

Sveučilište u Zagrebu
Filozofski fakultet
Odsjek za psihologiju

PSIHOMETRIJSKA VALIDACIJA TESTA ST-1

Diplomski rad

Ivana Tuđman

Mentor: Doc. dr. Damir Ljubotina

Zagreb, 2007.

SADRŽAJ

SAŽETAK	2
1. UVOD	4
<i>O inteligenciji</i>	4
<i>O specijalnim sposobnostima</i>	6
<i>O testovima inteligencije</i>	7
<i>O specijalnim testovima</i>	10
2. CILJ ISTRAŽIVANJA	12
3. METODOLOGIJA	13
4. REZULTATI I RASPRAVA	15
<i>Analiza Testa ST-1</i>	15
<i>Analiza zadataka</i>	17
<i>Faktorska analiza</i>	21
<i>Skraćenje testa</i>	24
5. ZAKLJUČAK	30
6. LITERATURA	31
7. PRILOZI	33
<i>Prilog 1</i>	33
<i>Prilog 2</i>	34

Psihometrijska validacija Testa ST-1
The Psychometric Validation of the Test ST-1

Ivana Tudman

SAŽETAK

Iskustva psihologa u praksi pokazala su da sadržaj testova s vremenom postaje poznat, pa postoji stalna potreba za razvojem novih testova inteligencije. Cilj ovog rada je provesti analizu 35 zadataka preliminarne forme Testa ST-1, koji ispituje prostorno razmišljanje, te predložiti njegovu skraćenu verziju.

Test je primijenjen na 212 ispitanika visoke stručne spreme, srednje stručne spreme, polaznika gimnazija i fakulteta različitih usmjerenja, u trajanju od 40 minuta. Dobivena je normalna distribucija rezultata, a prosječni indeks lakoće ($p=0,630$) govori da je test sudionicima bio nešto prelagan. Cronbachov koeficijent ukazuje na visoku pouzdanost testa ($\alpha=0,874$). Faktorskom analizom dobiva se veći broj faktora, no veličine njihovih karakterističnih korijena ukazuju na postojanje jednog glavnog faktora.

Na osnovi rezultata predložena je skraćena verzija testa od 32 zadatka. Pritom se očuvala normalna distribucija rezultata, a metrijske karakteristike tek su se neznatno promijenile. Pouzdanost testa nešto je porasla ($\alpha=0,876$), a u faktorskoj strukturi još više do izražaja dolazi generalni faktor. Možemo zaključiti da je skraćivanjem Test ST-1 dobio na ekonomičnosti, uz povećanu kvalitetu.

U daljnjim bi istraživanjima trebalo provjeriti rezultate dobivene u ovom radu na reprezentativnom uzorku, kao i funkcioniranje kraće verzije u praksi, te provesti standardizaciju. Budući da je ovaj test osmišljen s ciljem testiranja ispitanika putem računala, preporuča se i provjeriti mijenja li takva primjena njegove metrijske karakteristike.

Ključne riječi: *inteligencija, specijalne sposobnosti, test specijalnih sposobnosti, Test ST-1, analiza čestica*

SUMMARY

Practical experience in psychology has shown that the contents of tests gradually become publically known, which generates a need for a constant development of new intelligence tests. The aim of this study is to analyze 35 items of the preliminary ST-1 spatial abilities test and to propose its shortened version.

The test was administered to 212 subjects including university graduates, high school graduates, high school students and university students of different orientations. The allotted time was 40 minutes. The results were normally distributed, and the average

difficulty index ($p = 0.630$) shows that the test may have been too easy. The Cronbach Coefficient Alpha points to a high degree of reliability of the test ($\alpha = 0.874$). The factor analysis results in a number of factors, but the size of their eigenvalues points to the existence of one general factor.

Based on the results obtained, a shorter version of the test is proposed, containing 32 items. Normal distribution of results is preserved, with only slight changes in metric characteristics. Reliability has increased ($\alpha=0.876$), while the factor structure makes the general factor even more prominent. To conclude, the shortened ST-1 Test features improved efficiency, as well as quality.

Further research should verify the results obtained in this study on a representative sample, apply the shorter version of the test in practice and conduct standardization. Since this test was designed to test examinees via computers, it is recommended to check whether its metric characteristics are changed through this means of test application.

Keywords: intelligence, spatial abilities, spatial abilities test, ST-1 Test, item analysis

1. UVOD

O inteligenciji

Inteligencija je jedan od najznačajnijih čimbenika za niz ljudskih aktivnosti, te ima vrlo složene veze s konativnom, emocionalnom i motivacijskom strukturom i funkcioniranjem ličnosti (Zarevski, 2000). Univerzalna definicija inteligencije zapravo i ne postoji, jer među vodećim psiholozima postoji neslaganje o tome što je uopće inteligencija pa različiti autori nude brojne različite definicije, ovisno o pristupu kojeg koriste. Psihologijski rječnik (Petz, 1992) inteligenciju definira kao "jedan od najvažnijih psihologijskih konstrukata koji nije jednoznačno definiran. Najčešće označava svojstvo uspješnog snalaženja jedinke u novim situacijama, u kojima ne pomaže stereotipno nagonsko ponašanje, a niti učenjem stečene navike, vještine i znanja. Za razliku od ovih neplastičnih i reproduktivnih oblika koji su korisni samo u nekim prilikama, inteligencija je svojstvo jedinke da pronalazi nove prilagođene reakcije u prilikama bilo koje vrste."

Suvišno je raspravljati o tome koja je definicija inteligencije točna, jer svaka obuhvaća samo manji ili veći dio onoga što čini inteligenciju, pa je i svaka definicija manje ili više točna (Petz, 2001). Budući da nije izgledno da će se u skorije vrijeme pojaviti sveobuhvatna i svima prihvatljiva definicija inteligencije, možda bi bilo najbolje prikloniti se Boringovom (1923, prema Zarevski, 2000) operacionalnom shvaćanju i njegovoj tvrdnji da je "inteligencija ono što mjere testovi inteligencije."

Priroda inteligencije može se proučavati sa strukturalističkog ili funkcionalističkog stajališta. Funkcionalistički pristup bavi se odvijanjem kognitivnih funkcija, dok strukturalistički pomoću faktorske analize nastoji utvrditi kakva je struktura kognitivnih sposobnosti.

Jedno od glavnih pitanja strukturalističkog pristupa je da li je inteligencija jedinstvena opća sposobnost ili se ipak sastoji od više posebnih sposobnosti (Vasta i sur., 1997). Najraniji zagovornik stajališta o općoj inteligenciji je Charles Spearman koji je predložio dvofaktorsku teoriju inteligencije. On je u proučavanju inteligencije prvi upotrijebio faktorsku analizu, i zaključio da postoji jedan generalni faktor inteligencije

(g) koji predstavlja opću mentalnu energiju i brojni specifični faktori (s), podložni vježbi i obrazovanju.

Drugi su teoretičari zastupali shvaćanje o postojanju više faktora. Tako je Thurstone utvrdio postojanje sedam tzv. primarnih mentalnih sposobnosti: verbalnog razumijevanja, verbalne fluentnosti, pamćenja, perceptivne brzine, specijalnih sposobnosti, numeričkog faktora i rezoniranja. On smatra da su te sposobnosti uglavnom neovisne i podjednako važne. Guilford je išao tako daleko u raščlanjivanju kognitivnih sposobnosti da je dobio čak 120 nezavisnih faktora.

Gardner u svojoj teoriji višestrukih inteligencija predlaže postojanje sedam potpuno nezavisnih vrsta inteligencije: jezik, logičko-matematička inteligencija, specijalne vještine, tjelesno-kinestetička nadarenost, muzička nadarenost, interpersonalne vještine i intrapersonalne vještine (kasnije je dodao i prirodoslovnu inteligenciju, kakvu je na primjer imao Darwin). O svakoj od njih govori kao o "inteligenciji" jer se one međusobno mogu jako razlikovati. Gardner također smatra da svaka od njih ima svoju neurološku osnovicu u različitim dijelovima mozga (Rathus, 2000).

Cattell (1971, prema Petz, 2001) dijeli inteligenciju na dvije različite "vrste": fluidnu (koja predstavlja sposobnost stjecanja novih znanja i spoznaja i rješavanja novih problema) i kristaliziranu (koja označava korištenje postojećeg nagomilanog znanja).

Prvi empirijski hijerarhijski model ponudio je Vernon (1950, prema Zarevski, 2000). Njegov model uključuje četiri razine: na prvoj je generalna kognitivna sposobnost, na drugoj su verbalno-edukacijska (v:ed) i specijalno-mehanička (k:m) sposobnost, a na trećoj se v:ed grana na verbalnu i numeričku sposobnost, dok se k:m dijeli na specijalnu, manualnu i sposobnost korištenja mehaničkih informacija. Najniža razina uključuje niz specifičnih faktora.

Neovisno o modelu i shvaćanju inteligencije, činjenica je da se pojedinci razlikuju prema stupnju razvijenosti kognitivnih sposobnosti, a njihove individualne razlike relativno su stabilne i mogu se mjeriti pomoću odgovarajućih testova inteligencije.

O spacijalnim sposobnostima

Prostorna sposobnost može se definirati kao sposobnost generiranja, zadržavanja i transformiranja mentalnih reprezentacija predmeta u prostoru, te njihovog odnosa s drugim predmetima i položajima u tom prostoru (Sternberg, 1982). Mnogi poznati umovi poput Alberta Einsteina, Michaela Faradaya, Hermana Von Helmholtza, Nikole Tesle, Benjamina Franklina i Francisa Galtona ustvrdili su da je prostorna sposobnost imala važnu ulogu u njihovim najvećim postignućima. U nekih pojedinaca spacijalne sposobnosti mogu biti nevjerojatno razvijene. Primjer je Nikola Tesla, koji je sebi mogao predočiti neki stroj do njegovog zadnjeg detalja. Tvrdio je da čak može u glavi testirati svoje izume (Svetličić, 2004).

Vizuo-spacijalno funkcioniranje, najčešće ispitivan vid spacijalnih sposobnosti, je mentalni proces povezan s nastojanjem da se interpretiraju vizualne informacije. Pritom je ključna sposobnost točnog percipiranja vizualnog svijeta, ali i mogućnost predočavanja vizualnog iskustva, čak i kada relevantan podražaj nije prisutan.

Prostorno razmišljanje sastavni je dio triju već spomenutih modela, Gardnerovog, Thurstoneovog i Vernonovog. Gardner kaže da se spacijalna inteligencija odnosi na sposobnost zamjećivanja vidnih i prostornih informacija, mogućnost njihove preobrazbe i oblikovanja, te mogućnost vizualizacije bez pomoći vanjskih vidnih podražaja. To je sposobnost shvaćanja "cjeline" predmeta, "Gestalt" organizacije, koja se razlikuje od logičko-matematičke organizacije (Rafi i sur., 2005). Za nju je ključna sposobnost zamišljanja slika u tri dimenzije te sposobnost pokretanja i zaokretanja tih predodžbi. Thurstone pak dijeli spacijalne sposobnosti u tri skupine: sposobnost prepoznavanja predmeta kada se promatra iz različitih kuteva, sposobnost mentalnog premještanja dijelova nekog sklopa i sposobnost zamišljanja prostornih odnosa kada je sam promatrač dio prostornog problema. Iako je on originalno tretirao te sposobnosti kao nezavisne, pokazalo se da postoji stanovita korelacija među njima. Vernon, kako je već spomenuto, u svom hijerarhijskom modelu spacijalno razmišljanje vidi kao dio šire spacijalno-mehaničke sposobnosti (Zarevski, 2000).

Povijesni razvoj istraživanja spacijalnih sposobnosti može se podijeliti u četiri razdoblja (<http://www.ul.ie/~mearsa/9519211/newpage6.htm>). Pionirski radovi iz tog

područja javili su se tridesetih godina prošlog stoljeća kada se ukazala potreba za razvijanjem testova koji neće biti pristrani u korist verbalnih sposobnosti. Psiholozi su počeli uviđati da na akademski uspjeh ipak ne utječu samo vještine čitanja i pisanja. Iako je u tom razdoblju Thurstone već ponudio svoj model višestrukih faktora (jedan od kojih su bile i spacijalne sposobnosti), neki su, poput Termana, još uvijek odbijali neverbalne testove kao takve, budući da nisu mjerili ono što su oni smatrali inteligencijom.

Drugo razdoblje, koje je trajalo do početka 1960-ih, obilježeno je brojnim istraživanjima koja su nastojala razdijeliti spacijalni faktor u više subfaktora, ali i razvojem brojnih novih spacijalnih testova. Tako su Goodman (1943), Whittenborn (1945) i Slater (1940) razvili testove koji su potvrdili postojanje dva subfaktora (spacijalnih odnosa i vizualizacije), dok je Thurstone u svojim istraživanjima ustanovio spacijalni faktor 1 i spacijalni faktor 2, pri čemu je prvi povezan sa statičnim, a drugi s dinamičnim slikama.

Treće je razdoblje obilježeno fluktuacijom interesa za spacijalne sposobnosti, ali i drugačijom prirodom tog interesa – u središtu interesa više nije bilo utvrditi prirodu spacijalnih sposobnosti već ispitati izvore varijance na spacijalnim testovima. Tako je Thurstone proučavao rodne, a Piaget dobne razlike u spacijalnim sposobnostima.

Posljednjih 15 do 20 godina uglavnom se nastavlja ono što je prije započeto, te se neprestano javljaju nove klasifikacije i definicije. Razvojem novih tehnologija omogućena su nova inovativna istraživanja koja svakodnevno dovode do novih spoznaja.

O testovima inteligencije

Test u psihologiji predstavlja "standardizirani postupak kojim se izaziva određena aktivnost čija se veličina mjeri i vrednuje radi lakše usporedbe individualnih rezultata i smještanja istih u skupinu rezultata" (Petz, 1992). Testovi se u psihologiji koriste u dijagnostičke (određivanje veličine neke psihičke karakteristike) ili u prognostičke svrhe (predviđanje budućeg uspjeha u određenim aktivnostima na temelju testovnih rezultata). Ukupni rezultat na testu uglavnom se formira zbrajanjem njegovih čestica, a nulta točka

ne postoji jer najmanji rezultat ne znači odsustvo mjerene osobine, već samo nemogućnost testa da izmjeri toliko malu zastupljenost iste u pojedincu.

Testovi se mogu podijeliti s obzirom na prirodu zadatka (testovi maksimalnog učinka i testovi tipičnih ponašanja), način primjene (individualni ili grupni) ili relevantnost vremena (testovi brzine i testovi snage). Često se testovi dijele i prema predmetu mjerenja, pa tako postoje testovi sposobnosti (senzornih, psihomotornih ili mentalnih), testovi znanja (provjera efekta učenja) te testovi ličnosti (analitički ili sintetički) (Petz, 1992).

Standardizirano testiranje inteligencije, koje se provodi već stotinjak godina, smatra se jednim od najvećih uspjeha psihologije (Benson, 2003), a iako postoje neslaganja o naravi inteligencije, psiholozi i nastavnici svakodnevno primjenjuju tisuće testova inteligencije (Rathus, 2000).

Prvi test inteligencije sastavio je 1906. godine Alfred Binet, danas poznat kao "otac testova", a zadatak mu je bio izdvojiti onu školsku djecu koja nisu u stanju pratiti školsku nastavu. Iako je njegov test iz današnje perspektive sadržavao premalo zadataka, a bio je i previše zasićen školskim znanjem (Zarevski, 2000), neke od njegovih osnovnih ideja zadržane su do danas.

Mnogi su osporavali testiranje inteligencije tvrdeći da je ono rasno, rodno i kulturalno pristrano, te da umanjuje važnost kreativnosti, ličnosti i praktičnoga znanja, i da propagira ideju kako se ljudi rađaju s nepromjenjivim intelektualnim potencijalom koji određuje njihov uspjeh u životu. Stoga istraživači inteligencije, počevši od 1970-ih, nastoje doskočiti ovakvim pristranostima. Mnoge pristranosti koje su navodili kritičari u novije vrijeme su u testovima smanjene, pa danas postoje mnogi pouzdani testovi kognitivnih sposobnosti koji su se u praksi pokazali valjanima i prognostički vrijednima.

No iskustva psihologa u praksi pokazala su da sadržaj testa s vremenom postaje poznat, a norme za interpretaciju zastarjele, pa postoji stalna potreba za razvojem novih testova inteligencije. Vježba dakle može utjecati na rezultat, a učinci vježbe mogu se proizvesti na tri načina:

- a) višekratnim zadavanjem određenog testa ili njegovih paralelnih formi
- b) analizom pogrešaka u rješavanju testova

- c) širokom raspravom o načelima rješavanja određenih mentalnih zadataka, tj. svojevrsnim "treningom inteligencije" ili tzv. vođenjem (engl. coaching) (Zarevski, 2000).

Korištenje adekvatnog psihologijskog instrumentarija iznimno je važno jer danas široko korištena disciplina upravljanja ljudskim potencijalima, osobito izbor i selekcija kadrova u području gospodarstva, te nekih područja obrazovanja i zdravstva, počiva na procesu psihodijagnostike. Testovi koji su težinski prilagođeni potrebama korisnika te standardizirani na reprezentativnim uzorcima omogućuju podizanje kvalitete psihodijagnostike te optimalizaciju procesa donošenja odluka.

Psihometrijska evaluacija čestica kompozitnih mjernih instrumenata jedan je od središnjih postupaka pri izradi novih ili reviziji postojećih psihologijskih instrumenata (Nunnally i Bernstein, 1994), pa tako i testova inteligencije. Izbor čestica za konačnu formu nekog mjernog instrumenta ovisi u prvom redu o njihovoj osjetljivosti, te diskriminativnoj valjanosti, a posljedice tog izbora odnose se na valjanost, pouzdanost i osjetljivost ukupnih testovnih rezultata. Da bi testovi poslužili svojoj svrsi, oni moraju biti pouzdani, osjetljivi i valjani. Pouzdanost testa odnosi se na stupanj u kojem on daje konzistentne rezultate. Ako je test visoko pouzdan, rezultat se može protumačiti kao mjera uratka pojedinca u testu, a ne slučajnih faktora koji nemaju veze s razvijenošću ispitanika kod pojedinca. Osjetljivost je karakteristika testa ili procedure mjerenja koja se odnosi na sposobnost testa da razlikuje ispitanike prema predmetu mjerenja. Valjanost testa odnosi se na ideju da se testom uistinu mjeri ono što se njime namjerava mjeriti. Ako test zaista mjeri ono što bi trebao mjeriti, na temelju uratka u testu možemo izvoditi zaključke o prilikama izvan testovne situacije. Pouzdanost i valjanost međusobno su povezane. Pouzdanost je nužan preduvjet valjanosti: nemoguće je znati mjeri li test ono za što se tvrdi da mjeri ako su rezultati dobiveni tim testom nepouzdan. No obrnuto ne vrijedi, pouzdanost sama po sebi ne garantira valjanost (Nunnally i Bernstein, 1994).

O specijalnim testovima

Postoji mnogo testova koji ispituju prostorno procesiranje, a neki od njih prilagođeni su primjeni na našim prostorima i mogu se pronaći u bazi testova Centra za psihodijagnostičke instrumente u Zagrebu. Oni mjere sposobnost prostornog procesiranja na nekoliko različitih načina.

Na primjer, PP-test, Test specijalnog rezoniranja (Reuchlin, M i Valin, E., prema www.cpi.ffzg.hr), obuhvaća zadatke koji se sastoje od zadanog trodimenzionalnog prikaza nekog tijela i četiri ponuđena dvodimenzionalna prikaza među kojima treba prepoznati zadano tijelo gledano iz određenog kuta.

Slične zadatke sadrži i drugi subtest u Kalifornija testu mentalne zrelosti (Sullivan, Clark, i Tiegs, 1957, prema Koren, I. i Mijić, V., 1976). U njemu su prikazani trodimenzionalni likovi koje treba mentalno rotirati i među ponuđenim tijelima odabrati ono koje odgovara zadanom tijelu gledanom iz drugačijeg kuta. Razlika od Testa PP je u tome što su ovdje i ponuđeni odgovori trodimenzionalni.

Test shvaćanja zapreminskih odnosa (Pr) iz baterije testova sposobnosti BTI (Bele-Potočnik i Zajec, 1982) po konstrukciji zadataka vrlo je sličan Testu ST-1. Štoviše, ideje za čestice Testa ST-1 proizašle su iz Testa Pr, samo što su zadaci obrnuti: za razliku od Testa ST-1, ovdje je zadan plošni prikaz a od ispitanika se traži da među ponuđenim odgovorima odabere ono tijelo koje se može sastaviti od tog plošnog prikaza.

Isto se od ispitanika traži i u subtestu S iz baterije DAT (Differential Aptitude Test; Bennett, Seashore i Wesman, 1952, prema Momirović i Bukvić, 1969). Baterija je 1970. prilagođena za područje bivše Jugoslavije i određene su norme. U njoj se nalazi i subtest PFB (shvaćanje odnosa među površinama), koji je dodan u verziji prilagođenoj našim prostorima, a čiji su autori Likert i Quasch. On ispituje sposobnost razlikovanja dvodimenzionalnih geometrijskih crteža i mentalnih operacija s njima. U svakom zadatku zadani su dijelovi geometrijskog lika, a među ponuđenim odgovorima treba odabrati onaj koji točno prikazuje lik kojeg se može sastaviti od zadanih dijelova.

Nekoliko subtestova koji obuhvaćaju specijalni faktor može se naći i u priručniku Ekstroma, Frencha, Harmana, i Dermena (1976), *Manual for Kit of Factor-Referenced Cognitive Tests*. To se primarno odnosi na subtestove Prostorna orijentacija (Spatial

Orientation-S), Vizualizacija (Visualization-VZ), te Fleksibilnost baratanja oblicima (Figural Flexibility-XF), no ti testovi nisu prevedeni na hrvatski i za njih ne postoje odgovarajuće nacionalne norme.

Dakle, testovi koji mjere specijalne sposobnosti postoje, no očito je da su njihove norme zastarjele a u nekim slučajevima one uopće ne postoje. Zbog zastarjelosti samih testova, možemo pretpostaviti da je i njihov sadržaj već postao poznat. Stoga je jasna potreba za razvojem novog testa.

Test ST-1 razvili su 2006. godine Marijan Pleština i Josip Radić iz tvrtke LANAP d.o.o. u suradnji s prof. Vladimirom Takšićem i Centrom za psihodijagnostičke instrumente u sklopu Filozofskog fakulteta u Zagrebu. O samom testu bit će više rečeno nešto kasnije.

2. CILJ ISTRAŽIVANJA

Glavni je cilj istraživanja provesti kompletnu analizu svih 35 zadataka Testa ST-1 i provjeriti konstruktnu valjanost testa. U slučaju da se pokaže kako je moguće skratiti test a da se pritom očuvaju jednako dobre metrijske karakteristike, dat će se prijedlog skraćene verzije testa, koji bi onda imao i kraću primjenu. Naime, test je u ovom radu primjenjivan u trajanju od 40 minuta, što je gotovo cijeli jedan školski sat. Budući da se u praksi redovito primjenjuje baterija različitih testova, bilo bi poželjno napraviti što kraću verziju testa, čime bi se povećala njegova ekonomičnost.

Dakle, problemi ovog rada su sljedeći:

1. provjeriti psihometrijske karakteristike 35 zadataka preliminarne forme Testa ST-1, što uključuje izračunavanje indeksa lakoće i koeficijenta diskriminativne valjanosti, te empirijskih karakterističnih krivulja zadataka;
2. provjeriti deskriptivnu statistiku i psihometrijske karakteristike ukupnih rezultata na Testu ST-1: pouzdanost tipa unutarnje konzistencije, osjetljivost, te konstruktnu valjanost;
3. na temelju analize zadataka i ukupnih testovnih rezultata sastaviti konačnu verziju Testa ST-1, te provjeriti kako se izbacivanjem zadataka mijenjaju deskriptivna statistika i psihometrijske karakteristike testa.

3. METODOLOGIJA

Test ST-1 preliminarno je primijenjen na 10 ispitanika, osmero studenata i dvije osobe visoke stručne spreme. Cilj preliminarne primjene je bio utvrditi eventualne nejasnoće i odrediti trajanje primjene testa. Na temelju povratnih informacija izmijenjena su tri zadatka u kojima nije postojao točan odgovor zbog loše grafičke izvedbe. Svaki je ispitanik također morao mjeriti vrijeme koje mu je bilo potrebno za rješavanje testa, te je na temelju toga zaključeno da bi 40 minuta većini ispitanika trebalo biti dovoljno da prođu kroz cijeli test.

Uzorak je u primjeni ovog testa bio prigodan, sudjelovalo je 212 ispitanika u dobi od 18 do 38 godina ($M=24,34$, $SD=2,77$). Test je primijenjivan grupno, na skupinama od 5 do 20 ljudi, u razdoblju od mjesec dana. U uzorku je veća zastupljenost žena (62,7%) nego muškaraca (37,3%). Sudionici su bili različitog stupnja obrazovanja iz različitih područja: 20 učenika gimnazija, 58 sudionika sa srednjom stručnom spremom, različitih struka, 40 s visokom stručnom spremom, te 94 studenata različitih fakulteta u okviru Zagrebačkog i Splitskog Sveučilišta: Filozofski fakultet (smjerovi psihologija (6), jezici (11)), Kineziološki fakultet (2), Fakultet elektronike i računarstva (10), Agronomski fakultet (smjer uređenje krajobraza) (23), Ekonomski fakultet (11), Arhitektonski fakultet (smjer dizajn) (7), Građevinski fakultet (6), Edukacijsko-rehabilitacijski fakultet (smjer logopedija) (2), Prehrambeno-biotehnološki fakultet (1), Matematičko-prirodoslovni fakultet (4), Pravni fakultet (8) te Fakultet strojarstva i brodogradnje (3). Budući da metrijske karakteristike testa ovise o uzorku na kojem se on primijenjuje, za analizu čestica testa bio bi poželjan heterogeni uzorak. Naime, homogeni uzorak dovodi do matrice interkorelacija specifičnih za taj uzorak, pa se test neopravdano može činiti nediskriminativnim.

Sam test ST-1 razvijen je s ciljem da bi se provjerila sposobnost predočavanja prostornih odnosa. Kao što je već spomenuto, ideje za čestice preuzete su iz Testa shvaćanja zapreminskih odnosa iz baterije testova sposobnosti BTI. Test se dakle sastoji od 35 zadataka, a svaki zadatak sadržava prikaz jednog geometrijskog tijela i šest plošnih prikaza. Ispitanik treba odabrati jedan plošni prikaz koji odgovara tijelu sa slike, tj. od

kojega je moguće sačiniti zadano tijelo, i točan odgovor označiti u posebnom listu za odgovore. Prije rješavanja testa ispitanicima je dana usmena i pismena uputa, te im je ponuđen jedan primjer koji se ne boduje (vidi primjer u Prilogu 1).

Sudionicima se u uputi kaže da je trajanje primjene testa ograničeno, te da rade što brže i što točnije. Pretpostavlja se da uradak u testu ne ovisi o prethodnom znanju, iako su radi veće heterogenosti uzorka odabrani studenti kojima je u okviru obrazovanja važnije baratanje prostornim odnosima (npr. dizajn, uređenje krajobraza), i oni kojima je to manje važno (npr. psihologija, pravo, ekonomija), bez obzira na to da li određeno obrazovanje razvija njihovu sposobnost prostornog razmišljanja ili su oni upravo zbog razvijenosti istih odabrali svoj fakultet. Međutim, u obradi se pokazalo da razlika među njima nije statistički značajna ($t=0,635$, $df=47$, $p>0,05$).

4. REZULTATI I RASPRAVA

Analiza testa ST-1

Podaci prikupljeni Testom ST-1 obrađeni su osobnim računalom pomoću programskog paketa SPSS for Windows verzija 14.0. Osnovni deskriptivni podaci prikazani su u Tablici 1.

Tablica 1. Deskriptivna statistika originalne verzije Testa ST-1

N	min	max	M	SD	α	prosj r_{ij}	min r_{ij}	max r_{ij}	prosj p	prosj r_{iu}
212	3	35	22,04	6,861	0,874	0,163	-0,106	0,567	0,630	0,431

LEGENDA:

N – broj ispitanika
min – najmanji dobiveni rezultat
max – najveći dobiveni rezultat
M – aritmetička sredina
SD – standardna devijacija
 α – Cronbachov alfa koeficijent pouzdanosti
 r_{ij} – interkorelacija zadataka
p – indeks lakoće zadatka
 r_{iu} – koeficijent diskriminativnosti zadatka

Ukupni rezultat formiran je kao jednostavna linearna kombinacija 35 zadataka koji čine Test ST-1. Distribucija tako dobivenih ukupnih rezultata ne razlikuje se značajno od normalne (Kolmogorov-Smirnovljevi $Z=1,254$, $p>0,05$), što je važno za daljnju obradu i interpretaciju rezultata. Ispitanici su u prosjeku točno rješavali 22 zadatka ($M=22,04$, $SD=6,861$), a raspon rezultata svih ispitanika kretao se od 3 do 35. Budući da se test sastoji od binarnih zadataka, što znači da se za netočan odgovor dobiva 0, a za točan odgovor 1 bod, rezultat 35 ujedno je i najveći mogući rezultat u ovom testu.

Muškarci su u prosjeku postizali nešto bolji rezultat ($M=23,23$, $SD=7,30$) nego žene ($M=21,47$, $SD=6,49$), no ta razlika nije statistički značajna ($t=1,815$, $df=208$, $p>0,05$), iako bi ona bila očekivana budući da su u području prostornog razmišljanja

rodne razlike, i to u korist muškaraca, izraženije nego u svim ostalim kognitivnim sposobnostima (McGlone i Aronson, 2006).

Koeficijent unutarnje konzistencije, Cronbachov α koeficijent pouzdanosti, iznosi $\alpha=0,874$, što znači da se test može smatrati visoko pouzdanim. Ovaj koeficijent pouzdanosti temelji se na interkorelacijama među pojedinim dijelovima istog testa, odnosno među zadacima koji čine test. On pokazuje veličinu djelovanja nesistematskih varijabilnih faktora u trenutku kada je test primijenjen. Budući da se djelovanje nesistematskih varijabilnih faktora od trenutka do trenutka mijenja, ne smijemo pretpostaviti da će Cronbachov α u sljedećoj primjeni biti iste veličine. On nam dakle ništa ne govori o stabilnosti rezultata u vremenu.

Prosječni indeks lakoće zadataka iznosi $p=0,630$, što u principu govori da je ispitanicima test bio nešto lakši od primjerenog (u kojem bi indeks lakoće bio $p=0,5$). Međutim, treba uzeti u obzir da najveći dio uzorka čine ispitanici visoke stručne spreme, studenti i gimnazijalci, za koje možemo pretpostaviti da su iznadprosječnih kognitivnih sposobnosti, pa je takav indeks očekivan. Drugim riječima, kada bismo ovaj test primijenili na uzorku koji je reprezentativan za opću populaciju, očekujemo da bi prosječni indeks lakoće bio manji, odnosno da bi tim ispitanicima test bio teži.

Prosječna interkorelacija zadataka iznosi $r_{ij}=0,163$, što znači da su zadaci unutar testa u prosjeku nisko povezani, no to je u ovakvim testovima očekivano. Naime, binarni zadaci imaju smanjene varijance pa se zbog toga u ovakvim testovima u pravilu javljaju niske prosječne korelacije. Međutim, niske interkorelacije zadataka mogu upućivati i na unutarnju heterogenost testa koja je poželjna kada se test koristi u prognostičke, ali ne i u dijagnostičke svrhe (Krković, 1978). Mogli bismo dakle pretpostaviti i da se faktorska struktura sastoji od više nepovezanih faktora koji objašnjavaju male proporcije ukupne varijance. To će se u nastavku provjeriti faktorskom analizom.

Diskriminativna valjanost označava mogućnost razlikovanja po-općem-uspjehu-u-testu iznadprosječnih od ispodprosječnih ispitanika na osnovi njihova uratka u zadatku (Krković, 1978). Ta se valjanost mjeri korelacijom između uratka u zadatku i uratka u cijelom testu, a orijentirana je na homogenizaciju testovnih rezultata. Obično se navodi da koeficijent korelacije od 0,0 do 0,2 ukazuje na nikakvu ili vrlo slabu povezanost, od 0,2 od 0,4 na slabu povezanost, od 0,4 do 0,7 na srednju povezanost, a vrijednosti iznad

0,7 na veliku povezanost među varijablama (Petz, 1992). Prosječna diskriminativna valjanost zadataka ovog testa iznosi $r_{iu}=0,431$, što bi dakle spadalo u područje srednje povezanosti.

Prije daljnjih obrada i prijedloga treba upozoriti na selekcioniranost uzorka na kojem je primijenjena preliminarna verzija testa. Naime, uzorak na kojem je test primijenjen nije reprezentativan. Većinom ga čine ispitanici visoke stručne spreme, studenti i gimnazijalci, te tek 58 sudionika različitih zanimanja i struka, sa srednjom stručnom spremom. Ispitanici su ujednačeni po kronološkoj dobi (18 do 38 godina), a različita je zastupljenost spolova (62,7% ženskih naprema 37,3% muških ispitanika). Zbog takvog uzorka ne možemo tvrditi da dobivene karakteristike testa vrijede i za opću populaciju.

Važno je napomenuti i da je Test ST-1 razvijen u svrhu testiranja prostornog razmišljanja pomoću računala, te će tako u budućnosti biti primjenjivan. Ovdje je ipak korištena papir-olovka verzija testa zbog jednostavnije primjene. Stoga bi bilo dobro provjeriti metrijske karakteristike testa primijenjenog putem računala, jer neki autori navode da ispitanici u tako primijenjenim testovima postižu slabije rezultate zbog korištenja različitih strategija rješavanja i različitog stava prema testiranju (npr. Swain, 1998). No budući da druga istraživanja nisu pokazala razliku u rezultatima računalne i papir-olovka primjene testa, ili su razlike išle u korist primjene na računalima (npr. Pommerich, 2004), čemu bi razlog mogla biti smanjena anksioznost ispitanika (Fox i Schwartz, 2001), možemo pretpostaviti da su te razlike vrlo individualne i da u prvom redu ovise o tome koliko je ispitanik upoznat s radom na računalu.

Analiza zadataka

Za svaki zadatak izračunat je indeks lakoće (što u slučaju binarnih zadataka odgovara aritmetičkoj sredini odgovora ispitanika na zadatak), te standardna devijacija kao mjera raspršenja rezultata. Isto tako utvrđena je i diskriminativna valjanost svakog zadatka, odnosno njegova korelacija s ukupnim rezultatom. Navedene vrijednosti prikazane su u Tablici 2.

Tablica 2. Prikaz analize 35 zadataka Testa ST-1

zadatak	M (p)	SD	r_{iu}	α kad se zadatak izbac	zadržavanje zadatka*	zasićenje s 1. faktorom
1	,9811	,1364	,198	,874	-	,218
2	,8632	,3444	,315	,872	+	,315
3	,7594	,4284	,293	,873	+	,273
4	,8160	,3884	,408	,871	+	,406
5	,7170	,4515	,280	,874	+	,258
6	,8208	,3845	,351	,872	+	,349
7	,8255	,3805	,507	,869	+	,523
8	,8349	,3721	,361	,872	+	,367
9	,6698	,4714	,445	,870	+	,449
10	,5613	,4974	,395	,872	+	,384
11	,6462	,4793	,444	,870	+	,431
12	,7453	,4367	,452	,870	+	,443
13	,6557	,4763	,451	,870	+	,446
14	,6321	,4834	,569	,867	+	,580
15	,8160	,3884	,470	,870	+	,477
16	,6226	,4859	,304	,874	+	,270
17	,6226	,4859	,523	,868	+	,526
18	,6651	,4731	,522	,868	+	,527
19	,5708	,4961	,407	,871	+	,402
20	,4198	,4947	,497	,869	+	,492
21	,5377	,4998	,497	,869	+	,482
22	,3019	,4602	,376	,872	+	,348
23	,1745	,3805	,232	,874	-	,203
24	,3821	,4870	,224	,875	-	,203
25	,2925	,4560	,378	,872	+	,350
26	,4906	,5011	,539	,868	+	,542
27	,6274	,4847	,577	,867	+	,593
28	,5142	,5010	,426	,871	+	,419
29	,4292	,4961	,418	,871	+	,410
30	,5755	,4954	,643	,865	+	,663
31	,7123	,4538	,480	,869	+	,497
32	,5755	,4954	,585	,867	+	,599
33	,7547	,4313	,479	,869	+	,509
34	,7075	,4560	,496	,869	+	,519
35	,7170	,4515	,544	,868	+	,570

LEGENDA:

M – aritmetička sredina
 p – indeks lakoće zadatka
 SD – standardna devijacija
 r_{iu} – indeks diskriminativnosti zadatka

α – Cronbachov alfa koeficijent pouzdanosti
"+" - zadatak ulazi u konačnu verziju testa
"- " - zadatak ne ulazi u konačnu verziju testa

*kao kriterij zadržavanja zadatka određen je $r_{iu} > 0,25$

Pregledom indeksa lakoće svih 35 zadataka možemo utvrditi da je najlakši zadatak u testu 1. ($p=0,9811$), dok je najteži 23. ($p=0,1745$). U praksi se testovi obično formiraju tako da zadaci budu poredani od najlakšeg prema najtežem. Iako postoji mogućnost da se u tom slučaju ispitanik "istroši" na laganim zadacima i potroši vrijeme na njih pa ako i stigne do težih nema snage i vremena rješavati ih, u prilog takvom rasporedu ide argument da ispitanik na početku testa zbog laganih zadataka dobije osjećaj uspjeha, a može se i provjeriti do koje je težine zadataka ispitanik u stanju rješavati test. U ovoj preliminarnoj verziji Testa ST-1 zadaci nisu poredani po težini, pa bi njihov redoslijed u konačnoj verziji trebalo izmijeniti, no o tome više nešto kasnije.

Težinska primjerenost testa određena je težinskom primjerenošću pojedinih zadataka. Prosječni indeks lakoće određuje se kao aritmetička sredina indeksa lakoće svih zadataka. Ako je prosječni $p < 0,5$, test je neprimjeren težak, pa je distribucija rezultata pozitivno asimetrična a osjetljivost mjerenja je smanjena u području brojčano niskih rezultata. Ako je pak prosječni $p > 0,5$, test je ispitanicima neprimjeren lagan, pa je distribucija rezultata negativno asimetrična čime je smanjena mogućnost razlikovanja ispitanika u području brojčano visokih rezultata. Da bi test bio težinski primjeren ($p=0,5$) a distribucija rezultata normalna, poželjno je odabrati zadatke svih stupnjeva lakoće, ali više onih približno prosječno teških, manji broj lakih i teških, a najmanje vrlo lakih i vrlo teških zadataka (Krković, 1978).

Ispitanike najbolje diskriminira 30. zadatak ($r_{iu}=0,643$), te je primjer najbolje postavljenog zadatka jer najbolje razlikuje uspješne od neuspješnih ispitanika. Najlošiji koeficijent diskriminativnosti ima 1. zadatak ($r_{iu}=0,198$), a pregledom indeksa lakoće možemo utvrditi da je razlog tome činjenica da njega gotovo svi ispitanici točno rješavaju. Stoga na temelju uratka u tom zadatku ne možemo razlikovati ispitanike koji test rješavaju dobro od onih koji test rješavaju loše.

Određivanje odnosa između pojedinih zadataka i mjerenog konstrukta (onoga što bi zadaci u osnovi trebali mjeriti), jedan je od ključnih problema pri konstrukciji nekog

mjernog instrumenta. Stoga je jedna od dodatnih metoda provjere valjanosti zadatka pregled empirijskih karakterističnih krivulja zadataka. Niti jedan konstrukt (u ovom slučaju spacijalna inteligencija) ne reflektira se savršeno u nekom konačnom broju zadataka. Drugim riječima, bodovi na nekom testu nikad u potpunosti ne odgovaraju stvarnoj veličini konstrukta.

Karakteristične krivulje zadataka ili linije traga zadataka povezuju mjereni atribut i vjerojatnost točnog odgovora na postavljeni zadatak. Na apscisi takvih prikaza iskazuje se izraženost mjerenog konstrukta, što je u ovom slučaju bio ukupni rezultat na Testu ST-1 arbitrarno kategoriziran prema veličini u šest razreda. Na ordinatu se uvrštava vjerojatnost točnog rješavanja datog zadatka, koja ovdje zapravo predstavlja indeks lakoće zadatka. Tako iz krivulje možemo vidjeti kolika je vjerojatnost da će osobe različitog stupnja razvijenosti mjerenog konstrukta točno riješiti zadatak. Očekuje se naravno da će s porastom veličine mjerenog konstrukta rasti i vjerojatnost točnog odgovora na zadatak.

Moderni modeli mjerenja pretpostavljaju sigmoidne linije traga, koje odgovaraju kumulativnoj normalnoj distribuciji. Kako u normalnoj distribuciji većina rezultata pada u područje srednjih vrijednosti, a prema krajevima distribucije sve je manje rezultata, sigmoidna krivulja je strmija na svojem srednjem dijelu, a na krajevima više zakrivljena. Linije su stoga u svom srednjem dijelu diskriminativnije, jer veća strmina znači i veću korelaciju između zadatka i ukupnog uratka.

Linije traga mogu biti vrlo korisne pri izradi lakše i teže verzije nekog testa. Ako linija ima strmiji nagib u području najboljih rezultata, zadatak bolje razlikuje nadprosječne ispitanike, pa je on primjereniji za težu verziju testa. Ako pak krivulja u početku ima strmi nagib no vrlo brzo postiže plato, zadatak bolje razlikuje ispodprosječne ispitanike te je primjereniji u lakšim verzijama testa.

Prije pregleda karakterističnih krivulja 35 zadataka Testa ST-1 treba upozoriti na jednu stvar. Naime, rezultati su umjetno podijeljeni u šest razreda, od kojih svaki obuhvaća raspon od šest rezultata. No budući da je distribucija rezultata blago negativno asimetrična (iako smo pokazali da se statistički značajno ne razlikuje od normalne) jer su zbog karakteristika uzorka zadaci ispitanicima bili nešto prelagani, raspodjela rezultata po razredima nije ujednačena. U prvi razred je tako svrstan samo jedan ispitanik, pa su

moguće vrijednosti na ordinati bile isključivo 0 i 1. Taj je ispitanik riješio samo tri zadatka (2, 12, 24), pa u ta tri slučaja krivulje nisu rastuće. Međutim, ti se prikazi ne mogu smatrati realnima jer se rezultat 1 u prvom razredu temelji isključivo na tom jednom ispitaniku. Izuzev problema s vrijednostima prvog razreda u ova 3 zadatka, linije traga pokazuju najveće odstupanje od sigmoidnih rastućih krivulja upravo u 1., 23. i 24. zadatku (odstupanje u 24. zadatku postoji bez obzira na vrijednost u prvom razredu) koji su izbačeni iz konačne verzije testa na temelju niskih diskriminativnih valjanosti. Odstupanje od sigmoidne krivulje vidljivo je i u 31. zadatku, no budući da on inače pokazuje dobre karakteristike ($r_{ii}=0,480$), činjenicu da njegova linija traga nije rastuća možemo pripisati umjetno stvorenim razredima rezultata i različitom broju rezultata u svakom razredu.

Prikazi karakterističnih krivulja svih 35 zadataka Testa ST-1 izloženi su u Prilogu 2.

Faktorska analiza

Faktorska analiza predstavlja skup statističko-matematičkih postupaka pomoću kojih se analiziraju interkorelacije između većeg broja manifestnih varijabli. Njezin cilj je utvrditi što manji broj latentnih varijabli (tzv. faktora) pomoću kojih se može objasniti kovarijabilitet između što više manifestnih varijabli (empirijski dobivenih podataka), odnosno veliki broj manifestnih varijabli zamijeniti što manjim brojem latentnih varijabli. Zatim se određuje stupanj povezanosti svake manifestne varijable s tim utvrđenim zajedničkim faktorima. Ovdje nam je dakle cilj provjeriti faktorsku strukturu testa, tj. njegovu faktorsku valjanost. Kao osnova za faktorsku analizu poslužile su međusobne korelacije 35 zadataka koji čine Test ST-1.

Prije provedbe faktorske analize proveden je Bartlettov test ($\chi^2=1850,058$, $df=595$, $p<0,01$) i izračunat Kaiser-Meyer-Olkinov indeks ($KMO=0,815$) koji su pokazali da je korelacijska matrica pogodna za provedbu faktorske analize.

Provedena je faktorska analiza testa pod modelom glavnih komponenata, te su dobiveni rezultati prikazani u Tablici 3. Uz Guttman-Kaiserov kriterij, prema kojemu se zadržavaju samo oni faktori čiji su karakteristični korijeni veći od 1, dobiveno je ukupno

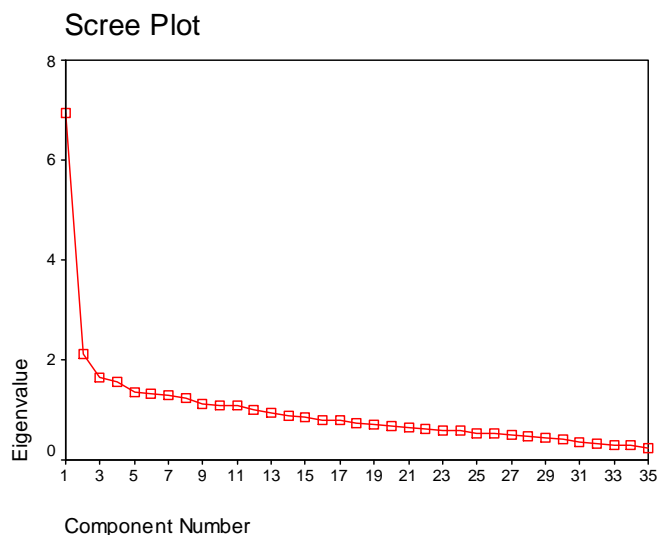
11 faktora. Ipak, veličine njihovih karakterističnih korijena ukazuju na jedan glavni faktor, koji objašnjava 19,873 % varijance, dok se vrijednosti ostalih faktora vrlo malo razlikuju. Prvi faktor dakle ne objašnjava velik dio ukupne varijance, no on ipak brojčano znatno iskače među ostalim faktorima. Zasićenja zadataka s glavnim faktorom prikazana su u Tablici 2. Prema očekivanjima, najveće zasićenje pokazuju zadaci najveće diskriminativne valjanosti.

Tablica 3. Ekstrahirani faktori i njihovi karakteristički korijeni pod modelom glavnih komponenata za Test ST-1

komponenta	karakteristični korijeni	% ukupne varijance objašnjen pojedinim faktorom
1	6,955	19,873
2	2,117	6,047
3	1,643	4,694
4	1,561	4,459
5	1,354	3,867
6	1,321	3,775
7	1,286	3,673
8	1,233	3,521
9	1,121	3,203
10	1,100	3,143
11	1,077	3,077

Pri konstrukciji testova odluke o zadržavanju faktora često se donose i na temelju izgleda tzv. *screepplot* prikaza. To je grafički prikaz na čijoj je apscisi redni broj ekstrahiranog faktora, a na ordinati vrijednost pripadajućeg karakterističnog korijena. Kada na prikazu postane vidljivo da su se karakteristični korijeni prestali međusobno razlikovati i da je njihov pad postao linearan, prekida se broj faktora koji će se zadržati. Pregledom *screepplot* prikaza Testa ST-1 može se uočiti da pad krivulje postaje linearan tek nakon drugog faktora (iako se strmina pada krivulje uvelike smanjuje već nakon prvog faktora), što bi moglo značiti da treba zadržati i drugi faktor.

Međutim, treba podsjetiti da su zadaci u Testu ST-1 binarni, što znači da je njihova varijanca smanjena. Zbog toga su njihove međusobne korelacije niže nego što bi bile na nekoj metrički superiornoj skali, pa dolazi do hiperfaktorizacije, odnosno do pojavljivanja većeg broja faktora s karakterističnim korijenom iznad 1 (Fulgosi, 1979).



Slika 1. Grafički prikaz vrijednosti karakterističnih korijena (*screeplot*) za faktore u Testu ST-1

Na temelju ovih podataka može se zaključiti da, unatoč tome što karakteristični korijeni i *screeplot* prikaz ukazuju na mogućnost postojanja većeg broja faktora, Test ST-1 u svojoj podlozi zapravo ima jedan glavni faktor. To naravno treba provjeriti na većem i reprezentativnijem uzorku. Dakle, faktorsku strukturu treba još jednom ispitati i nakon izbacivanja nediskriminativnih zadataka i skraćanja testa.

Skraćenje testa

Zadaci koji neće ući u konačnu verziju testa odabrani su prvenstveno na temelju koeficijenta diskriminativne valjanosti. Taj koeficijent nam govori koliko zadatak može razlikovati ispitanike dobrog i lošeg općeg uspjeha u testu – ako je on visok, to znači da

zadatak dobro razlikuje uspješne od neuspješnih ispitanika, a ako je nizak, znači da je slabo pokazuje koliko je kod ispitanika izražena sposobnost koja se mjeri testom. Pri odabiru zadataka za završnu verziju testa treba obratiti pozornost i na zasićenje zadataka glavnim faktorom, te provjeriti kako se izbacivanjem svakog zadatka mijenja pouzdanost čitavog testa (svi ovi podaci vidljivi su u Tablici 2).

Na temelju dobivenih podataka izbačeni su zadaci 1, 23 i 24. Ta tri zadatka imaju najniže koeficijente diskriminativne valjanosti, ispod arbitrarno određenog kriterija od 0,25. Oni dakle najlošije diskriminiraju po-općem-uspjehu-u-testu dobre od loših ispitanika. Očekivano, ti zadaci imaju i najmanja zasićenja s 1. faktorom. Pojedinačno gledano, izbacivanjem ovih zadataka ne smanjuje se pouzdanost testa – izbacivanjem 1. ili 23. zadatka Cronbachov α koeficijent pouzdanosti ostaje nepromijenjen ($\alpha=0,874$), dok se izbacivanjem 24. zadatka on čak neznatno povećava ($\alpha=0,875$).

Deskriptivna statistika skraćenog testa od 32 zadatka prikazana je u Tablici 4.

Tablica 4. Deskriptivna statistika Testa ST-1 skraćenog na 32 zadatka

N	min	max	M	SD	α	prosj r_{ij}	min r_{ij}	max r_{ij}	prosj p	prosj r_{iu}
212	2	32	20,50	6,660	0,876	0,180	-0,026	0,567	0,641	0,453

LEGENDA:

N – broj ispitanika
 min – najmanji dobiveni rezultat
 max – najveći dobiveni rezultat
 M – aritmetička sredina
 SD – standardna devijacija
 α – Cronbachov alfa koeficijent pouzdanosti
 r_{ij} – interkorelacija zadataka
 p – indeks lakoće zadatka
 r_{iu} – koeficijent diskriminativnosti zadatka

Izbacivanjem tri zadatka promijenile su se, logično, aritmetička sredina i standardna devijacija testa. One sada iznose $M=20,50$, $SD=6,660$. Raspon rezultata također je smanjen. Oblik distribucije nije se bitno promijenio, te se ona i dalje statistički značajno ne razlikuje od normalne (Kolmogorov-Smirnovljevi $Z=1,338$, $p>0,05$).

Od izbačenih zadataka, jedan je najlakši u čitavom testu (1. zadatak), dok dva spadaju među najteže (štoviše, 23. zadatak je zaista prema indeksu lakoće najteži u testu). To ne začuđuje, jer zadaci koji su vrlo lagani i vrlo teški uistinu ne mogu dobro razlikovati ispitanike jer ih u jednom slučaju rješavaju gotovo svi ispitanici, dok ih u drugom slučaju ne rješava gotovo nitko. Zadaci srednje težine dakle imaju najveću varijancu, pa je i očekivano da će oni najviše korelirati s ukupnim rezultatom. Zbog takvih osobina izbačenih zadataka ne čudi da se prosječni indeks lakoće zadatka povećao, te sada iznosi $p=0,641$. Takav porast nije poželjan, no on je vrlo malen a treba još jednom podsjetiti i na osobine nereprezentativnog uzorka kojem je ovaj test, kako se pokazalo, bio prelagan. Kako Test ST-1 nije namijenjen samo populaciji za koju se očekuje da je nadprosječno inteligentna, možemo pretpostaviti da bi na reprezentativnom uzorku opće populacije taj indeks ipak bio manji.

Prosječna interkorelacija zadataka skraćene verzije Testa ST-1 iznosi $r_{ij}=0,180$ te je veća nego u originalnoj verziji. Iako je ona još uvijek niska, njezin porast upućuje na povećanje dijagnostičke valjanosti testa i njegovo faktorsko čišćenje. Na homogenizaciju testovnih rezultata upućuje i povećanje, iako vrlo maleno, prosječnog koeficijenta diskriminativne valjanosti ($r_{iu}=0,453$).

Povećanje interkorelacije zadataka zapravo znači smanjenje osjetljivosti testa, no budući da male interkorelacije znače i manju dijagnostičku valjanost, u praksi je običaj odabirati zadatke nejednake težine ali sa što većim međusobnim korelacijama (Krković, 1978). Skraćenom verzijom testa postignut je dakle pomak u tom smjeru.

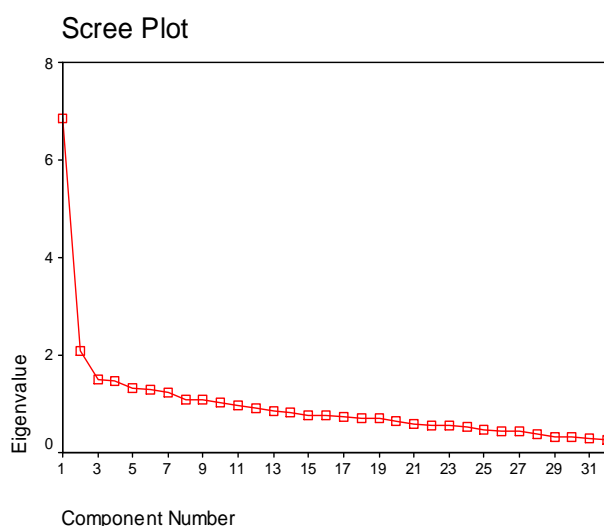
Cronbachov α koeficijent pouzdanosti neznatno je porastao ($\alpha=0,876$), što znači da je pouzdanost skraćene verzije testa čak nešto veća nego u originalnoj verziji. Budući da se Cronbachov koeficijent računa na temelju interkorelacije zadataka, možemo pretpostaviti da su izbačeni zadaci bili u niskim korelacijama s drugim zadacima.

Kako bismo provjerili da li je skraćivanjem testa došlo do promjene u njegovoj faktorskoj strukturi, provedena je faktorska analiza na 32 zadatka nove verzije testa. Prije toga je primijenjen Bartlettov test sferičnosti ($\chi^2=1713,826$, $df=496$, $p<0,01$) i izračunat Kaiser-Meyer-Olkinov indeks ($KMO=0,830$), koji su potvrdili prikladnost korelacijske matrice za provođenje faktorske analize. Rezultati faktorske analize prikazani su u Tablici 5.

Tablica 5. Ekstrahirani faktori na skraćenoj verziji Testa ST-1 i njihovi karakteristični korijeni pod modelom glavnih komponenta

komponenta	karakteristični korijeni	% ukupne varijance objašnjen pojedinim faktorom
1	6,841	21,379
2	2,097	6,552
3	1,489	4,652
4	1,463	4,572
5	1,327	4,148
6	1,298	4,056
7	1,228	3,837
8	1,099	3,436
9	1,092	3,412
10	1,026	3,207

Ukupan broj faktora smanjen je s 11 na 10, a i dalje se zadržala istaknutost jednog glavnog faktora (vidi Sliku 2). Kao i u originalnoj verziji, prvi faktor objašnjava najveći dio varijance, čak nešto veći nego u dužoj verziji, a iako je na *screeplot* prikazu opet uočljiv linearan pad krivulje tek nakon drugog faktora, treba još jednom podsjetiti da Test ST-1 sadrži binarne zadatke čija je varijanca smanjena, pa dolazi do hiperfaktorizacije.



Slika 2. Grafički prikaz vrijednosti karakterističnih korijena (*screeplot*) za faktore u skraćenoj verziji Testa ST-1

Na temelju faktorske analize provedene na 32 zadatka skraćene verzije Testa ST-1 možemo dakle zaključiti o postojanju jednog glavnog faktora koji objašnjava nešto više od 21% varijance. Faktorska struktura je u odnosu na originalnu verziju ostala uglavnom nepromijenjena.

Sve u svemu, metrijske karakteristike skraćene verzije Testa ST-1 u odnosu na originalnu verziju uglavnom su ostale nepromijenjene. Iako porasti koeficijenta pouzdanosti, prosječne diskriminativne valjanosti zadataka i prosječne interkorelacije zadataka idu u prilog skraćenoj verziji testa, razlike su zaista neznatne. Takav nalaz ne iznenađuje ako se sjetimo da je skraćena verzija dobivena izbacivanjem tek tri zadatka iz originalnog Testa ST-1.

Budući da je običaj da se u psihologijskim testovima, zbog već navedenih razloga, zadaci kreću od najlakšeg prema najtežem, predlažem da se u sljedećim primjenama Testa ST-1 redosljed zadataka izmijeni, u skladu s indeksima lakoće dobivenima ovim preliminarnim ispitivanjem. U Tablici 5 prikazan je predloženi novi redosljed zadataka.

Tablica 5. Prijedlog novog redosljeda zadataka za Test ST-1

redosljed zadatka	postojeća oznaka zadatka
1.	zadatak 2
2.	zadatak 8
3.	zadatak 7
4.	zadatak 6
5.	zadatak 4
6.	zadatak 15
7.	zadatak 3
8.	zadatak 33
9.	zadatak 12
10.	zadatak 5
11.	zadatak 35
12.	zadatak 31
13.	zadatak 34
14.	zadatak 9
15.	zadatak 18

16.	zadatak 13
17.	zadatak 11
18.	zadatak 14
19.	zadatak 16
20.	zadatak 17
21.	zadatak 27
22.	zadatak 30
23.	zadatak 32
24.	zadatak 19
25.	zadatak 10
26.	zadatak 21
27.	zadatak 28
28.	zadatak 26
29.	zadatak 29
30.	zadatak 20
31.	zadatak 22
32.	zadatak 25

Kako se vidi iz tablice, redosljed zadataka uvelike je izmijenjen, što znači da se u ovoj primjeni zadaci nisu kretali od lakših prema težima. Ipak, takav redosljed zadataka bio bi koristan jer bi se u tom slučaju ovaj test mogao primjenjivati i kao test brzine. Takva bi primjena bila ekonomičnija, jer je i u akademskom i u poslovnom okruženju, gdje se najčešće koriste čitave baterije testova, 40 minuta predugo za primjenu samo jednog testa. Ovisno o namjeri ispitivanja, tj. ovisno o tome želimo li bolje razlikovati nadprosječne ili ispodprosječne ispitanike, vrijeme primjene može se skraćivati, odnosno produljivati. No metrijske karakteristike Testa ST-1 kao testa brzine tek treba ispitati.

Za buduće primjene Testa ST-1 predlaže se ipak zadržavanje prvog zadatka koji se ne bi bodovao. Naime, iako se on po svojim karakteristikama na našem uzorku nije pokazao dobrim, neupitno je da je on ispitanicima vrlo lagan. To je u smislu osjetljivosti loše, no korisno je za stvaranje psihičkog olakšanja kod ispitanika pri početku rješavanja testa (Bosanac, 2006). Tako bi ispitanici na početku rješavanja stekli osjećaj uspjeha i opuštenije krenuli na druge zadatke, a nebodovanjem tog zadatka uklonio bi se problem njegovih loših metrijskih karakteristika.

Kao glavni sljedeći korak u razvoju konačne verzije Testa ST-1 preporuča se njegova ponovna primjena na većem i reprezentativnom uzorku i njegova standardizacija s obzirom na rod, dob, stupanj obrazovanja i ostale demografske osobine. Na prikupljenim podacima trebalo bi ponovno provjeriti funkcioniranje skraćene verzije, ili odmah primijeniti kraći oblik testa.

Kako smo već naveli, Test ST-1 razvijen je s ciljem primjene putem računala. U novije vrijeme sve je više takvih testova jer se sve više uviđaju njihove prednosti nad testovima papir-olovka. Oni štede vrijeme, eliminiraju pogreške ocjenjivanja, omogućuju veću standardizaciju i prikupljanje informacija o mnogim varijablama koje nam inače nisu dostupne, poput latencije odgovora, redoslijeda odgovaranja, broja ispravaka i mnogih drugih (Fox i Schwartz, 2001). Drugim riječima, omogućuju prikupljanje odgovora ispitanika u strogo kontroliranim eksperimentalnim uvjetima. Stoga bi bilo dobro da se sljedeća primjena Testa ST-1 provede putem računala, i zatim provjere metrijske karakteristike tako primijenjenog testa.

5. ZAKLJUČAK

Na uzorku od 212 ispitanika provedena je analiza 35 zadatka preliminarne verzije Testa ST-1 koji ispituje specijalno rezoniranje. Distribucija ukupnih rezultata statistički se značajno ne razlikuje od normalne (Kolmogorov-Smirnovljevi $Z=1,254$; $p>0,05$). Cronbachov koeficijent pouzdanosti iznosi $\alpha=0,874$, pa možemo reći da se radi o visoko pouzdanom testu. Prosječna korelacija zadatka s ukupnim rezultatom iznosi $r_{iu}=0,431$. Prosječni indeks lakoće je $p=0,630$, što govori da test ispitanicima nije bio pretjerano težak. No, većinom se radi o ispitanicima za koje očekujemo da su u prosjeku više inteligencije od prosjeka populacije.

Prosječna interkorelacija zadataka je prilično niska, $r_{ij}=0,163$, no budući da je faktorska analiza ukazala na postojanje jednog glavnog faktora, to možemo pripisati smanjenoj varijanci binarnih zadataka prije nego unutarnjoj heterogenosti testa.

Na temelju analize čestica predložena je i skraćena verzija Testa ST-1. Nakon izbacivanja 3 zadatka s indeksom diskriminativnosti manjim od 0,25, metrijske karakteristike testa ostale su uglavnom nepromijenjene. Normalna distribucija rezultata ostala je očuvana. Cronbachov α koeficijent neznatno je porastao ($\alpha=0,876$), što znači da je pouzdanost skraćene verzije čak nešto veća nego u originalnom testu. Zbog izbacivanja dva vrlo teška zadatka a samo jednog vrlo laganog, prosječni indeks lakoće nešto je porastao ($p=0,641$). Faktorska struktura testa također se nije bitno promijenila. Prema Kaiser-Guttmanovom kriteriju dobiven je jedan faktor manje nego u originalnoj verziji, a prvi faktor objašnjava čak nešto veći dio varijance (21,38% naspram 19,87%). I ovdje nam karakteristični korijeni i *screeplot* prikaz dozvoljavaju zaključak o postojanju jednog glavnog faktora.

Za daljnja istraživanja preporuča se primjena Testa ST-1 na reprezentativnom uzorku, kako bi se provjerili rezultati dobiveni u ovom radu i provela standardizacija testa. Budući da je test osmišljen s ciljem primjene putem računala, dobro bi bilo provjeriti i da li se takvom primjenom mijenjaju njegove metrijske karakteristike.

6. LITERATURA

Bele-Potočnik, Ž. i Zajec, M. (1982). *BTI – Priručnik za internu upotrebu*. Zavod SR Slovenije za produktivnost dela (Center za psihodiagnostična sredstva): Ljubljana.

Benson, E. (2003). Intelligent intelligence testing. *Monitor on Psychology*, 34 (2), 48-51.

Bosanac, I. (2006). *Analiza zadatka dviju paralelnih formi Problemnog testa*. Diplomski rad. Filozofski fakultet: Zagreb.

Ekstrom, R.B., French, J.W., Harman, H.H. i Dermen, D. (1976): *Manual for Kit of Factor-Referenced Cognitive Tests*. Educational testing service: Princeton, New Jersey.

Fox, S. i Schwartz, D (2001). Social desirability and controllability in computerized and paper-and-pencil personality questionnaires. *Computers in Human Behavior*, 18 (4), 389-410.

Fulgosi, A. (1979). *Faktorska analiza*. Školska knjiga: Zagreb.

Koren, I. i Mijić, V (1976). *Kalifornija test mentalne zrelosti – standardizacija*. Republička samoupravna interesna zajednica za zapošljavanje: Zagreb.

Krković, A. (1978). *Elementi psihologije I*. Filozofski fakultet: Zagreb.

McGlone, M. S. i Aronson, J. (2006). Stereotype threat, gender identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, 27, 486-493.

Momirović, K. i Bukvić, A. (1969). *Baterija testova sposobnosti (DAT) – priručnik*.

Nunnally, J. C., Bernstein, I. H. (1994). *Psychometric Theory*. McGraw-Hill: New York.

Petz, B. (1992). *Psihologijski rječnik*. Prosvjeta: Zagreb.

Petz, B. (2001). *Uvod u psihologiju*. Naklada Slap: Jastrebarsko.

Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for Passage-Based Tests. *Journal of Technology, Learning, and Assessment*, 2 (6), 1-30.

Rafi, A., Anuar, K., Samad, A., Hayati, M. i Mahadzir, M. (2005). Improving spatial ability using a web-based virtual environment. *Automation in Construction*, 14 (6), 707-715.

Rathus, S. A. (2000). *Temelji psihologije*. Naklada Slap: Jastrebarsko.

Sternberg, R.J. (1982). *Handbook of Human Intelligence*. Cambridge University Press: New York.

Svetličić, L. (2004). *Utjecaj glazbe na prostorno zaključivanje*. Diplomski rad. Filozofski fakultet: Zagreb.

Swain, C. (1998). Testing strategies for computerized tests: What do we know? *FETConnections*, 1, 4.

Vasta, R., Haith, M. M. i Miller, S. A. (1997). *Dječja psihologija*. Naklada Slap: Jastrebarsko.

Zarevski, P. (2000). *Struktura i priroda inteligencije*. Naklada Slap: Jastrebarsko.

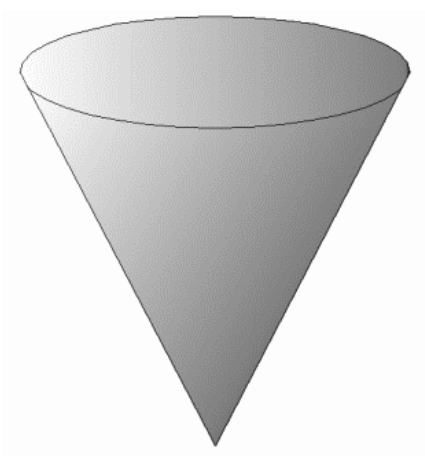
<http://www.cpi.ffzg.hr>

<http://www.ul.ie/~mearsa/9519211/newpage6.htm>

7. PRILOZI

PRILOG 1 Primjer zadatka u Testu ST-1

Odaberite plašt koji odgovara tijelu sa slike.



A)



B)



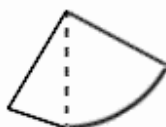
C)



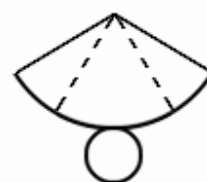
D)



E)

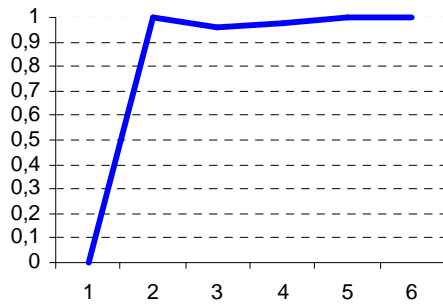


F)

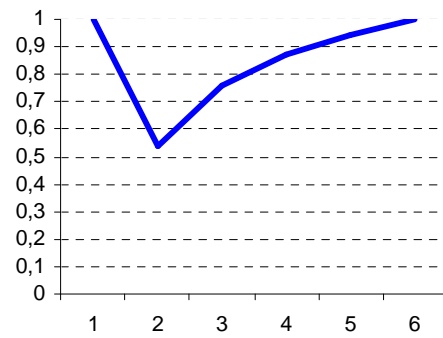


PRILOG 2 Karakteristične krivulje 35 zadataka Testa ST-1

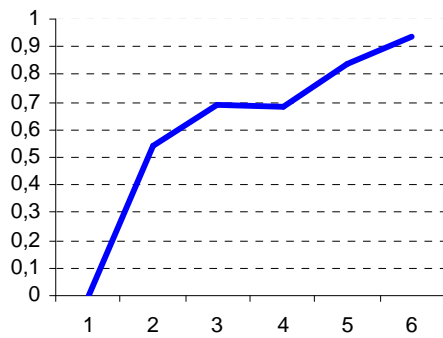
z1



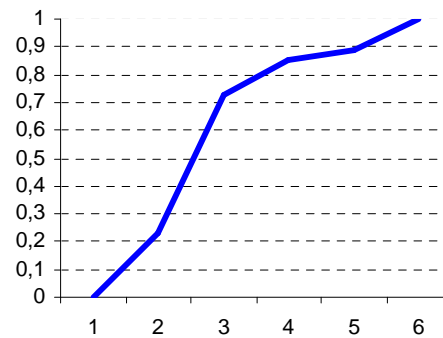
z2



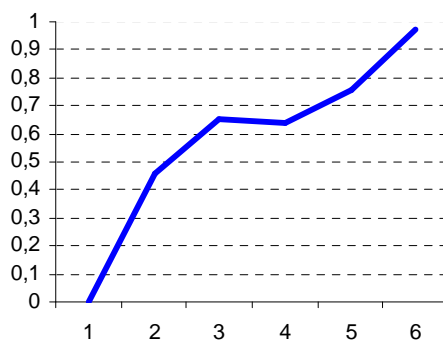
z3



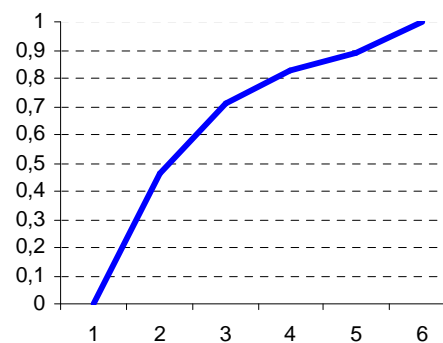
z4



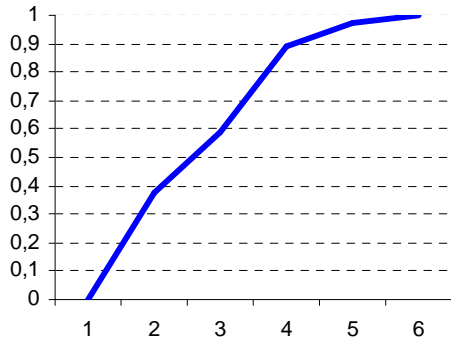
z5



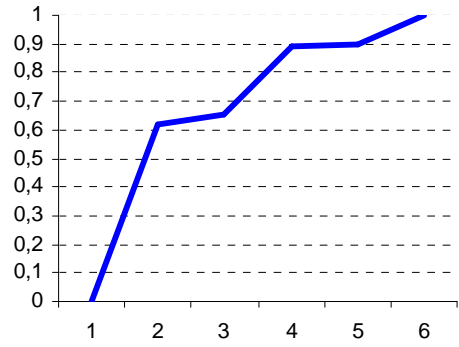
z6



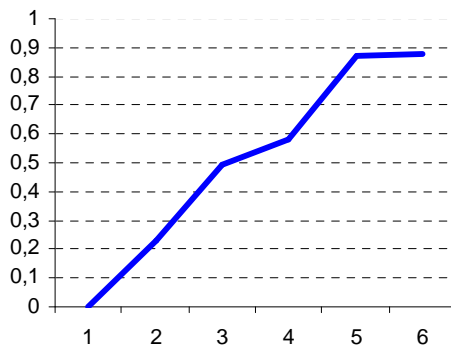
z7



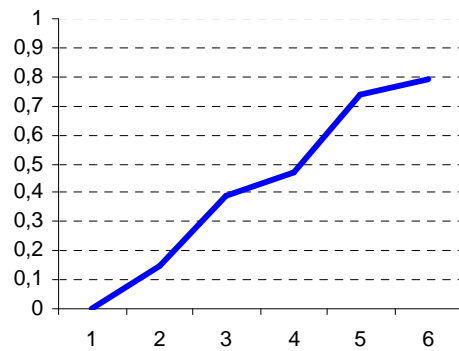
z8



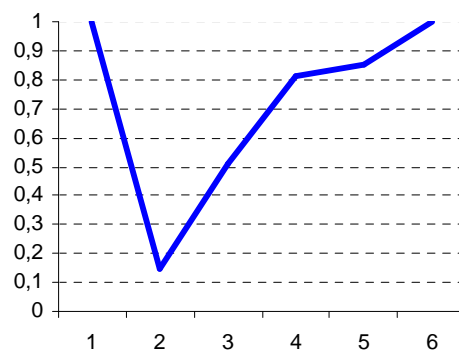
z9



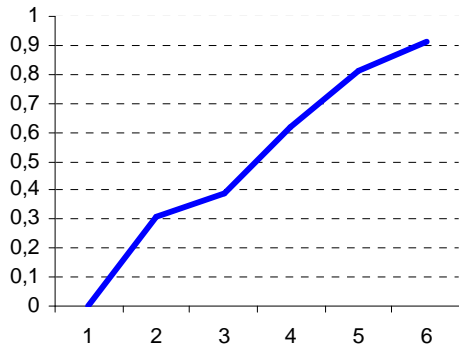
z10



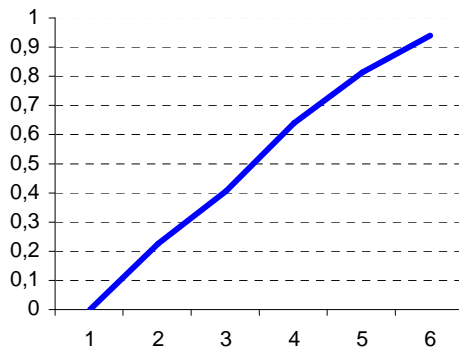
z12



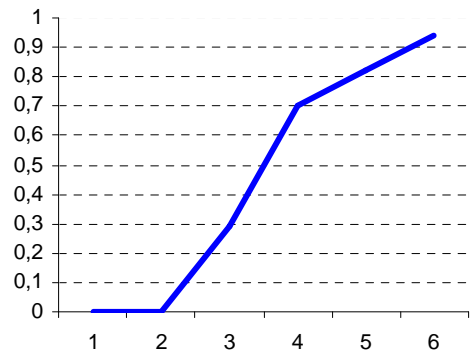
z11



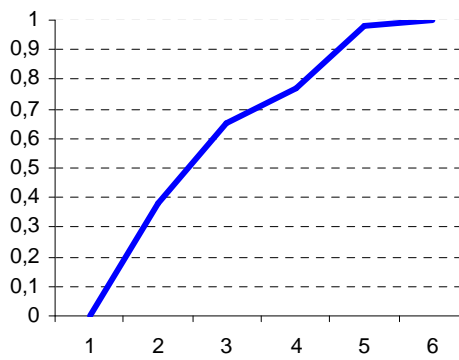
z13



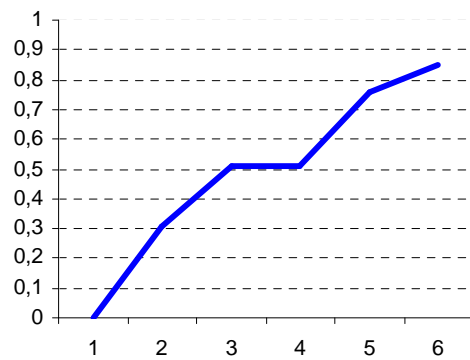
z14



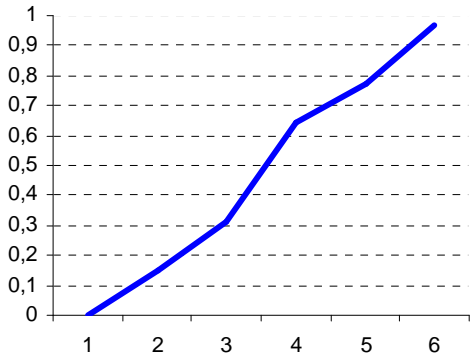
z15



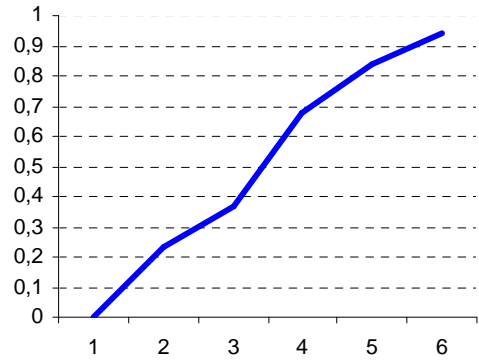
z16



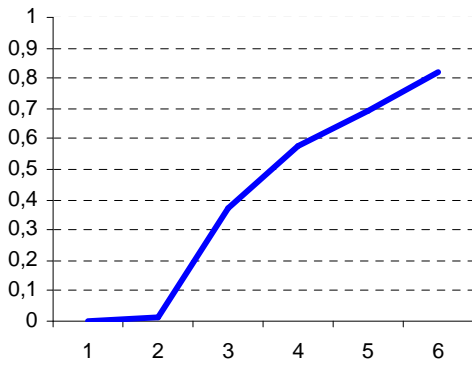
z17



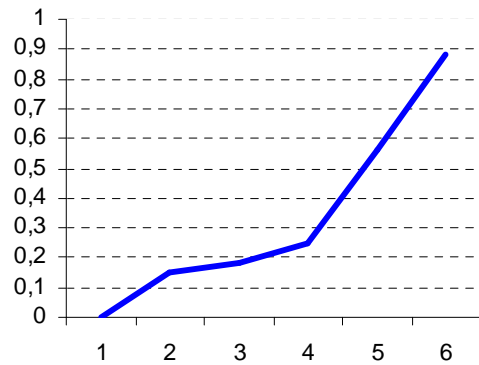
z18



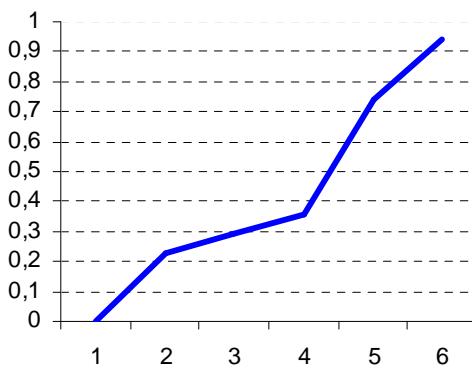
z19



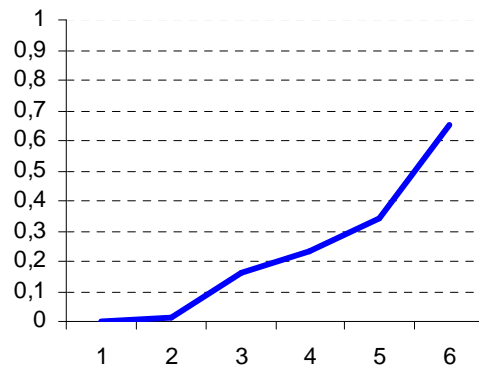
z20



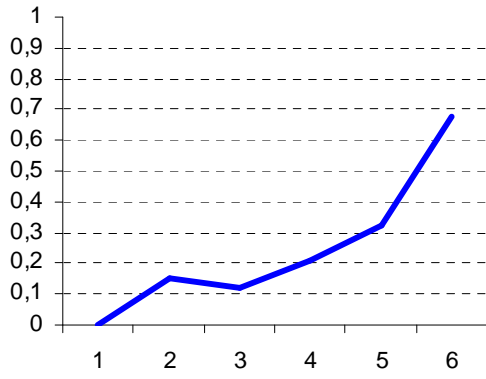
z21



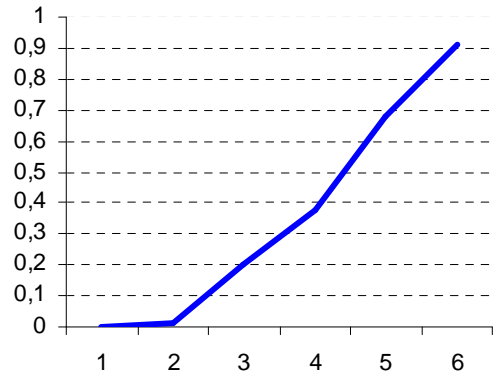
z22



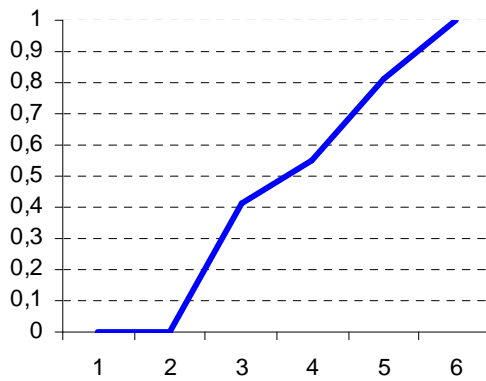
z25



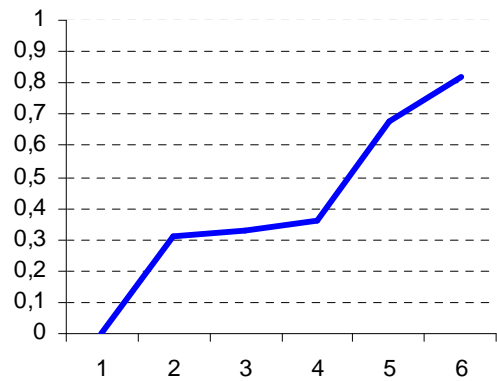
z26



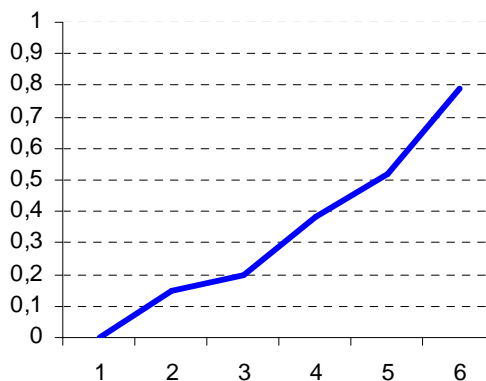
z27



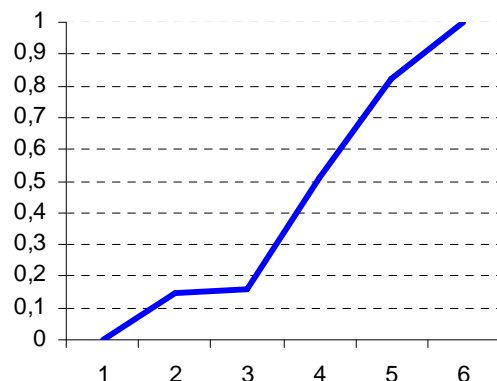
z28



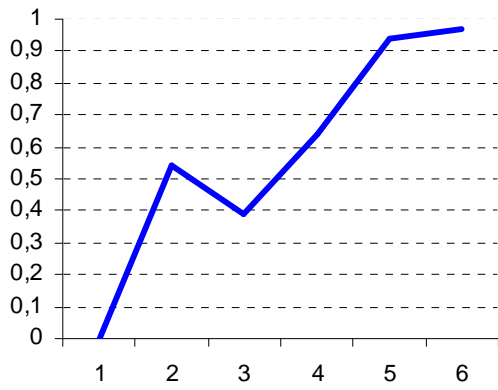
z29



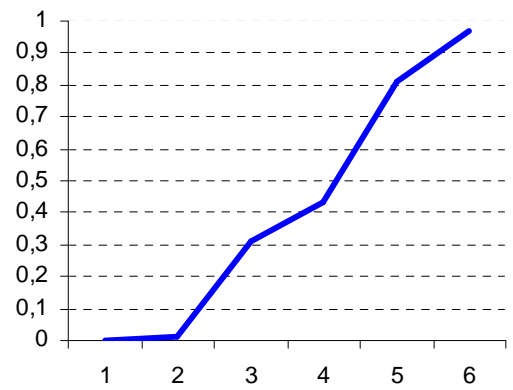
z30



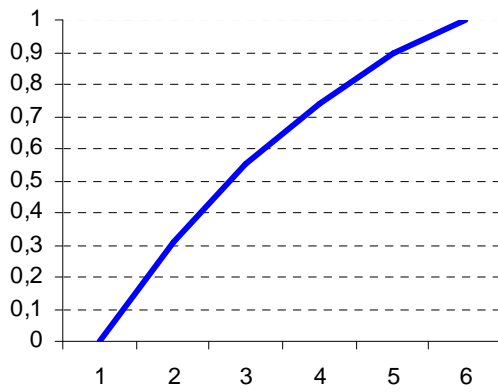
z31



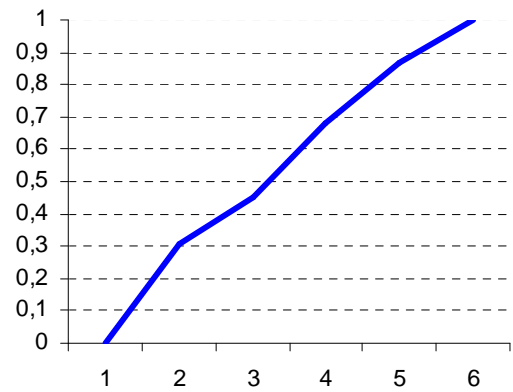
z32



z33



z34



z35

