



Psihometrijska analiza ispita probne državne mature

Školska godina 2008./ 2009.

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:

Natalija Ćurković, dipl. psiholog

Josip Šabić, 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 probne državne mature u travnju i svibnju 2009. 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), Talijanski materinski jezik (osnovna razina), Mađarski jezik te Grčki jezik.

2. OBILJEŽJA ISPITA PROBNE 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 probne državne mature s obzirom na ove tri kategorije.

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

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

	Likovna umjetnost Logika
Težak	Informatika Biologija Kemija Fizika Vjeronauk Psihologija Sociologija Geografija Povijest

S obzirom da prilikom primjene ispita probne državne mature učenici uglavnom nisu bili motivirani za postizanje maksimalnoga učinka jer postignuti rezultati nisu imali nikakve posljedice za učenike, bilo je realnije očekivati da će primijenjeni ispiti imati tendenciju biti teži nego lakši. Može se zaključiti da bi ispiti koji su prilikom ove primjene bili teški, u uvjetima motiviranosti za postizanje maksimalnoga učinka (npr. državna matura) bili prosječne težine, dok bi prosječno teški ispiti bili lagani. Ispiti koji su prilikom ove primjene bili lagani, u tom slučaju vjerojatno bi bili vrlo lagani i kao takvi imali bi vrlo nisku diskriminativnu valjanost, odnosno malu mogućnost razlikovanja učenika.

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. Međutim, postoji i drugo moguće objašnjenje, a to je nedostatak motivacije učenika zbog čega se nisu dovoljno potrudili u postizanju maksimalnoga učinka.

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

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 probne državne mature iz sljedećih predmeta možemo smatrati pouzdanim instrumentima: Matematika (viša razina), Srpski jezik, Španjolski jezik (viša razina), Talijanski jezik (viša i osnovna razina), Francuski jezik (viša razina), Njemački jezik (viša razina), Engleski jezik (viša razina), Talijanski materinski jezik (viša razina), Biologija, Kemija, Likovna umjetnost, Psihologija, Geografija i Povijest.

Najveće odstupanje od zadanoga kriterija prisutno je kod ovih ispita: Hrvatski jezik (osnovna razina, 0,78) i Logika (0,75).

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 zadatcima 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: Engleski jezik (osnovna razina) – 11,5%, Biologija - 13,3%, Kemija – 16,9%, Fizika - 17,5%, Etika - 14,3% i Psihologija – 11,1%.

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) – 25%, Hrvatski jezik

(osnovna razina) – 39,6%, Srpski jezik – 22,6%, Latinski jezik (viša razina) – 33,3%, Talijanski materinski jezik (viša razina) – 24,6%, Kemija – 20,8%, Etika – 47,6%, Likovna umjetnost – 30%, Logika – 48,6%, Sociologija – 33,9% i Geografija – 20,7%.

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, Latinski jezik (viša i osnovna razina), Španjolski jezik (viša razina), Talijanski jezik (osnovna razina), Francuski jezik (viša razina), Talijanski materinski jezik (viša razina), Etika i Glazbena umjetnost. Stoga kod ovih ispita nije bilo moguće provesti faktorsku analizu.

Kod nekih ispita utvrđeno je da mjere dva predmeta mjerenja. To su ispiti iz sljedećih predmeta: Talijanski jezik (viša razina), Kemija i Likovna umjetnost. Ispiti iz Hrvatskoga jezika (viša i osnovna razina) mjere tri predmeta mjerenja. 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 PROBNE 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 probne državne mature iz Matematike (viša razina), Francuskoga jezika (viša razina), Njemačkoga jezika (viša i osnovna razina) i Povijesti. Raspodjela ispita prema razini valjanosti prikazana je u tablici 2.

Tablica 2. Raspodjela ispita probne državne mature prema razini valjanosti

RAZINA VALJANOSTI	PREDMET IZ KOJEGA SE PISAO ISPIT PROBNE DRŽAVNE MATURE
Visoka	Matematika (viša razina) Francuski jezik (viša razina) Njemački jezik (viša razina) Njemački jezik (osnovna razina) Povijest
Zadovoljavajuća	Matematika (osnovna razina) Srpski jezik Latinski jezik (viša razina) Latinski jezik (osnovna razina) Španjolski jezik (viša razina) Talijanski jezik (viša razina) Talijanski jezik (osnovna razina) Engleski jezik (viša razina) Engleski jezik (osnovna razina) Talijanski materinski (viša razina) Informatika Biologija Fizika Glazbena umjetnost Likovna umjetnost Logika Sociologija Geografija Vjeronauk Psihologija
Niska	Hrvatski jezik (viša razina) Hrvatski jezik (osnovna razina) Kemija Etika

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 probne državne mature iz Francuskoga jezika (osnovna razina), Talijanskoga materinskoga jezika (osnovna razina), Španjolskoga jezika (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 probne državne mature

Broj zadataka - broj zadataka u ispitu

O - broj zadataka otvorenoga tipa

Z - broj zadataka zatvorenoga tipa

PO - broj zadataka poluotvorenoga tipa (nadopunjavanje, kratki odgovori i sl.)

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 probne državne mature iz Hrvatskoga jezika i Matematike

Predmet	Hrvatski jezik (viša razina)	Hrvatski jezik (osnovna razina)	Matematika (viša razina)*	Matematika (osnovna razina)**
N	22591	10532	14955; 4785 (32%)	16936; 11219 (66%)
Broj zadataka	92	91	45	33
O	11	11	30	17
Z	81	80	15	16
PO	0	0	0	0
Raspon	0 - 157	0 – 154	0 – 60	0 – 40
MAX	160	154	60	40
M	93,6	84,5	27,0	18,5
SD	27,74	31,37	12,49	7,20
SPM	11,76	14,71	3,75	3,08
IT<0,1	0	1	1	1
IT>0,9	5	0	0	0
IT%	5,4	1,1	2,2	3,0
IP(α)	0,82	0,78	0,91	0,82
KD%	25,0	39,6	0	15,2
FA	3	3	1	1

* - iz psihometrijskih analiza ispita iz Matematike (viša razina) izostavljeni su rezultati učenika koji na drugom dijelu ispita nisu ponudili odgovor na više od 10 zadataka. Ukupni broj učenika je bio 14955, a analize su provedene na rezultatima 4785 učenika, tj. na 32% rezultata.

** - iz psihometrijskih analiza ispita iz Matematike (osnovna razina) izostavljeni su rezultati učenika koji na drugom dijelu ispita nisu ponudili odgovor na više od 8 zadataka. Ukupni broj učenika je bio 16936, a analize su provedene na rezultatima 11219 učenika, tj. na 66% rezultata.

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

Predmet	Srpski jezik	Latinski jezik (viša razina)	Latinski jezik (osnovna razina)	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)
N	156	70	47	60	433	81	110	1811	2700	15896	11906	100
Broj zadataka	84	60	50	64	64	52	64	69	52	69	52	61
O	12	0	0	4	4	2	4	4	2	4	2	11
Z	54	30	25	60	52	40	60	57	44	57	40	50
PO	18	30	25	0	8	10	0	8	6	8	10	0
Raspon	0 - 95	23 - 90	16 - 89	20,7 – 95,8	1,9 – 100	9 – 100	0 – 95,8	0 – 96,9	0 – 100	0 – 98,2	0 – 100	0 – 94
MAX	112	118	95	100	100	100	100	100	100	100	100	100
M	57,3	53,3	50,7	60,4	56,7	65,6	60,4	53,3	51,2	65,0	69,4	60,5
SD	19,79	17,55	16,34	18,90	22,19	21,00	19,78	23,60	22,52	19,17	21,93	17,00
SPM	6,25	7,05	6,03	5,98	5,87	6,64	5,60	5,28	7,80	5,07	7,91	5,10
IT<0,1	3	3	1	0	1	0	1	1	0	0	0	0
IT>0,9	3	1	1	6	2	3	1	4	1	6	6	1
IT%	7,1	6,7	4,0	9,4	4,7	5,8	3,1	7,2	1,9	8,7	11,5	1,6
IP(α)	0,90	0,84	0,86	0,90	0,93	0,90	0,92	0,95	0,88	0,93	0,87	0,91
KD%	22,6	33,3	16,0	15,6	4,7	13,5	9,4	0	1,9	7,2	1,9	24,6
FA	-	-	-	-	2	-	-	1	1	1	1	-

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

Predmet	Informatika	Biologija	Kemija	Fizika
N	2190	7612	2283	7162
Broj zadataka	32	128	77	40
O	5	0	32	16
Z	24	64	43	24
PO	3	64	2	0
Raspon	0 - 31	0 - 124	0 – 98	0 – 77
MAX	32	128	100	80
M	12,7	36,3	31,9	23,6
SD	5,50	20,68	15,61	12,76
SPM	2,27	4,14	4,41	4,76
IT<0,1	3	17	13	7
IT>0,9	0	0	0	0
IT%	9,4	13,3	16,9	17,5
IP(α)	0,83	0,96	0,92	0,86
KD%	12,5	9,4	20,8	15,0
FA	1	1	2	1

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

Predmet	Etika	Glazbena umjetnost	Likovna umjetnost	Vjeronauk	Logika	Psihologija	Sociologija	Geografija	Povijest
N	42	138	2354	266	342	3602	1411	2028	3324
Broj zadataka	45	61	90	41	35	63	56	116	119
O	15	0	20	7	0	9	3	6	0
Z	20	57	60	24	24	18	44	72	79
PO	10	4	10	10	11	36	9	38	40
Raspon	0 – 40	32 – 98	15 - 137	4 - 48	0 – 55	3 - 104	0 – 61	0 – 184	0 - 111
MAX	60	100	160	54	60	117	72	200	119
M	26,4	68,0	83,7	21,1	27,4	39,8	28,8	81,9	49,7
SD	9,70	14,73	23,74	8,47	7,57	18,23	8,95	27,2	19,45
SPM	3,22	5,10	6,80	2,90	3,78	5,16	3,78	7,7	4,35
IT<0,1	5	0	5	3	0	7	5	8	10
IT>0,9	1	6	1	0	1	0	0	0	1
IT%	14,3	9,8	6,7	7,3	2,9	11,1	8,9	6,9	9,2
IP(α)	0,89	0,88	0,92	0,88	0,75	0,92	0,82	0,92	0,95
KD%	47,6	19,7	30,0	12,2	48,6	7,9	33,9	20,7	11,8
FA	-	-	2	1	1	1	1	1	1

4. OCJENA KVALITETE SVAKOGA POJEDINOGA ISPITA PROBNE DRŽAVNE MATURE

4.1. Hrvatski jezik (viša razina)

Analize ispita probne državne mature iz Hrvatskoga jezika za višu razinu provedene su na rezultatima 22591 učenika. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 93,6$; $MAX = 160$). Od ukupno 92 zadatka pet je zadataka bilo izrazito lagano, s indeksom težine većim od 0,9, što znači da ih je točno riješilo preko 90% učenika. Pouzdanost izražena Cronbachovim α koeficijentom iznosi 0,82. Iz ovoga rezultata zaključujemo da je ispitni materijal osrednje homogenosti. Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 11,76. 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 $\pm 11,76$ bodova. Faktorskom analizom ispita utvrđeno je postojanje tri faktora, gdje prvi faktor objašnjava oko 11% ukupne varijance ispita, drugi faktor oko 4% ukupne varijance ispita, a treći faktor oko 3% ukupne varijance ispita. Ova tri faktora mogli bismo imenovati kao faktor čitanja i razumijevanja teksta na temelju kojeg se piše esej, faktor pismenog izražavanja i uporabe hrvatskoga jezika u eseju, te faktor poznavanja pravopisa. Ovi faktori su zapravo artefakt različitih tipova zadataka u ispitu. Ispit je sadržavao dvije cjeline. Prva se cjelina sastojala od zadataka višestrukoga izbora, a druga cjelina od zadataka otvorenoga tipa. Niska međusobna povezanost dviju cjelina ($r = 0,39$) ukazuje na potrebu preispitivanja kriterija i načina vrjednovanja eseja kao i na potrebu sadržajne dorade zatvorenih zadataka.

Ni jedan ispitanik nije postigao rezultat viši od 157 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 25,0%. Ovaj nalaz u skladu je s trofaktorskom strukturom ispita, tj. budući da se ispit sastoji od triju skupina zadataka, koje mjere različite predmete mjerenja, nije moguće očekivati visoku korelaciju uspjeha na pojedinim zadacima s uspjehom na cijelome ispitu.

Zaključno se može reći da na ispitu probne državne mature iz Hrvatskoga jezika za višu razinu nije postignut puni raspon bodova te da ispit mjeri minimalno tri različita predmeta mjerenja, odnosno ispit 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 probne državne mature iz Hrvatskoga jezika za osnovnu razinu provedene su na rezultatima 10532 učenika. Ispit je bio prosječne težine ($M = 84,5$; $MAX = 154$). Od ukupno 91 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. Cronbachov α koeficijent ukazuje na nisku razinu homogenosti ispitnoga materijala ($\alpha = 0,78$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 14,71. 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 $\pm 14,71$ bodova. Faktorskom analizom ispita utvrđeno je postojanje tri faktora, gdje prvi faktor objašnjava oko 12% ukupne varijance ispita, drugi faktor oko 6% ukupne varijance ispita, a treći faktor oko 3% ukupne varijance ispita. Ova tri faktora mogli bismo imenovati kao faktor pismenoga izražavanja, faktor poznavanja gramatike, te faktor poznavanja pravopisa. Ovi faktori su zapravo artefakt različitih tipova zadataka u ispitu. Ispit je sadržavao dvije cjeline. Prva se cjelina sastojala od zadataka višestrukoga izbora, a druga cjelina od zadataka otvorenoga tipa. Niska međusobna povezanost dviju cjelina ($r = 0,31$) ukazuje na potrebu preispitivanja kriterija i načina vrjednovanja eseja kao i na potrebu sadržajne dorade zatvorenih zadataka.

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.

Vrlo velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 39,6%. Jedan zadatak je imao negativan koeficijent diskriminativnosti. Ovaj nalaz u skladu je s trofaktorskom strukturom ispita, tj. budući da se ispit sastoji od triju skupina zadataka, koje mjere različite predmete mjerenja, nije moguće očekivati visoku korelaciju uspjeha na pojedinim zadacima s uspjehom na cijelome ispitu.

Zaključno se može reći da ispit probne državne mature iz Hrvatskoga jezika za osnovnu razinu mjeri minimalno tri različita predmeta mjerenja, odnosno ispit nema zadovoljavajuću konstruktivnu valjanost.

4.3. Matematika (viša razina)

Analize ispita probne državne mature iz Matematike za višu razinu provedene su na rezultatima 4785 učenika. Ispit je rješavalo 14955 učenika, no veliki je broj učenika predao ispite u kojima većinu zadataka nisu ni pokušali riješiti. Ispiti tih učenika su izostavljeni iz analiza. Ispit je sadržavao ukupno 45 zadataka od čega ih je 15 bilo zatvorenog, a 30 otvorenog tipa. Ispit je bio prosječne težine ($M = 27,0$; $MAX = 60$). Od ukupno 45 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. Cronbachov α koeficijent ukazuje na visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,91$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 3,75. 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 $\pm 3,75$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 23% 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.

Iako bi se iz onoga što je prethodno rečeno o ovome ispitu moglo zaključiti da ispit probne 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, valja naglasiti da je potreban oprez pri donošenju čvrstih zaključaka zbog problema nedovoljne motivacije za rješavanje ispita probne državne mature iz Matematike koja je postojala kod velikoga broja učenika.

4.4. Matematika (osnovna razina)

Analize ispita probne državne mature iz Matematike za osnovnu razinu provedene su na rezultatima 11219 učenika. Ispit je rješavalo 16936 učenika, no veliki je broj učenika predao ispite u kojima većinu zadataka nisu ni pokušali riješiti. Ispiti tih učenika su izostavljeni iz analiza. Ispit je sadržavao ukupno 33 zadatka od čega ih je 16 bilo zatvorenog, a 17 otvorenog tipa. Ispit je bio prosječne težine ($M = 18,5$; $MAX = 40$). Od ukupno 33 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. Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,82$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 3,08. 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 $\pm 3,08$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 17% 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 15,2% ih nije imalo zadovoljavajući koeficijent diskriminativnosti.

Iako bi se iz onoga što je prethodno rečeno o ovome ispitu moglo zaključiti da ispit probne 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, valja naglasiti da je potreban oprez pri donošenju čvrstih zaključaka zbog problema nedovoljne motivacije za rješavanje ispita probne državne mature iz Matematike koja je postojala kod velikoga broja učenika.

4.5. Srpski jezik

Analize ispita probne državne mature iz Srpskoga jezika provedene su na rezultatima 156 učenika. Ispit je sadržavao ukupno 84 zadatka od čega ih je 54 bilo zatvorenog, 18 poluotvorenog i 12 otvorenog, esejskog tipa. Ispit je bio prosječne težine ($M = 57,3$; $MAX = 112$). Od ukupno 84 zadatka 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 visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,90$). 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 6,25. 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 $\pm 6,25$ bodova.

Velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 22,6%.

Ispit probne državne mature iz Srpskoga jezika ima solidna metrijska obilježja, tj. riječ je o pouzdanome i valjanome instrumentu, no ima previše zadataka niske diskriminativnosti. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.6. Latinski jezik (viša razina)

Analize ispita probne državne mature iz Latinskoga jezika za višu razinu provedene su na rezultatima 70 učenika. Ispit je sadržavao ukupno 60 zadataka od čega ih je 30 bilo zatvorenog, a 30 poluotvorenog tipa. Ispit je bio prosječne težine ($M = 53,3$; $MAX = 118$). Od ukupno 60 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. Jedan zadatak je 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 ($\alpha = 0,84$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 7,05. 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 $\pm 7,05$ bodova.

Vrlo velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 33,3%. Jedan zadatak je imao negativan koeficijent diskriminativnosti.

Ispit probne državne mature iz Latinskoga jezika za višu razinu jezika ima solidna metrijska obilježja, no ima previše zadataka niske diskriminativnosti. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.7. Latinski jezik (osnovna razina)

Analize ispita probne državne mature iz Latinskoga jezika za osnovnu razinu provedene su na rezultatima 47 učenika. Ispit je sadržavao ukupno 50 zadataka od čega ih je 25 bilo zatvorenog, a 25 poluotvorenog tipa. Ispit je bio prosječne težine ($M = 50,7$; $MAX = 95$). Od ukupno 50 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 zadatak je 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 6,03. 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 $\pm 6,03$ bodova.

Od ukupnoga broja zadataka 16,0% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Jedan zadatak je imao negativan koeficijent diskriminativnosti.

Ispit probne državne mature iz Latinskoga jezika za osnovnu razinu ima solidna metrijska obilježja, no ima previše zadataka niske diskriminativnosti. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.8. Španjolski jezik (viša razina)

Analize ispita probne državne mature iz Španjolskoga jezika za višu razinu provedene su na rezultatima 60 učenika. Ispit je sadržavao ukupno 64 zadataka od čega ih je 60 bilo zatvorenog, a 4 otvorenog, esejskog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 60,4$; $MAX = 100$). Od ukupno 64 zadataka šest 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 pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,98. 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 $\pm 5,98$ bodova.

Od ukupnoga broja zadataka 15,6% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Dva zadatka su imala negativan koeficijent diskriminativnosti.

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

4.9. Talijanski jezik (viša razina)

Analize ispita probne državne mature iz Talijanskoga jezika za višu razinu provedene su na rezultatima 433 učenika. Ispit je sadržavao ukupno 64 zadatka od čega ih je 52 bilo zatvorenog, 8 poluotvorenog i 4 otvorenog, esejskog tipa. Ispit je bio prosječne težine ($M = 56,7$; $MAX = 100$). Od ukupno 64 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. 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 ($\alpha = 0,93$). Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,87. 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 $\pm 5,87$ bodova. Faktorskom analizom ispita utvrđeno je postojanje dva faktora, gdje prvi faktor objašnjava oko 21% ukupne varijance ispita, a drugi faktor oko 11% ukupne varijance ispita. Ova dva faktora mogli bismo imenovati kao faktor razumijevanja pročitanaoga teksta i faktor razumijevanja slušanaoga teksta. Dijelovi ispita kojima se ispitivalo razumijevanje pročitanaoga teksta i razumijevanje slušanaoga teksta međusobno nisko koreliraju ($r = 0,30$).

Ni jedan ispitanik nije postigao rezultat niži od 1,9 bodova. S obzirom na vrlo velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 4,7% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Jedan zadatak je imao negativan koeficijent diskriminativnosti.

Ispit probne državne mature iz Talijanskoga jezika za višu razinu mjeri minimalno dva različita predmeta mjerenja, odnosno ispit nema zadovoljavajuću konstruktnu valjanost.

4.10. Talijanski jezik (osnovna razina)

Analize ispita probne državne mature iz Talijanskoga jezika za osnovnu razinu provedene su na rezultatima 81 učenika. Ispit je sadržavao ukupno 52 zadatka od čega ih je 40 bilo zatvorenog, 10 poluotvorenog i 2 otvorenog, esejskog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 65,6$; $MAX = 100$). Od ukupno 52 zadatka 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 visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,90$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 6,64. 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 $\pm 6,64$ bodova.

Od ukupnoga broja zadataka 13,5% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Dva zadatka su imala negativan koeficijent diskriminativnosti.

Ispit probne 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.

4.11. Francuski jezik (viša razina)

Analize ispita probne državne mature iz Francuskoga jezika za višu razinu provedene su na rezultatima 110 učenika. Ispit je sadržavao ukupno 64 zadatka od čega ih je 60 bilo zatvorenog, a 4 otvorenog, esejskog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 60,4$; $MAX = 100$). Od ukupno 64 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 zadatak je 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 ($\alpha = 0,92$). Standardna pogreška mjerenja

koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,60. 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 $\pm 5,60$ bodova.

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

Ispit probne 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. Ovaj ispit ima dobre metrijske karakteristike. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.12. Njemački jezik (viša razina)

Analize ispita probne državne mature iz Njemačkoga jezika za višu razinu provedene su na rezultatima 1811 učenika. Ispit je sadržavao ukupno 69 zadataka od čega ih je 57 bilo zatvorenog, 8 poluotvorenog i 4 otvorenog, esejskog tipa. Ispit je bio prosječne težine ($M = 53,3$; $MAX = 100$). Od ukupno 69 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. 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 5,28. 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 $\pm 5,28$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 26% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat viši od 96,9 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.

Svi su zadatci imali zadovoljavajuće koeficijente diskriminativnosti.

Ispit probne državne mature iz Njemačkoga jezika za višu razinu ima zadovoljavajuća metrijska obilježja, tj. riječ je o pouzdanome i valjanome instrumentu.

4.13. Njemački jezik (osnovna razina)

Analize ispita probne državne mature iz Njemačkoga jezika za osnovnu razinu provedene su na rezultatima 2700 učenika. Ispit je sadržavao ukupno 52 zadatka od čega ih je 44 bilo zatvorenog, 6 poluotvorenog i 2 otvorenog, esejskog tipa. Ispit je bio prosječne težine ($M = 51,2$; $MAX = 100$). Od ukupno 52 zadatka 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 ($\alpha = 0,88$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 7,80. 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 $\pm 7,80$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 23% 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.

Samo jedan zadatak nije imao zadovoljavajući koeficijent diskriminativnosti.

Ispit probne državne mature iz Njemačkoga jezika za osnovnu razinu ima zadovoljavajuća metrijska obilježja, tj. riječ je o pouzdanome i valjanome instrumentu.

4.14. Engleski jezik (viša razina)

Analize ispita probne državne mature iz Engleskoga jezika za višu razinu provedene su na rezultatima 15896 učenika. Ispit je sadržavao ukupno 69 zadataka od čega ih je 57 bilo zatvorenog, 8 poluotvorenog i 4 otvorenog, esejskog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 65,0$; $MAX = 100$). Od ukupno 69 zadatka šest 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 ($\alpha = 0,93$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,07. 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 $\pm 5,07$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 20% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat viši od 98,2 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 7,2% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Nijedan zadatak nije imao negativan koeficijent diskriminativnosti.

Ispit probne 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.

4.15. Engleski jezik (osnovna razina)

Analize ispita probne državne mature iz Engleskoga jezika za osnovnu razinu provedene su na rezultatima 11906 učenika. Ispit je sadržavao ukupno 52 zadatka od čega ih je 40 bilo zatvorenog, 10 poluotvorenog i 2 otvorenog, esejskog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 69,4$; $MAX = 100$). Od ukupno 52 zadatka šest 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 ($\alpha = 0,87$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 7,91. 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 $\pm 7,91$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 23% 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.

Samo jedan zadatak nije imao zadovoljavajući koeficijent diskriminativnosti. Nijedan zadatak nije imao negativan koeficijent diskriminativnosti.

Ispit probne 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.

4.16. Talijanski materinski jezik (viša razina)

Analize ispita probne državne mature iz Talijanskoga materinskoga jezika za višu razinu provedene su na rezultatima 100 učenika. Ispit je sadržavao ukupno 61 zadatak od čega ih je 50 bilo zatvorenog, a 11 otvorenog, esejskog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je prelagan ($M = 60,5$; $MAX = 100$). Od ukupno 61 zadatka 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 5,10. 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 $\pm 5,10$ bodova.

Velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 24,6%. Četiri su zadatka imala negativan koeficijent diskriminativnosti.

Ispit probne državne mature iz Talijanskoga materinskoga 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. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.17. Informatika

Analize ispita probne državne mature iz Informatike provedene su na rezultatima 2190 učenika. Ispit je sadržavao ukupno 32 zadatka od čega ih je 24 bilo zatvorenog, 3 poluotvorenog i 5 otvorenog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je pretežak ($M = 12,7$; $MAX = 32$). Od ukupno 32 zadatka tri su zadatka bila izrazito teška, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika. Cronbachov α koeficijent ukazuje na osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,83$). Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti

i standardnu devijaciju iznosi 2,27. 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 $\pm 2,27$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 19% 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 12,5% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Nijedan zadatak nije imao negativan koeficijent diskriminativnosti.

Ispit probne državne mature iz Informatike sadrži previše teških zadataka. Bilo bi dobro da je ispit duži, tj. da ima više zadataka. Trebalo bi promijeniti omjer broja zadataka prosječne težine te lakih i teških zadataka tako da zadataka prosječne težine bude najviše.

4.18. Biologija

Analize ispita probne državne mature iz Biologije provedene su na rezultatima 7612 učenika. Ispit je sadržavao ukupno 128 zadataka od čega ih je 64 bilo zatvorenog, a 64 poluotvorenog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je pretežak ($M = 36,3$; $MAX = 128$). Od ukupno 128 zadataka sedamnaest 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 4,14. 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 $\pm 4,14$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 18% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat viši od 124 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 9,4% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Dva zadatka su imala negativan koeficijent diskriminativnosti.

Ispit probne državne mature iz Biologije 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. Kemija

Analize ispita probne državne mature iz Kemije provedene su na rezultatima 2283 učenika. Ispit je sadržavao ukupno 77 zadataka od čega ih je 40 bilo zatvorenog, a 37 otvorenog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je pretežak ($M = 31,9$; $MAX = 100$). Od ukupno 77 zadataka trinaest 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 visoku razinu homogenosti ispitnoga materijala ($\alpha = 0,92$). 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 4,41. 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 $\pm 4,41$ bodova. Faktorskom analizom ispita utvrđeno je postojanje dva faktora, gdje prvi faktor objašnjava oko 19% ukupne varijance ispita, a drugi faktor oko 4% ukupne varijance ispita. Drugi faktor je zapravo artefakt jer su njime saturirani samo najteži zadatci.

Ni jedan ispitanik nije postigao rezultat viši od 98 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.

Velik broj zadataka nije imao zadovoljavajući koeficijent diskriminativnosti, njih 20,8%. Jedan zadatak je imao negativan koeficijent diskriminativnosti.

Ispit probne 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.

4.20. Fizika

Analize ispita probne državne mature iz Fizike provedene su na rezultatima 7162 učenika. Ispit je sadržavao ukupno 40 zadataka od čega ih je 24 bilo zatvorenog, a 16 otvorenog tipa. Ispit težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je pretežak ($M = 23,6$; $MAX = 80$). Od ukupno 40 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 osrednju razinu homogenosti ispitnoga materijala ($\alpha = 0,86$). Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,76. 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 $\pm 4,76$ bodova. Faktorskom analizom ispita utvrđeno je postojanje jednoga faktora koji objašnjava oko 22% ukupne varijance ispita.

Ni jedan ispitanik nije postigao rezultat viši od 77 bodova iako je maksimalan broj bodova na ispitu bio 80. S obzirom na vrlo velik broj ispitanika, bilo je očekivano da će biti postignut puni raspon bodova.

Od ukupnoga broja zadataka 15,0% ih nije imalo zadovoljavajući koeficijent diskriminativnosti. Nijedan zadatak nije imao negativan koeficijent diskriminativnosti.

Ispit probne državne mature iz Fizike 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.

4.21. Etika

Analize ispita probne državne mature iz Etike provedene su na rezultatima 42 učenika. Ispit je sadržavao ukupno 45 zadataka od čega ih je 20 bilo zatvorenog, 10 poluotvorenog i 15 otvorenog, esejskog tipa. Iako prosječni postignuti rezultat na ispitu ukazuje na to da je on bio prosječne težine ($M = 26,4$; $MAX = 60$), on je zapravo bio težak. Većina učenika ostvarila je manje od polovice maksimalnog mogućeg broja bodova, a ukupni raspon bodova varirao je od 0 do 40. Prema tome, distribucija rezultata bila je pozitivno asimetrična. Od ukupno 45 zadataka 5 je bilo izrazito teško, s indeksom težine manjim od 0,1, što znači da ih je točno

riješilo manje od 10% učenika. Nasuprot tome, bio je samo jedan izrazito lagani zadatak kojeg je rješavalo preko 90% učenika.

Pouzdanost ispita izražena Cronbachovim α koeficijentom iznosi 0,89. Standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 3,2. 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 $\pm 3,2$ boda.

Koeficijenti diskriminativnosti zadataka kreću se kod gotovo polovice zadataka (47,6%) u okvirima manjim od prihvatljivih vrijednosti, odnosno ispod 0,3.

Zaključno se o ispitu iz Etike može reći da je on učenicima bio težak te nitko od učenika nije uspio postići maksimalni broj bodova. Velik broj zadataka učenicima je bio težak ili vrlo težak, a tek jedan izrazito lagan. Diskriminativnost zadataka je niska. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.22. Glazbena umjetnost

Analize ispita probne državne mature iz Glazbene umjetnosti provedene su na rezultatima 138 učenika. Ispit je sadržavao ukupno 61 zadatak od čega ih je 57 bilo zatvorenog, 4 poluotvorenog i niti jedan otvorenog tipa. Prosječni postignuti rezultat na ispitu ukazuje na to da je on bio nešto lakši ($M = 68,0$; $MAX = 100$). Tako je minimalni postignuti rezultat bio 32, a maksimalni 98. Prema tome, može se reći da su učenici velikim dijelom savladali gradivo koje je bilo ispitivano ispitom. U skladu s time je i podatak koji govori da je 6 zadataka učenicima bilo izuzetno lagano te ih je rješavalo preko 90% učenika, dok im niti jedan zadatak nije bio izrazito težak.

Pouzdanost ispita izražena Cronbachovim α koeficijentom iznosi 0,88, a standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,1. 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 $\pm 5,1$ boda.

Prosječni koeficijent diskriminativnosti zadataka iznosi 0,31 tj. na granici je prihvatljivog iznosa. Iako postoji velik broj zadataka čiji koeficijenti ukazuju na visoku

diskriminativnost, postoji čak oko 20% zadataka s vrlo niskim ili čak negativnim koeficijentima zbog kojih je prosječni koeficijent tek na graničnoj vrijednosti.

Zaključno se o ispitu iz Glazbene umjetnosti može reći da je on učenicima bio lagan. Iako je diskriminativnost zadataka većinom bila prihvatljiva, kod čak 20% zadataka vrijednosti koeficijentata diskriminativnosti bile su vrlo niske ili čak negativne. Zbog malog broja učenika koji su rješavali ovaj ispit, složenije analize nisu mogle biti provedene.

4.23. Likovna umjetnost

Analize ispita probne državne mature iz Likovne umjetnosti provedene su na rezultatima 2354 učenika. Ispit je sadržavao ukupno 90 zadataka od čega ih je 60 bilo zatvorenog, 10 poluotvorenog i 20 otvorenog tipa. Prosječni postignuti rezultat na ispitu ukazuje na to da je on bio prosječne težine ($M = 68,0$; $MAX = 100$). Međutim, iako su učenici postizali prosječne rezultate na ispitu, nitko nije postigao maksimalni broj bodova, odnosno raspon bodova varirao je između 15 i 137. To govori da je ispit u određenoj mjeri ipak bio učenicima težak. Tako je učenicima 5 zadataka bilo izuzetno teško te ih je riješilo manje od 10% učenika dok im je samo 1 zadatak bio izuzetno lagan i riješili su ga gotovo svi učenici.

Pouzdanost ispita izražena Cronbachovim α koeficijentom iznosi 0,92. Budući da ispit ima veliki broj zadataka, izračunata je i vrijednost koeficijenta α za slučaj kada bi ispit imao standardnu dužinu od 40 zadataka. Tada bi α iznosila 0,83. Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 6,8. 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 $\pm 6,8$ bodova.

Prosječni koeficijent diskriminativnosti zadataka iznosi 0,3 tj. na granici je prihvatljivog iznosa. Iako postoji velik broj zadataka čiji koeficijenti ukazuju na visoku diskriminativnost, postoji čak oko 30% zadataka s vrlo niskim ili čak negativnim koeficijentima zbog kojih je prosječni koeficijent tek na graničnoj vrijednosti.

Faktorskom analizom ispita utvrđeno je postojanje dva faktora koji objašnjavaju tek 19% varijance iz čega se može zaključiti da ispit mjeri heterogene sadržaje koji nisu međusobno visoko povezani. Ova dva faktora odnose se na prvi i drugi dio ispita. Budući da

su oni međusobno visoko povezani, može se govoriti o postojanju jednog nadređenog faktora. Pritom je zanimljivo da su ova dva dijela ispita međusobno negativno povezana, odnosno učenici koji postižu bolje rezultate u prvom dijelu, postižu slabije rezultate u drugom dijelu i obrnuto što je potpuno neočekivano i govori da se u ova dva dijela očito radi o mjerenju različitih koncepata.

Zaključno se o ispitu iz Likovne umjetnosti može reći da su učenici uglavnom postizali polovične rezultate na ispitu. Iako je diskriminativnost zadataka većinom bila prihvatljiva, kod čak 30% zadataka vrijednosti koeficijenata diskriminativnosti bile su vrlo niske ili čak negativne. Ispit se dijeli na dva faktora koja su međusobno negativno povezana.

4.24. Vjeronauk

Analize ispita probne državne mature iz Vjeronauka provedene su na rezultatima 266 učenika. Ispit je sadržavao ukupno 41 zadatak od čega ih je 24 bilo zatvorenog, 10 poluotvorenog i 7 otvorenog tipa. Prosječni postignuti rezultat na ispitu ukazuje na to da je on učenicima bio težak ($M = 21,1$; $MAX = 54$). Raspon rezultata varirao je između 4 i 48. Od svih zadataka u ispitu, učenicima su 3 bila iznimno teška te ih je riješilo manje od 10%, dok im niti jedan zadatak nije bio izuzetno lagana.

Pouzdanost ispita izražena Cronbachovim α koeficijentom iznosi 0,88, a standardna pogriješka mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 2,9. 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 $\pm 2,9$ bodova

Faktorskom analizom ispita utvrđeno je postojanje jednog faktora, koji objašnjava 20% ukupne varijance. Iz toga se može zaključiti da je ispit prilično homogen s obzirom na ispitivani sadržaj.

Prosječni koeficijent diskriminativnosti zadataka iznosi 0,38 te ukazuje na visoku diskriminativnost zadataka. Ipak, postoji 7 zadataka (12,2%) čiji indeksi su ispod prihvatljive vrijednosti od 0,3 pri čemu 2 zadatka imaju gotovo nulte vrijednosti koeficijenta.

Zaključno se o ispitu iz Vjeronauka može reći da je on učenicima bio prilično težak, odnosno u ispitu su dominirali teški zadatci. Zadatci su pretežno bili visoko diskriminativni, a sadržaj ispita je homogen.

4.25. Logika

Analize ispita probne državne mature iz Logike provedene su na rezultatima 342 učenika. Ispit je sadržavao ukupno 35 zadataka od čega ih je 24 bilo zatvorenog i 11 poluotvorenog tipa dok otvorenih zadataka s produženim odgovorima nije bilo. Prosječni postignuti rezultat na ispitu ukazuje na to da je on težinom bio primjeren populaciji koja je njime ispitana, odnosno bio je prosječne težine ($M = 27,4$; $MAX = 60$). Raspon postignutih rezultata varirao je od 0 do 55, odnosno nitko od učenika nije postigao maksimalni mogući broj bodova. Među zadatcima su dominirali oni srednje težine, odnosno oni koje rješava između 40% i 60% učenika te lagani zadatci koje rješava između 61% i 80% učenika. Nadalje, niti jedan zadatak nije bio izrazito težak, a jedan je bio izrazito lagan. Prema tome, u ukupnom broju zadataka, samo je jedan imao ekstremnu vrijednost težine.

Pouzdanost ispita izražena Cronbachovim α koeficijentom iznosi 0,75 i ona govori o nešto manjoj homogenosti ispitnog materijala nego kod ostalih ispita. Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 3,8. 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 $\pm 3,8$ boda.

Faktorskom analizom dobiva se jedan faktor, ali on objašnjava svega 12% varijance i upućuje na heterogenost ispitnog materijala gdje uratci na različitim zadatcima nisu međusobno visoko povezani.

Prosječni koeficijent diskriminativnosti zadataka iznosi 0,23 i upućuje na nisku diskriminativnost. Ispod ove očekivane vrijednosti od 0,3 nalazi ih se čak 48,6%, odnosno gotovo polovica. Pritom većina od tih zadataka ima vrijednosti koeficijenta niže od 0,2 pri čemu je jedan zadatak čak i negativno diskriminativan. Vrijednosti koeficijenta prihvatljive su uglavnom samo u politomnim zadatcima koji „prirodno“ imaju bolju diskriminativnost od

onih dihotomnih jer je na temelju više različitih rezultata koje je na tim zadacima moguće postići osigurano bolje razlikovanje među učenicima.

Zaključno se o ispitu iz Logike može reći da je on učenicima bio prosječne težine, ali nitko od učenika nije uspio postići maksimalni broj bodova. Diskriminativnost zadataka je niska. Ispitivanje konstruktne valjanosti ispita pokazalo je postojanje jednog predmeta mjerenja, ali on ukazuje na heterogenost ispitnog materijala, odnosno rezultati koje su učenici postizali na različitim dijelovima ispita prilično su nepovezani.

4.26. Psihologija

Analize ispita probne državne mature iz Psihologije provedene su na rezultatima 3602 učenika. Ispit je sadržavao ukupno 63 zadatka od čega ih je 36 bilo zatvorenog, 18 poluotvorenog tipa i 9 otvorenih zadataka s produženim odgovorima. Prosječni postignuti rezultat na ispitu ukazuje na to da on težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je vrlo težak ($M = 39,8$; $MAX = 117$). Raspon postignutih rezultata varirao je od 3 do 104, odnosno nitko od učenika nije postigao maksimalni mogući broj bodova. Među zadacima su dominirali oni vrlo teški i teški, odnosno oni koje rješava između 0% i 20% te između 21% i 40% učenika. Nadalje, niti jedan zadatak nije bio izrazito lagan, odnosno nije bio takav da ga je moglo riješiti više od 90% učenika, dok ih je čak 7 bilo izrazito teško što znači da ih je riješilo manje od 10% učenika.

Pouzdanost ispita izražena Cronbachovim α koeficijentom iznosi 0,92 i ona govori o zadovoljavajućoj homogenosti ispitnog materijala. Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 5,2 bodova. 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 $\pm 5,2$ boda.

Faktorskom analizom dobiva se jedan faktor, ali on objašnjava 18% varijance i upućuje na određenu razinu heterogenosti ispitnog materijala.

Prosječni koeficijent diskriminativnosti zadataka iznosi 0,37 i upućuje na vrlo zadovoljavajuću razinu diskriminativnosti zadataka. Ispod očekivane vrijednosti od 0,3 nalazi se svega 7,9% zadataka što znači da velika većina zadataka ima željene koeficijente

diskriminativnosti. Međutim, valja naglasiti da je u ovom ispitu većina zadataka politomna, a politomni zadatci sami po sebi imaju bolju diskriminativnost od onih dihotomnih jer je na temelju više različitih rezultata koje je na tim zadacima moguće postići osigurano bolje razlikovanje među učenicima. Tako su niži koeficijenti diskriminativnosti dobiveni upravo na dihotomnim zadacima gdje jedan ima čak i negativnu vrijednost koeficijenta diskriminativnosti.

Zaključno se o ispitu iz Psihologije može reći da je on učenicima bio težak te nitko od učenika nije uspio postići maksimalni broj bodova. Diskriminativnost zadataka je visoka. Ispitivanje konstruktne valjanosti ispita pokazalo je postojanje jednog predmeta mjerenja, ali on ukazuje na određenu heterogenost ispitnog materijala.

4.27. Sociologija

Analize ispita probne državne mature iz Sociologije provedene su na rezultatima 3411 učenika. Ispit je sadržavao ukupno 56 zadataka od čega ih je 44 bilo zatvorenog, 9 poluotvorenog tipa i 3 otvorena zadataka s produženim odgovorima. Prosječni postignuti rezultat na ispitu ukazuje na to da on težinom nije bio posve primjeren populaciji koja je njime ispitana, odnosno bio je nešto teži ($M = 28,8$; $MAX = 72$). Raspon postignutih rezultata varirao je od 0 do 61, odnosno nitko od učenika nije postigao maksimalni mogući broj bodova. Među zadacima su dominirali oni vrlo teški i teški, odnosno oni koje rješava između 0% i 20% te između 21% i 40% učenika. Nadalje, niti jedan zadatak nije bio izrazito lagan, odnosno nije bio takav da ga je moglo riješiti više od 90% učenika, dok ih je 5 bilo izrazito teško što znači da ih je riješilo manje od 10% učenika. Pritom jedan zadatak nitko nije riješio.

Pouzdanost ispita izražena Cronbachovim α koeficijentom iznosi 0,82 i ona govori o nešto manjoj homogenosti ispitnog materijala. Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 3,8 bodova. 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 $\pm 3,8$ bodova.

Faktorskom analizom dobiva se jedan faktor, ali on objašnjava svega 11% varijance i upućuje na heterogenost ispitnog materijala.

Prosječni koeficijent diskriminativnosti zadataka iznosi 0,26 i upućuje na nižu razinu diskriminativnosti zadataka. Ispod očekivane vrijednosti od 0,3 nalazi se čak 34% zadataka, odnosno jedna trećina.

Zaključno se o ispitu iz Sociologije može reći da je on učenicima bio nešto teži te nitko od učenika nije uspio postići maksimalni broj bodova. Diskriminativnost zadataka nije zadovoljavajuća. Ispitivanje konstruktne valjanosti ispita pokazalo je postojanje jednog predmeta mjerenja, ali on ukazuje na heterogenost ispitnog materijala.

4.28. Geografija

Analize ispita probne državne mature iz Geografije provedene su na rezultatima 2028 učenika. Ispit je sadržavao ukupno 116 zadataka od čega su 72 bila zatvorenog, 38 poluotvorenog, a 6 otvorenog tipa. Prosječni postignuti rezultat na ispitu ukazuje na to da on težinom nije bio posve primjeren populaciji koja je njime ispitana, odnosno bio je malo pretežak ($M = 81,9$; $MAX = 200$). Od ukupno 116 zadataka 8 bilo izrazito teško, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika. Nasuprot tome, nije postojao niti jedan zadatak koji je bio izrazito lagan, odnosno kojeg je rješavalo preko 90% učenika. U ispitu primjerene težine, broj izrazito laganih i izrazito teških zadataka trebao bi biti podjednak, i na ukupni broj zadataka u ovom ispitu (116) trebalo bi ih biti po maksimalno 5 do 6. Budući da je ispit učenicima bio težak, niti jedan učenik nije postigao maksimalni broj bodova te je raspon postignutih rezultata varirao između 0 i 184 boda.

Pouzdanost ispita izražena Cronbachovim α koeficijentom iznosi 0,92. Međutim, kako veličina ovog koeficijenta ovisi o broju zadataka na način da veći broj zadataka znači i veći koeficijent, onda ovako visok koeficijent treba uzeti s rezervom budući da ispit ima veliki broj zadataka. Ukoliko bi ispit imao manji broj zadataka, a da njegova pouzdanost ostane nepromijenjena, koeficijent α iznosio bi 0,80 što govori da učenici nisu bili pretjerano dosljedni prilikom rješavanja ovog ispita. Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 7,7. To znači da bi s vjerojatnošću od 68% (2/3) u ponovljenom mjerenju rezultati učenika varirali u odnosu na dobivene $\pm 7,7$ boda.

Faktorskom analizom ispita utvrđeno je postojanje jednog faktora, ali on objašnjava svega 11% ukupne varijance. Iz toga se može zaključiti da ispit mjeri heterogene sadržaje koji nisu međusobno visoko povezani.

Prosječni koeficijent diskriminativnosti zadataka iznosi oko 0,25 što je na granici prihvatljivih vrijednosti. Međutim, u ovom prosječnom iznosu nalazi se čak 21% zadataka koji imaju manje vrijednosti od željene, a to je 0,3.

Zaključno se o ispitu iz Geografije može reći da je on učenicima bio suviše težak te nitko od učenika nije niti približno uspio postići maksimalni broj bodova. Velik broj zadataka učenicima je bio težak ili vrlo težak, a niti jedan izrazito lagan. Nadalje, diskriminativnost zadataka je na donjoj granici prihvatljivosti. Ispitivanje konstruktne valjanosti ispita pokazalo je postojanje jednog predmeta mjerenja, ali on je vrlo heterogen te su rezultati koji su učenici postizali na različitim dijelovima ispita prilično nepovezani.

4.29. Povijest

Analize ispita probne državne mature iz Povijesti provedene su na rezultatima 3324 učenika. Ispit je sadržavao ukupno 119 zadataka od čega ih je 79 bilo zatvorenog i 40 poluotvorenog tipa dok otvorenih zadataka s produženim odgovorima nije bilo. Prosječni postignuti rezultat na ispitu ukazuje na to da on težinom nije bio primjeren populaciji koja je njime ispitana, odnosno bio je pretežak ($M = 49,7$; $MAX = 119$). Od ukupno 119 zadataka 10 je bilo izrazito teško, s indeksom težine manjim od 0,1, što znači da ih je točno riješilo manje od 10% učenika. Nasuprot tome, bio je samo jedan izrazito lagani zadatak kojeg je rješavalo preko 90% učenika. U ispitu primjerene težine, broj izrazito laganih i izrazito teških zadataka trebao bi biti podjednak, i na ukupni broj zadataka u ovom ispitu (119) trebalo bi ih biti po maksimalno 6. Budući da je ispit učenicima bio težak, niti jedan učenik nije postigao maksimalni broj bodova te je raspon postignutih rezultat varirao između 0 i 111 bodova.

Pouzdanost ispita izražena Cronbachovim α koeficijentom iznosi 0,95. Međutim, kako veličina ovog koeficijenta ovisi o broju zadataka na način da veći broj zadataka znači i veći koeficijent, onda ovako visoki koeficijent treba uzeti s rezervom budući da ispit ima veliki broj zadataka. Ukoliko bi ispit imao manji broj zadataka, a da njegova pouzdanost ostane

nepromijenjena, koeficijent α iznosio bi oko 0,80 što govori da učenici nisu bili pretjerano dosljedni prilikom rješavanja ovog ispita, odnosno da ispitni materijal nije bio homogen. Standardna pogreška mjerenja koja objedinjuje koeficijent pouzdanosti i standardnu devijaciju iznosi 4,4. 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 $\pm 4,4$ boda.

Faktorskom analizom ispita utvrđeno je postojanje jednog faktora, ali on objašnjava svega 15% ukupne varijance. Iz toga se može zaključiti da ispit mjeri heterogene sadržaje koji nisu međusobno visoko povezani.

Prosječni koeficijent diskriminativnosti zadataka kreću se pretežno u okvirima prihvatljivih vrijednosti, odnosno iznad 0,3, a ispod ove očekivane vrijednosti nalazi ih se 11,8%.

Zaključno se o ispitu iz Povijesti može reći da je on učenicima bio težak te nitko od učenika nije uspio postići maksimalni broj bodova. Velik broj zadataka učenicima je bio težak ili vrlo težak, a tek jedan izrazito lagan. Diskriminativnost zadataka je prihvatljiva. Ispitivanje konstruktne valjanosti ispita pokazalo je postojanje jednog predmeta mjerenja, ali on je vrlo heterogen te su rezultati koji su učenici postizali na različitim dijelovima ispita prilično nepovezani.

5. ODREĐIVANJE PRAGOVA OCJENA

Budući da je Probna državna matura bila prvi primijenjeni ispit tog tipa u Hrvatskoj, bilo je potrebno osmisliti način određivanja ocjena na svakom ispitu. Kako bi kriteriji za postizanje pojedinih ocjena na ispitima bili objektivniji, pristup određivanju pragova ocjena 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 obrazovnog sustava u kojem se ispiti

primjenjuju. Za probnu 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 NACIONALNIH ISPITA

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 pogreška mjerenja

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

SPM – standardna pogreška 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

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

KD – koeficijent diskriminativnosti

p_{ki} – bodovi učenika i na zadatku k

\bar{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