

FAKULTET PRIRODOSLOVNO-MATEMATIČKIH ZNANOSTI
SVEUČILIŠTE U SPLITU

Martina Banovac

VREDNOVANJE INTELIGENTNIH TUTORSKIH SUSTAVA

DIPLOMSKI RAD

SPLIT, 2008.

Studijska grupa: MATEMATIKA I INFORMATIKA
Predmet: PRIMJENA RAČUNALA U NASTAVI

VREDNOVANJE INTELIGENTNIH TUTORSKIH SUSTAVA

DIPLOMSKI RAD

Student:
Martina Banovac

Mentor:
prof.dr.sc. Slavomir Stankov
Neposredni voditelj:
mr.sc. Ani Grubišić

SPLIT, 2008.

VREDNOVANJE INTELIGENTNIH TUTORSKIH SUSTAVA

EVALUATING INTELLIGENT TUTORING SYSTEMS

SAŽETAK

S obzirom da su inteligentni tutorski sustavi sve više se uključuju u obrazovanje, kao posljedica toga, potreba za opreznom i sistematičnom kontrolom ovih sustava postalo je važno pitanje. Zbog činjenice da se ti utječu na sam proces učenja i poučavanja, smatramo da se u nastavnom procesu mogu koristiti samo oni sustavi čija je učinkovitost ispitana.

Na temelju spoznaja iz literature dali smo pregled metoda vrednovanja inteligentnih tutorskih sustava, s obzirom na to što se želi vrednovati, te smo izdvojili eksperiment kao najpouzdaniju metodu vrednovanja. Budući da je glavni cilj obrazovnog sustava poučavanje, naveli smo što je to veličina učinka, te smo naveli kriterije koji se koriste za procjenu obrazovnog učinka.

Sa ciljem utvrđivanja učinkovitosti korištenja sustava xTEx-Sys u primarnom obrazovanju proveli smo eksperiment uz parcijalna ispitivanja u OS Bol, Split nad učenicima petih razreda iz područnog znanja iz matematike (geometrija). U eksperimentu su sudjelovala dva peti razreda. Ukupno 48 učenika petih razreda u razdoblju od tri tjedna, od 11.5 do 1.6.2007. godine, je prošlo postupak učenja i poučavanja na sustavu xTEx-Sys. Također su ispunjavali dvije ankete, te su pisali inicijalni test, dva međutesta i završni test.

Ključne riječi:

Inteligentni tutorski sustavi, metode vrednovanja, eksperiment, veličina učinka, učenje i poučavanje

Key words:

Intelligent tutoring systems, evaluation methods, experiment, effect size, programming learning and teaching

Ovaj diplomski rad izrađen je na Fakultetu prirodoslovno-matematičkih znanosti, Sveučilište u Splitu.

Ovom prigodom posebno zahvaljujem sljedećim osobama koje su pridonijele ostvarenju ovog diplomskog rada:

- prof. dr. sc. Slavomiru Stankovu na nesebičnoj pomoći tijekom cijelog studiranja*
- mr. sc. Ani Grubišić na pomoći i savjetima tijekom izrade diplomskog rada*
- ravnatelju i djelatnicima OŠ Bol, Split na podršci i pomoći prilikom realizacije eksperimenta*
- svim dragim i bliskim ljudima koji su mi pružali podršku*
- te najveću zahvalu upućujem mojoj obitelji na uvijek prisutnoj podršci i bezgraničnom strpljenju.*

Sadržaj

1. UVOD.....	1
2. TEHNIKE VREDNOVANJA	3
2.1. FORMATIVNO I SUMATIVNO VREDNOVANJE	3
2.2. PREGLED METODA VREDNOVANJA	4
2.3. KLASIFIKACIJA METODA VREDNOVANJA.....	8
2.4. EKSPERIMENT KAO METODA VREDNOVANJA.....	9
2.4.1. Pogreške u eksperimentu.....	10
2.4.2. Vrste eksperimentalnih istraživanja	10
2.5. VREDNOVANJE OVISNO O ARHITEKTURI.....	14
3. VREDNOVANJE OBRAZOVNOG UČINKA.....	19
3.1. VELIČINA UČINKA.....	19
3.2. MJERE POSTIGNUĆA.....	20
3.3. MJERE VEZANE ZA REZULTATE	21
3.4. MJERE AFEKTA UČENJA	22
4. UČENJE I POUČAVANJE UZ POMOĆ SUSTAVA XTEX-SYS.....	24
4.1. CILJ, SUDIONICI I TIJEK ISTRAŽIVANJA	24
4.2. TIJEK ISTRAŽIVANJA EKSPERIMENTALNE GRUPE.....	25
4.3. ANALIZA REZULTATA ISTRAŽIVANJA	27
4.3.1. Analiza testova	27
4.3.2. Analiza anketa eksperimentalne grupe.....	30
5. ZAKLJUČAK.....	41
6. LITERATURA	42
7. PRILOZI.....	43

1. Uvod

Prirodna je situacija da suvremena nastava prati razvoj tehnologije, te nastoji u obrazovni proces uvesti nova nastavna sredstva kako bi se učenicima približili nastavni sadržaji, motiviralo ih se na rad, poboljšalo razumijevanje, otkrivanje i usvajanje pojmova, pojava i zakonitosti. Kao što su tijekom prošlih godina u nastavni proces kao pomagala ušli grafoskopi, dijaskopi i dr., tako smo danas svjedoci sve češćeg poučavanja i učenja uz pomoć računala, te učenja od računala.

Utjecaj informacijske i komunikacijske tehnologije (eng. *information and communication technology*, ICT) na proces stjecanja znanja, sposobnosti i vještina su velike. Uz potporu ove tehnologije u obrazovanju se sve više koriste različiti programski sustavi za realizaciju klasične nastave. Posebna vrsta takvih sustava su inteligentni tutorski sustavi (eng. *Intelligent Tutoring System*, ITS). Primjena tih programskih paketa u središte stavlja učenika, dok se kod klasične nastave u učionici u središtu nalazi sam učitelj. Iako se ti sustavi razvijaju kao potpora nastavi, mali je broj onih koji su vrednovani upotrebom neke metodologija za vrednovanje. Vrednovanje sustava, a pogotovo vrednovanje njegovog učinka, zahtijeva dugotrajan angažman nastavnika koji koristi te sustave u nastavi, ali isto tako i učenika koji prisustvuju toj nastavi.

Donedavno je mala pozornost bila usmjerena na vrednovanje ITS-a. Većina ITS istraživanja odnosila su se na predviđanja o mogućnostima ITS-a i razmatranje rezultata primjene uključenih u konstruiranje aktualnih komponenti i sustava. Bitna istraživanja kao što su prikazivanje znanja, istraživanje i planiranje, pridonose dizajnu i primjeni ITS-a. Inteligentni tutorski sustavi sve više se uključuju u obrazovanje i kao posljedica toga, potreba za opreznom i sistematičnom kontrolom ovih sustava postalo je važno pitanje. Do danas su se razvili mnogi inteligentni tutorski sustavi. Međutim, malo se istraživalo na razvoju metoda vrednovanja inteligentni tutorskih sustava. Vrednovanjem učinka tih sustava se želi utvrditi koliko učinkovito učenici uče uz pomoć tog sustava, te koliko novi načini upotrebe informacijske i komunikacijske tehnologije utječu na proces učenja i poučavanja.

Prikladno vrednovanje može poslužiti kao pomoćno sredstvo za ubrzavanje razvoja istraživanja na način da se daju prijedlozi za potpuno unapređenje građe i ponašanja ITS-ova. Kada bi takvi sustavi bili dostupni u budućnosti u obrazovnim institucijama, vrednovanje će morati dati odgovore na pitanja o koristi tih sustava, odnosno njihovoj mogućnosti da potiču i unaprjeđuju učenje. Vrednovanje može utjecati na to što i kako učenici uče. Može pružiti informaciju o građi sustava i utjecaju svog ponašanja na korisnika koji može poslužiti kao povratna informacija sustavu i na taj način pripomoći u poboljšanju sustava. Na ovaj način, vrednovanje pomaže u određivanju koliko neki sustav zadovoljava potrebe i da otkrije svoju vrijednost istraživanja, kao i snage i nedostatke. Naposljetku, vrednovanje može u konačnici utjecati na izbor treba li ili ne koristiti određeni inteligentni tutorski sustav.

Izbor metodologije vrednovanja prvenstveno ovisi o pitanjima na koja želimo dobiti odgovor, jer ne postoji jedinstveni način vrednovanja svih sustava. Zatim se postave hipoteze koje treba biti moguće testirati, te potvrditi ili opovrgnuti na osnovi određenih uvjeta i rezultata istraživanja. Zatim slijedi definiranje metodologije za vrednovanje kojom će se ispitati istinitost nul-hipoteze. Nakon provedenog istraživanja potrebno je analizirati podatke. Ako rezultati ne potvrđuju hipotezu, bilo bi dobro od istraživača, da iznese moguća objašnjenja za dobivene rezultate.

Prilikom vrednovanja složenih sustava može se vrednovati cjelokupni sustav ili samo neke njegove komponente. Tehnike vrednovanja koje su prikladne za vrednovanje cjelokupnog sustava nisu prikladne za vrednovanje samo nekih njegovih komponenti. Metoda vrednovanja ovisi o svrsi istraživanja. Prvo, treba usporediti koliko je učinkovit sustav u odnosu na tradicionalni način podučavanja, i drugo, treba odrediti značajke sustava koje su bitne za vrednovanje. Ako želimo utvrditi učinak sustava, tada moramo vrednovati cjelokupni sustav. Najprikladnija metoda za vrednovanje sustava je eksperiment jer pomoću nje možemo najpouzdanije otkriti uzroke javljanja određene pojave.

Eksperiment je znanstvenoistraživački postupak za otkrivanje uzročno–posljedičnih veza među pojavama. Glavna obilježja eksperimentalnog istraživanja su sistematsko i namjerno mijenjanje uvjeta u kojima se neka pojava javlja. Uvjeti koji se namjerno mijenjaju u eksperimentalnom istraživanju nazivaju se nezavisnim varijablama, a pojava na koju se djeluje naziva se zavisnom varijablom.

U svrhu našeg istraživanja proveli smo eksperiment s paralelnim skupinama. Kod eksperimenta s paralelnim skupinama imamo dvije grupe ispitanika, gdje je svaka skupina nosilac svog eksperimentalnog faktora. Kontrolna skupina se poučava na tradicionalni način, a eksperimentalna skupina uz pomoć određenog inteligentnog tutorskog sustava. Kod osnovnog modela eksperimenta s paralelnim skupinama provodi se inicijalno i završno ispitivanje kako bi se izmjerilo inicijalno i završno znanje ispitanika. U našem slučaju smo imali i dva međutesta gdje smo također mjerili razinu znanja ispitanika. Takvu vrstu eksperimenta nazivamo *eksperiment s paralelnim skupinama uz parcijalno ispitivanje stanja* [GRUB2007]. Za vrednovanje učinka sustava koristili smo programski alat EVEDIN (EVALuation of EDucational INfluence) [GRUB2007]. EVEDIN omogućuje automatsko upravljanje kontrolnom i eksperimentalnom grupom, izradu i provedbu svih testova, njihovo ispravljanje i ocjenjivanje, te računanje veličine učinka eksperimentalnog faktora.

2. Tehnike vrednovanja

Postoji nekoliko osnovnih načela o vrednovanju ekspertnih sustava koji se mogu primjeniti i na vrednovanje ITS-a. Kompleksni sustav kao, ITS, se može razmatrati u granicama potpunog i čitavog sustava ili u granicama specifičnih obilježja. Tehnike koje su pogodne za vrednovanje cjelokupnog ITS-a, nisu prikladne za vrednovanje pojedinih ITS komponenti ili osobina. Također, različite tehnike vrednovanja su korisne ovisno o njihovim interesima. Različite metode vrednovanja su pogodne za različite ciljeve, s različitim komponentama ili čitavim sustavom. Naime, treba imati na umu da je razvijanje vrednovanja kompleksan proces.

2.1. Formativno i sumativno vrednovanje

Vrednovanje je proces u kojem su važni podaci prikupljeni i transformirani u informacije za donošenje odluka [MARK1993]. Frye, Littman i Soloway [FRYE1988] su napravili razliku između formativnog i sumativnog vrednovanja instruktivskih materijala, koji su u skladu s namjenom inteligentnih tutorskih sustava. *Formativno vrednovanje* je vrednovanje koje se javlja tijekom dizajniranja i ranog razvijanja projekta i usmjereno je na trenutne potrebe osoba koje se bave poboljšanjem dizajna i ponašanja sustava. To često dovodi do pitanja "Koja je veza između arhitekture i ponašanja ITS-a?". Suprotno tome, *sumativno vrednovanje* odnosi se na vrednovanje cjelokupnog sustava i definiranje bitnih zahtjeva o sustavu. Sumativno vrednovanje potiče na još jedno bitno pitanje o vrednovanju ITS-a. "Koji je obrazovni učinak ITS-a na učenike? "

Formativno vrednovanje se često uzima kao dio metodologije računalnog programiranja, koje karakterizira razvijanje u okviru oblikovanja, provedbe i formativne procjene [MARK1993]. Formativno vrednovanje se koristi za postizanje detaljnih informacija koje se mogu koristiti za izmjenu i poboljšanje rada ITS-a. Osnovno pitanje, kada planirati formativno vrednovanje je: "Koje su najvažnije vrste podataka koje treba pribaviti i koji je glavni izvor tih podataka?". Ovo pitanje se odnosi na izjednačavanje bitnih stvari, od detalja o oblikovanju sučelja do ciljeva sustava. Instruktivski projektanti preporučuju da bi se formativno vrednovanje trebalo ranije razvijati i nastaviti tijekom razvijanja sustava. Budući da se ovi rezultati koriste kao baza za daljnje razvijanje, neformalne ili „ad hoc“ metode su općenito prihvaćene za formativno vrednovanje.

Suprotno tome, sumativnim vrednovanjem se pokušavaju potvrditi ili opovrgnuti neki bitni zahtjevi o sustavu ili tehnikama koje su korištene u sustavu [MARK1993]. Općenito, zahtijeva povezivanje ciljeva sustava. Specifična pitanja mogla bi biti: "Što posebno izvršava ITS?", "Je li ITS udovoljava zahtjevima za koje je on dizajniran?", "Kakav je učinak jedne vrste sustava ili komponenti koje se odnose na sličan sustav ili komponente?". Budući da se to odnosi na bitne zahtjeve, sumativno vrednovanje očekuje mnogo rigoroznije metodologijske standarde nego formativno vrednovanje.

Osim formativnog i sumativnog vrednovanja, javljaju se još neki modeli vrednovanja. Interno vrednovanje koje se odnosi na vrednovanje neke komponente sustava, eksterno vrednovanje koje se odnosi na vrednovanje cjelokupnog sustava, eksploratorno vrednovanje zahtijeva dubinsku analizu sustava u prirodnom okruženju koristeći višestruke izvore podataka, dol eksperimentalno vrednovanje koristi eksperimente za sistematično mijenjanje nezavisnih varijabli i istovremeno mjerenje zavisnih varijabli. Slučajnim odabirom potrebno je podijeliti sudionike u kontrolnu i eksperimentalnu grupu, te je potrebno da te grupe ne budu statistički značajno različite.

Postoji nekoliko dogovorenih standarda unutar zajednice istraživača ITS-a koji žele vrednovati sustav. Međutim, druga područja su razvila tehnike vrednovanje koje bi mogle biti prikladne za ITS. Računalno najbliže ITS-u je ekspertni sustav, dok je obrazovno najbliži CAI (eng. *Computer-Assisted instruction*) sustav. ITS istraživanja također mogu biti korisna za razvoj polja obrazovanja i psihologije.

2.2. Pregled metoda vrednovanja

Metode vrednovanja inteligentnih tutorskih sustava uključuju mnoštvo faktora koje treba promatrati i upravo zbog toga je vrednovanje ITS-a teška zadaća. Sjedeće metode koriste istraživači iz područja ekspertnih sustava, obrazovanja, računalnih znanosti i psihologije (prema [IQBA1999]). Prve četiri metode su pogodne za interno vrednovanje obzirom na testiranje komponenti sustava, a ne cjelokupnu učinkovitost sustava. Metode 6 i 7 usredotočene su na cjelokupnu učinkovitost sustava i pogodna su za eksterno vrednovanje. Metoda 5 može se koristiti za oboje, za testiranje komponenti ili čitavog sustava, stoga je pogodna i za interno i za eksterno vrednovanje. Preostale metode, od 8 do 20, uključuju istraživanja sustava u stvarnom kontekstu koristeći višestruke izvore podataka, najčešće tamo gdje je uzorak mali i područje nedovoljno razumljivo. Metode od 8 do 12 su pogodne za interno vrednovanje s obzirom da su usmjerene na vrednovanje dijela sustava, a ne čitavog sustava. Metode od 15 do 20 su pogodne za eksterno vrednovanje, a 13 i 14 za oba načina vrednovanja.

1. Dokaz ispravnosti (eng. *proof of correctness*)

U provjeravanju programa, sustav je vrednovan u ovisnosti je li ili nije njegovo ponašanje ispunilo određene zahtjeve i ciljeve, odnosno je li postoji korespodencija između njegove strukture i njegovih specifikacija, tj ispituje se radi li ono što treba. Ova metoda vrednuje unutarnje komponente sustava Naime, ova metoda ne odgovara za sve programe iz umjetne inteligencije koji rade s analitički neuhvatljivim problemima predstavljenim nepotpuno određenim funkcijama. Ova kategorija uključuje i ITS-ove. Što znači da je ova metoda neprihvatljiva za vrednovanje inteligentnih tutorskih sustava.

2. Usporedbe aditivnog eksperimentalnog dizajna (eng. *additive experimental design*)

Ova metoda koristi se za vrednovanje utjecaja velikih komponenti sustava koje se mogu eksperimentalno izmijeniti i izbaciti. Prema tome one pripadaju eksperimentalnom istraživanju i prikladne su samo za interno vrednovanje. Ovaj pristup zahtijeva velik broj učenika i nije isplativ ako komponenta nije od velike važnosti. Prednost aditivnog dizajna uključuje mogućnost manipuliranja individualnim aspektima sustava da bi se direktno odredio njihov učinak i važnost.

3. Dijagnostička točnost (eng. *diagnostic accuracy*)

Dijagnostičku točnost sustava je moguće odrediti pomoću procedura koje naglašavaju kvalitetu mikro-teorija koje koristi sustav. Ova metoda omogućava kontrolu modela učenika kroz zahtjeve i pripada u eksperimentalna istraživanja.

4. Povratne informacije / kvalitete poučavanja (eng. *feedback / instruction quality*)

Direktni učinak povratne informacije/kvalitete poučavanja na učenika može se eksperimentalno procijeniti sekvencijalnim analitičkim procedurama koje su prikladne za interno vrednovanje. Jedan od načina mjerenja ovog utjecaja je da se izračuna udio određenih akcija od ukupnog broja akcija koje se javljaju unutar određenog okvira koje vodi određenoj ciljanoj akciji. Korist ovakvog mjerenja je u tome što se računanje može odnositi na različitu razinu detalja povratne informacije sustava ili na različite tipove povratnih informacija.

5. Analiza osjetljivosti (eng. *sensitivity analysis*)

Ispituje komponentu ili sustav da se ispita na koji način ponašanje odgovara na različite informacije koje sustav prima. To je osobito bitno za vrednovanje inteligentnih tutorskih sustava, koji se trebaju prilagođavati različitim karakteristikama individualnih učenika. Osjetljivost sustava na različite karakteristike bi mogla ukazati na ta to da li bi sustav trebalo poboljšati. Sustav koji pokazuje slične odgovore na znatno različite informacije može se smatrati manje poželjan od onoga koji pokazuje veće razlike. Ova se metoda može primijeniti za sumativno vrednovanje ako se mogu precizno odrediti mjere osjetljivosti i ako im se može dodijeliti značenje (koja ponašanja su poželjna i kada ih treba pozvati), a inače se primjenjuje za formativno vrednovanje. [MARK1993].

6. Eksperimentalno istraživanje (eng. *experimental research*)

Ovo je uobičajena metoda u psihologiji i obrazovanju. Omogućava otkrivanje veze između akcija koje se poduzimaju u poučavanju i odgovarajućih rezultata koje učenici postižu u odnosu na te akcije, a također se može odrediti i koliko su takve veze značajne, i stoga je posebno pogodan za ispitivanje učinka procesa učenja i poučavanja. Eksperiment se kod formativnog vrednovanja može koristiti za uspoređivanje učinka nekih značajki sustava. Češće se koristi za sumativno vrednovanje kod kojeg su poželjniji sveukupni zaključci, nego skupljanje informacija. Prema [MARK1993] postoje različita eksperimentalna istraživanja kao što su eksperimenti s jednom grupom, eksperimenti s kontrolnom grupom, te kvazi-eksperimentalna istraživanja. Najprije je potrebno odrediti što se želi istražiti, pa se zatim odrede hipoteze (npr. da će nakon nekog načina poučavanja značenje razlike među grupama). Hipotezu treba biti moguće testirati, te provoditi ili opovrgnuti na osnovi određenih uvjeta i rezultata istraživanja. Zatim se određuje oblik istraživanja kako bi istraživač mogao ispitati hipotezu. Istraživač nakon provedenog istraživanja analizira podatke. Ako rezultati ne potvrđuju hipotezu bilo bi dobro da istraživači, ako je moguće, predlože moguća objašnjenja za svoje rezultate. Eksperimentalno istraživanje je prikladnije za eksterno vrednovanje jer daje sveukupne zaključke.

7. Vrednovanje proizvoda (eng. *product evaluation*)

Vrednovanje proizvoda je vrednovanje širokog opsega koje za cilj ima utvrditi procedure vrednovanja koje mogu ispitati opravdanost daljnjeg razvoja sustava. Obrazovna učinkovitost inteligentnih tutorskih sustava, živih učitelja i tradicionalnih metoda poučavanja mora biti uspoređena na osnovi stvarnih podataka. Trebaju se vrednovati samo važne aplikacije sustava. Također je potrebna velika grupa subjekata za preciznu procjenu učinkovitosti sustava. Ova metoda pripada eksperimentalnom istraživanju.

8. Pregled stručnjaka (eng. *expert inspection*)

Ova metoda se koristi da bi se utvrdilo da li sustav doseže standardni nivo performansi. Ova vrsta vrednovanja se najčešće koristi za razvoj ekspertnih sustava i vrlo dobar primjer je Turingov test

koji uspoređuje ponašanje čovjeka (učitelja) i računala. Ovakvo ispitivanje je korisno za vrednovanje karakterističnih komponenti sustava što je pogodno za interno vrednovanje. Ova metoda se može klasificirati kao eksploratorna jer se ne ispituje nikakva kontrolirana hipoteza.

9. Nivo dogovora (eng. *level of agreement*)

Nivo dogovora između stručnjaka područnog znanja se odnosi na kontrolu znanja i prikladan je za interno vrednovanje. Ova metoda koristi stručnjake i korelacijske metode da bi se procijenila konzistencija znanja i uvjerenja i uvjerenja, te zbog toga ova metoda spada u eksploratorno istraživanje.

10. Čarobnjak iz Oza (eng. *Wizard of Oz*)

Eksperiment Čarobnjak iz Oza koristi čovjeka za simulaciju ponašanja predloženog sustava tako da čovjek može testirati aspekte sustava prije same implementacije. Ova metoda je prikladna za interno vrednovanje. Teškoća u otkrivanju kako će se čovjek ponašati onemogućava donošenje odluka unaprijed, na način kako bi se parametri trebali proučavati. Ova metoda spada u eksploratorno istraživanje.

11. Metrike performansi (eng. *performance metrics*)

Ovo su metode koje koriste kvantitativne rezultate da bi se istražili individualni faktori ili značajke sustava, pa zbog toga spadaju u kategoriju internog vrednovanja. One su važne kod eksploratornih postavki kada ima malo podataka da bi se mogli donijeti statistički značajni zaključci. Može doći do pojave lažnih pozitivnih i lažnih negativnih rezultata. Ova metoda obuhvaća tri podpodručja: dijagnostička ispravnost (eng. *diagnostic accuracy*), gdje se mjera uspješnosti komponenti sustava ocjenjuje bilježenjem "pogođenim i promašenim stavkama" učenika, mogućnost korištenja (eng. *feature usage*), koja ocjenjuje kako korisnici koriste različite mogućnosti tutora; kognitivna promjena (eng. *cognitive change*) koja opisuje razlike populacije s jedne strane i vremenskog razdoblja s druge strane.

12. Unutrašnje vrednovanje (eng. *internal evaluation*)

Metode unutrašnjeg vrednovanja proučavaju odnos između arhitekture i ponašanja sustava, te obično uključuje detaljnu eksploratornu analizu akcija i struktura podataka. Ova metoda obuhvaća sljedeće podkategorije: analiza razine znanja (eng. *knowledge level analysis*) za vrednovanje baza znanja; analiza procesa (eng. *process analysis*) za vrednovanje algoritama; odnošenje i zamjena eksperimenata (eng. *ablation and substitution experiments*) za promatranje dostignuća sustava zamjenom dijelova sustava s primitivnijim verzijama.

13. Metoda zasnovana na kriterijima (eng. *criterion-based*)

U drugom pristupu vrednovanja, sustav se smatra uspješnim ili adekvatnim ako ne pokazuje velike nedostatke unutar okoline koja je namijenjena za njegovu primjenu.

Osoba koja vrednuje sustav mora odrediti koji su to "veliki nedostaci", s obzirom na to kako sustav ispunjava svoje zahtjeve i specifikacije. Kako je to subjektivno mišljenje upitno je na temelju čega su odabrani takvi kriteriji. Teško odrediti određeni kriterij na temelju objektivnosti za sustav baze znanja, usmjeravajući da je ovaj tip vrednovanja djelomično koristan za ITS (McGraw i Harbison-Briggs, prema [MARK1993]). Prikladniji je za formativno vrednovanje, pogotovo u ranijoj fazi gdje se radi o općenitim karakteristikama, a rjeđe o preciznim detaljima ili točno određenim aspektima sustava, kao oblikovanje sučelja, gdje kriterij može biti precizno izmjereno.

Općenite smjernice za prosuđivanje konstrukcije programa, ponašanja i karakteristika se ponekad preporučuju. Popularno za CAI, ovaj pristup se također koristio za ITS. Smjernice mogu podići interes i izazov vezano za dizajn i uporabu okruženja za učenje. Međutim, uporaba kontrolnih popisa i smjernica u određivanju uspješnosti se ozbiljno kritizira. Procjene nisu međusobno u korelaciji, niti su u korelaciji sa učenikovim postignućem.

14. Pilot testiranje (eng. *pilot testing*)

Koristi se za ispitivanje ponašanja sustava na neočekivane rezultate do kojih može doći prilikom korištenja sustava. Ova metoda je vrlo važna za otkrivanje problema u sustavu. Ako autori sustava mogu pretpostaviti da će korisnici biti slične stručnosti i iskustva kao i oni, onda mogu izraditi sustav prema svojim potrebama i biti prilično sigurni da će to odgovarati i korisnicima. Međutim, jasno je da se korisnici mogu razlikovati prema svom znanju o tutorskom poučavanju i tutorima, te osnovnim spoznajnim sposobnostima. Stvarni sustav bi se trebao ispitati na stvarnim korisnicima kako bi se otkrili eventualni problemi i neočekivani rezultati. Gagné i Golas (prema [MARK1993]) govore o tri vrste pilot testiranja: jedan prema jedan testiranje, testiranje malih grupa, testiranje u stvarnim uvjetima (eng. *field testing*). Kod jedan prema jedan testiranja promatraju se interakcije učenika s materijalom za poučavanje koje se razvija (otkrivaju se neodgovarajuća očekivanja; nejasni smjerovi, pitanja i informacije), a obično se provodi u ranim fazama razvoja. Testiranje malih grupa se obično provodi na malim grupama učenika koji predstavljaju ciljnu populaciju. Izvodi se u kasnijim fazama razvoja, kad se program i njegov sadržaj stabiliziraju. Testiranje u stvarnim uvjetima ispituje upotrebu sustava sa stvarnim predavačima i učenicima u stvarnom okruženju kada je sustav gotovo završen. Nastoji se otkriti kakvi se problemi događaju kad se sustav uvodi u okolinu gdje će se zapravo koristiti (Hoecker i Elias, prema [MARK1993]).

15. Certifikacija (eng. *certification*)

Zasniva se na metodama za identificiranje od strane komponentnih živih učitelja. Takva procedura ima koristi od autoritativne potpore ispravnosti sustava i stoga spada u eksploratorno istraživanje. Prikladnost pristupa za cijeli sustav čini ga prikladnim za eksterno vrednovanje. Ova metoda omogućava dobivanje povratne informacije o prednostima i nedostacima sustava kao dio formativnog vrednovanja i dobivanja ocijene prikladnosti kao dio sumativnog vrednovanja. No, ova metoda nameće pitanje standarda i kriterija za vrednovanje sustava i njegovih komponenti, te točnosti kojom ljudi mogu identificirati učinkovite sustave.

16. Vanjsko vrednovanje (eng. *outside assessment*)

Vanjsko vrednovanje, odnosno mišljenje stručnjaka ili velikog broja potencijalnih korisnika sustava, može biti značajna ako postoji dogovor. Ova metoda radi na cjelokupnom sustavu i zasnovana je na proučavanju mišljenja korisnika, pa je stoga prikladna za eksploratorno eksterno vrednovanje.

17. Postojanje dokaza (eng. *existence proofs*)

Metoda zasniva svoje zaključke na uspješnoj implementaciji sustava i predlaže novu arhitekturu sustava. Kako se metoda odnosi na cijeli sustav, prikladna je za eksterno vrednovanje. Ova metoda prikladna je jedino ako istraživač opiše ciljeve i pretpostavke istraživanja, moguće promjene u dizajnu, te dokumentira moguća iznenađenja i neuspjehe. Korisna je u početku eksploratornog istraživanja za identifikaciju ključnih problema.

18. Promatranje i kvalitativna klasifikacija fenomena (eng. *observation and the qualitative classification of phenomena*)

Ova metoda ima za cilj identificirati klase fenomena, modele i trendove u interakciji čovjeka i sustava. Ovi podaci se prikupljaju za eksploratorno istraživanje primarno kroz promatranje, od neposrednih do izmišljenih situacija i to kao dio eksternog vrednovanja.

19. Strukturne zadaće i kvantitativna klasifikacija fenomena (eng. *structured task and the quantitative classification of phenomena*)

Ovo je pristup koji se fokusira na skup podataka ograničavajući ili organizirajući odgovore subjekta unutar eksploratornog istraživanja. Kvantitativni podaci su sakupljeni u planiranim situacijama

unutar eksternog vrednovanja, pomoću razgovora ili upitnika. Prednosti ovog pristupa su u tome što se veliki i precizni podaci mogu sakupiti uporabom planiranih zadaća.

20. Studije usporedbe (eng. *comparison studies*)

Studije usporedbe su opća klasa metoda koje bilježe sličnosti i razlike između ponašanja i oblikovanja promatranog sustava i nekog drugog sustava. Ove metode su prikladne za eksterno vrednovanje i primjenjuju se na cjelokupne performanse sustava u eksploratornom istraživanju.

Mark i Greer [MARK1993] uz dokaz ispravnosti, metodu zasnovanu na kriterijima, certifikaciju, analizu osjetljivosti, pilot testiranje i eksperimentalno istraživanje kao tehnike vrednovanja spominju još i **znanje i ponašanje stručnjaka** (eng. *expert knowledge and behaviour*). Ova metoda se koristi kao eksplicitni standard za vrednovanje sustava (Gaschnig i drugi, 1983, prema [MARK1993]). Stručnjak pregledava sve relevantne aspekte površinskog ponašanja tijekom formativnog vrednovanja. Pregled stručnjaka je moguć jedino kada je ponašanje dosljedno i predvidivo. Ova metoda ima ograničene primjene za vrednovanje ITS-a, zbog složenog i dinamičnog ponašanja. Pregled je ponekad usmjeren na ITS komponente, što može biti vrlo korisno za formativno vrednovanje (Hofmeister, 1986, prema [MARK1993]). Najpoznatiji primjer takvog načina vrednovanja je Turingov test (Parry i Hofmeister, prema [MARK1993]). Međutim, Turingov test neće dati pravedne rezultate, jer se živi učitelj i sustav razlikuju u onom što nude učenicima, a sam proces učenja i poučavanja je sam po sebi nedovoljno jasan.

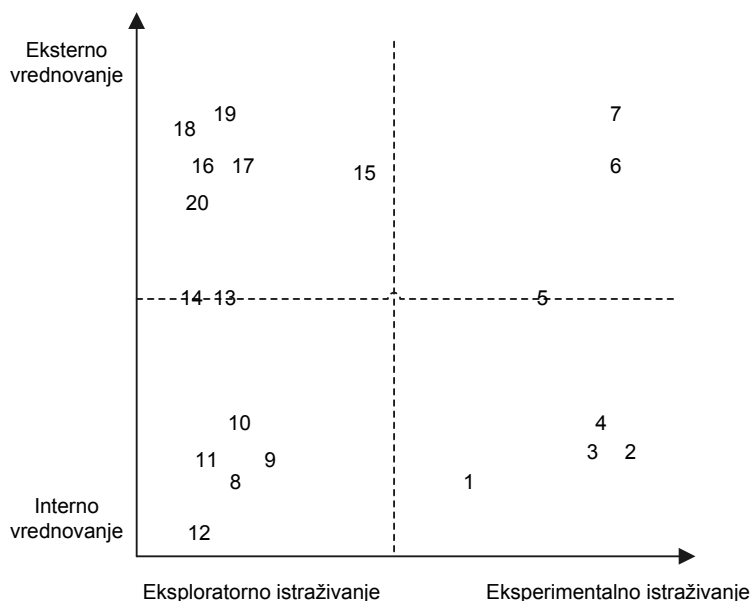
2.3. Klasifikacija metoda vrednovanja

Zbog raznolikosti metoda vrednovanja, veoma je teško odlučiti koja od metoda je prikladna u određenom kontekstu. Kod razmatranja zahtijeva vrednovanja pojavljuju se dva bitna pitanja s kojima se susreće osoba koja želi koristiti inteligentni tutorski sustav i nakon toga ga vrednovati [IQBA1999]:

1. Što se vrednuje – cijeli sustav ili jedan njegov dio?
2. Ima li dovoljno učenika i jesu li ispunjeni ostali uvjeti da bi se provelo vrednovanje zasnovano na eksperimentu?

Ova dva pitanja se mogu iskoristiti za klasifikaciju različitih metoda, tako da se jedna metoda može razlikovati od drugih prema stupnju prikladnosti za eksterno vrednovanje (odnosi se na cijeli sustav), odnosno interno vrednovanje (odnosi se na dio sustava). Nadalje, metoda se može klasificirati prema stupnju prikladnosti za eksploratorno istraživanje, odnosno eksperimentalno istraživanje. Metode eksploratornog istraživanja bi se trebale primjenjivati na malim uzorcima učenika i tamo gdje je područje nedovoljno razumljivo. Metode eksperimentalnog istraživanja se koriste manipulacijom varijabli i zahtijevaju statistički značajne grupe.

Na Slika 2.1 je prikazana klasifikacija metoda vrednovanja. Metode koje se nalaze na donjoj lijevoj strani grafa prikladne su za vrednovanje komponenti sustava i ne zahtijevaju velike uzorke učenika ili rigorozne statističke analize. Metode koje zahtijevaju rad s kontrolnom grupom i statističke procedure nalaze se na gornjoj desnoj strani grafa. Gornji dio grafa prikazuje metode koje vrednuju cjelokupne performanse sustava. Metode na desnoj strani zahtijevaju statističku analizu, dok metode na lijevoj strani ne zahtijevaju.



1. Dokaz ispravnosti 2. Usporedbe aditivnog eksperimentalnog dizajna 3. Dijagnostička točnost 4. Povratne informacija/kvalitete poučavanja 5. Analiza osjetljivosti 6. Eksperimentalno istraživanje 7. Vrednovanje proizvoda 8. Pregled stručnjaka 9. Nivo dogovora 10. Čarobnjak iz Oza 11. Metrike performansi 12. Unutrašnje vrednovanje 13. Vrednovanje zasnovano na kriterijima 14. Pilot testiranje 15. Certifikacija 16. Vanjsko vrednovanje 17. Postojanje dokaza 18. Promatranje i kvalitativna klasifikacija 19. Strukturne zadaće i kvantitativna klasifikacija 20. Studije usporedbe

Slika 2.1 Klasifikacijski graf metoda vrednovanja [IQBA1999]

Metoda vrednovanja ovisi o svrsi istraživanja. Prvo treba usporediti koliko je učinkovit sustav u odnosu na tradicionalni način poučavanja i drugo treba odrediti značajke sustava koje su važne za vrednovanje. Budući da je istraživanje usmjereno na vrednovanje cjelokupnog sustava, a ne samo jednog njegovog dijela, radi se o eksternom vrednovanju. Usporedbe tehnika poučavanja zahtijevaju testiranje hipoteze da bi se dobio odgovor na pitanje "Koji je način poučavanja učinkovitiji?" Ovo dovodi do zaključka da bi se trebao koristiti eksperiment.

2.4. Eksperiment kao metoda vrednovanja

Često se kao karakteristike eksperimentalnog istraživanja navode namjerno izazivanje pojave koja se želi ispitivati, te ponavljanje pojave i kontrola uvjeta u kojima se ispitivana pojava javlja [ROTN1978]. Glavna obilježja eksperimentalnog istraživanja su sistematsko i namjerno mijenjanje uvjeta u kojima se neka pojava javlja. Uvjet koji se namjerno mijenja u eksperimentalnom istraživanju naziva se *eksperimentalnom ili nezavisnom varijablom*, a pojava na koju se djeluje naziva se *zavisnom varijablom*. Zahvaljujući upravo ovoj karakteristici, eksperimentalnim istraživanjem moguće je najpouzdanije utvrditi zakonitosti javljanja neke pojave.

Eksperimentalno istraživanje je najsigurnija metoda istraživanja jer pomoću nje možemo najpouzdanije otkriti uzroke javljanja određene pojave [ROTN1978]. Ono je najpogodniji znanstveno-istraživački postupak za otkrivanje uzročno-posljedičnih veza među pojavama.

2.4.1. Pogreške u eksperimentu

Prilikom provođenja eksperimentalnih istraživanja potrebno je neutralizirati ili kontrolirati utjecaj svih nezavisnih varijabli osim jedne, eksperimentalne, kako bi se mogao izdvojiti njen utjecaj kao uzrok. Ako se to ne uspije postići, djelovanje tih ostalih (parazitarskih) varijabli može dovesti do pogreške u rezultatu eksperimenta. Te se **pogreške u eksperimentu** svrstavaju u tri skupine:

Pogreške tipa S – proizlaze iz djelovanja razlika u osobnostima subjekata eksperimenta npr. učenika, učitelja i sl. To na primjer može biti predznanje, mentalne sposobnosti, interes, stupanj motiviranosti, spol, dob.

Pogreške tipa G – proizlaze iz djelovanja onih faktora koji su zajednički za svaku eksperimentalnu grupu kao jedinstvenu cjelinu. One se odnose na:

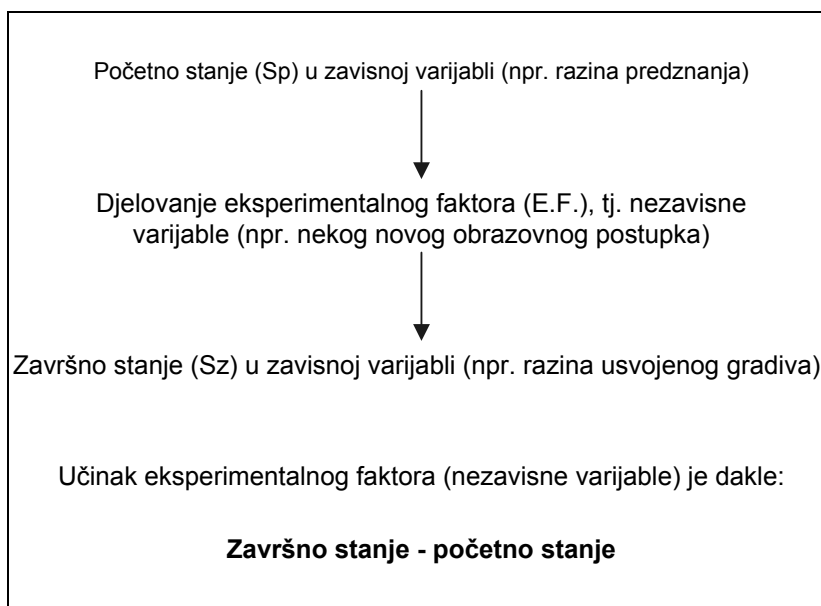
- ✓ Razlike u vezi učitelja:
 - metodičko umijeće
 - dob
 - spol
 - odnos prema učeniku
 - mišljenje učitelja o vrijednosti eksperimentalnog faktora koji se unosi
- ✓ Razlike u vezi s vremenom:
 - doba dana kada se vrši nastava
 - redni broj sata
 - dan u tjednu
- ✓ Razlike u vezi s prostorijama:
 - prozračnost
 - osvjetljenje
 - veličina učionice u odnosu na broj učenika

Pogreške tipa R – proizlaze iz ponavljanja u vremenskom smislu i u smislu mjesta. Odnosno, ako se eksperiment vrši na raznim mjestima.

Mužić u [MUŽI1999] navodi da su ovo pogreške koje se javljaju prilikom svakog eksperimenta. Da bi se te pogreške smanjile prati se veći broj nezavisnih varijabli kao i njihovo međudjelovanje, a isto tako i praćenje većeg broja zavisnih varijabli. Potrebni su dovoljno valjani i pouzdani postupci mjerenja tih varijabli kao i odgovarajući eksperimentalni uvjeti.

2.4.2. Vrste eksperimentalnih istraživanja

Budući da je glavni cilj eksperimenta ustanoviti djelovanje neke nezavisne varijable na zavisnu varijablu, logično je da se pritom mora uzeti u obzir početno stanje, tj. ono koje postoji prije samog djelovanja nezavisne varijable. To možemo iskazati pomoću sheme koja je prikazana na Slika 2.2.



Slika 2.2 Opća shema eksperimenta

Ta se shema obično iskazuje u kraticama prikazanim u Tablica 2.1:

Tablica 2.1 Kratice za eksperimentalnu shemu

S_p	S_i	početno stanje
E.F.	F	eksperimentalni faktor
S_z	S_f	završno stanje
$S_z - S_p$	$S_f - S_i$	učinak

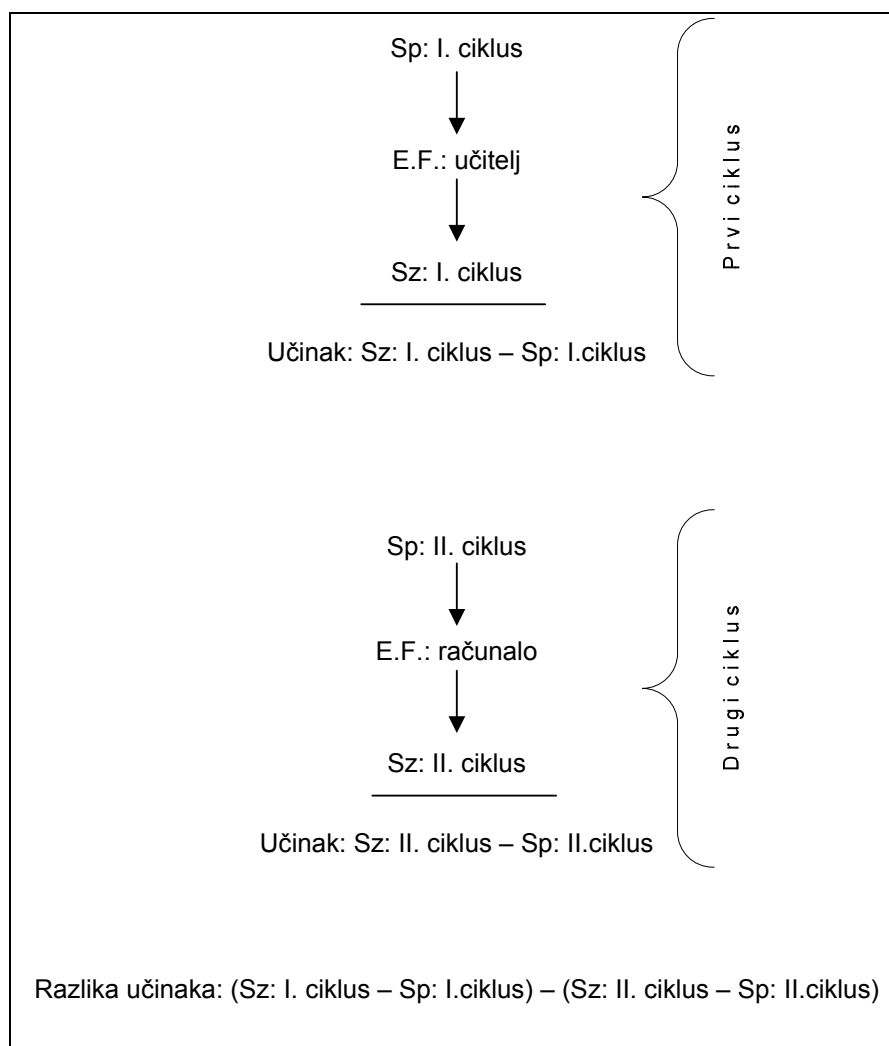
Obično se vrste eksperimentalnog istraživanja dijele u tri skupine [MUŽI1999]:

1. *Eksperiment s jednom skupinom* – jedan ili više faktora se uvode uvijek u istu skupinu ispitanika
2. *Eksperiment s paralelnim skupinama* – postoje dvije ili više skupina od kojih je svaka nosioc svog eksperimentalnog faktora
3. *Eksperimenti s rotacijom eksperimentalnih faktora* – faktori se rotiraju među skupinama, tj. svaki postupak kojeg ispituje ulazi kao eksperimentalni faktor sad u jednu, sad u drugu skupinu

EKSPERIMENT S JEDNOM SKUPINOM

U eksperimentu s jednom skupinom u istoj se skupini, npr. u istom razredu postepeno unose eksperimentalni čimbenici čiji se učinak uspoređuje. Valjanost takvog eksperimenta ovisi o tome jesu li nastavni sadržaji koji se poučavaju u prvom i drugom ciklusu međusobno povezani ili ne. Naime, bilo

bi posve neprihvatljivo da se u prvom ciklusu (unošenje prvog faktora) uči zbrajanje prirodnih brojeva, a u drugom (unošenje drugog faktora) bude oduzimanje prirodnih brojeva, jer usvajanje oduzimanja ovisi o znanju zbrajanja. Međutim, ako su sadržaji međusobno nepovezani, npr. zbrajanje prirodnih brojeva i geometrijski likovi, takav eksperiment može imati nekog opravdanja. Taj eksperiment prikazuje Slika 2.3

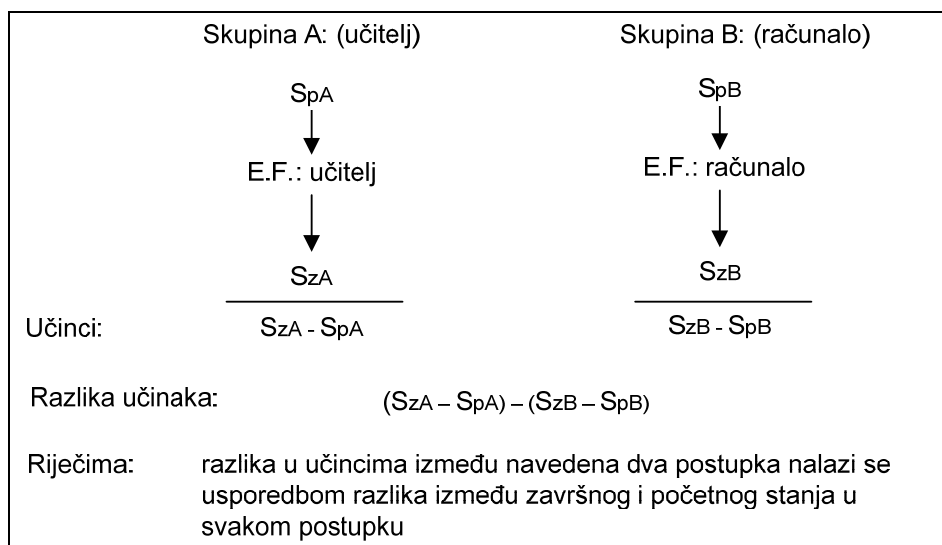


Slika 2.3 Shema eksperimenta s jednom skupinom

EKSPERIMENT S PARALELNIM SKUPINAMA

Kod eksperimenta s paralelnim skupinama formiramo dvije ili više grupa, pa kod jedne ne mijenjamo određeni uvjet, a kod drugih grupa mijenjamo. Rijetko se kod takvih eksperimenata uspoređuju rezultati više od triju grupa. Skupinu kod koje se mijenja određeni uvjet nazivamo eksperimentalna skupina, a ona grupa kod koje se radi na uobičajeni način naziva se kontrolna skupina.

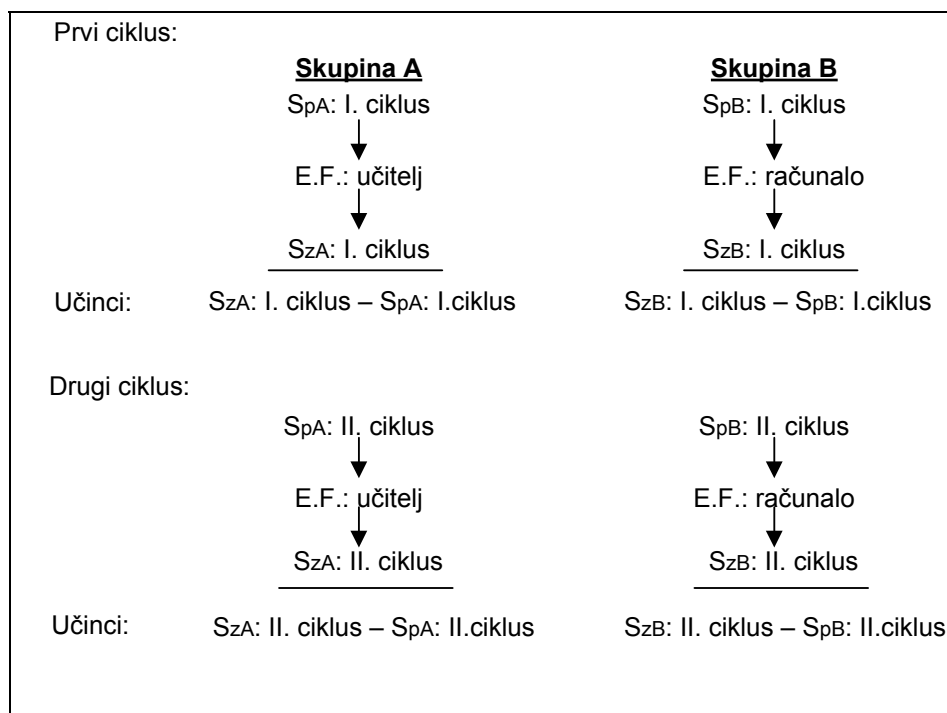
Na Slika 2.4 nalazi se primjer eksperimenta s paralelnim skupinama u kojim se uspoređuje učinkovitost usvajanja nekog radnog umijeća u klasičnoj nastavi i njezinog usvajanja uz pomoć računala kao stroja koji interaktivno instruiru učenika.



Slika 2.4 Shema eksperimenta s paralelnim skupinama

EKSPERIMENT S ROTACIJOM EKSPERIMENTALNIH FAKTORA

Eksperiment s rotacijom faktora predstavlja kombinaciju eksperimenta s paralelnim skupinama i eksperimenta s jednom skupinom, tj. u skupinama se izmjenjuju eksperimentalni faktori koji se uspoređuju, što je prikazano na Slika 2.5. Ako se u oba ciklusa, time i u obje skupine i pri usvajanju obaju sadržaja, ustanovi prednost istog faktora, tada ima više razloga za zaključak o većoj učinkovitosti tog faktora.



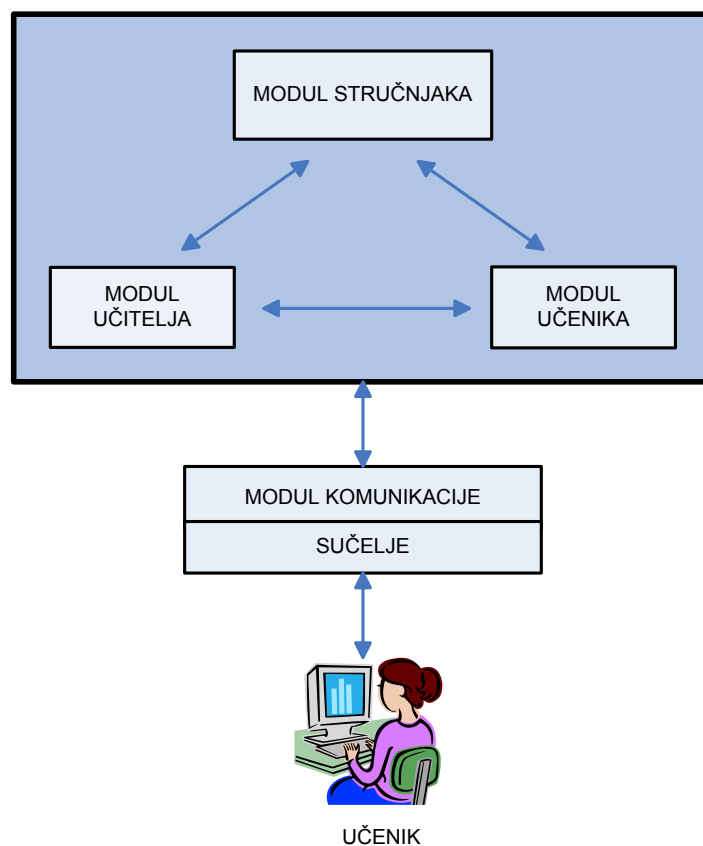
Slika 2.5 Shema eksperimenta s rotacijom eksperimentalnih faktora

2.5. Vrednovanje ovisno o arhitekturi

Budući da su ITS-ovi složeni sustavi, razne komponente ITS-a mogu zahtijevati različite pristupe vrednovanja. Važno je razmotriti prikladnost primjene tehnika vrednovanja na komponente ITS-a. Opća arhitektura o kojoj je ovdje riječ sastoji se od 4 modula ITS-a [BURN1988].

1. *Modul stručnjaka* se bavi pohranjivanjem i vladanjem znanjem nekog određenog područja.
2. *Modul učitelja* jedinica je za vođenje procesa stjecanja znanja i vještina učenika
3. *Modul učenika* (znanje učenika) opisuje razumijevanje stavova o učenikovu znanju i potrebama kroz dijagnoze, kao što su razumijevanje učenikova ponašanja i kroz stalni uvid u znanje tog učenika
4. *Modul komunikacije* predstavlja informacije učeniku i prikuplja " povratnu informaciju" od učenika

Slika 2.6 prikazuje tipičnu arhitekturu inteligentnih tutorskih sustava [BURN1988].



Slika 2.6 Moduli ITS-a

1. Modul stručnjaka

U inteligentnim tutorskim sustavima, informacija o nekoj temi ili području može se učeniku predstaviti kroz tekstualne opise, primjere, probleme koje treba riješiti, pitanja na koja treba odgovoriti, aktivne i interaktivne grafičke prikaze i povratnu informaciju o učenikovu znanju. Modul stručnjaka ITS-a sadržava znanje o određenom području koje će se podučavati. ITS koristi svoju domenu znanja za razmišljanje i rješavanje problema ili pitanja koje su postavljena ili su postavljena od strane učenika. Za ovu namjeru, znanje mora biti predstavljeno na način da podupire zaključivanje koje obuhvaća proces rješavanja problema unutar modula stručnjaka [MARK1993]. Mogu se zahtijevati različita predstavljanja znanja unutar iste domene da bi se poduprla primjena alternativnih strategija poučavanja.

Može se tražiti dodatno područno znanje da bi se riješile učenikove pogreške. Model domene možda će morati pružiti znanje da bi se ispravila uobičajena pogreška shvaćanja ili informacije koje nedostaju, a koje učenik može pokazati tijekom interakcije podučavanja.

Ekspertiza uključuje sve procese potrebne da bi se napravio sadržaj određenog područnog znanja. Takav proces može uključivati stjecanje materijala podučavanja sljedeći zahtjeve sustava za onim materijalima koji će se koristiti prilikom obrade sljedeće teme ili za ispravljanjem učenikove pogreške [SIEM1998]. Također, sustav može zahtijevati drugačiju prezentaciju ili pogled istog materijala podučavanja tražeći od procesa područnog znanja da upravlja ili popravlja područno znanje.

Ovim modulom se kontrolira predstavljanje i vladanje određenim područnim znanjem. Idealno, točnost ovog modula bi trebala biti osigurana prije završetka tog sustava [MARK1993]. Formativno vrednovanje je od velike važnosti za one koji su razvili inteligentne tutorske sustave. Sumativne tvrdnje su interesantne jedino ako je ta domena netočna.

Moguće metode za formativno vrednovanje modula stručnjaka, tj. područnog znanja su *stručna kontrola* i *Turing test*. Briga o tome jesu li primjenjivi na cjelokupno vrednovanje sustava također se mogu primijeniti na područno znanje, ovisno o domeni, o načinu na koji je predstavljena i svrsi za koju područno znanje služi. Lakše je provjeriti činjenice i mehanizme u manje kompleksnim i manje razumljivim domenama. Slično tome, ako se osnova područno znanje može procijeniti, mora se također utvrditi je li se standardi onoga što treba biti naučeno podudaraju s tumačenjem ili ne. Ako domena nije dovoljno razumljiva ili je teško sastaviti ispitna pitanja ili druge načine utvrđivanja točnosti domene ili ako je teško kontrolirati tumačenje znanja, onda stručna kontrola možda i ne bi bila poželjna.

2. Modul učitelja

Modul učitelja oblikuje i vodi sat prema instrukcijskim metodama. Cilj onih koji su razvili inteligentni tutorski sustav jest da sagrađe sustav koji će stvoriti individualizirane upute prilagođene učenikovim karakteristikama kao što su predznanje, stupanj kognitivnog razvoja i način učenja (Ackerman, Sternberg i Glaser, prema [MARK1993], da bi se što uspješnije promoviralo njihovo učenje.

ITS bi trebao predstaviti različite osobine podučavanja. Svaka lekcija predstavlja određeni cilj učitelja. Modul učitelja mora imati kontrolu odabira i slijeda materijala koji će se obrađivati, mora biti u mogućnosti da odgovori na učenikovo pitanje o određenoj temi i mora primijeniti načine kako odrediti kada je učeniku potrebna pomoć i kakva pomoć mu je potrebna. Zato ovaj modul mora uključiti različite strategije podučavanja (O'Neil, D.A. Slawson i E.L. Baker, prema [SIEM1998]).

Metode poučavanja koriste se da bi se predstavio materijal i ovisi o temi poučavanja, te ciljevima ITS-a. Strategija podučavanja određuje prirodu materijala koji će se koristiti da bi se učenika vodilo kroz nastavni proces i da bi se učeniku, ako je potrebno, pružila pomoć ukoliko se pojavi problem [SIEM1998]. Prema tome, u različitim situacijama, mogu biti korištene različite metode. Široka paleta metoda nudi veću fleksibilnost i nastavni proces se može bolje prilagoditi učeniku. Didaktika je zadužena za odabir sljedećeg obrazovnog cilja kojeg je potrebno zadovoljiti, te određivanje prikladne metoda za postizanje ovog cilja. Metoda se može odabrati prema posebnosti situacije kao što su učenikove potrebe i naklonjenosti, njegovo iskustvo i domena razgovora.

Cjelokupna kontrola sustava koordinira znanje o ovim modulima znanja da bi se uspostavilo poučavanje usmjereno ka učeniku, tj. da bi se prilagodio nastavni proces učenikovim potrebama i interesima [SIEM1998]. Postoje mnogobrojne prilagodbe zadataka koje zahtijevaju usklađivanje cjelokupne kontrole sustava. ITS, na primjer, mora primjenjivati prikladne metode podučavanja za pojedini predmet i cjelinu ako je to potrebno odabirući oblik koji je najkorisniji učeniku za određenu situaciju. Isto tako, sustav mora računati i na pogreške učenika do kojih može doći tijekom podučavanja. Svaka uočena pogreška zahtijeva proces ispravke koji je prilagođen individualnim potrebama učenika. Odabir prikladne metode podučavanja i njezina prezentacija mogu zahtijevati znanje o učenikovim potrebama, željama pohranjenima u modelu učenika.

Sljedeće pitanje je pitanje o proaktivnom i reaktivnom podučavanju. ITS može omogućiti da se prilagodi učeniku tako što se pruži obostrana pomoć koja može biti zatražena ili od strane sustava ili od strane učenika. Napredni učenik može prepoznati da mu je potrebna pomoć i može odlučiti hoće li aktivirati sustav pomoći. Ipak, kada je učenik nov ili kada je domena znanja šira, pomoć zatražena od strane sustava čini se prikladnijom. Učeniku možda bude potrebna intervencija kada napravi nesvjesnu pogrešku ili kada učenik ne zna kako nastaviti.

Standardi prema kojima se poučavanje može usporediti su teorija obrazovanja i stručni učitelji [MARK1993]. Poučavanje nije nužno dobro razumljivo ili opisano što ga čini težim za vrednovanje. Formativne tehnike vrednovanja jedva se primjenjuju u vrednovanju ITS podučavanja. Ako se razvijaju, standardi procjenjivanja značenja ITS podučavanja, oni mogu utjecati na određene odluke kao što su niz obrazovnih metoda, koje program nudi, stupanj do kojeg se program može prilagoditi učeniku i stupanj obrazovanja temeljen na obrazovnim i psihološkim istraživanjima u podučavanju. Kako se takav kriterij može procijeniti, formativno ili sumativno, još je nepoznato.

Analiza osjetljivosti i certifikacija mogu biti jedni od postupaka za istraživanje. *Turing testovi i eksperimentalne tehnike* su teško primjenjivi u komponenti izoliranog podučavanja. Ponašanje komponente učitelja ovisi o učeniku i domeni znanja [MARK1993]. Može biti teško predočiti stvarno ponašanje modula učitelja bez nepodudaranja ostalih modula. Moguće je eksperimentalno usporediti ITS-ove komponente podučavanja dok se ostale komponente drže istovjetnima (Mark i Greer, prema [MARK1993]).

Kad bi ITS mogao promatrati svoje radnje kao što bi to učinio učitelj koristeći informacije o svojim radnjama i ponašanju učenika da bi mjerio svoje poučavanje koji bi se značaj mogao pridodati vrednovanju ITS-a. Vrednovanje komponenti koje su direktno ili indirektno "pogođene" sistemskim učenjem ne mogu biti uzete u obzir kao pokazatelji mogućnosti sustava ni u koje drugo vrijeme nego u ono kada su dobivene. Bilo bi nužno proučiti promjene ponašanja sustava kroz vrijeme da bi se odredilo je li komponenta učenja zbilja poboljšala sustav ili ne, ne samo na temelju kriterija koje je ova komponenta koristila kao bazu učenja nego i na temelju drugih kriterija relevantnih za ostale komponente sustava.

3. Modul učenika

Da bi se provelo pametno poučavanje, učitelj mora imati razumijevanja prema učeniku kojeg uči. Zato ITS koristi ovaj modul da bi pokazao učenikovo znanje i vještine o određenoj temi. ITS koristi ovaj modul da bi analizirali učenikovo uključivanje u nastavni proces [SIEM1998]. Učenikov doprinos može biti odgovaranje na pitanja koja postavi ITS. Prema tome, moderniji modul učenika može sadržavati detaljnije informacije o učeniku. Ova informacija ne mora se odnositi samo na znanje koje je učenik usvojio tijekom nastavnog procesa. Informacija o učeniku, kao što je zanimanje za određeni predmet, prošlo iskustvo u stjecanju znanja i njegov napredak mogu biti važni za prilagodbu nastavnog procesa učeniku. Nadalje, modul učenika može zabilježiti sve pogreške koje neki učenik napravi. Znanje učenika pokazuje relativne jače i slabije strane učenika [SIEM1998]. Da bi se došlo do ove informacije, tj. da bi se održao model studenta, proces znanja studenta analizira ponašanje učenika.

ITS-ov modul učenika može izvesti dijagnozu i oblikovanje (Ohlsson, prema [MARK1993]). Dijagnoza proučava ponašanje učenika da bi iz njega dobila značajnu informaciju. Oblikovanje povezuje prošle i sadašnje informacije o učeniku na značajne načine, te zadržava te informacije za korištenje od strane sustava. Standard prema kojem se uspoređuje ITS-ovo znanje učenika je stvarni učenik čije znanje može biti promjenjivo, kontradiktorno i teško za promatrati.

Modul učenika je analogan obrazovnom i psihološkom instrumentu testiranja i na taj način pokušava odrediti karakteristike učenika [MARK1993]. Uobičajeni instrumenti testiranja se prosuđuju na temelju četiri osnovne karakteristike: valjanost (eng. *validity*), pouzdanost (eng. *reliability*), objektivnost (eng. *objectivity*) i standardizacija (eng. *standardization*) (Mayer, prema [MARK1993]). Test je *valjan* ako dokaz pokaže da mjeri ono što je svrha mjerenja. *Pouzdan* je ako su rezultati određenog područja povezani. *Objektivan* je ako je sproveden i ocijenjen na isti način za svakog učenika. *Standardiziran* je ako se rezultati testa mogu prevesti u značajan opis učenikovog postignuća. Normativna referenca opisuje učenikovo postignuće s obzirom na postignuća ostalih učenika. Domena reference opisuje učenikovo postignuće u smislu točno određenih standarda postignuća za određeno područje.

Dok su ciljevi modula učenika slični ciljevima instrumenta testiranja, postoje važne razlike u njihovom korištenju i ciljevima koji ograničavaju domet do kojih se standardi za vrednovanje primjenjuju komponenti znanja učenika [MARK1993]. Osobine instrumenata testiranja koje su najvažnije za vrednovanje komponente dijagnostičiranja su valjanost i pouzdanost. Metoda je valjana ako mjeri ono za što je najavljena, a pouzdana je ako daje konzistentne rezultate uspoređenih vrednovanja učenika. Objektivnost i standardizacija su manje primjenjive. Objektivnost se ne slaže s ITS-ovim ključnim ciljevima predstavljanja učenja "jedan na jedan". Također, budući da je briga ITS-a određeni učenik, normativna referenca je teže primjenjivo na ITS [MARK1993]. Neki oblik domene reference je često obuhvaćen u građi i funkciji djelovanja komponente dijagnostike, budući da se ona bavi interpretiranjem ponašanja učenika u načinima korištenja domene znanja. Ipak, ovo ne znači da će komponenta dijagnostike proizvesti analizu učenika koja će se podudarati sa standardima referentne domene. Tipovi informacija relevantni za individualno poučavanje mogu se razlikovati od tipova informacija relevantnih za domenu reference [MARK1993].

Modul učenika je valjan u onoj mjeri koja točno odražava učenika kroz određeno vrijeme. U ovome se modul učenika razlikuje od većine instrumenata testiranja koji se ne bave promjenama kroz određeno vrijeme osim u smislu kada se one mogu koristiti kao diskretna mjerenja učenika u intervalima. Budući da je model učenika baziran na ideji promatranja promjena u ponašanju i razumijevanja učenika, kriterij pouzdanosti nije primjenjiv na ovu metodu. Objektivnost nije prikladna za vrednovanje ove metode zbog istih razloga zbog kojih je neprikladna za vrednovanje

dijagnostike. Budući da je važno pitanje ove komponente pojedinačno poučavanje, standardizacija u smislu primjenjivanja skupa standarda za analizu uspješnosti, također nije primjenjiva na oblikovanje učenika [MARK1993].

Valjanost i pouzdanost su najuspješnije proučavane kroz *eksperimentalne studije* [MARK1993]. Prilikom određivanja dijagnostičke valjanosti, dijagnostička informacija dobivena od sustava se uspoređuje s onom dobivenom neovisno o njemu. Slično tome, model učenika može biti proglašen valjanim ako se vili je li se ponavljana mjerenja podudaraju sa pojedinačnim mjerenjima. Da bi se proučila pouzdanost, informacije se mogu proučiti da bi se odredilo je li sustav daje pouzdane rezultate.

4. Komunikacijski modul

Bez obzira na kvalitetu dizajna računalnog programa i projektantskog znanja ITS će biti od male pomoći ako učenici ne mogu razumjeti informacije koje predstavlja, ili ako krivo protumače upute ili naprave greške zbog njima nepoznatog sučelja [MARK1993]. Obilježja sučelja koja su trenutno najčešće prezentirana u obrazovnim računalnim sustavima su grafički prikazi i pripremljen test, dok su najuobičajeniji načini unošenja informacija grafički prikaz i izbornici koji se otvaraju klikom miša i unošene određene notacije putem tipkovnice. Upute i osvrti mogu biti od koristi prilikom ranog razvoja sustava. Ipak, *pilot testiranje* kod članova populacije za koje je sustav dizajniran će prije identificirati probleme za formativno vrednovanje sučelja (England, prema [MARK1993]). Drugi pristup prikladan za formativno ili sumativno vrednovanje sučelja jest da se *eksperimentalno uspoređivanje verzija sučelja* da bi se vidjelo koja je uspješnija.

3. Vrednovanje obrazovnog učinka

Budući da je glavni cilj obrazovnog sustava podučavati, najveći test je jesu li učenici iz njega uspješno naučili. Dakle potrebno je procijeniti veličinu obrazovnog učinka. Kriterij koji se koristi za procjenu obrazovnog učinaka ITS-a uključuju postignuće i loše utjecaje. Postignuće i mjerenja vezana za njih bave se primanjem, razumijevanjem, prezentiranjem, pamćenjem i prijenosom učeničkih znanja i vještina. Mjere afekta se bave stavovima i osjećajima koji imaju utjecaja na učenikovo korištenje i učenje od ITS-a. Obrazovni učinci se ponekad ispituju kroz formativna vrednovanja ITS-a i od veće su važnosti u sumativnom vrednovanju kad se zahtijeva dokaz za formalne tvrdnje je li ITS postiže svoje ciljeve poučavanjem ili ne. Formativno vrednovanje prikuplja kvalitativne ili kvantitativne procjene obrazovnih učinaka kroz pilot testiranje dok su sumativni zahtjevi bazirani na eksperimentalnom testiranju.

3.1. Veličina učinka

Veličina učinka (eng. *effect size*) govori o relativnoj jačini eksperimentalnog faktora i omogućava prijelaz s jednostavnog pitanja istraživanja „Je li nešto dobro ili ne?“ na složenije pitanje „Koliko je nešto dobro?“. Veličina učinka je standardni način za usporedbu rezultata dva eksperimenta. Ona određuje veličinu razlike između dviju grupa i zbog toga se smatra stvarnom mjerom značajnosti te razlike. Može biti pozitivna ili negativna. Pozitivna je kada je eksperimentalna grupa bolja, a negativna je kada je kontrolna grupa bolja.

Veličina učinka se može izračunati pomoću različitih formula i pristupa i njena vrijednost se može zbog toga uvelike razlikovati. Za interpretaciju veličine učinka mnogi statističari koriste Cohen-ov način interpretacije [COHE1969]. Cohen smatra da su veličine učinka koje su manje od 0.1 beznačajne, one između 0.1 i 0.3 male, veličine učinka između 0.3 i 0.5 umjerene, a one veličine učinka veće od 0.5 smatra velikima. U Tablica 3.1 se nalazi interpretacija veličine učinka prema Cohen-u.

Glavni način vrednovanja učinka inteligentnih tutorskih sustava je ispitivanje uče li učenici učinkovito pomoću tog sustava (prema [MARK1993]). Kriteriji koji se koriste za vrednovanje utjecaja inteligentnih tutorskih sustava na obrazovanje uključuju: postignuće (eng. *achievement*) i afekt (eng. *affect*). Postignuće se odnosi na stjecanje, razumijevanje, zadržavanje i prijenos učenikovog znanja i vještina (Haertel i Calfee, prema [MARK1993]). Mjere afekta se odnose na stavove i emocije koje mogu utjecati na način kako učenici koriste inteligentne tutorske sustave i uče od njih (Malone, prema [MARK1993]). Za postignuće su također važne i mjere vezane za rezultate.

Tablica 3.1 Interpretacija veličine učinka prema Cohen-u

Cohen's Standard	Veličina učinka	Postotak osoba iz kontrolne grupe čiji je uspjeh ispod uspjeha prosječne osobe iz eksperimentalne grupe	Postotak nepreklapanja distribucija grupa
	2.0	97.7	81.1%
	1.9	97.1	79.4%
	1.8	96.4	77.4%
	1.7	95.5	75.4%
	1.6	94.5	73.1%
	1.5	93.3	70.7%
	1.4	91.9	68.1%
	1.3	90	65.3%
	1.2	88	62.2%
	1.1	86	58.9%
	1.0	84	55.4%
	0.9	82	51.6%
LARGE	0.8	79	47.4%
	0.7	76	43.0%
	0.6	73	38.2%
MEDIUM	0.5	69	33.0%
	0.4	66	27.4%
	0.3	62	21.3%
SMALL	0.2	58	14.7%
	0.1	54	7.7%
	0.0	50	0%

3.2. Mjere postignuća

Vrednovanje obrazovnih postignuća, kao utjecaja na ITS, uključuje određivanje kako dobro sustav podučava bitna znanja i vještine. To uključuje i prikupljanje i razumijevanje te predstavljanje učenikovog znanja [SIEM1998].

Znanje se može kategorizirati u tipove, na osnovu stupnja svjesnosti o znanju i namjenama za koje se ono koristi. Znanje se može podijeliti na kategorije: referencijalno (eng. *referential*), činjenično (eng. *factual*) i proceduralno (eng. *procedural*) [MARK1993].

Poznavanje simbola i značenja može se nazvati *referencijalno* znanje, verbalna informacija ili semantičko znanje. *Činjenično* znanje je određeno znanje o objektima i odnosima između objekata u svijetu. Mayer činjenično znanje uključuje u kategoriju semantičkog znanja, dok ga McGraw i Harbison-Briggs klasificiraju kao dio veće kategorije deklarativnog znanja koje se lako da opisati (prema [MARK1993]).

Proceduralno znanje je znanje o tome kako nešto napraviti. Može se podijeliti na eksplicitno i implicitno znanje. Eksplicitno proceduralno znanje se može opisati na osnovu algoritama, pravila ili procedura i može se koristiti kao vodič (Mayer, prema [MARK1993]). Ovakav sistem se još naziva intelektualna vještina i deklarativno znanje. Implicitno proceduralno znanje je znanje o tome kako nešto napraviti što nije moguće verbalno opisati. Mayer ovo naziva vještinama (prema [MARK1993]). McGraw i Harbison-Briggs (prema, [MARK1993]) koriste termin proceduralno znanje jedino kad se misli samo na implicitno proceduralno znanje.

Ciljevi učenja također se mogu uzeti u obzir prilikom procjenjivanja obrazovnih učinka. Bloom je razradio klasifikaciju kognitivnih obrazovnih ciljeva, opće kategorije rezultata učenja. On određuje 6

glavnih tipova ciljeva kojih on naziva *znanjem, razumijevanjem, primjenom, analizom, sintezom i vrednovanjem* (Bloom, prema [MARK1993]). Na Bloomovom *stupnju znanja* učenik se može sjetiti nekog izraza, činjenice ili postupka, ali ga ne može primijeniti ili razumjeti. *Razumjevanje* (eng. *comprehension*) podrazumijeva da se učenik može koristiti materijalima do nekih granica, dovoljno da da definiciju ili izvuče zaključak. *Stupnjem primjene* (eng. *application*) učenik započinje koristiti znanje u konkretnim situacijama. *Analiza* (eng. *analysis*) nalaže da učenik ne samo da može odrediti ideje, već ih također može proučavati, te razglabati o njihovim povezanostima. *Sinteza* (eng. *synthesis*) nalaže da učenik organizira predočene materijale s ciljem stvaranja novih ideja. *Vrednovanje* (eng. *evaluation*) uključuje mogućnost prosudbe vrijednosti znanja.

Mark koristi Mayer-ove i Bloom-ove kategorije tipova znanja kao radni okvir za određivanje mjera za procjenu postignuća. Referencijalno, činjenično, proceduralno i metakognitivno znanje su termini koji označavaju kategorije znanja. Prisjećanje, razumijevanje, primjena, analiza, sinteza i procjena su termini koji određuju rezultate znanja ili stupnjeve razumijevanja koje ljudi pokazuju. Ove kategorije znanja zajedno tvore taksonomiju tipova ponašanja koji pokazuju je li učenik naučio što je podučavano ili ne. Takva kategorizacija može pomoći pri određivanju mjera postignuća da bi se procijenio uspjeh. U idealnim okolnostima, mjere postignuća trebale bi biti valjani, pouzdani i korisni ciljevi, mjerljivi rezultati [MARK1993].

Dominantan pristup da bi se utvrdilo obrazovno postignuće učenika s prijašnjim sustavima podučavanja, kao što je učenje pripomognuto računalima, određuje je li se učenik ispravno ponaša. Ipak, takvo vrednovanje koje se bazira na ispravnim i neispravnim odgovorima ne reflektira mentalne procese iz pitanja [SIEM1998]. Uspostavom i pojavom ITS-a došlo se do potrebe da se utvrde razlozi zašto učenici daju točne ili netočne odgovore tako što se određuje kako dobro sustav podučava učenike znanju i vještinama koje podupiru mentalne procese koji se traže da bi se riješili problemi u domeni podučavanja.

Model učenika može poslužiti da bi se procijenilo kako ITS podučava znanje o rješavanju problema u domeni. Littman i Soloway su prvi predložili upotrebu tehnika oblikovanja učenika da bi poduprli novi pristup vanjskog vrednovanja. Oni predlažu da uz pomoć njihovih sposobnosti da oblikuju učenika, ITS bi mogao napraviti razne probleme koje bi učenik trebao biti u mogućnosti riješiti. Ovi problemi, potom, mogu služiti u testiranju učenika. Postotak uspješnosti učenika je mjerilo uspješnosti učenja. Ispravno rješenje učenikovog problema, pokazuje da su se procesi ili znanje i vještine uspješno podučavali od strane sustava. Nadalje, rezultat takvog testa ne mora samo pokazati je li učenik sposoban riješiti određene probleme, već ujedno može dati razloge zašto je učenik bio sposoban ili zašto nije bio sposoban riješiti problem. Tehnike učenikovog modeliranja temeljeno na modelima procesa mogu poslužiti da pretpostave proces kroz kojeg učenik mora proći da bi riješio probleme dok tehniku modeliranja učenika koje nisu bazirane na modelima procesa mogu biti korištene da se odrede neka znanja i vještine koje učenik mora koristiti da bi riješio probleme.

3.3. Mjere vezane za rezultate

Ostali kriteriji vezani za postignuća koji su učitelji cijenili korisnim u procjeni utjecaja na obrazovanje uključuju transfer, pamćenje, vrijeme učenja i postotak dovršenosti. Prijenos, memoriranje i vrijeme učenja mogu se proučavati eksperimentalno, dok se dovršenost promatra tijekom terenskog pokusa.

Prijenos (eng. *transfer*) je mogućnost primjene informacija ili vještina naučenih u jednom kontekstu, u novom i nepoznatom kontekstu gdje je to znanje važno (Mayer, prema [MARK1993]). Prijenos označava apstraktno razumijevanje materijala i mogućnost određivanja činjenica relevantnih za nove situacije. Prijenos je briga učitelja kada je važno generaliziranje znanja, kao u metakognitivnom i učenju proceduralnih vještina koje se moraju primijeniti u mnogo različitih situacija. Također je važno kada učenici moraju vještinu naučenu u simuliranom okruženju primijeniti u stvarnosti. Prijenos se općenito proučava testiranjem učenika u situaciji učenja i u situaciji kada se vjeruje da je naučeni materijal prenosiv (Johnson, prema [MARK1993]).

Memoriranje (pamćenje) (eng. *relention*) je mogućnost zadržavanja naučenog kroz vrijeme (Mayer, prema [MARK1993]). Što se učenik može duže prisjećati i koristiti naučenim informacijama i vještinama, to znači da učenik ima bolju usvojenost znanja. Pamćenje se procjenjuje usporednim mjerenjima istog fenomena kroz dulje vrijeme. Na primjer dva dijela matematičkog ispita za koje se zna da su usporedivi, mogu se provesti u različito vrijeme da bi se odredilo je li se znanje s vremenom promijenilo, neovisno o podučavanju. Prilikom demonstracije učenja, učitelji mogu poželjeti procijeniti memoriranje naučenog kao dodatan dokaz koristi njihovog pristupa.

Vrijeme za učenje (eng. *time*) je vrijeme koje je potrebno da bi se nešto naučilo [MARK1993]. Kao kriterij za sustave vrednovanja, vremenske vrijednosti variraju. Usporedbe vremena potrebnog za učenje kod učenika koji koristi različite sustave, ponekad se koriste u prepirkama da je jedan sustav bolji od drugog. Ipak, vrijeme učenja se mora promatrati u smislu postignuća i motivacije. Jedna situacija gdje je vrijeme učenja prikladan kriterij za vrednovanje je učenje s provjeravanjem (eng. *mastery learning*) u kojem učenik mora pokazati predodređene stupnjeve razumijevanja teme ili vještine prije učenja novih materijala.

Brzina završavanja (eng. *completion rate*) ili *odustajanje* (eng. *drop-out*) se promatra iz terenskih pokusa kada učenik ima opciju nastavljanja ili prekidanja obrazovnog programa (Hocker i Elias, prema [MARK1993]). To može biti značajno kada se razmatraju obrazovni učinci sustava. Na primjer ITS koji je težak za uporabu ili koji podučava na neprikladnom nivou može obeshrabriti učenika i navesti ih da odustanu. Ovo može biti loš izbor za upotrebu unutar učionice čak ako oni učenici koji ga koriste i postignu dobre rezultate, jer učenička populacija kao cjelina ne nalazi to jednostavnim ili poželjnim rješenjem.

3.4. Mjere afekta učenja

Afekt procesa podučavanja bavi se stavovima i emocijama koje je prouzročio inteligentni tutorski sustav [SIEM1998]. Kada stavovi i osjećaji posreduju u učenju, važni su za učinkovitost ITS-a. Često se daje argument u korist korištenja računala u obrazovanju da su računala intrinzično motivirajuća. Ovaj se izraz koristi da pokrije razne faktore koji potiču i sprečavaju pojedinačno sudjelovanje u određenim ponašanjima (Ellis i Sabornie, prema [MARK1993]). Određuje ljudsku želju da budu aktivni i uključeni.

Učitelji se mogu baviti: *akademsom motivacijom*, zanimanjem za uključivanjem u akademske aktivnosti, *motivacijom postignuća*, zanimanjem za zadržavanjem standarda izvrsnosti ili nekog drugog

poželjnog završetka, *stavovima* prema obrazovanju, računalima ili određenim programima (Thomas, prema [MARK1993]). Motivacija u smislu učenja može se gledati kao učenikova želja da bude alternativan i uključen u proces podučavanja i zbog toga se priznaje kao važan faktor učenja [SIEM1998]. Najuočajenija metoda procjene motivacije jest da se pita učenika da ocijeni svoje slaganje s određenim stavovima, vrednovanjima i aktivnostima (Moore, prema [MARK1993]). Usporedbe vremena provedenog na zadaćama povezanim i nepovezanim aktivnostima kroz nastavni proces su drugi pokazatelji zainteresiranosti. Također i stopa "ispadanja", tj. ukupno vrijeme provedeno na sat podučavanja pokazuje stupanj zainteresiranosti učenika. Konačno vrijeme se ponekad koristi kao pokazatelj motivacije gdje je aktivnost dobrovoljna kao što je to i odustajanje i dovršavanje programa. Mjerenje motivacije pokazuje što učenik osjeća prema određenom sustavu. Motivacija učenika koji rade s određenim sustavom pokazuje hoće li sustav biti prihvaćen i korišten. Istraživači su također predložili da uporaba računala može "hraniti" samopoštovanje što može potaknuti obrazovno postignuće (Emihovich i Miller, prema [MARK1993]). Samopoštovanje se može procijeniti koristeći specijalno izrađenih mjera samopoštovanja koje su dostupne iz komercijalnih i akademskih izvora.

Međutim, dok mjerenje afekta učenja mogu pokazati kako je učenikovo mišljenje o sustavu (Jacobson i Smith, prema [MARK1993]), ne odražavaju se točno na postignuća ili kriterija vezanog za postignuće (Corbet i Anderson, prema [MARK1993]). Prilikom procjenjivanja utjecaja ITS-a ili temeljna činjenica na njima, važno je razmatrati može li biti prisutno više od jedne interpretacije dokaza o lošem utjecaju. Često se o značenju afekta učenja može raspravljati i u svakom slučaju se mora odmjeriti s drugim faktorima u vrednovanju. Mjerenja loših utjecaja mogu biti korisna pri identificiranju problema ili briga tijekom formativnog vrednovanja, ali oni nisu dovoljno pouzdani da bi osigurali da postojeći problemi budu identificirani. Mogu predložiti hoće li ITS biti prihvaćen i korišten ili ne [MARK1993]. Također, mogu biti zanimljiva kao mjerenja dugoročnih utjecaja ITS-a na stavove i obrazovanje. Konačno, mogu biti poželjna da nadopune mjerenja postignuća u sumativnim vrednovanjima. Ipak bi trebalo imati na umu da su rijetko uvjerljivi dokazi o učinkovitosti ako se razmatraju izolirano (ili u izoliranoj okolini).

4. Učenje i poučavanje uz pomoć sustava xTEx-Sys

Sustav xTEx-Sys (*eXtended Tutor-Expert System*) [STAN2005] je računalni sustav čija je svrha poboljšati proces učenja i poučavanja učenika, te povećati učinkovitost rada učitelja. Glavna razlika u odnosu na ostale načine učenja i poučavanja je da se ovom sustavu može pristupiti s bilo koje lokacije i u bilo koje vrijeme. xTEx-Sys je Web orijentirana inteligentna autorska ljuška, tj. specijalizirano okruženje prilagođeno potrebama stručnjaka za razvoj baza područnog znanja, učiteljima za izgradnju courseware-a i učenicima za učenje, poučavanje i testiranje znanja [STAN2005]. Sudionici sustava su, dakle, učenici, učitelji, stručnjaci za područno znanje, te administrator sustava. Stručnjak oblikuje znanje, tj. gradi baze područnog znanja. Znanje je prikazano uz pomoć semantičke mreže koja se sastoji od čvorova i veza. Čvorovi prikazuju neke objekte ili pojmove, a veze odnose među tim objektima, odnosno pojmovima. Učitelj gradi courseware koji predstavlja didaktički oblikovan nastavni sadržaj nekog područnog znanja za izvođenje na računalu [STAN2005]. Elementi nastavnog sadržaja su nastavne cjeline, nastavne teme, nastavne jedinice, nastavni pojmovi i elementi za testiranje i ocjenjivanje znanja učenika.

4.1. Cilj, sudionici i tijek istraživanja

Testiranje je provedeno za područno znanje iz matematike. Cilj testiranja je bilo utvrđivanje učinka korištenja sustava xTEx-Sys, te postignuća učenika u procesu učenja i poučavanja.

U eksperimentu je sudjelovalo 48 učenika petih razreda OŠ "Bol", Split koji su bili podijeljeni u dvije grupe. Zbog organizacijskih problema unaprijed su definirane kontrolna i eksperimentalna grupa. U kontrolnoj grupi su bili učenici jednog razreda, a u eksperimentalnoj učenici drugog razreda. U obje grupe je bilo po 24 učenika. Unaprijed određena pripadnost grupi nije predstavljala ograničenje provedenom eksperimentu, jer se inicijalnim testom pokazalo da nema statistički značajne razlike među grupama i bile su prihvatljiv uzorak za statističku analizu i interpretaciju učinkovitosti.

Kontrolna grupa je sudjelovala u tradicionalnom procesu učenja i poučavanja, dok je eksperimentalna grupa koristila sustav xTEx-Sys. Učenici su bili uključeni u odgovarajuće oblike učenja i poučavanja četiri sata tjedno u razdoblju od 11.05.2007.g. do 1.06.2007.g., odnosno 14 sati. Tijek istraživanja je prikazan u Tablica 4.1.

Obje grupe su pisale inicijalni test, po dva međutesta, te završni test, s tim da je eksperimentalna grupa ispunjavala anketu očekivanja i anketu mišljenja o sustavu. Inicijalni test je dao informaciju o postojanju statistički značajne razlike u predznanju učenika. Završni test je dao informaciju o postojanju statistički značajnih razlika između grupa s obzirom na različit način poučavanja.

Tablica 4.1 Tijek istraživanja

Aktivnost	Datum
Inicijalni test	11.05.2007.
učenje i poučavanje Kontrolna grupa na tradicionalni način Eksperimentalna grupa uz pomoć sustava xTeX-Sys	od 12.05.2007. do 18.05.2007.
Prvi međutest	18.05.2007.
učenje i poučavanje Kontrolna grupa na tradicionalni način Eksperimentalna grupa uz pomoć sustava xTeX-Sys	od 19.05.2007. do 25.05.2007.
Drugi međutest	25.05.2007.
učenje i poučavanje Kontrolna grupa na tradicionalni način Eksperimentalna grupa uz pomoć sustava xTeX-Sys	od 16.05.2007. do 1.06.2007.
Završni test	1.06.2007.

Na završetku eksperimenta je napravljena analiza prisustvovanja testovima u okviru eksperimenta. Rezultati su sljedeći:

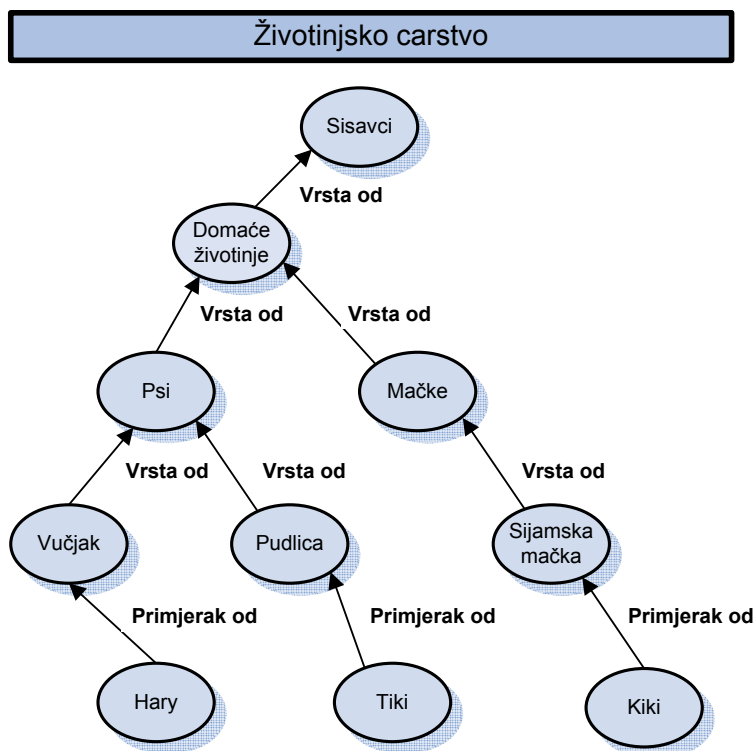
1. Inicijalni test (Prilog A) su pisali svi učenici obje grupa
2. Prvi međutest (Prilog B) su pisala 23 učenika iz kontrolne grupe i 22 učenika iz eksperimentalne grupe
3. Drugi međutest (Prilog C) su pisala 22 učenika iz kontrolne grupe i 23 učenika iz eksperimentalne grupe
4. Završni test (Prilog D) su pisala 24 učenika iz kontrolne grupe i 23 učenika iz eksperimentalne grupe

Na kraju smo dobili da 20 učenika iz kontrolne grupe ima sva četiri testa, te 21 učenika iz eksperimentalne grupe koji također imaju sva četiri testa.

4.2. Tijek istraživanja eksperimentalne grupe

U prvom tjednu učenicima je prikazana prezentacija o različitim načinima učenja i poučavanja, od klasičnih pa sve do onih najnovijih tehnologija, odnosno učenja pomoću računala. Također, istaknuta je razlika između pojmova "stjecanje znanja o računalu" i "stjecanje znanja učenjem s računalom i poučavanjem od računala".

Kako je gradivo prikazano skupom međusobno povezanih pojmova (čvorovima), učenicima je objašnjen pojam semantičke mreže pomoću Slika 4.1 koja se odnosi na područno znanje o životinjama.



Slika 4.1 Primjer semantičke mreže

Također, učenici su upoznati i sa programskim sustavom xTEx-Sys, koji će koristiti za učenje i poučavanje znanja iz geometrije u tjednima koji su slijedili. Zatim je učenicima prezentiran sadržaj iz geometrije, ali samo ono što učenici znaju iz prethodnih razreda, da bi se učenici upoznali s radom sustava. Nakon toga su učenici ispunili anketu očekivanja, čiji će rezultati biti analizirani u sljedećim poglavljima.

U drugom tjednu su učenicima dani korisničko ime i lozinka, te im je objašnjen način prijave na sustav, te pojedini odnosi među pojmovima iz odabranog nastavnog sadržaja. Poučavali su se tri školska sata iz jednog dijela nastavnog sadržaja, te im je naglašeno da se mogu prijaviti na sustav od kuće, te tako učiti. Posljednji sat toga tjedna učenici su pisali prvi međutest. U trećem tjednu učenici su se također poučavali tri sata. Učili su drugi dio nastavnog sadržaja, te posljednji, četvrti, sat toga tjedna učenici su pisali drugi međutest. Tijekom četvrtog tjedna učenici su ponavljali cjelokupni nastavni sadržaj, također u trajanju od tri sata, a četvrti sat su pisali završni test. U ovom dijelu učenici su ispunjavali i anketni upitnik o njihovom mišljenju o sustavu xTEx-Sys, te načinu učenja i poučavanja na taj način, čiji je formular dan u Prilog E i Prilog F. Rezultati ove ankete također će biti analizirani u sljedećim poglavljima, kao i rezultati testova.

4.3. Analiza rezultata istraživanja

U ovom poglavlju biti će analizirani testovi, te anketni upitnici koje su ispunjavali učenici.

4.3.1. Analiza testova

Ukupno 41 učenika je pisao sve testove vezane za poznavanje gradiva iz *Skupovi točaka u ravnini*. Od toga je 20 učenika iz kontrolne grupe, a 21 učenik iz eksperimentalne grupe. Također treba napomenuti da su u eksperimentalnoj grupi bila dva učenika koji rade po individualiziranom programu, dok u kontrolnoj grupi nije bilo takvih učenika.

U Tablica 4.2 i Tablica 4.3 se nalaze rezultati, tj. broj bodova koje su učenici postigli u testovima.

Tablica 4.2 Rezultati testova eksperimentalne grupe

	Inicijalni test Bodovi	Prvi međutest bodovi	Drugi međutest bodovi	Završni test bodovi
učenik1	100	88	100	100
učenik2	92	81	94	100
učenik18	98	88	100	100
učenik20	100	94	94	100
učenik21	92	94	100	98
učenik13	94	81	100	92
učenik7	91	63	94	88
učenik22	100	75	88	88
učenik3	64	75	88	80
učenik4	94	56	63	75
učenik8	60	75	75	75
učenik14	52	68	81	75
učenik24	100	56	68	75
učenik17	85	75	88	73
učenik12	58	63	94	64
učenik23	64	68	75	64
učenik6	61	75	63	63
učenik9	80	56	44	63
učenik5	25	50	50	50
učenik19	66	50	50	47
učenik15	55	31	25	34

Tablica 4.3 Rezultati testova kontrolne grupe

	Inicijalni test Bodovi	Prvi međutest bodovi	Drugi međutest bodovi	Završni test Bodovi
učenik42	100	94	100	100
učenik44	100	92	100	100
učenik29	98	88	94	94
učenik38	100	69	88	92
učenik25	63	81	63	88
učenik28	82	75	81	81
učenik47	86	69	75	81
učenik33	94	56	63	75
učenik37	82	69	44	75
učenik45	81	81	75	75
učenik46	75	50	63	75
učenik26	83	56	63	64
učenik30	81	63	44	64
učenik36	64	68	56	63
učenik27	56	56	56	56
učenik40	46	50	56	50
učenik48	44	44	31	47
učenik34	82	63	31	44
učenik39	66	56	31	44
učenik43	23	44	38	25

Inicijalni test je dan u obje grupe kako bi se ispitalo je li postoji statistički značajna razlika između grupa. Iako su grupe bile podjeljene po razredima, nakon analize inicijalnog testa utvrđeno je da ne postoji statistički značajna razlika među grupama, što je vidljivo i u Tablica 4.4.

Tablica 4.4 Podaci potrebni za ispitivanje statistički značajne razlike

	Kontrolna grupa	Eksperimentalna grupa
\bar{X}	75.3	77.67
$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$	21.0015	21.2399
$s_{\bar{X}_c - \bar{X}_e} = \sqrt{\frac{\sigma_c^2}{N_c} + \frac{\sigma_e^2}{N_e}}$	6.5982	
$t = \frac{\bar{X}_c - \bar{X}_e}{s_{\bar{X}_c - \bar{X}_e}}$	-0.3587	

Broj stupnjeva slobode računamo po formuli $(N_c - 1) + (N_e - 1)$, a u našem slučaju je taj broj 39. Granična vrijednost t na razini značajnosti 5% uz 39 stupnjeva slobode iznosi 1,684. Kako je apsolutna vrijednost dobivene t-vrijednosti manja od granične, to znači da su grupe statistički ekvivalentne tj. ne postoji statistički značajna razlika između grupa s obzirom na rezultate inicijalnog testa.

Za svaki provedeni test definiraju se nul-hipoteze. Nul-hipoteza nam govori da "nema razlike" među pojavama koje mjerimo. U eksperimentu koji se ovdje obrađuje definiraju se nul-hipoteze za prvi i drugi međutest, te jedna hipoteza koja se odnosi na završni test. Da bi hipoteze mogli prihvatiti ili odbaciti potrebno je izračunati razliku između rezultata testova i inicijalnog testa za kontrolnu i eksperimentalnu grupu, te t-vrijednost da bi se moglo potvrditi ili opovrgnuti odgovarajuća nul-hipoteza. Račun proveden nad razlikama u rezultatima prikazan je za svaku hipotezu posebno.

HIPOTEZA H1

H1: „Ne postoji statistički značajna razlika između kontrolne i eksperimentalne grupe kod rezultata prvog međutesta.“

Tablica 4.5 Podaci potrebni za prihvaćanje ili odbacivanje nul-hipoteze H1

	Kontrolna grupa	Eksperimentalna grupa
\bar{X}	-9.1	-8.0476
$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$	15.2035	18.6748
$s_{\bar{X}_c - \bar{X}_e} = \sqrt{\frac{\sigma_c^2}{N_c} + \frac{\sigma_e^2}{N_e}}$	5.307	
$t = \frac{\bar{X}_c - \bar{X}_e}{s_{\bar{X}_c - \bar{X}_e}}$	-0.1983	

Pošto je apsolutna vrijednost dobivene t-vrijednosti manja od granične t-vrijednosti (1,684), znači da razlika nije statistički značajna. Dakle, nul-hipoteza H1 se prihvaća. Negativni predznak t-vrijednosti govori da je eksperimentalna grupa statistički neznačajno bolja od kontrolne grupe .

HIPOTEZA H2

H2: „Ne postoji statistički značajna razlika između kontrolne i eksperimentalne grupe kod rezultata drugog međutesta“

Tablica 4.6 Podaci potrebni za prihvaćanje ili odbacivanje nul hipoteze H2

	Kontrolna grupa	Eksperimentalna grupa
\bar{X}	-12.7	0.1429
$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$	17.463	20.5312
$s_{\bar{X}_c - \bar{X}_e} = \sqrt{\frac{\sigma_c^2}{N_c} + \frac{\sigma_e^2}{N_e}}$	20.5312	
$t = \frac{\bar{X}_c - \bar{X}_e}{s_{\bar{X}_c - \bar{X}_e}}$	-2.1609	

Pošto je apsolutna vrijednost dobivene t-vrijednosti veća od granične t-vrijednosti (1,684), znači da je razlika statistički značajna, odnosno da smo opovrgnuli nul-hipotezu H2. Kako je ponovno dobivena negativna vrijednost t-testa, to pokazuje da je eksperimentalna grupa je statistički značajno bolja od kontrolne grupe.

HIPOTEZA H3

H3: „Ne postoji statistički značajna razlika između kontrolne i eksperimentalne grupe kod rezultata završnog testa“

Tablica 4.7 Podaci potrebni za prihvaćanje ili odbacivanje nul-hipoteze H3

	Kontrolna grupa	Eksperimentalna grupa
\bar{X}	-5.65	-1.2857
$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N-1}}$	12.8689	14.3567
$s_{\bar{X}_c - \bar{X}_e} = \sqrt{\frac{\sigma_c^2}{N_c} + \frac{\sigma_e^2}{N_e}}$	4.2539	
$t = \frac{\bar{X}_c - \bar{X}_e}{s_{\bar{X}_c - \bar{X}_e}}$	-1.0259	

Pošto je apsolutna vrijednost dobivene t-vrijednosti manja od granične t-vrijednosti (1,684), znači da razlika nije statistički značajna. Dakle, nul-hipoteza H3 se prihvaća. Negativni predznak t-vrijednosti govori da je eksperimentalna grupa statistički neznačajno bolja od kontrolne grupe.

Veličinu učinka izračunali smo na temelju razlike između aritmetičkih sredina kontrolne i eksperimentalne grupe, podijeljene standardnom devijacijom. Tako smo kod navedenih hipoteza dobili redom sljedeće rezultate $0,069 \sigma$, $0,7354 \sigma$ i $0,3391 \sigma$. Dobivena ukupna veličina učinka je $0,2859 \sigma$.

Analizirajući uzroke ovakvih rezultata primijetili smo nekoliko okolnosti koje su mogle utjecati na veličinu učinka. Naime, učenici eksperimentalne grupe su se, tek nekoliko sati prije prvog međutesta, po prvi put susreli i upoznali s radom inteligentnog tutorskog sustava xTeX-Sys, te su u vrlo kratkom vremenu trebali razumjeti princip učenja putem semantičkih mreža, zbog čega je veličina učinka vrlo mala. Također, trebali su znati rukovati s računalom jednako kao i s knjigom. Razlog većoj veličini učinka nakon provedenog drugog međutesta je ta što su učenici ozbiljnije pristupili radu na sustavu, nakon što su vidjeli rezultate prethodnog testa. U većoj mjeri su pristupali sustavu i sa kućnog računala, te su se bolje upoznali s radom sustava. Što se pak tiče završnog testa, veličina učinka je umjerena što se i moglo očekivati, budući da je bio kraj školske godine, te se nastava održavala u kasnim poslijepodnevnim satima.

4.3.2. Analiza anketa eksperimentalne grupe

Anketne upitnike ispunilo je ukupno 24 učenika, od čega 7 djevojčica i 17 dječaka. U prvoj anketi željela sam prikupiti njihova očekivanja o primjeni računala u nastavi, odnosno, o primjeni sustava

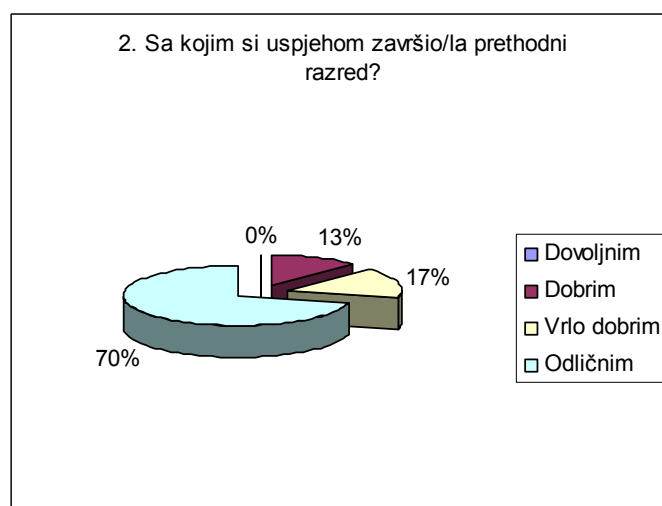
xTEx-Sys, a u drugoj, završnoj anketi, njihovo mišljenje o sustavu. Analize anketa nisu rađene posebno za djevojčice i posebno za dječake zbog malog broja djevojčica.

4.3.2.1. Analiza ankete očekivanja

Ovdje sam prije svega željela saznati posjeduju li učenici računalo kod kuće, koliko se služe s njim i u koji svrhu, te njihovo očekivanje o učenju uz pomoć sustava xTEx-Sys.

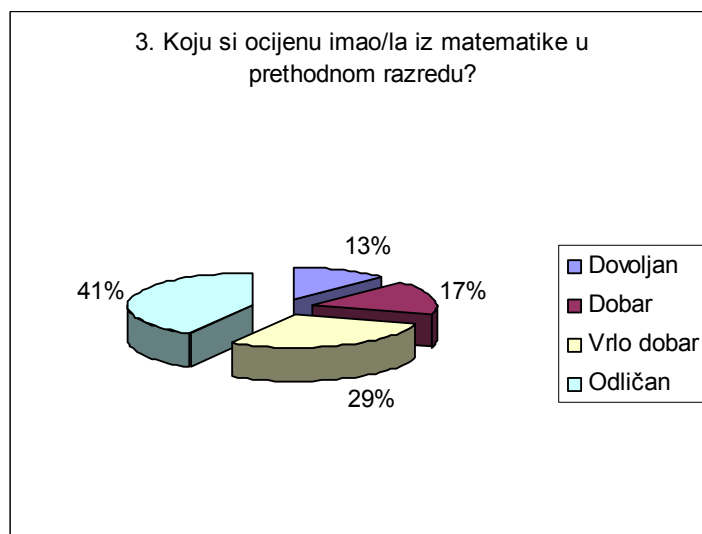


Graf 4.1

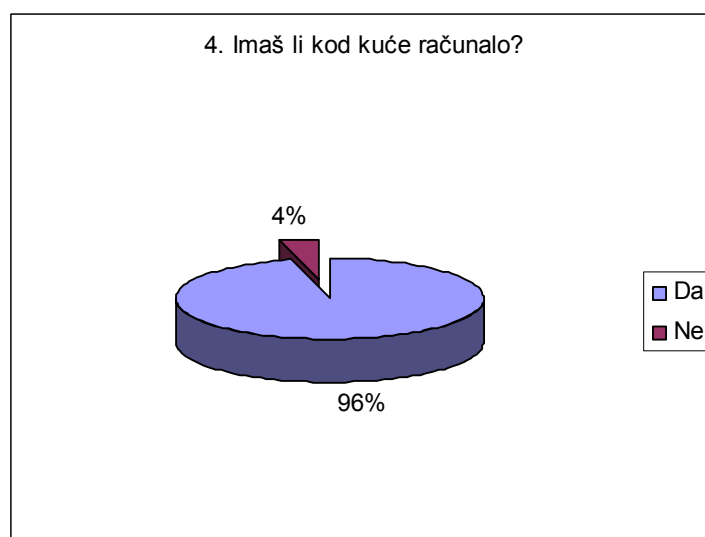


Graf 4.2

Kao što je vidljivo iz Graf 4.2, većina učenika je prethodni razred završila s odličnim uspjehom. Ukupno njih 70% je završilo prethodni razred s odličnim uspjehom, sa vrlo dobrim je prošlo 17%, te sa dobrim 13% učenika. Što se tiče ocjene koju su učenici imali u prethodnom razredu iz matematike, iz Graf 4.3 je vidljivo da je 41% učenika imalo odličan uspjeh, 29% učenika imalo je vrlodobar uspjeh, 17% dobar i 13% dovoljan uspjeh.

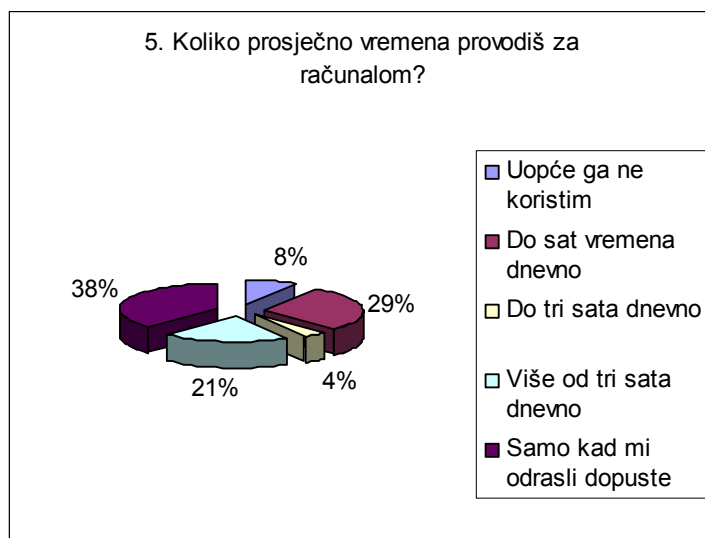


Graf 4.3



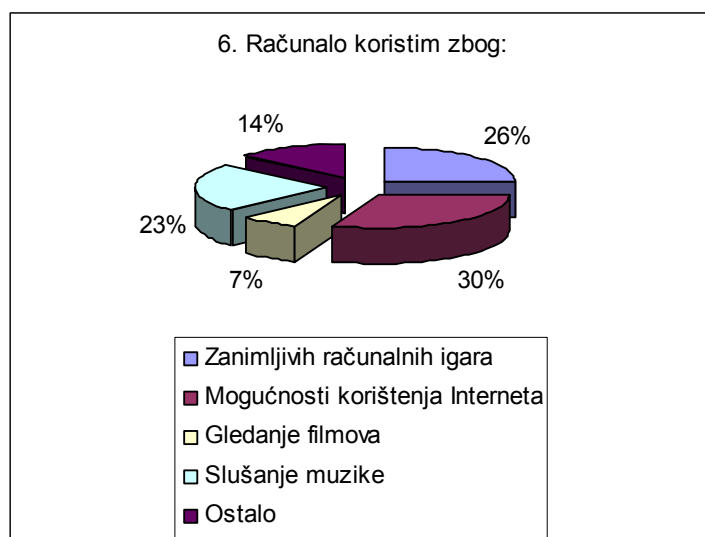
Graf 4.4

Na pitanje „Imaš li kod kuće računalo?“ potvrdno je odgovorilo 96% učenika. Vidljivo je da jako mali broj (svega 4%) učenika kod kuće nema računalo, dakle većina ih je mogla od kuće pristupiti sustavu i poučavati se.



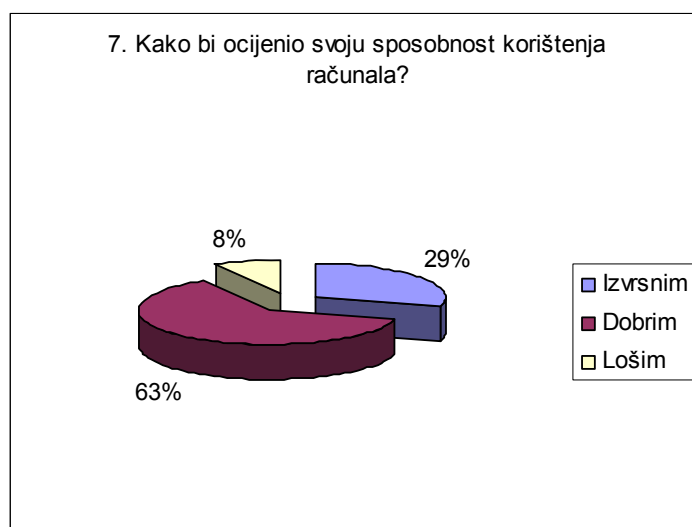
Graf 4.5

Na 5. pitanje učenici su mogli zaokružiti više odgovora. Kao što je prikazano u Graf 4.5 38% učenika koristi računalo samo kada im odrasli dozvole, što je bio i najčešći odgovor. Iznenadujući je podatak da čak 25% učenika koristi računalo tri ili više sati dnevno. Također, iz grafikona je vidljivo da 29% učenika koristi računalo do sat vremena dnevno, a svega 8% učenika ga uopće ne koristi.



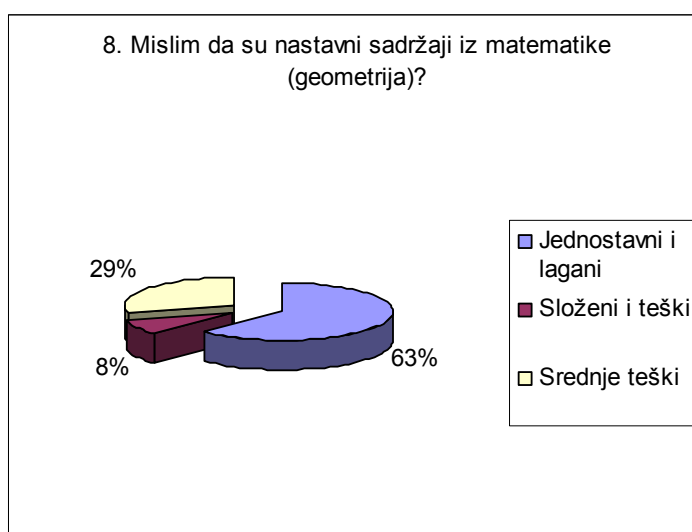
Graf 4.6

I na 6. pitanje učenici su mogli dati više odgovora. Iz Graf 4.6 vidljivo je da najveći broj učenika, njih 30%, koristi računalo zbog mogućnosti korištenja Interneta, no ipak ih 26% koristi zbog računalnih igara. Razdvoje li se odgovori po spolu dolazi se do zaključka da dječaci većinom koriste računalo zbog računalnih igara (54%), zatim zbog mogućnosti korištenja Interneta (26%), dok djevojčice uglavnom koriste računalo zbog pristupa Internetu (51%), a zatim zbog računalnih igara.



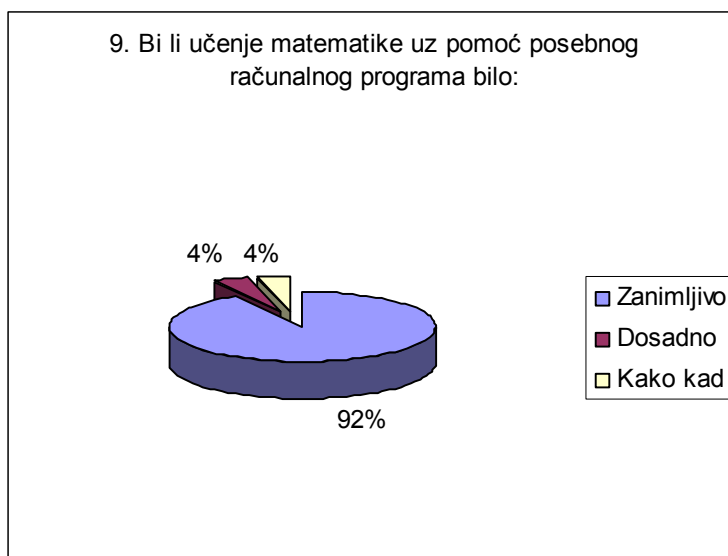
Graf 4.7

Iz Graf 4.7 je vidljivo, da što se tiče njihove procjene samih sebe o sposobnostima i znanju korištenja računalom većina učenika je dala odgovor b) dobrim, što je i bilo za očekivati. Samo dva učenika su priznala da imaju loše znanje o korištenju računala, a to su upravo učenici koji su rekli da kod kuće na posjeduju računalo, a ujedno i ne idu na izbornu nastavu informatike.

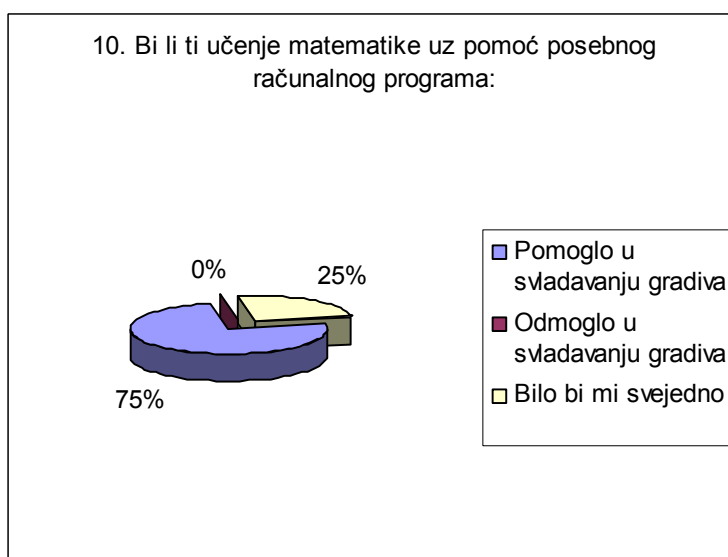


Graf 4.8

Nastavne sadržaje matematike (geometrija) je, iznenađujuće, čak 63% učenika ocijenilo jednostavnim i laganim, dok samo 8% učenika smatra te sadržaje složenim i teškima što prikazuje Graf 4.8.



Graf 4.9



Graf 4.10

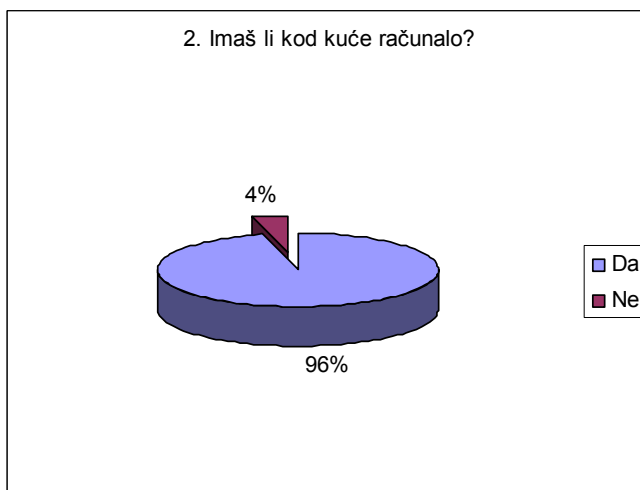
Na posljednja dva pitanja o tome kako bi im bilo da uče matematiku pomoću posebnog računalnog programa, dobili smo vrlo pozitivne rezultate. Naime, čak 92% učenika smatra da bi im to bilo zanimljivo, dok 75% učenika smatra da bi im to pomoglo u svladavanju novih nastavnih sadržaja. Broj učenika kojima bi učenje uz pomoć posebnog računalnog programa bilo dosadno ili svejedno je podjednak (4%). Rezultati su vidljivi u Graf 4.9 i Graf 4.10.

4.3.2.2. Analiza završne ankete

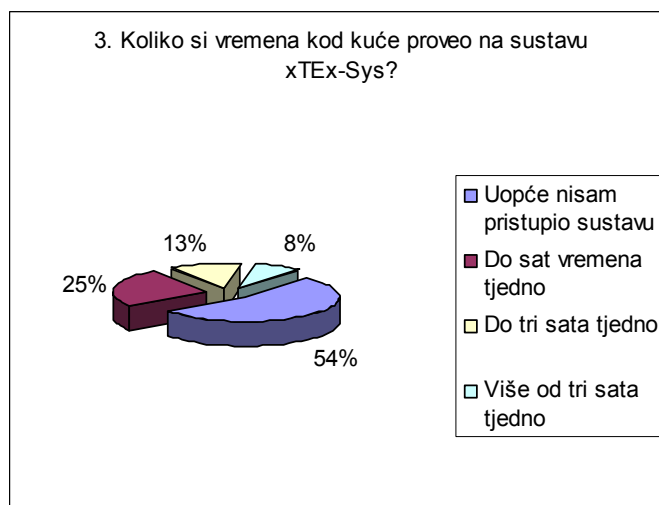
Cilj ove ankete bio je saznati kakve su reakcije učenika na sustav xTeX-Sys, jer je način na koji su učili bio bitno drugačiji od klasičnog načina učenja i poučavanja. Željelo se saznati je li im se i koliko promijenilo mišljenje s obzirom na anketu očekivanja. Zanimalo me njihovo mišljenje o sustavu xTeX-Sys, jesu li razumjeli nastavne sadržaje, bi li željeli nastaviti učiti pomoću sustava xTeX-Sys, te koliko im je pomogao ovaj sustav.



Graf 4.11

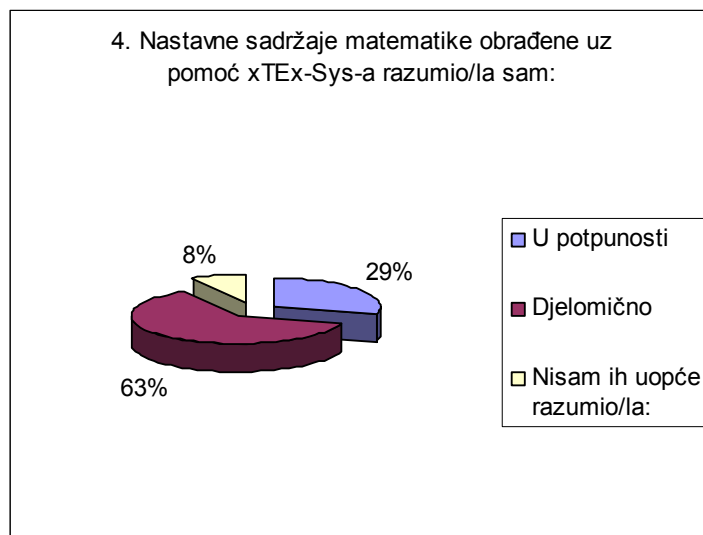


Graf 4.12



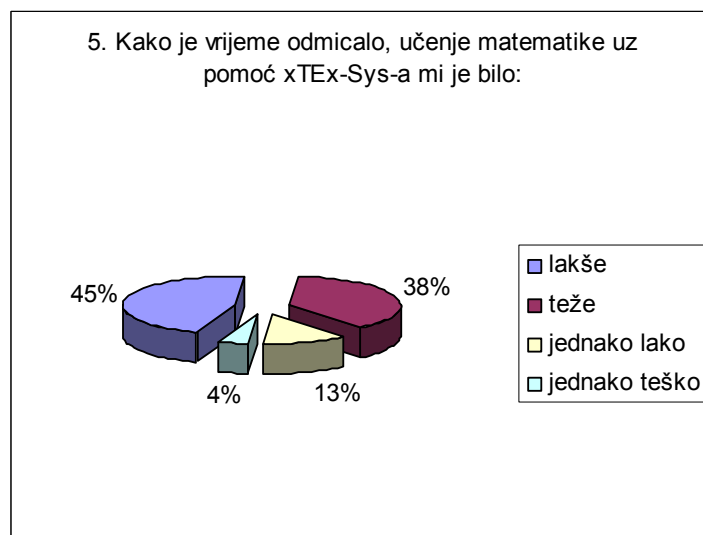
Graf 4.13

Iako se čak 96% učenika izjasnilo da kod kuće ima računalo, 54% ih uopće nije pristupilo sustavu kao što je vidljivo u Graf 4.13. Od onih koji su pristupili sustavu, 25% ih je provelo do sat vremena tjedno na sustavu, 13% do tri sata tjedno i 8% više od tri sata tjedno.



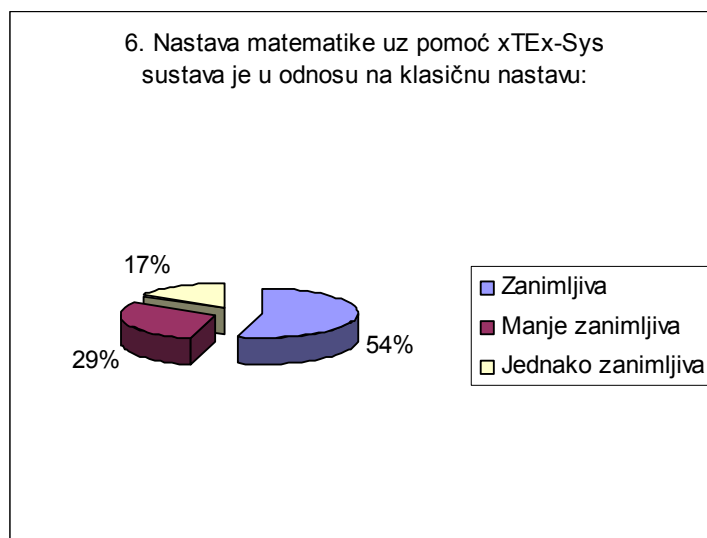
Graf 4.14

Sljedeće pitanje odnosilo se na razumijevanje nastavnog gradiva obrađenog uz pomoć sustava xTeX-Sys. Kao što se vidi u Graf 4.14, vrlo malen broj učenika ih ocjenjuje nejasnima, svega 8%, dok se većina učenika, njih 63%, izjasnila da su nastavne sadržaje razumjeli djelomično.



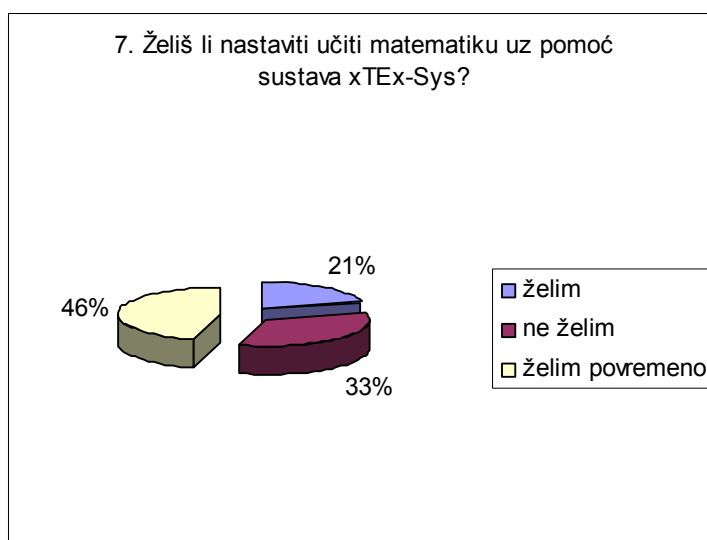
Graf 4.15

Na peto pitanje 45% učenika je odgovorilo da im je učenje matematike uz pomoć xTeX-Sys-a postajalo sve lakše kako je vrijeme odmicalo (Graf 4.15). Njih ukupno 58% navodi da im nije bilo teško učiti na ovakav način, dok je 38% učenika bilo teže i kompliciranije.

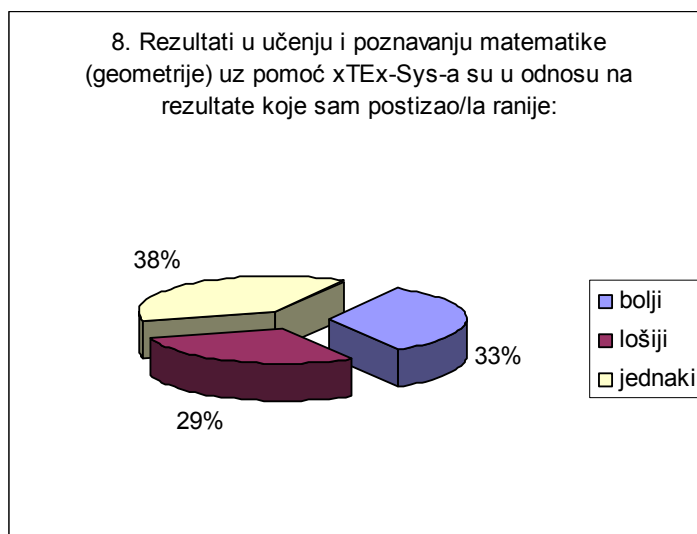


Graf 4.16

Promatrajući odgovore na pitanja u kojima smo željeli saznati njihov interes za daljnjim učenjem matematike uz pomoć sustava xTEx-Sys dobili smo sljedeće rezultate: 54% učenika smatra zanimljivijom nastavu onu uz pomoć sustava xTEx-Sys u odnosu na klasičnu nastavu, dok ih 29 smatra manje zanimljivom, a 17% jednako zanimljivom što je vidljivo i u Graf 4.16. Međutim, u odnosu na takve odgovore, iznenađujuće je da samo 21% učenika želi nastaviti učenje nastavnih sadržaja iz matematike uz pomoć sustava, dok ih 33% to ne želi. Od ukupnog broja učenika, gotovo polovici učenika (46%), je svejedno, što se vidi iz Graf 4.17.

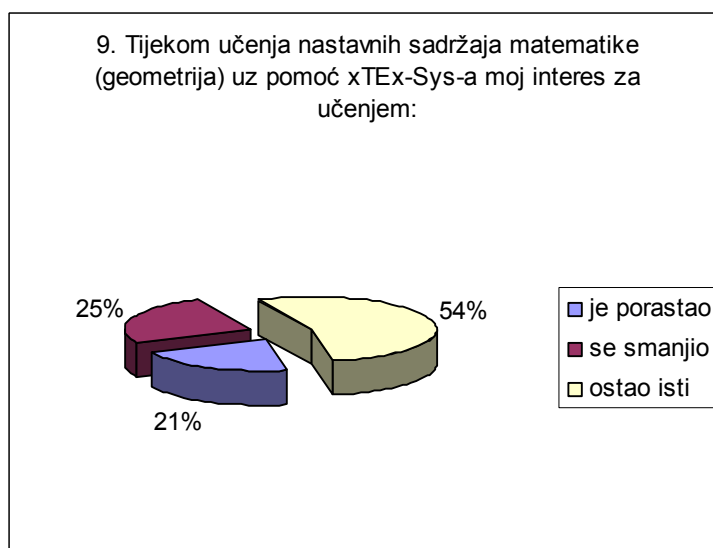


Graf 4.17



Graf 4.18

Iz Graf 4.18 je vidljivo da su rezultati u učenju i poznavanju informatike uz pomoć xTEx-Sys-a u odnosu na rezultate koje su postizali ranije sljedeći: kod 29% učenika rezultati su lošiji, 38% učenika postiglo je jednake, a 33% ih je postiglo jednake rezultate kao i ranije.



Graf 4.19

Što se tiče interesa za učenje matematike, tijekom učenja uz pomoć sustava xTEx-Sys je uglavnom ostao isti kao što pokazuje Graf 4.19. Kod 21% učenika interes je porastao, a kod 25% učenika se smanjio.

Posljednje pitanje u anketi je bilo otvorenog tipa. U njemu se od učenika tražilo da daju svoje mišljenje o sustavu xTEx-Sys, te o načinu na koji im je sustav prezentiran. Ovdje su navedeni samo neki od odgovora:

- „Ovaj sustav je baš super, zanimljivije nego u učionici. Mogu učiti ono šta mi nije jasno. I dobila sam bolje ocjene.“
- „xTeX-Sys je dobar, ali malo dosadan, iako je profesorica uvijek provala učiniti to zanimljivijim.“
- „Meni je draže kad učimo normalno u učionici, jer mogu pitati profesoricu kad mi nešto nije jasno. Nebi baš volio dalje ovako učiti.“
- „Profesorica nam je baš super objasnila te čvorove i podčvorove i veze. Sviđa mi se program i baš bi bilo dobro da još nešto učimo pomoću njega.“
- „Mislim da je dobro, ali malo teže nego kad učim iz knjige. A od kuće mi je bilo teško za učiti jer mi kompjuter nije baš najbolji i spor je. Ipak bi radije da učimo iz knjige.“
- „Bilo mi je OK! U početku je bilo malo teško dok nisam shvatio, ali kada mi je učiteljica još malo objasnila bilo je super. I dobio sam sve petice.“
- „Sustav je super i zanimljiv. I sviđa mi se kako nam je profesorica objasnila sve pomoću životinjskog carstva.“
- „xTeX-Sys je malo kompliciran, pa mi je trebalo malo vremena da ga razumim. Poslje je bilo baš zanimljivo.“
- „Mislim da bi bilo super kad bi malo radili u učionici, a malo ovako preko xTeX-Sys-a. Prvo da nam učiteljica objasni, a onda da mi sami učimo.“
- „Učenje je dobro, ali ja nemam kući kompjuter pa nemogu učiti doma, a moram pisati ispit. To mi nije baš pošteno.“
- „Profesorica nam je sve super objasnila i više puta ako nam nije bilo jasno. A sustav je isto super.“
- „Učiteljica je sve super objasnila, ali ovako mi je teže učiti, pa nebi više to želio.“
- „Ma ja neznam šta će meni ti čvorovi i veze ikad trebati u životu.“
- „Baš je bilo zanimljivo. Volila bi i dalje učiti pomoću xTeX-Sys-a.“

5. Zaključak

Inteligentni tutorski sustavi su u širokoj primjeni u poučavanju i učenju. Obećavaju obogaćenje mogućnosti učenja tako što pojedinom učeniku nude podršku u podučavanju i učenju. Vrednovanje ovih obrazovnih sustava postaje važno. Ono može služiti kao alat za daljnji razvoj istraživanja na polju ITS-a tako što se daju prijedlozi za cjelokupno poboljšanje arhitekture i ponašanja ovih sustava.

Iako se inteligentni tutorski sustavi sve više koriste u nastavnom procesu, mali broj tih sustava je vrednovan. Upravo zbog toga javlja se sve veća potreba za njihovim vrednovanjem. Vrednovanje sustava, a pogotovo njegovog učinka zahtjeva dugotrajni angažman učitelja koji upotrebljava taj sustav, ali i učenika koji sudjeluju u nastavi. Kako inteligentni tutorski sustavi utječu na sam proces, te postignuća učenika potrebno ih je vrednovati prije njihove primjene.

U ovom radu posebna je pažnja usmjerena na različitost metoda vrednovanja, u ovisnosti o tome što se želi vrednovati. Na temelju spoznaja iz literature dali smo također pregled i klasifikaciju metoda vrednovanja, te smo posebno opisali eksperiment kao metodu vrednovanja, budući da se ta metoda ponajviše koristi u psihološkim i pedagoškim istraživanjima. Objasnili smo što je veličina učinka i na koji način se najčešće računa.

Proveli smo eksperimentalno istraživanje da bismo vrednovali učinkovitost sustava xTEx-Sys. Sudjelovanjem u ovom istraživanju, kao budući učitelj uvidjela sam koliko ovakav način organizacije može pomoći u radu. Jednom napravljena baza znanja može se uvijek koristiti i nadograđivati. Potreban je određen vremenski period da se nauči izgraditi bazu znanja i oblikovati nastavne sadržaje na sustavu, no što je poznavanje računala bolje potrebno je manje vremena. Osim što učitelju pomaže organizirati gradivo, ovakvim načinom organizacija nastave odnos učenika i učitelja je kvalitetniji. Učenici učitelja smatraju osobom koja će im pomoći u radu i poticati ga na bolje rezultate, te se ne ustručavaju komunicirati s učiteljem. Nemoguće je gradivo, vrijeme i ispite predviđeno za određene sadržaje prilagoditi svakom učeniku u tradicionalnoj nastavi, a sustav xTEx-Sys se prilagođava individualnim potrebama učenika, te su učenici sami primijetili da je učenje uz pomoć ovog sustava kao da svaki od njih ima svog učitelja.

Mišljenja sam da bi na ovakav način trebalo organizirati dodatno obrazovanje učitelja, te ih upoznati s novim tehnologijama i načinima održavanja nastave, jer dojmovi su isključivo pozitivni, a iz komentara učenika u anketi možemo zaključiti da su jako dobro prihvatili sustav xTEx-Sys i pozitivno reagirali na drugačiji načina nastave, te da bi željeli ponovno učiti uz pomoć ovog sustava. Iz naših anketa, te analize rezultata možemo zaključiti da se xTEx-Sys može uspješno koristiti u izgradnji baza znanja, te da su učenici postigli bolje rezultate korištenjem ovog sustava.

Dakle, možemo zaključiti da je svrha ovog rada sistematizacija metoda vrednovanja koje se koriste za vrednovanje inteligentnih tutorskih sustava, primjena jedne od metoda, točnije, eksperimenta za vrednovanje učinkovitosti inteligentne autorske ljsuke xTEx-Sys, te interpretacija sustava za automatsko vrednovanje učinka inteligentnih tutorskih sustava.

6. Literatura

- [DEMP2004] Dempster, J. (2004) *CAR e-Learning guides: Evaluating E-Learning*, Centre of Academic Excellence, University of Warwick, www.warwick.ac.uk/go/cap/resources/eguides
- [GRUB2007] Grubišić, A., *Vrednovanje učinka inteligentnih sustava e-učenja*, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva. URL: http://www.pmfst.hr/~ani/radovi/Ani_Grubisic_mr_Vrednovanje_ucinka_inteligentnih_sustava_e_ucenja_2007.pdf, 2007
- [IQBA1999] Iqbal A., Oppermann R., Patel A. And Kinshuk *A Classification of Evaluation Methods for Intelligent Tutoring Systems*, Software Ergonomie '99 – Design von Informationswelten (Eds. U. Arend, E. Eberleh & K. Pitschke), B. G. Teubner Stuttgart, Leipzig, 1999, pp. 169-181
- [MARK1993] Mark M.A. and Greer J.E. *Evaluation methodologies for intelligent tutoring systems*, Journal of Artificial Intelligence and Education, 4 (2/3), 1993, pp. 129-153.
- [PHIL2002] Phillips, R.A., & Gilding, A. *Approaches to evaluating the effect of ICT on student learning*. Learning and Teaching Support Network, 2002.
- [SHUT1996] Shute, V. J., and Psotka, J., (1996) *Intelligent tutoring systems: Past, present and future*. In D. Jonasses (Ed.), *Handbook of Research for Educational Communications and Technology* (pp. 570-600). New York, NY: Macmillan.
- [SIEM1998] Siemer, J., Angelides, M.C. (1998) *A comprehensive method for the evaluation of complete intelligent tutoring system*, Decision Support Systems 22, pp. 85-102
- [STAN2005] Stankov, S., Grubišić, A., Žitko, B., Krpan, D. (2005) *Vrednovanje učinkovitosti procesa učenja i poučavanja u sustavima za e-učenje*, Školski Vjesnik-časopis za pedagoška i školska pitanja, 54, 1-2; 21-31
- [TWID1993] Twidale, M. B. (1993) *Redressing the balance: The advantages of informal evaluation techniques for Intelligent Learning Environments*. Journal of Artificial Intelligence In Education, 4(2/3):155-178.
- [MUŽI1999] Uvod u metodologiju istraživanja odgoja i obrazovanja, Educa, Zagreb
- [OLIV2000] Oliver, M., (2000). *An introduction to the Evaluation of Learning Technology*, London: University of North London.
- [PHIP1999] Phipps, R., Associate, S., and Merisotis (1999), *A Review of Contemporary Research on the Effectiveness of Distance Learning in Higher Education*, The Institute for Higher Education Policy
- [ROTN1978] Rot, N., *Opšta psihologija*, Zavod za udžbenike i nastavna sredstva, Beograd, 1978

7. Prilozi

- A. Inicijalni test
- B. Prvi međutest
- C. Drugi međutest
- D. Završni test
- E. Anketa očekivanja
- F. Završna anketa

1. Nacrtaj dužinu duljine 5cm.

2. Što je pravac?

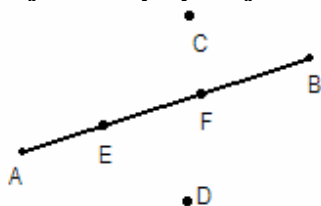
Pravac je _____

3. Nacrtaj dva okomita pravca i označi ih.

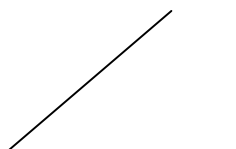
4. Nacrtaj polupravac p s početnom točkom T.

5. Koja je oznaka za dužinu?

6. Koje točke pripadaju dužini sa slike?



7. Kako zovemo kut sa slike?



8. Nabroji vrste trokuta s obzirom na stranice!

9. Izračunaj opseg trokuta ako su mu stranice duge 4cm, 60mm i 3cm!

10. Nacrtaj jedan pravokutnik i označi mu vrhove i stranice!

11. Izračunaj opseg kvadrata duljine stranice 15cm!

ZAOKRUŽI JEDAN ILI VIŠE PONUĐENIH ODGOVORA

1. **Kako zovemo pravac koji prolazi polovištem dužine?**
 - a) polukrug
 - b) dijagonale paralelograma
 - c) simetrala
 - d) paralelni pravci
2. **Što od navedenog ima DULJINU?**
 - a) pravac
 - b) polupravac
 - c) dužina
 - d) ravnina
3. **Koja je formula za opseg pravokutnika?**
4. **Što je od navedenog dio kruga?**
 - a) kružni luk
 - b) pravac
 - c) polukrug
 - d) kružni isječak
5. **Kome su sve stranice jednake duljine?**
 - a) kvadratu
 - b) pravokutniku
 - c) paralelogramu
 - d) ništa od navedenog
6. **Što je temeljna značajka KVADRATA?**
 - a) Ima dva para paralelnih stranica
 - b) Određen je kružnicom
 - c) Sve stranice su jednake duljine
 - d) Prolazi središtem dužine
7. **Što je od navedenog primjer DUŽINE?**
 - a) dijagonale paralelograma
 - b) simetrala
 - c) kružni odsječak
 - d) tetiva
8. **Kome su dijagonale međusobno jednake duljine?**
 - a) pravokutniku
 - b) rombu
 - c) kvadratu
 - d) paralelogramu
9. **Što je temeljna značajka POLUPRAVCA?**
 - a) na obje svoje strane proteže se u beskonačnost
 - b) na jednu svoju stranu proteže se u beskonačnost
 - c) najkraća spojnica dviju točaka
 - d) ništa od navedenog
10. **Kružni luk je dio od:**
 - a) kružnice
 - b) pravca
 - c) dužine
 - d) kružnog isječka

ZAOKRUŽI JEDAN ILI VIŠE PONUĐENIH ODGOVORA

1. Susjedni kutevi čine:

- a. pravi kut
- b. ispruženi kut
- c. puni kut
- d. tupi kut

2. Koji kut je veći od 90° ?

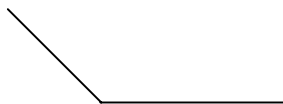
- a. tupi
- b. pravi
- c. izbočeni
- d. šiljasti

3. Veličinu kuta izražavamo u:

- a. kutnim sekundama
- b. milimetrima
- c. kutnim stupnjevima
- d. centimetrima

4. Koji kut se nalazi na slici?

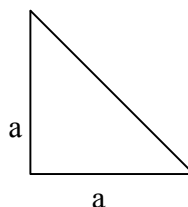
- a. pravi
- b. tupi
- c. šiljasti
- d. izbočeni

5. Kolika je površina pravokutnog trokuta ako su mu duljine stranica $a = 30\text{cm}$, $b = 6\text{dm}$?

- a. 9 cm^2
- b. 36 dm^2
- c. 9 dm^2
- d. 36 cm^2

6. Što je na slici?

- a. raznostranični trokut
- b. jednakokrani trokut
- c. tupokutni trokut
- d. pravokutni trokut



7. Što je od navedenog osnosimetrični lik?

- a. dužina
- b. raznostranični trokut
- c. krug
- d. paralelogram

8. Koliko osi simetrije ima pravokutnik?

- a. 1
- b. 2
- c. 3
- d. 4

9. Kolika je duljina osnovice jednakokrannog trokuta kojem je opseg 28cm , a duljina kraka 10cm ?

- a. na obje svoje strane proteže se u beskonačnost
- b. na jednu svoju stranu proteže se u beskonačnost
- c. najkraća spojnica dviju točaka
- d. ništa od navedenog

10. Koje su vrste trokuta s obzirom na stranice

1. Simetrala dužne je:

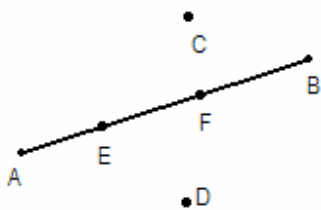
- a. dužina
- b. pravac
- c. polupravac

2. Što je od navedenog dio kruga?

- a. polukrug
- b. pravac
- c. kružni odsječak
- d. polupravac

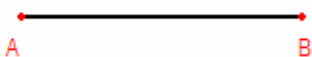
3. Što se nalazi na slici:

- a. kružni vijenac
- b. polukrug
- c. kružni isječak
- d. kružni odsječak

4. Izračunaj površinu pravokutnika kojem je $a = 5 \text{ cm}$, $b = 30 \text{ mm}$?**5.**

Točke koje pripadaju dužini su: _____

Točke koje ne pripadaju dužini su: _____

6. Točke A i B zovu se :

- a. Početne točke
- b. Krajnje točke

7. Izbočeni kut je

8. Koji su od navedenih kuteva manji od ispruženog kuta?

- a. Tupi kut
- b. izbočeni kut
- c. šiljasti kut
- d. puni kut

9. Koja je osnovna mjerna jedinica za veličinu kuta?

- a. kutne sekunde
- b. kutne minuta
- c. kutni stupnjevi
- d. ništa od navedenog

10. Kako zovemo kut koji ima 360° ?

- a. pravi kut
- b. ispruženi kut
- c. puni kut

11. Kako zovemo kuteve čiji je zbroj veličina 180° ?

- a. vršni kutevi
- b. suplementarni kutevi

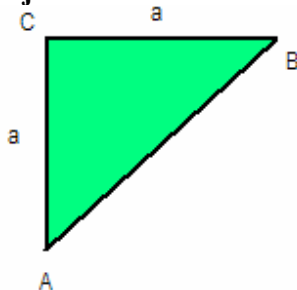
12. Koje vrste trokuta imamo s obzirom na veličine njegovih kuteva?

13. Najdulja stranica pravokutnog trokuta zove se:

- a. osnovica
- b. kateta
- c. hipotenuza
- d. krak

14. Izračunaj opseg jednakostraničnog trokuta kojem je duljina stranice 7 cm.

15. Što je na slici?



- a. pola kvadrata
- b. jednakokrani trokut
- c. jednakostranični trokut
- d. raznostranični trokut

16. Što je osnosimetrična slika trokuta?

- a. kvadrat
- b. paralelogram
- c. trokut
- d. pravac

17. Koliko osi simetrije ima kvadrat?

- a. 1
- b. 2
- c. 3
- d. 4

18. Što je od navedenog primjer osnosimetričnog lika?

- a. romb
- b. raznostranični trokut
- c. krug
- d. polupravac

19. Ako je veličina kuta $\beta = 38^\circ$, kolika je veličina njemu suplementarnog kuta α ?

20. Opseg jednakokrakog trokuta je 32 dm, a duljina osnovice je 12 dm. Kolika je duljina kraka?

- a. 12 dm
- b. 10 dm
- c. 6 dm
- d. 8 dm

OŠ Bol, Split
Split: 11.05.2007

ANKETNI UPITNIK

PRED TOBOM JE ANKETNI UPITNIK.

MOLIMO TE DA ISKRENO ODGOVORIŠ NA PITANJA, JER NITKO NEĆE ZNATI DA SI UPRAVO TI ISPUNJAVAO/LA OVAJ UPITNIK!

AKO NEKO PITANJE NIJE JASNO, SVAKAKO ZATRAŽI POMOĆ!

ZAHVALJUJEMO SE NA SURADNJI I STRPLJENJU!

1. Kojeg si spola?
 - a) Muško
 - b) Žensko

2. Sa kojim si uspjehom završio/la prethodni razred?
 - a) Dovoljnim
 - b) Dobrim
 - c) Vrlodobrim
 - d) Odličnim

3. Koju ocjenu si imao/la iz matematike u prethodnom razredu?
 - a) Dovoljan
 - b) Dobar
 - c) Vrlodobar
 - d) Odličan

4. Imaš li kod kuće računalo?
 - a) Da
 - b) Ne

5. Koliko prosječno vremena provedeš za računalom?
 - a) Uopće ga ne koristim
 - b) Do sat vremena dnevno
 - c) Do tri sata dnevno
 - d) Više od tri sata dnevno
 - e) Samo kada mi odrasli dopuste

6. Računalo koristim zbog:
 - a) Zanimljivih računalnih igara
 - b) Mogućnosti korištenja Interneta
 - c) Gledanje filmova
 - d) Slušanja muzike
 - e) _____

7. Kako bi ocjenio svoju sposobnost korištenja računala?
 - a) Izvrsnim
 - b) Dobrim
 - c) Lošim

8. Mislim da su nastavni sadržaji iz matematike (geometrija):
- a) Jednostavni i lagani
 - b) Složeni i teški
 - c) _____
9. Bi li učenje matematike (geometrije) uz pomoć posebnog računalnog programa bilo:
- a) Zanimljivo
 - b) Dosadno
 - c) _____
10. Bi li ti učenje matematike (geometrije) uz pomoć posebnog računalnog programa:
- a) Pomoglo u svladavanju gradiva
 - b) Odmoglo u svladavanju gradiva
 - c) Bilo bi mi svejedno
 - d) _____

OŠ Bol, Split
Split: 01.06.2007

ANKETNI UPITNIK

PRED TOBOM JE ANKETNI UPITNIK.

MOLIMO TE DA ISKRENO ODGOVORIŠ NA PITANJA, JER NITKO NEĆE ZNATI DA SI UPRAVO TI ISPUNJAVAO/LA OVAJ UPITNIK!

AKO NEKO PITANJE NIJE JASNO, SVAKAKO ZATRAŽI POMOĆ!

ZAHVALJUJEMO SE NA SURADNJI I STRPLJENJU!

1. Kojeg si spola?
 - a. Muško
 - b. Žensko

2. Imaš li kod kuće računalo?
 - a. Da
 - b. Ne

3. Koliko si vremena kod kuće proveo na sustavu xTeX-Sys?
 - a. Uopće nisam pristupio sustavu
 - b. Do sat vremena tjedno
 - c. Do tri sata tjedno
 - d. Više od tri sata tjedno
 - e. _____

4. Nastavne sadržaje matematike obrađene uz pomoć xTeX-Sys-a razumio/la sam:
 - a. u potpunosti
 - b. djelomično
 - c. nisam ih uopće razumio/la

5. Kako je vrijeme odmicalo, učenje matematike uz pomoć xTeX-Sys-a mi je bilo:
 - a. teže (kompliciranije)
 - b. lakše (jednostavnije)
 - c. jednako teško
 - d. jednako lako

6. Nastava matematike uz pomoć xTeX-Sys sustava je u odnosu na klasičnu nastavu:
 - a. zanimljivija
 - b. manje zanimljiva
 - c. jednako zanimljiva

7. Želiš li nastaviti učiti matematiku uz pomoć sustava xTeX-Sys?
 - a. želim
 - b. ne želim
 - c. želim povremeno

8. Rezultati u učenju i poznavanju matematike (geometrije) uz pomoć xTeX-Sys-a su u odnosu na rezultate koje sam postizao/la ranije:

- a. bolji
- b. lošiji
- c. jednaki

9. Tijekom učenja nastavnih sadržaja matematike (geometrija) uz pomoć xTeX-Sys-a moj interes za učenjem:
- a. je porastao
 - b. se smanjio
 - c. ostao isti

10. Svojim riječima opiši što misliš o učenju matematike uz pomoć xTeX-Sys-a?