0	0	30	17
Z	80	78	15	16
E	11	11	0	0
Raspon	0 – 159	2 – 154	0 – 60	0 – 40
MAX	160	154	60	40
M	113,0	102,1	30,5	20,6
SD	18,50	18,80	13,18	8,30
SPM	5,85	6,78	3,95	2,99
IT<0,1	0	0	0	2
IT>0,9	10	9*	1	1
IT%	11,0	10,1	2,2	9,1
IP(α)	0,90	0,87	0,91	0,87
KD%	20,9	39,3*	4,4	0
FA	2	3	1	1

* - u ovaj broj su uključena i 2 zadatka na kojima su svim pristupnicima dodijeljeni bodovi (Hrvatski jezik – osnovna razina)

Tablica 4. Prikaz psihometrijskih obilježja ispita državne mature iz stranih jezika

Predmet	Srpski jezik	Španjolski jezik (viša razina)	Talijanski jezik (viša razina)	Talijanski jezik (osnovna razina)	Francuski jezik (viša razina)	Njemački jezik (viša razina)	Njemački jezik (osnovna razina)	Engleski jezik (viša razina)	Engleski jezik (osnovna razina)	Talijanski materinski jezik (viša razina)	Talijanski materinski jezik (osnovna razina)
N	138	87	422	99	102	1557	3193	16651	15635	66	78
Broj zadataka	84	64	64	52	64	69	52	69	52	61	61
O	18	0	0	5	0	8	6	8	10	0	0
Z	54	60	60	45	60	57	44	57	40	50	50
E	12	4	4	2	4	4	2	4	2	11	11
Raspon	10 – 96,75	5 – 79	7,62 – 98,67	16 – 98,67	44 – 80	0 – 100	0 – 100	3,33 – 100	0 – 100	31 – 95	29 – 91
MAX	100	80	100	100	100	100	100	100	100	100	100
M	62,6	60,4	68,9	74,3	67,6	69,2	60,1	75,3	72,2	74,1	67,6
SD	16,70	12,79	17,78	20,90	9,46	21,53	22,14	15,26	19,92	12,38	14,03
SPM	5,01	4,61	5,03	7,24	2,99	4,81	7,98	4,32	6,30	4,29	5,25
IT<0,1	2	0	0	0	0	0	0	0	0	0	0
IT>0,9	2	33	14	18	29	4	7	9	10	14	9
IT%	4,8	51,6	21,9	34,6	28,4	5,8	13,5	13,0	19,2	23,0	14,8
IP(α)	0,91	0,87	0,92	0,88	0,90	0,95	0,87	0,92	0,90	0,88	0,86
KD%	19,0	31,3	17,2	9,6	26,6	1,4	7,7	10,1	0	32,8	44,3
FA	-	-	1	-	-	2	1	1	1	-	-

Tablica 5. Prikaz psihometrijskih obilježja ispita državne mature iz prirodoslovnih predmeta

Predmet	Informatika	Biologija	Kemija	Fizika
N	2945	7925	3446	9395
Broj zadataka	32	128	100	36
O	8	64	60	12
Z	24	64	40	24
Raspon	0 – 31	2 – 127	4 – 99	0 – 80
MAX	32	128	100	80
M	15,7	52,2	38,0	38,4
SD	6,21	27,64	19,12	18,57
SPM	2,32	4,79	3,82	5,25
IT<0,1	1	7	11	0
IT>0,9	1	0	1*	0
IT%	6,3	5,5	12,0	0
IP(α)	0,86	0,97	0,96	0,92
KD%	9,4	5,5	7,0*	0
FA	1	1	1	1

* - u ovaj broj je uključen i 1 zadatak na kojem je svim pristupnicima dodijeljen bod (Kemija)

Tablica 6. Prikaz psihometrijskih obilježja ispita državne mature iz društvenih predmeta

Predmet	Etika	Glazbena umjetnost	Likovna umjetnost	Vjeronauk	Logika	Psihologija	Sociologija	Geografija	Povijest	Politika i gospodarstvo	Filozofija
N	85	112	2208	353	284	4560	889	1125	2565	8640	538
Broj zadataka	45	61	90	41	49	63	56	114	119	60	55
O	10	4	10	17	24	45	12	42	40	35	14
Z	20	57	60	24	25	18	44	72	79	25	23
E	15	0	20	0	0	0	0	0	0	0	18
Raspon	4 – 47	30 – 96	0 – 97	7 – 50	10 – 56	0 – 114	9 – 69	0 – 191	7 – 115	0 – 58	7 – 82
MAX	60	100	105	54	60	117	72	200	119	60	100
M	26,7	72,5	61,7	24,9	27,7	56,9	39,5	86,7	55,2	32,9	34,9
SD	8,13	15,99	15,75	9,00	9,55	27,03	10,75	28,91	21,41	8,47	15,23
SPM	3,15	5,06	4,17	2,98	4,05	5,41	3,72	7,65	4,28	3,05	4,03
IT<0,1	7	0	1	0	1	1	0	8	3	4	5
IT>0,9	2	6	1	0	1	1	0	1	3	7	1*
IT%	20	9,8	2,2	0	4,1	3,2	0	7,9	5,0	18,3	10,9
IP(α)	0,85	0,90	0,93	0,89	0,82	0,96	0,88	0,93	0,96	0,87	0,93
KD%	42,2	9,8	14,4	12,2	12,2	3,2	17,9	14,9	8,4	31,7	16,4
FA	-	-	1	1	1	1	1	3	1	1	2

* - u ovaj broj je uključen i 1 zadatak na kojem je svim pristupnicima dodijeljen bod (Filozofija)

4. OCJENA KVALITETE SVAKOGA POJEDINOGA ISPITA DRŽAVNE MATURE

4.1. Hrvatski jezik (viša razina)

Analize ispita državne mature iz Hrvatskoga jezika za višu razinu provedene su na rezultatima 21207 učenika. Ispit je sadržavao ukupno 91 zadatak od čega ih je 80 bilo zatvorenog tipa, a 11 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 113,0$; $MAX = 160$). Od ukupno 91 zadatka deset je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo preko 90% učenika. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,90$), no ovaj je pokazatelj ovisan o velikom broju zadataka u ispitu. Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,85. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 5,85$ bodova. Faktorskom analizom ispita utvrđeno je postojanje 2 faktora, gdje prvi faktor objašnjava oko 11% ukupne varijance ispita, a drugi faktor oko 4% ukupne varijance ispita. Ovakva faktorska struktura nije u skladu s visokim Cronbachovim α koeficijentom i ukazuje na heterogenost ispitnoga materijala.

Ni jedan ispitanik nije postigao rezultat viši od 159 bodova iako je maksimalan broj bodova na ispitu bio 160. S obzirom na vrlo velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 20,9%. Ovaj nalaz u skladu je s dvofaktorskom strukturom ispita.

Zaključno se može reći da na ispitu državne mature iz Hrvatskoga jezika za višu razinu nije postignut puni raspon bodova te da nema zadovoljavajuću konstruktivnu valjanost. Također, ispit sadrži previše laganih i vrlo laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše.

4.2. Hrvatski jezik (osnovna razina)

Analize ispita državne mature iz Hrvatskoga jezika za osnovnu razinu provedene su na rezultatima 12269 učenika. Ispit je sadržavao ukupno 89 zadataka od čega ih je 78 bilo zatvorenog tipa, a 11 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 102,1$; $MAX = 154$). Od ukupno 89 zadatka devet je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo preko 90% učenika. Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,87$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 6,78. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 6,78$ bodova. Faktorskom analizom ispita utvrđeno je postojanje tri faktora, gdje prvi faktor objašnjava oko 9% ukupne varijance ispita, drugi faktor oko 3% ukupne varijance ispita, a treći faktor oko 3% ukupne varijance ispita. Ovakva faktorska struktura nije u skladu s visokim Cronbachovim α koeficijentom i ukazuje na heterogenost ispitnoga materijala.

Cijeli je raspon rezultata u ispitu pokriven, tj. bilo je ispitanika koji su postigli 2 boda (na dva su zadatka svim pristupnicima pridijeljeni bodovi), kao i ispitanika koji su postigli maksimalan mogući broj bodova.

Vrlo velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 39,3%. Ovaj nalaz u skladu je s trofaktorskom strukturom ispita. Jedan je zadatak imao negativan koeficijent diskriminativnosti.

Zaključno se može reći da ispit državne mature iz Hrvatskoga jezika za osnovnu razinu nema zadovoljavajuću konstruktnu valjanost. Također, ispit sadrži previše laganih i vrlo laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše.

4.3. Matematika (viša razina)

Analize ispita državne mature iz Matematike za višu razinu provedene su na rezultatima 9626 učenika. Ispit je sadržavao ukupno 45 zadataka od čega ih je 15 bilo zatvorenog, a 30 otvorenog tipa. Ispit je bio prosječne težine ($M = 30,5$; $MAX = 60$). Od ukupno 45 zadataka jedan je zadatak bio izrazito lagan, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,91$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 3,95. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 3,95$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 25% ukupne varijance ispita.

Cijeli je raspon rezultata u ispitu pokriven, tj. bilo je ispitanika koji su postigli 0 bodova, kao i ispitanika koji su postigli maksimalan mogući broj bodova.

Dva zadatka nisu imala zadovoljavajuće koeficijente diskriminativnosti.

Ispit državne mature iz Matematike za višu razinu zadovoljava psihometrijske kriterije te da ga se može smatrati pouzdanim i valjanim instrumentom za mjerenje poznavanja gradiva koje je u njemu zastupljeno.

4.4. Matematika (osnovna razina)

Analize ispita državne mature iz Matematike za osnovnu razinu provedene su na rezultatima 23934 učenika. Ispit je sadržavao ukupno 33 zadatka od čega ih je 16 bilo zatvorenog, a 17 otvorenog tipa. Ispit je bio prosječne težine ($M = 20,6$; $MAX = 40$). Od ukupno 33 zadatka dva su zadatka bila izrazito teška, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika, a jedan je zadatak bio izrazito lagan, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika.. Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,87$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 2,99. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 2,99$ bodova.

Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 22% ukupne varijance ispita.

Cijeli je raspon rezultata u ispitu pokriven, tj. bilo je ispitanika koji su postigli 0 bodova, kao i ispitanika koji su postigli maksimalan mogući broj bodova.

Svi su zadatci imali zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Matematike za osnovnu razinu zadovoljava psihometrijske kriterije te da ga se može smatrati pouzdanim i valjanim instrumentom za mjerenje poznavanja gradiva koje je u njemu zastupljeno.

4.5. Srpski jezik

Analize ispita državne mature iz Srpskoga jezika provedene su na rezultatima 138 učenika. Ispit je sadržavao ukupno 84 zadatka od čega ih je 54 bilo zatvorenog tipa, 18 otvorenog tipa, a 12 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 62,6$; $MAX = 100$). Od ukupno 84 zadatka dva su zadatka bila izrazito teška, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika. Dva su zadatka bila izrazito lagana, s indeksom težine većim od 0,9, što znači da ih je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,91$). Visoka vrijednost ovoga koeficijenta očekivana je zbog velikoga broja zadataka jer što je veći broj zadataka, veća je pouzdanost. Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,01. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 5,01$ bodova.

Velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 19,0%. Dva zadatka su imala negativan koeficijent diskriminativnosti.

Ispit državne mature iz Srpskoga jezika ima solidna metrijska obilježja, no ima relativno velik broj zadataka niske diskriminativnosti. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.6. Španjolski jezik (viša razina)

Analize ispita državne mature iz Španjolskoga jezika za višu razinu provedene su na rezultatima 87 učenika. Ispit je sadržavao ukupno 64 zadatka od čega ih je 60 bilo zatvorenog tipa, a 4 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 60,4$; $MAX = 80$). Od ukupno 64 zadatka trideset i tri su zadatka bila izrazito lagana, s indeksom težine većim od 0,9, što znači da ih je točno riješilo preko 90% učenika. Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,87$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,61. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 4,61$ bodova.

Vrlo velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 31,3%. Dva zadatka su imala negativan koeficijent diskriminativnosti.

Ispit državne mature iz Španjolskoga jezika za višu razinu sadrži previše laganih i vrlo laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše. Ispit sadrži velik broj zadataka niske diskriminativnosti. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.7. Talijanski jezik (viša razina)

Analize ispita državne mature iz Talijanskoga jezika za višu razinu provedene su na rezultatima 422 učenika. Ispit je sadržavao ukupno 64 zadatka od čega ih je 60 bilo zatvorenog tipa, a 4 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 68,9$; $MAX = 100$). Četrnaest je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,92$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,03. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na

dobivene $\pm 5,03$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 18% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat niži od 7,62 bodova niti viši od 98,67 bodova. S obzirom na vrlo velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 17,2% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Jedan zadatak je imao negativan koeficijent diskriminativnosti.

Ispit državne mature iz Talijanskoga jezika za višu razinu sadrži previše laganih i vrlo laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše.

4.8. Talijanski jezik (osnovna razina)

Analize ispita državne mature iz Talijanskoga jezika za osnovnu razinu provedene su na rezultatima 99 učenika. Ispit je sadržavao ukupno 52 zadatka od čega ih je 45 bilo zatvorenog tipa, 5 otvorenog tipa, a 2 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 74,3$; $MAX = 100$). Od ukupno 52 zadatka osamnaest je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo preko 90% učenika. Cronbachov α koeficijent ukazuje na solidnu razinu homogenosti ispitnoga materijala ($\alpha = 0,88$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 7,24. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 7,24$ bodova.

Od ukupnoga broja zadataka 9,6% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Jedan je zadatak imao negativan koeficijent diskriminativnosti.

Ispit državne mature iz Talijanskoga jezika za osnovnu razinu sadrži previše laganih i vrlo laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.9. Francuski jezik (viša razina)

Analize ispita državne mature iz Francuskoga jezika za višu razinu provedene su na rezultatima 102 učenika. Ispit je sadržavao ukupno 64 zadatka od čega ih je 60 bilo zatvorenog tipa, a 4 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 67,6$; $MAX = 100$). Dvadeset i devet zadataka je bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,90$). Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 2,99. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 2,99$ bodova.

Velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 26,6%. Jedan zadatak je imao negativan koeficijent diskriminativnosti.

Ispit državne mature iz Francuskoga jezika za višu razinu sadrži previše laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše. Ispit sadrži velik broj zadataka niske diskriminativnosti. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.10. Njemački jezik (viša razina)

Analize ispita državne mature iz Njemačkoga jezika za višu razinu provedene su na rezultatima 1557 učenika. Ispit je sadržavao ukupno 69 zadataka od čega ih je 57 bilo zatvorenog tipa, 8 otvorenog tipa, a 4 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 69,2$; $MAX = 100$). Od ukupno 69 zadataka četiri zadatka su bila izrazito lagana, s indeksom težine većim od 0,9, što znači da ih je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na vrlo visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,95$). Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,81. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom

mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 4,81$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 26% ukupne varijance ispita.

Cijeli je raspon rezultata u ispitu pokriven, tj. bilo je ispitanika koji su postigli 0 bodova, kao i ispitanika koji su postigli maksimalan mogući broj bodova.

Od ukupnoga broja zadataka 1,4% ih nije imalo zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Njemačkoga jezika za višu razinu sadrži previše laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše. Osim toga, ispit državne mature iz Njemačkoga jezika za višu razinu ima zadovoljavajuća metrijska obilježja, tj. riječ je o pouzdanome i valjanome instrumentu.

4.11. Njemački jezik (osnovna razina)

Analize ispita državne mature iz Njemačkoga jezika za osnovnu razinu provedene su na rezultatima 3193 učenika. Ispit je sadržavao ukupno 52 zadatka od čega ih je 44 bilo zatvorenog tipa, 6 otvorenog tipa, a 2 su bili kriteriji na esejskom dijelu ispita. Ispit je bio prosječne težine ($M = 60,1$; $MAX = 100$). Od ukupno 52 zadatka sedam je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,88$). Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 7,98. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 7,98$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 21% ukupne varijance ispita.

Cijeli je raspon rezultata u ispitu pokriven, tj. bilo je ispitanika koji su postigli 0 bodova, kao i ispitanika koji su postigli maksimalan mogući broj bodova.

Od ukupnoga broja zadataka 7,7% ih nije imalo zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Njemačkoga jezika za osnovnu razinu sadrži previše laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše. Osim toga, ispit državne mature iz Njemačkoga jezika za osnovnu razinu ima zadovoljavajuća metrijska obilježja, tj. riječ je o pouzdanome i valjanome instrumentu.

4.12. Engleski jezik (viša razina)

Analize ispita državne mature iz Engleskoga jezika za višu razinu provedene su na rezultatima 16651 učenika. Ispit je sadržavao ukupno 69 zadataka od čega ih je 57 bilo zatvorenog tipa, 8 otvorenog tipa, a 4 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 75,3$; $MAX = 100$). Od ukupno 69 zadatka devet je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo preko 90% učenika. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,92$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,32. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 4,32$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 17% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat manji od 3,33 bodova. Maksimalni mogući broj bodova je postignut. S obzirom na vrlo velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 10,1% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Nijedan zadatak nije imao negativan koeficijent diskriminativnosti.

Ispit državne mature iz Engleskoga jezika za višu razinu sadrži previše laganih i vrlo laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše.

4.13. Engleski jezik (osnovna razina)

Analize ispita državne mature iz Engleskoga jezika za osnovnu razinu provedene su na rezultatima 15635 učenika. Ispit je sadržavao ukupno 52 zadatka od čega ih je 40 bilo zatvorenog tipa, 10 otvorenog tipa, a 2 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 72,2$; $MAX = 100$). Od ukupno 52 zadatka deset je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo preko 90% učenika. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,90$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 6,30. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 6,30$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 21% ukupne varijance ispita.

Cijeli je raspon rezultata u ispitu pokriven, tj. bilo je ispitanika koji su postigli 0 bodova, kao i ispitanika koji su postigli maksimalan mogući broj bodova.

Svi su zadatci imali zadovoljavajuće koeficijente diskriminativnosti.

Ispit državne mature iz Engleskoga jezika za osnovnu razinu sadrži previše laganih i vrlo laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše. Osim toga, ispit državne mature iz Engleskoga jezika za osnovnu razinu ima zadovoljavajuća metrijska obilježja, tj. riječ je o pouzdanome i valjanome instrumentu.

4.14. Talijanski materinski jezik (viša razina)

Analize ispita državne mature iz Talijanskoga materinskoga jezika za višu razinu provedene su na rezultatima 66 učenika. Ispit je sadržavao ukupno 61 zadatak od čega ih je 50 bilo zatvorenog tipa, a 11 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 74,1$; $MAX = 100$). Od ukupno 61 zadatka četrnaest je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,88$). Standardna pogreška mjerenja

koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,29. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 4,29$ bodova.

Velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 32,8%. Tri su zadatka imala negativan koeficijent diskriminativnosti.

Ispit državne mature iz Talijanskoga materinskoga jezika za višu razinu sadrži previše laganih i vrlo laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.15. Talijanski materinski jezik (osnovna razina)

Analize ispita državne mature iz Talijanskoga materinskoga jezika za osnovnu razinu provedene su na rezultatima 78 učenika. Ispit je sadržavao ukupno 61 zadatak od čega ih je 50 bilo zatvorenog tipa, a 11 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 67,6$; $MAX = 100$). Od ukupno 61 zadatka devet je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,86$). Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,25. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 5,25$ bodova.

Vrlo velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 44,3%. Pet je zadataka imalo negativan koeficijent diskriminativnosti.

Ispit državne mature iz Talijanskoga materinskoga jezika za osnovnu razinu sadrži previše laganih i vrlo laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše. Ispit sadrži vrlo velik broj zadataka niske diskriminativnosti. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.16. Informatika

Analize ispita državne mature iz Informatike provedene su na rezultatima 2945 učenika. Ispit je sadržavao ukupno 32 zadatka od čega ih je 24 bilo zatvorenog, a 8 otvorenog tipa. Ispit je bio prosječne težine ($M = 15,7$; $MAX = 32$). Od ukupno 32 zadatka jedan je zadatak bio izrazito težak, s indeksom težine manjim od 0,1, što znači da ga je točno riješilo manje od 10% učenika. Jedan je zadatak bio izrazito lagan, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,86$). Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 2,32. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 2,32$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 21% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat viši od 31 bodova iako je maksimalan broj bodova na ispitu bio 32. S obzirom na vrlo velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 9,4% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Nijedan zadatak nije imao negativan koeficijent diskriminativnosti.

Ispit državne mature iz Informatike ima zadovoljavajuća metrijska obilježja, tj. riječ je o pouzdanome i valjanome instrumentu. Bilo bi dobro da je ispit duži, tj. da ima više zadataka.

4.17. Biologija

Analize ispita državne mature iz Biologije provedene su na rezultatima 7925 učenika. Ispit je sadržavao ukupno 128 zadataka od čega ih je 64 bilo zatvorenog, a 64 otvorenog tipa. Ispit je bio prosječne težine ($M = 52,2$; $MAX = 128$). Od ukupno 128 zadataka sedam je zadataka bilo izrazito teško, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika. Cronbachov α koeficijent ukazuje na vrlo visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,97$). Visoka vrijednost ovoga koeficijenta očekivana je zbog velikoga broja zadataka jer što je veći broj zadataka, veća je pouzdanost. Standardna

pogrješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,79. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 4,79$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 25% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat niži od 2 boda niti viši od 127 boda iako je maksimalan broj bodova na ispitu bio 128. S obzirom na vrlo velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 5,5% ih nije imalo zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Biologije sadrži previše teških zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše. Ispit državne mature iz Biologije ima zadovoljavajuća metrijska obilježja, tj. riječ je o pouzdanome i valjanome instrumentu.

4.18. Kemija

Analize ispita državne mature iz Kemije provedene su na rezultatima 3446 učenika. Ispit je sadržavao ukupno 100 zadataka od čega ih je 40 bilo zatvorenog, a 60 otvorenog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je pretežak ($M = 38,0$; $MAX = 100$). Od ukupno 100 zadataka jedanaest je zadataka bilo izrazito teško, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika. Cronbachov α koeficijent ukazuje na vrlo visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,96$). Visoka vrijednost ovoga koeficijenta očekivana je zbog velikoga broja zadataka jer što je veći broj zadataka, veća je pouzdanost. Standardna pogrješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 3,82. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 3,82$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 23% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat niži od 4 boda niti viši od 99 bodova iako je maksimalan broj bodova na ispitu bio 100. S obzirom na vrlo velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 7,0% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Jedan zadatak je imao negativan koeficijent diskriminativnosti.

Ispit državne mature iz Kemije sadrži previše teških i vrlo teških zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše.

4.19. Fizika

Analize ispita državne mature iz Fizike provedene su na rezultatima 9395 učenika. Ispit je sadržavao ukupno 36 zadataka od čega ih je 24 bilo zatvorenog, a 12 otvorenog tipa. Ispit je bio prosječne težine ($M = 38,4$; $MAX = 80$). Od ukupno 36 zadataka niti jedan zadatak nije bio izrazito težak ili izrazito lagan. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,92$). Standardna pogrješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,25. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 5,25$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 28% ukupne varijance ispita.

Cijeli je raspon rezultata u ispitu pokriven, tj. bilo je ispitanika koji su postigli 0 bodova, kao i ispitanika koji su postigli maksimalan mogući broj bodova.

Svi su zadatci imali zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Fizike zadovoljava psihometrijske kriterije te da ga se može smatrati pouzdanim i valjanim instrumentom za mjerenje poznavanja gradiva koje je u njemu zastupljeno. Bilo bi dobro da se u ispitu nalazi još koji vrlo lagan i vrlo težak zadatak.

4.20. Etika

Analize ispita državne mature iz Etike provedene su na rezultatima 85 učenika. Ispit je sadržavao ukupno 45 zadataka od čega ih je 20 bilo zatvorenog tipa, 10 otvorenog tipa, a 15 su bili kriteriji na esejskom dijelu ispita. Ispit je bio prosječne težine ($M = 26,7$; $MAX = 60$). Od ukupno 45 zadatka sedam je zadataka bilo izrazito teško, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika. Dva su zadatka bila izrazito lagana, s indeksom težine većim od 0,9, što znači da ih je točno riješilo više od 90% učenika.

Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,85$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 3,15. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 3,15$ bodova.

Vrlo velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 42,2%. Sedam je zadataka imalo negativan koeficijent diskriminativnosti.

Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.21. Glazbena umjetnost

Analize ispita državne mature iz Glazbene umjetnosti provedene su na rezultatima 112 učenika. Ispit je sadržavao ukupno 61 zadatak od čega ih je 57 bilo zatvorenog, a 4 otvorenog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 72,5$; $MAX = 100$). Šest zadataka je bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,90$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,06. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 5,06$ bodova.

Od ukupnoga broja zadataka 9,8% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Jedan zadatak je imao negativan koeficijent diskriminativnosti.

Ispit državne mature iz Glazbene umjetnosti sadrži previše laganih i vrlo laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.22. Likovna umjetnost

Analize ispita državne mature iz Likovne umjetnosti provedene su na rezultatima 2208 učenika. Ispit je sadržavao ukupno 90 zadataka od čega ih je 60 bilo zatvorenog tipa, 10 otvorenoga tipa, a 20 su bili kriteriji na esejskom dijelu ispita. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 61,7$; $MAX = 105$). Jedan je zadatak bio izrazito lagan, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika. Jedan je zadatak bio izrazito lagan, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,93$). Visoka vrijednost ovoga koeficijenta očekivana je zbog velikoga broja zadataka jer što je veći broj zadataka, veća je pouzdanost. Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,17. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 4,17$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 14,5% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat viši od 97 bodova iako je maksimalan broj bodova na ispitu bio 105. S obzirom na vrlo velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 14,4% ih nije imalo zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Likovne umjetnosti sadrži previše laganih zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše.

4.23. Vjeronauk

Analize ispita državne mature iz Vjeronauka provedene su na rezultatima 353 učenika. Ispit je sadržavao ukupno 41 zadatak od čega ih je 24 bilo zatvorenog, a 17 otvorenog tipa. Ispit je bio prosječne težine ($M = 24,9$; $MAX = 54$). Od ukupno 41 zadataka niti jedan zadatak nije bio izrazito težak ili izrazito lagan. Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,89$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 2,98. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 2,98$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 20% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat niži od 7 bodova niti viši od 50 bodova iako je maksimalan broj bodova na ispitu bio 54. S obzirom na velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 12,2% ih nije imalo zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Vjeronauka zadovoljava psihometrijske kriterije te da ga se može smatrati pouzdanim i valjanim instrumentom za mjerenje poznavanja gradiva koje je u njemu zastupljeno.

4.24. Logika

Analize ispita državne mature iz Logike provedene su na rezultatima 284 učenika. Ispit je sadržavao ukupno 49 zadataka od čega ih je 25 bilo zatvorenog tipa, a 24 otvorenog tipa. Ispit je bio prosječne težine ($M = 27,7$; $MAX = 60$). Od ukupno 49 zadataka jedan je zadatak bio izrazito težak, s indeksom težine manjim od 0,1, što znači da ga je točno riješilo manje od 10% učenika. Jedan je zadatak bio izrazito lagan, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika.

Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,82$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,05. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi

da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 4,05$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 14% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat niži od 10 bodova niti viši od 56 bodova iako je maksimalan broj bodova na ispitu bio 60. S obzirom na velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Vrlo velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 34,7%. Tri su zadatka imala negativan koeficijent diskriminativnosti.

4.25. Psihologija

Analize ispita državne mature iz Psihologije provedene su na rezultatima 4560 učenika. Ispit je sadržavao ukupno 63 zadataka od čega ih je 18 bilo zatvorenog tipa, a 45 otvorenog tipa. Ispit je bio prosječne težine ($M = 56,9$; $MAX = 117$). Od ukupno 63 zadatka jedan je zadatak bio izrazito težak, s indeksom težine manjim od 0,1, što znači da ga je točno riješilo manje od 10% učenika. Jedan je zadatak bio izrazito lagan, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika.

Cronbachov α koeficijent ukazuje na vrlo visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,96$). Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,41. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 5,41$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 28% ukupne varijance ispita.

Ni jedan ispitanik nije postigao viši od 114 bodova iako je maksimalan broj bodova na ispitu bio 117. S obzirom na velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 3,2% ih nije imalo zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Psihologije zadovoljava psihometrijske kriterije te da ga se može smatrati pouzdanim i valjanim instrumentom za mjerenje poznavanja gradiva koje je u njemu zastupljeno.

4.26. Sociologija

Analize ispita državne mature iz Sociologije provedene su na rezultatima 889 učenika. Ispit je sadržavao ukupno 56 zadataka od čega ih je 44 bilo zatvorenog tipa, a 12 otvorenog tipa. Ispit je bio prosječne težine ($M = 39,5$; $MAX = 72$). Od ukupno 56 zadataka niti jedan zadatak nije bio izrazito težak ili izrazito lagan.

Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,88$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 3,72. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 3,72$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 15% ukupne varijance ispita.

Ni jedan ispitanik nije postigao niži od 9 bodova niti viši od 69 bodova iako je maksimalan broj bodova na ispitu bio 72. S obzirom na velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 17,9% ih nije imalo zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Sociologije zadovoljava psihometrijske kriterije te da ga se može smatrati pouzdanim i valjanim instrumentom za mjerenje poznavanja gradiva koje je u njemu zastupljeno.

4.27. Geografija

Analize ispita državne mature iz Geografije provedene su na rezultatima 1125 učenika. Ispit je sadržavao ukupno 114 zadataka od čega ih je 72 bilo zatvorenog tipa, a 42 otvorenog tipa. Ispit je bio prosječne težine ($M = 86,7$; $MAX = 200$). Od ukupno 114 zadataka osam je zadataka bilo izrazito teško, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo

manje od 10% učenika. Jedan je zadatak bio izrazito lagan, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika.

Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,93$). Visoka vrijednost ovoga koeficijenta očekivana je zbog velikoga broja zadataka jer što je veći broj zadataka, veća je pouzdanost. Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 7,65. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 7,65$ bodova. Faktorskom analizom ispita utvrđeno je postojanje tri faktora koji objašnjavaju redom oko 14%, 3% i 2% ukupne varijance ispita. Višefaktorska struktura ovoga ispita je zapravo artefakt međusobno ovisnih zadataka povezivanja i sređivanja.

Ni jedan ispitanik nije postigao viši od 191 bodova iako je maksimalan broj bodova na ispitu bio 200. S obzirom na velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 14,9% ih nije imalo zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Geografije bi trebao sadržavati malo manje teških i vrlo teških zadataka.

4.28. Povijest

Analize ispita državne mature iz Povijesti provedene su na rezultatima 2565 učenika. Ispit je sadržavao ukupno 119 zadataka od čega ih je 79 bilo zatvorenog tipa, a 40 otvorenog tipa. Ispit je bio prosječne težine ($M = 55,2$; $MAX = 119$). Od ukupno 119 zadataka tri su zadatka bila izrazito teška, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika. Tri su zadatka bila izrazito lagana, s indeksom težine većim od 0,9, što znači da ih je točno riješilo više od 90% učenika.

Cronbachov α koeficijent ukazuje na vrlo visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,96$). Visoka vrijednost ovoga koeficijenta očekivana je zbog velikoga broja zadataka jer što je veći broj zadataka, veća je pouzdanost. Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,28. To znači da se sa

sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 4,28$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 18% ukupne varijance ispita.

Ni jedan ispitanik nije postigao niži od 7 niti viši od 115 bodova iako je maksimalan broj bodova na ispitu bio 119. S obzirom na velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 8,4% ih nije imalo zadovoljavajući koeficijent diskriminativnosti.

Ispit državne mature iz Povijesti zadovoljava psihometrijske kriterije te da ga se može smatrati pouzdanim i valjanim instrumentom za mjerenje poznavanja gradiva koje je u njemu zastupljeno.

4.29. Politika i gospodarstvo

Analize ispita državne mature iz Politike i gospodarstva provedene su na rezultatima 8640 učenika. Ispit je sadržavao ukupno 60 zadataka od čega ih je 25 bilo zatvorenog tipa, a 35 otvorenog tipa. Ispit je bio prosječne težine ($M = 32,9$; $MAX = 60$). Od ukupno 60 zadataka četiri su zadatka bila izrazito teška, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika. Sedam je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo više od 90% učenika.

Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,87$). Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 3,05. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 3,05$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 13% ukupne varijance ispita.

Ni jedan ispitanik nije postigao viši od 58 bodova iako je maksimalan broj bodova na ispitu bio 60. S obzirom na velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Vrlo velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 31,7%.

4.30. Filozofija

Analize ispita državne mature iz Filozofije provedene su na rezultatima 538 učenika. Ispit je sadržavao ukupno 55 zadataka od čega ih je 23 bilo zatvorenog tipa, a 14 otvorenog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je pretežak ($M = 34,9$; $MAX = 100$). Od ukupno 55 zadataka pet je zadataka bilo izrazito teško, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika. Jedan je zadatak bio izrazito lagan, s indeksom težine većim od 0,9, što znači da ga je točno riješilo više od 90% učenika.

Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,93$). Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,03. To znači da se sa sigurnošću od oko 68% (2/3) može tvrditi da bi u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 4,03$ bodova. Faktorskom analizom ispita utvrđeno je postojanje dvaju faktora koji objašnjavaju oko 23% i 6% ukupne varijance ispita. Prvi faktor je saturiran kriterijima s esejskoga dijela ispita, a drugi faktor preostalim zadacima.

Ni jedan ispitanik nije postigao niži od 7 bodova niti viši od 82 boda iako je maksimalan broj bodova na ispitu bio 100. S obzirom na velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 16,4% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Jedan je zadatak imao negativan koeficijent diskriminativnosti.

Ispit državne mature iz Filozofije sadrži previše teških i vrlo teških zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše.

5. ODREĐIVANJE PRAGOVA OCJENA

Kako bi kriteriji za postizanje pojedinih ocjena na ispitima bili objektivniji, pristup određivanju pragova ocjena na ispitima državne mature bio je nešto drugačiji nego prilikom svakodnevnoga školskoga ocjenjivanja. Stoga je za svaki predmet iz kojeg su učenici pisali ispite, formirana skupina čiji je zadatak bio izraditi pragove ocjena. Svaka skupina bila je imenovana od strane Centra. Skupine su bile sastavljene od predstavnika sveučilišta i srednjoškolskih profesora.

Za određivanje pragova ocjena može se koristiti više različitih metoda koje su znanstveno utvrđene i provjerene. Odabir najprikladnije određene metode ovisi o vrsti i namjeni ispita te o karakteristikama cjelokupnoga obrazovnoga sustava u kojem se ispiti primjenjuju. Za državnu maturu odabir metode i točan slijed postupaka pri određivanju ocjena napravili su članovi Istraživačko-razvojnoga odjela u suradnji s prof. R. Hambletonom, ravnateljem Centra za edukacijska mjerenja Sveučilišta u Massachusettsu, Amherst, SAD. Pri tome je odabrana modificirana Angoffova metoda određivanja pragova prolaznosti.

5.1. Angoffova metoda određivanja pragova prolaznosti

U Angoffovoj metodi, prag prolaznosti na ispitu izračunava se iz očekivanih rezultata za pojedine zadatke (Assessment System Corporation, 2009). Ocjenjivač razmatra svaki zadatak i donosi prosudbu o vjerojatnosti da će pristupnik koji posjeduje znanje tek dovoljno za dobivanje prolazne ocjene odgovoriti točno na zadatak. Na temelju prosudbi svih zadataka iz ispita, donosi se sud o tome koliki minimalni broj bodova učenik treba postići kako bi položio ispit. Osim procjena zadataka od strane stručnjaka, može se koristiti i informacija o stvarnom postotku riješenosti pojedinih zadataka. Na taj način ocjenjivači mogu korigirati svoje procjene.

Na isti način se određuju i pragovi ostalih ocjena u ispitu. Primjerice, ocjenjivač razmatra svaki zadatak i donosi prosudbu o vjerojatnosti da će pristupnik koji posjeduje znanje tek dovoljno za ocjenu „vrlo dobar“ odgovoriti točno na zadatak.

Pri korištenju Angoffove metode, ključan je koncept učenika koji posjeduje „tek dovoljno znanje“ za dobivanje određene ocjene (Assessment System Corporation, 2009).

LITERATURA

- Assessment System Corporation (2009). *User's Manual for the Angoff Analysis Tool. An Excel Workbook for Analyzing Modified Angoff Studies*. St. Paul: Assessment System Corporation.
- Državni izpitni center. (2007). *Letno poročilo. Splošna matura 2007*. Ljubljana: Državni izpitni center.
- Field, A. (2005). *Discovering Statistics Using SPSS*. London: Sage.
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kehoe, J. (1997). *Basic Item Analysis for Multiple-Choice Tests*.
<http://www.ericdigestd.org/1997-1/basic.html>
- Norusis, M. J. (1998). *SPSS/PC+ Advanced Statistics V2.0 for the IBM PC/XT/AT and PS/2*. Chicago: SPSS Inc.
- Osterlind, S. J. (2001). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats*. Boston: Kluwer Academic Publishers.
- Petz, B. (2004). *Osnovne statističke metode za nematematičare*. Jastrebarsko: Naklada Slap.
- Petz, B. (2005). *Psihologijski rječnik*. Jastrebarsko: Naklada Slap.
- Tacq, J. (1997). *Multivariate Analysis Techniques in Social Science Research: From Problem to Analyse*. London: Sage.
- Tucker, S. (2007). *Using Remark Statistics for Test Reliability and Item Analysis*. Neobjavljeni rad. Baltimore: University of Maryland.

PRILOG I.

PRIKAZ FORMULA KORIŠTENIH U PSIHOMETRIJSKOJ ANALIZI ISPITA DRŽAVNE MATURE

Aritmetička sredina

$$M = \frac{\sum_{i=1}^N X_i}{N}$$

M – aritmetička sredina

X_i – individualni rezultati u varijabli X

N – broj rezultata u varijabli

Standardna devijacija

$$SD = \sqrt{\frac{\sum (X_i - M)^2}{N}}$$

SD – standardna devijacija

X_i – rezultati u varijabli X ($i=1, \dots, N$)

M – aritmetička sredina u varijabli

N – broj rezultata u varijabli

Cronbachov α koeficijent

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum V_z}{V_u} \right)$$

α – Cronbachov α koeficijent

k – broj zadataka u ispitu

V_z – varijanca pojedinog zadatka

V_u – varijanca cijelog ispita

Standardna pogriješka mjerenja

$$SPM = SD\sqrt{1-\alpha}$$

SPM – standardna pogriješka mjerenja

SD – standardna devijacija

α – Cronbachov α koeficijent

Aritmetička sredina zadatka

$$M_k = \frac{\sum_{i=1}^N k_i}{N}$$

M_k – aritmetička sredina zadatka

k_i – individualni rezultat na zadatku k

N – broj rezultata na zadatku

Težina zadatka

$$IT = \frac{M_k}{t_{k(\max)}}$$

IT – indeks težine zadatka

M_k – aritmetička sredina uratka na zadatku k

$T_{k(\max)}$ – maksimalni mogući broj bodova u zadatku k

Koeficijent diskriminativnosti

$$KD = \frac{\sum_{i=1}^n (p_{ki} - \overline{p_k}) (\sum_{j \neq k}^m p_{ji} - \sum_{j \neq k}^m \overline{p_j})}{n \cdot SD(p_k) \cdot SD(\sum_{j \neq k}^m p_{ji})}$$

KD – koeficijent diskriminativnosti

p_{ki} – bodovi učenika i na zadatku k

$\overline{p_k}$ – aritmetička sredina bodova na zadatku k

$SD(p_k)$ – standardna devijacija rezultata na zadatku

$\sum_{j \neq k}^m p_{ji}$ – ukupna suma bodova za sve zadatke na ispitu

$SD(\sum_{j \neq k}^m p_{ji})$ – standardna devijacija ukupnih rezultata na ispitu bez zadatka k

$\sum_{j \neq k}^m \bar{p}_j$ – aritmetička sredina ukupnih rezultata na ispitu bez zadatka k

n – broj učenika

m – broj zadataka