

FAKULTET PRIRODOSLOVNO-MATEMATIČKIH ZNANOSTI  
SVEUČILIŠTE U SPLITU

Natalija Zulim

# **Replikacija eksperimenata u vrednovanju sustava e-učenja**

DIPLOMSKI RAD

SPLIT, SRPANJ, 2008.

Studijska grupa: Matematika i informatika  
Predmet: Primjena računala u nastavi

# **Replikacija eksperimenata u vrednovanju sustava e-učenja**

DIPLOMSKI RAD

Student: Natalija Zulim

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

SPLIT, SRPANJ, 2008.

## **ZAHVALA**

*Hvala prof.dr.sc. Slavomiru Stankovu na savjetima i razumjevanju tijekom izrade ovog rada.*

*Veliko hvala mr.sc. Ani Grubišić na velikom strpljenju, uloženom trudu i podmoći koju mi je pružila u izradi ovog rada.*

*Hvala i ostalim profesorima, asistentima, prijateljima i kolegama na ukazanoj podršci.*

*Najveće hvala mojim roditeljima na razumjevanju i podršci tijekom studiranja.*

---

---

# SADRŽAJ

<b>1. UVOD</b> .....	<b>1</b>
<b>2. PROVOĐENJE EKSPERIMENTA I REPLIKACIJE</b> .....	<b>3</b>
2.1. HIJERARHIJSKI MODEL ZA OBLIKOVANJE USPJEŠNIH REPLIKACIJA.....	4
2.1.1. Matematički model .....	5
2.1.2. Proučavanje heterogenosti .....	8
2.1.2.1. Bayesovi faktori.....	10
2.1.3. Primjer.....	10
2.2. OBLIKOVANJE EKSPERIMENTA I REPLIKACIJE .....	12
2.2.1. Cilj istraživanja .....	12
2.2.2. Sudionici .....	13
2.2.3. Provedba.....	16
2.3. ANALIZA PODATAKA .....	17
2.3.1. Definiranje hipoteza .....	18
2.3.2. Veličina učinka .....	19
2.3.3. Statistička analiza.....	20
2.4. REZULTATI I INTERPRETACIJA .....	21
2.4.1. Valjanost.....	22
2.5. PRIMJER .....	24
2.5.1. Uvod .....	24
2.5.2. Postupak .....	25
2.5.3. Analiza podataka .....	27
2.5.4. Rezultati.....	28
2.5.4. Interpretacija.....	32
2.5.5. Određivanje valjanosti .....	33
<b>3. REPLIKACIJA EKSPERIMENTA ZA VREDNOVANJE UČINKOVITOSTI SUSTAVA xTeX-Sys</b> .....	<b>34</b>
3.1. E-UČENJE .....	34
3.2. xTeX-Sys.....	35
3.3. OPIS EKSPERIMENTA I REPLIKACIJE .....	36
3.3.1. Cilj eksperimenta i replikacije.....	36
3.3.2. Sudionici .....	36
3.3.3. Provedba eksperimenta i replikacije .....	37
3.3.3.1. Provjera ekvivalentnosti grupa za originalan eksperiment.....	37
3.3.3.2. Povjera ekvivalentnosti grupa za replikaciju eksperimenta .....	39
3.3.4. Analiza podataka i interpretacija .....	40
3.3.4.1. Analiza i interpretacija podataka za originalan eksperiment.....	41
3.3.4.2. Analiza i interpretacija podataka za replikaciju eksperimenta .....	42
3.3.5. Zaključak .....	44
<b>4. ZAKLJUČAK</b> .....	<b>46</b>
<b>5. LITERATURA</b> .....	<b>48</b>
<b>6. PRILOZI</b> .....	<b>50</b>

---

# 1.Uvod

Eksperimenti u područjima fizike, medicine, psihologije, sociologije, ekologije, ekonomije, itd... se često provode od strane različitih istraživačkih timova. Kod većine istraživanja je primijećeno da ukoliko se ne ponavljaju nisu od velikog interesa. Repliciranje eksperimenta se često proučava. S obzirom na raširenu uporabu čini se da nedostaje istraživanja i sistematskih rasprava o značenju i rezultatima replikacije. Nadalje, nije sasvim jasno što to znači "ponavljajući" eksperiment, koji su rezultati repliciranja eksperimenata ili što se mora uzeti kao "uspješna" replikacija (Tversky i Kahnemann, 1982, Utts, 1991, prema [BAYA1997]).

Eksperiment dolazi od latinske riječi experimentum što znači pokus, dokaz (grč. πειράω = iskušavam, ispitujem). Pojam je u srednjem vijeku i renesansi imao značenje iskustva. Prvi koji počinje koristiti pojam eksperiment u današnjem značenju bio je F. Bacon o čemu opširno govori u djelu Novum Organon (1622.-23.). Osobitu važnost eksperimentu u svom znanstvenom istraživanju daje G. Galilej [JALŠ2008].

Replikacija je pokušaj reprodukcije empirijske studije kako bi se dobila buduća vrijednost otkrića ili uspješan rezultat takvih pokušaja. Replicirani eksperiment se odnosi na "ispravnu" ili "identičnu" replikaciju [ALMQ2006]. Naravno, postizanje "uspjeha" u repliciranju ne znači ništa drugo nego postizanje cilja zbog kojeg se replikacija i izvodila (Shapin i Coly, 1985; Neuliep, 1990; Berry, 1996 prema [BAYA1997]). Stoga, kada oslikavaju replikaciju, znanstvenici moraju imati jasno definiranu svrhu replikacije.

Kod oblikovanja eksperimenta i replikacije treba voditi računa o ciljevima, odabiru sudionika, definiranju ekvivalentnih eksperimentalnih i kontrolnih grupa, postavljanju hipoteza, provođenju adekvatne analize podataka te na kraju pravilne interpretacije. Osnovni cilj eksperimenta i replikacije jest utvrditi da li postoji, koliki je i kakav je utjecaj određene nezavisne varijable na određenu zavisnu varijablu.

U zadnje vrijeme sve se više susrećemo sa pojmom e-učenje. E-učenje je pojam koji opisuje učenje prilikom kojeg se koristi računalo, koje je najčešće spojeno na neku vrstu mreže (lokalna ili Internet) te dozvoljava korisniku učenje gotovo bilo kada, bilo gdje. Pod e-učenjem uobičajeno se podrazumijeva izvođenje obrazovnog procesa uz pomoć informacijsko-komunikacijske tehnologije (korištenjem elektroničkih medija kao što su CD-ROM, DVD i internet). Posebna klasa inteligentnih sustava e-učenja su inteligentni tutorski sustavi (ITS). Inteligentni tutorski sustavi [STAN2003] su računalni sustava namijenjeni potpori i poboljšanju procesa učenja i poučavanja u odabranom područnom znanju, uzimajući u obzir pri tom individualnost onoga tko uči i tko se poučava.

U drugom poglavlju opisano je što to znači uspješna replikacija i koji su ciljevi uspješne replikacije. Dan je matematički model po kojem se oblikuje uspješan eksperiment i replikacija. Iznosen je primjer koji se temelji na danom matematičkom modelu. Zatim slijedi detaljniji opis kako provesti eksperiment i replikaciju. Kako je eksperiment i njegova replikacija istraživački proces koji ima više faza, opisane su pojedine faze u tom postupku: što su ciljevi, kako odabirati sudionike, postaviti hipoteze, kako adekvatno napraviti analizu podataka i na kraju pravilno ih

interpretirati. Opisan je primjer dvaput ponovljenog eksperimenta koji se bavi ocjenjivanjem nastavne učinkovitosti korištenja razvojno simulacijskog modela.

U trećem poglavlju prikazani su rezultati eksperimenta i replikacije sa studentima Kemijsko tehnološkog fakulteta (KTF) i Fakulteta prirodoslovno-matematičkih područja i kineziologije (PMF) o vrednovanju učinkovitosti učenja uz pomoć inteligentnog tutorskog sustava xTEx-Sys [STAN2003]. Nakon što smo odredili ciljeve eksperimenta i replikacije, odredili smo studente u kontrolnoj i eksperimentalnoj grupi tako da grupe budu ekvivalentne kako za eksperiment tako i replikaciju, nakon čega se može provoditi daljnja analiza podataka potrebna za prihvaćanje ili odbacivanje nul-hipoteza. Na kraju smo izračunali veličinu učinka za svaki parcijalni test i za završni test za eksperiment i replikaciju te prosječnu veličinu učinka za eksperiment i replikaciju koja nam služi za uspoređivanje učinkovitosti učenja uz pomoć xTEx-Sys-a u eksperimentu i replikaciji.

## 2. Provođenje eksperimenta i replikacije

Kako u eksperimentu istraživač namjerno izaziva određenu pojavu, on poznaje mjesto i vrijeme određene pojave, a to mu omogućuje da se unaprijed pripremi za točno opažanje. Poznavanje uvijeta pod kojim neka pojava nastaje pruža mogućnost i drugim istraživačima da repliciraju opažanja i da tako kontroliraju točnost rezultata. U kontroliranim uvjetima moguće je izazvanu pojavu u tolikoj mjeri stabilizirati da se ona može i izmjeriti, tj. njena veličina izraziti nekim suvislim sustavom numeričkih vrijednosti. Na stupnju izazivanja neke pojave vrši se kvalitativna analiza, na stupnju mjerenja prilika pod kojima neka pojava nastaje vrši se asocijalna analiza, a na stupnju mjerenja izazvane pojave vrši se kvantitativna analiza [BUJA1967].

Kad je eksperiment proveden tako da se opaža i mjeri zavisna varijabla jedanput uz prisustvo, a drugi put uz odsustvo nezavisne varijable, odnosno pod uvjetima dviju kvalitativno različitih situacija, tada se taj eksperiment zove *faktorijalni eksperiment*. Faktorijalni eksperiment daje odgovor na pitanje što se događa sa zavisnom varijablom pod utjecajem nezavisne varijable [MEJO2003].

Ako je pak nezavisna varijabla u eksperimentu uvijek prisutna, a mijenja se samo njen intezitet ili količina da bi se proučilo njeno djelovanje na zavisnu varijablu, tada se taj tip eksperimenta zove *funkcionalni eksperiment*. Rezultati funkcionalnog eksperimenta trebaju dati odgovor na pitanje kako utječe nezavisna varijabla na zavisnu varijablu [MEJO2003].

*Replikacija* je pokušaj reprodukcije empirijske studije kako bi se dobila buduća vrijednost otkrića ili uspješan rezultat takvih pokušaja [ALMQ2006]. Replicirani eksperiment se odnosi na "ispravnu" ili "identičnu" replikaciju. Kako joj ime govori, ispravna replikacija se određuje imitirajući što je moguće više originalni eksperiment. Cilj je obično nadati se jednakim zaključcima.

Replikaciju djelimo na:

- **blisku replikaciju (eng. close replication)** je replikacija koja pokušava zadržati sve poznate uvjete studije istim ili barem jako sličnima kakvi su bili u originalnom eksperimentu.
- **različita replikacija (eng. differentiated replication)** je replikacija koja uključuje namjerne ili barem poznate varijacije u većini prikladnih aspekata studije.

Postoje i druge podjele replikacija pa imamo:

- **vanjska replikacija (eng. external replication)** je neovisna replikacija izvedena od različitih istraživača od onih koji su bili u originalnom eksperimentu.
- **unutrašnja replikacija (eng. internal replication)** je ona koja nije vanjska.

## **2.1. Hijerarhijski model za oblikovanje uspješnih replikacija**

Replikacija eksperimenata se jako dugo provodi u istraživanjima. Kako god, s obzirom na raširenu uporabu, sistematske studije o rezultatima i motivima replikacija su rijetke. Kao posljedica toga se čini da nema preciznog zapisa o značenju replikacije. U radu koji opisuje hijerarhijski model uspješno oblikovanih replikacija su prikupljeni i sistematizirani neki mogući rezultati replikacija, što vodi ka različitim (ali preciznim) zapisima o uspjehu repliciranja. Stoga smatram da nam taj rad može uvelike pomoći kod oblikovanja replikacije, određivanja njezinih ciljeva kao i značenja uspješno oblikovane replikacije [BAYA1997].

Eksperiment je postupak kojim namjerno, u strogo kontroliranim uvjetima, izazivamo neku pojavu u svrhu opažanja ili/i mjerenja [BUJA1967]. Eksperiment je pokušaj oblikovan sa svrhom dokazivanja hipoteza koje su definirane ranije i iznesene u kontroliranim postavkama, u kojima se većina kritičkih faktora može kontrolirati ili pratiti [ALMQ2006]. Bayesov hijerarhijski model dopušta fleksibilno sjedinjenje pretpostavljene veze između eksperimenata u analizi. Bayesova predviđanja distribucije su prirodno oruđe za izračunavanje vjerojatnosti uspješne replikacije i stoga oblikuje replikaciju s većom vjerovatnošću uspjeha [BAYA1997].

Neki od temeljnih eksplicitnih ili implicitnih mogućih ciljeva replikacije [BAYA1997]:

**[g1] Redukcija slučajne pogreške** (eng. Reduction of random error). Za replikaciju eksperimenata najpotrebnije jednostavno skupljanje podataka kako bi se dobio što precizniji zaključak. U ovim situacijama replikacije su napravljene tako da se slažu što je moguće pažljivije s originalnim eksperimentom (ponekad se predstavlja kao "ispravna" replikacija) i zaključci su temeljeni na kombinaciji obje analize eksperimenta (originalna i replicirana).

**[g2] Potvrda zaključka** (eng. validation of conclusions). Ako je rezultat iz originalnog istraživanja znakovit ili iznenađujući, tada postaje interesantno vidjeti vjerojatnost da takvi rezultati imaju jedino mogućnosti za uspjeh. Situacija je slična onoj u [g1] osim da se replikacija obično analizira individualno (nema meta-analiza sa prijašnjim eksperimentima) i da "točno" slaganje uvjeta ne smatra presudnim.

**[g3] Proširivanje zaključka** (eng. extension of conclusions). U mnogo situacija, replikacije su napravljene tako da istražuju koliko vrijedi zaključak u originalnom eksperimentu kada počinju slabiti (ili mijenjati se) promjene u ko-varijablama (eng. covariable), eksperimentalnim uvjetima..., itd. Često su izvođene replikacije različitih populacija i interes leži u usporedbama kako se populacije mogu razlikovati s obzirom na karakteristike istraživanja.

**[g4] Uočavanje zabluda** (eng. bias detection). Zanimljiv cilj (iako ne jasno utemeljen) u nekim replikacijama je detekcija nekih zabluda za koje se pretpostavlja da su predstavljene u originalnom eksperimentu. U ovim slučajevima, replikacije su jako pažljivo oblikovane pa je logično pretpostaviti da su slobodne od pretpostavljenih sklonosti u originalnom eksperimentu.



### 2.1.1. Matematički model

Ovdje su promatrana dva eksperimenta, originalni  $\varepsilon_0$  i replikacija  $\varepsilon_r$ . Prva replikacija je opažena kao "kritična" (Lindsay i Ehrenberg, 1993 prema [BAYA1997]), njih je najteže oblikovati, podaci o vezi između eksperimenata su manjkavi i oblikovanje se teško oslanja na prethodnu informaciju (situacije koje uključuju više eksperimenata se trebaju promatrati na sličan način). Također, trebamo promatrati rezultate [g2] i [g3] koje su sa oblikovanim replikacijama određeni za vrednovanje originalnog zaključka ili za specijaliziranje heterogenosti između originalnog i repliciranog eksperimenta. Cilj [g1] je jednostavan, podrazumijeva se uzimanje podataka iz  $\varepsilon_0$  i  $\varepsilon_r$  pa tako oblikovanje  $\varepsilon_r$  smanjuje pronalaženje veličine uzorka replikacije,  $n_r$ , potrebne za dobiti propisanu točnost ili snagu.

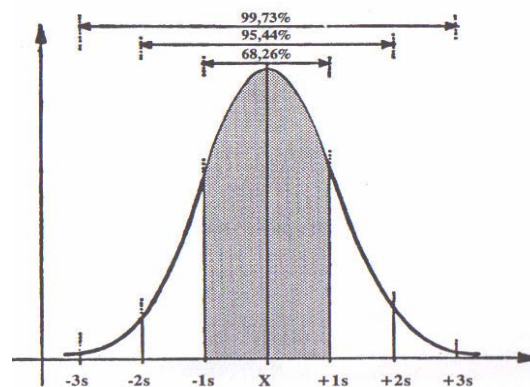
Kod oblikovanja i analize replikacije je korištena Bayesova metoda jer Bayesovo predviđanje distribucije povećava mogućnost da će replikacija biti "uspješna".

Bayesov hijerarhijski model ima tri nivoa. Prvi nivo hijerarhije će stvarati uvjetnu distribuciju promatrača (obično su dovoljne statističke vrijednosti, statistike testa, procjene ili možda p-vrijednosti, Blyth i Smith, 1993 prema [BAYA1997]), drugi nivo će stvarati vezu između dva eksperimenta, i treći nivo će odrediti distribuciju za hiper-parametre (eng. hyperparameters). Promatrana je replikacija standardnog eksperimenta u kojem je eksperimentalna grupa uspoređena sa kontrolom koristeći uobičajeni t-test.

U *originalnom eksperimentu*,  $\varepsilon_0$ , donji indeks E (C) identificira varijable za eksperimentalnu (kontrolnu) grupu, i treba odrediti kada je E učinkovit. Najčešća pretpostavka je da su glavni uzorci normalno distribuirani:

$$\bar{X}_E | \mu_E, \sigma^2 \approx N\left(\mu_E, \frac{\sigma^2}{n_E}\right), \bar{X}_C | \mu_C, \sigma^2 \approx N\left(\mu_C, \frac{\sigma^2}{n_C}\right) \quad (2.1)$$

gdje su  $n_E$  i  $n_C$  prihvaćene veličine uzorka. Normalna raspodjela je raspodjela kod koje često prilikom mjerenja neke pojave dobivamo rezultate koji pokazuju tendenciju grupiranja oko jedne središnje vrijednosti i tendenciju raspršivanja oko te središnje vrijednosti. Potpuno je određena s dva parametra: aritmetičkom sredinom koja određuje os simetrije i standardnom devijacijom podataka koja određuje širinu raspodjele i njen točan oblik.



Slika 2.1. Normalna krivulja

Krivulja koja prikazuje takvu raspodjelu naziva se normalna krivulja, a neki je po matematičaru Gaussu, koji ju je uz neke druge matematičare i definirao, nazivaju i Gaussovom krivuljom [PETZ2007]. Normalna krivulja je zvonastog oblika, unimodalna je i simetrična oko aritmetičke sredine što je prikazano na slici 2.1. Kod normalne raspodjele u intervalu od  $\bar{X} \pm 1s$  se nalazi 68,26 % svih rezultata, u intervalu  $\bar{X} \pm 2s$  95,44 % rezultata, a u intervalu  $\bar{X} \pm 3s$  99,73 % svih rezultata.

Obična analiza se testira ovako (nejednakost može biti obrnuta):

$$\begin{aligned} H_0 : \mu_E &\leq \mu_C \\ H_1 : \mu_E &> \mu_C \end{aligned} \quad (2.2)$$

Test je (i p-vrijednost je označena) poznatim t-testom:

$$T = \frac{\bar{X}_E - \bar{X}_C}{S \sqrt{\frac{1}{n_E} + \frac{1}{n_C}}} \quad (2.3)$$

gdje je  $S^2$  zajednički uzorak (eng. pooled sample) standardne devijacije,

$$S^2 = \frac{(n_E - 1)S_E^2 + (n_C - 1)S_C^2}{n_E + n_C - 2} \quad (2.4)$$

i  $S_E$ ,  $S_C$  su standardne devijacije eksperimentalne i kontrolne grupe. Također se izračunava veličina učinka (eng. effect size):

$$\delta = \frac{\mu_E - \mu_C}{\sigma} \quad (2.5)$$

Distribucija t-testa (2.3) ovisi o  $\mu_E$ ,  $\mu_C$  i  $\sigma$  samo između  $\delta$  i to je ne-središnji t (eng. non-central):

$$T|\delta \approx NCT(\theta, \nu) \quad (2.6)$$

s  $\nu = n_E + n_C - 2$  stupnjeva slobode i ne-središnjim parametrom  $\theta = \delta\sqrt{n^*}$ , gdje je

$$n^* = \left( \frac{1}{n_E} + \frac{1}{n_C} \right)^{-1} \quad (2.7)$$

(ako je  $n_E = n_C = n$  tada je  $n^* = n/2$ ). Naravno, kada je  $\mu_E = \mu_C$ , tada je  $\delta = 0 = \theta$ , i rezultirajuća distribucija (2.6) se tada smanjuje na poznatu Studentovu-t distribuciju sa  $\nu$  stupnjem slobode.

Ne-središnja t-raspodjela sa ne-središnjim parametrom  $\theta$  i  $\nu$  stupnjeva slobode je definirana razlomkom kako slijedi:

$$NCT(\theta, \nu) = \frac{z + \theta}{\chi_\nu^2 / \nu} \quad (2.8)$$

gdje su  $z$  i  $\chi^2_\nu$  nezavisne nasumično odabrane varijable raspodjeljene kako normalne i  $\chi^2$  sa  $\nu$  na stupnjeva slobode. Ako je  $\theta = 0$  distribucija se naziva Studentova  $t$  sa  $\nu$  stupnjeva slobode. Studentova  $t$ -raspodjela je slična normalnoj, tj. ona je također simetrična i ima zvonast oblik, ali je pri dnu šira od normalne raspodjele.

Za replikaciju određujemo sličan model. Uzimamo  $t_0, \delta_0, n_0^*, \nu_0$  i označavamo veličine definirane gore za originalni eksperiment,  $\varepsilon_0$ , i uzimamo  $t_r, \delta_r, n_r^*, \nu_r$  i označavamo kao odgovarajuće za *replirani eksperiment*,  $\varepsilon_r$ . Tada imamo prvi korak hijerarhije:

$$t_i | \delta_i \approx NCT\left(\delta_i \sqrt{n_i^*}, \nu_i\right), \quad i = 0, r \quad (2.9)$$

U drugom koraku smo izmjerili ovisnosti između  $\varepsilon_0$  i  $\varepsilon_r$  značenjem zajedničke distribucije  $p(\delta_0, \delta_r)$ . Zbog jednostavnosti je određena veza između  $\varepsilon_0$  i  $\varepsilon_r$  određujući  $\delta_0$  i  $\delta_r$  kao promjenjive, što je u povratku napravljeno određivanjem da su one uvjetno slučajne varijable. Obično se uzima normalna distribucija za njihove poznate distribucije. S obzirom na općenitost, ali osobito za postizanje rezultata, uzeta je Studentova- $t$  distribucija. Dakle, za drugi korak hijerarhije imamo:

$$\delta_i | \mu \approx St(\mu, \nu^2, f), \quad i = 0, r \quad (2.10)$$

U zadnjem koraku hijerarhije odredili smo prijašnje distribucije za hiper- parametre. Kao što vidimo, (2.9) i (2.10) koriste teži model. Srećom, oba tipa, ne-središnji i Student-  $t$ , se mogu lako izraziti kao kontinuirano miješanje normalnih distribucija sa inverznim-gama distribucijama. Neka je  $Y$  nasumično odabrana varijabla,  $Y > 0$  sa gama distribucijom i parametrima  $(a, b)$ . Tada vjerojatnost distribucije  $X = 1/Y$  zovemo inverzna-gama distribucija sa parametrima  $a$  i  $b$ . Kada koristimo ovo rastavljanje u (2.9) i (2.10) dobivamo hijerarhijski model koji ćemo koristiti:

$$\begin{aligned} \text{I.} \quad t_i | \delta_i, \sigma_i^2 &\approx N\left(\delta_i \sqrt{n_i^* \sigma_i^2}, \sigma_i^2\right), & i = 0, r \\ \text{II.} \quad \delta_i | \mu, \tau^2 &\approx N(\mu, \tau^2), & i = 0, r \\ \sigma_i^2 &\approx IGa(\nu_i / 2, \nu_i / 2), & i = 0, r \\ \text{III.} \quad p(\mu), \tau^2 &\approx IGa(a + 1, ak), \end{aligned} \quad (2.11)$$

gdje je  $\nu_i = n_E^i + n_C^i - 2$ , i  $n_i^* = \left[ (1/n_E^i) + (1/n_C^i) \right]^{-1}$ , za  $i = 0, r$ .

Distribucija za  $\sigma_i^2$  u koraku II. je potpuno specifična i ovisi samo o globalnoj veličini uzorka eksperimenta. Drugo, veza određena između originalne veličine učinka  $\delta_0$  i replicirane  $\delta_r$  je izmjerena kao normalna  $N(\mu, \tau^2)$  distribucija u koraku II. Tamo mora biti jaka prijašnja informacija koja uzima u obzir "super-populaciju" veličinu učinka  $\mu$  u određenom problemu. U koraku III. hiperparametar  $\tau^2$  izračunava koliko blizu ili daleko mogu biti  $\delta_0$  i  $\delta_r$ .

Parametar  $k$  je "pogođena" vrijednost za  $\tau^2$ , tako da mala vrijednost za  $k$  pokazuje da ne očekujemo da  $\delta_0$  i  $\delta_r$  budu puno dalje od ostalih.

Vrijednost  $a$  je pokazatelj snage našeg vjerovanja u izračunavanje distribucije. Nadalje,  $\text{Var}(\tau^2) = k^2 / (a - 1)$ , tako da je za fiksnu vrijednost  $k$ , veće  $a$  više koncentrirano na prijašnju distribuciju od  $\tau^2$ . Ova prijašnja distribucija je ekvivalentna za određivanje jednoličnog priora  $\mu$  u (2.10), dok su  $v^2$  i  $f$  određeni sa  $v^2 = ak / (a + 1)$  i  $f = 2(a + 1)$ .

Ovo je generalizacija poznate prakse u kojoj su  $\delta_0$  i  $\delta_r$  određene kao identične (ovo će biti ograničeni slučaj modela (2.11) kada  $\tau^2 \rightarrow 0$ , tako da  $\delta_0 = \delta_r = \mu$  i to će biti uobičajeni model za postizanje cilja [g1] u uvodu. Drugi ekstremni stav, zvanično onaj koji se odnosi na  $\varepsilon_0$  i  $\varepsilon_r$  kao totalno nepovezane, se može također vidjeti kao drugi ograničeni slučaj modela (2.11) kada je  $\tau^2 \rightarrow +\infty$ ).

Određivanje  $a$  i  $k$  ne može biti dopušteno. Ne možemo ništa znati o vezi između  $\varepsilon_0$  i  $\varepsilon_r$  prije nego je  $\varepsilon_r$  izvedena. Naravno, ako razmišljamo o replikaciji  $\varepsilon_0$ , koja je replicirana već nekoliko puta, i ako se naša replikacija smatra izmjenjivom sa izvedenima, tada možemo dalje koristiti te prošle replikacije da dopustimo ocjenjivanje  $a$  i  $k$ , (naravno, ovo možemo učiniti i ako se koristi model (2.11), za analizu  $\varepsilon_r$ , ali ne za svrhe oblikovanja).

Koristeći model (2.11), "točna" replikacija ima neke osobitosti koje možemo uključiti u model. Osobito:

- $\varepsilon_r$  će biti pažljivo oblikovana da bude što sličnija  $\varepsilon_0$ , što nas često navodi na pretpostavku da su veličine učinka identične ( $\delta_0 = \delta_r$ ). Ova tvrdnja je jako kritizirana (Lindsay i Ehrenberg, 1993 prema [BAYA1997]). Za "točnu" replikaciju pretpostaviti ćemo da su  $\delta_0$  i  $\delta_r$  jako slični, a zato je  $\tau^2$  u (2.11) formuli jako mala.
- $t_r$  je obično analiziran samo kako bi se vidjeli ili ne isti zaključci kao i u  $\varepsilon_0$ .

Sada ćemo objasniti oblikovanje replikacije u prethodnom modelu. Pretpostavimo da je test forma (2.6) korištena kod  $\varepsilon_0$  na  $\alpha$ -nivou, rezultirajući u značajnu vrijednost  $t_0$ . Tada je definicija uspjeha replikacije:  $\varepsilon_r$  je uspješna replikacija ako rezultira u  $t_r$  što je također značajno na  $\alpha$ -nivou.

Ostale mogućnosti "uspjeha" uključuju promatranje p-vrijednosti u  $\varepsilon_r$  barem da bude mala kao ona u  $\varepsilon_0$ , ili kombinacija oba eksperimenta i potraga za kombiniranim- $t$  koji će biti značajan na  $\alpha$ -nivou ili sa odgovarajuće malom p-vrijednosti.

## 2.1.2. Proučavanje heterogenosti

Događa se često da se neke karakteristike replicirane populacije razlikuju od onih iz  $\varepsilon_0$ , ili da su različiti eksperimentalni uvjeti. Što više, čest je slučaj da su replikacije eksplicitno oblikovane tako da su uvjeti u  $\varepsilon_0$  i  $\varepsilon_r$  prilično različiti, osnovni cilj replikacija tada postaje proširiti svrhu na

istraživanja  $\varepsilon_0$  ili proučavanje njihovih ograničenja. Ove različite replikacije su također jako važne u istraživanju (Lindsay i Ehrenberg, 1993 prema [BAYA1997]).

Kod različitih replikacija česta pomoć statističke analize je uspoređivanje obje populacije, u ovom scenariju, usporedba obje veličine učinka  $\delta_0$  i  $\delta_r$ . Također, u ovim slučajevima se moraju koristiti sve valjane informacije i  $\varepsilon_r$  neće biti analizirana sama. U pogledu specijalnih karakteristika različitih replikacija mogu se identificirati dvije prirodne mogućnosti definiranja što je to uspješna replikacija:

**D1:**  $\varepsilon_r$  je uspješna ako možemo napraviti dovoljno točan zaključak o razlikama veličinama učinka  $\delta_{0r} = \delta_r - \delta_0$ ;

**D2:**  $\varepsilon_r$  je uspješna ako možemo riješiti test  $H_0 : \delta_0 = \delta_r$  nasuprot  $H_1 : \delta_0 \neq \delta_r$ , sa dovoljno uvjerljivosti, jedan ili drugi način.

Specifično, pretpostavljamo da "uspjeh" znači "uspješno smanjivanje varijacije  $\delta_{0r} = \delta_r - \delta_0$ ", tako da je prirodan način za izračunati D1:

$$\Pr(\text{uspjeh}) = \Pr[\text{Var}(\delta_{0r}|t_0, t_r) \text{ je dovoljno mala}] \quad (2.12)$$

Kada je cilj repliciranja (D1), treba točno odrediti  $\delta_{0r} = \delta_r - \delta_0$  tada je replikacija  $\varepsilon_r$  uspješna kada je

$$\text{Var}(\delta_{0r}|t_0, t_r) \leq S \quad (2.13)$$

gdje je S specifični početak.

Za definiciju D2 trebamo pretpostaviti da je testiranje riješeno u terminima Bayesovih faktora pa je način izračunavanja vjerojatnosti uspjeha  $\varepsilon_r$

$$\Pr(\text{uspjeh}) = \Pr[\text{jedan od Bayesovih faktora, jedan u korist } H_1, \text{ ili u korist } H_0, \text{ je dovoljno velik}] \quad (2.14)$$

Kada treba izračunati vjerojatnosti uspjeha kada je cilj riješiti test  $H_0 : \delta_{0r} = 0$  nasuprot  $H_1 : \delta_{0r} \neq 0$ , sa  $\delta_{0r} = \delta_r - \delta_0$ . U ovom slučaju, zajedno sa (2.14)  $\varepsilon_r$  je uspješna replikacija ako

$$B_{01}(tr) > B \text{ ili } B_{10}(tr) > B \quad (2.15)$$

gdje je B specifičan početak. Dok je  $B_{10} = 1/B_{01}$ , (2.15) je ekvivalentno za dobivanje Bayesovog faktora  $B_{01}$ :

$$B_{01} > B \text{ ili } B_{01} < 1/B \quad (2.16)$$

Vjerojatnost uspjeha sada je na dva kraja područja između  $1/B$  i iznad B.

### 2.1.2.1. Bayesovi faktori

Na početku sa danim podacima D pretpostavljamo da imamo postavljene dvije hipoteze  $H_0$  i  $H_1$  uključujući gustoću vjerojatnosti  $pr(D|H_0)$  ili  $pr(D|H_1)$ . Dana je prethodna vjerojatnost (eng. a priori probabilities)  $pr(H_0)$  i vjerojatnost  $pr(H_1)=1- pr(H_0)$ , te iz podataka dobijemo kasniju vjerojatnost (eng. a posteriori probabilities)  $pr(H_0|D)$  i  $pr(H_1| D)=1- pr(H_0|D)$ . Kada je uključeno više hipoteza za izračun Bayesovog faktora onda ga označavamo  $B_{jk}$  za hipotezu  $H_j$  nasuprot hipotezi  $H_k$ . Bayesov faktor računamo na sljedeći način:

$$\frac{pr(H_1|D)}{pr(H_2|D)} = \frac{pr(D|H_1)pr(H_1)}{pr(D|H_2)pr(H_2)}, \text{ gdje je } B_{01} = \frac{pr(D|H_0)}{pr(D|H_1)} \quad (2.17)$$

Riječima opisano, kasnija vjerojatnost (eng. a posteriori probabilities) = Bayesov faktor \* prethodna vjerojatnost (eng. a priori probabilities).

U našem slučaju jedna je hipoteza označena kao nul-hipoteza  $H_0$ , dok je druga kao alternativna hipoteza  $H_1$ , stavljajući vjerojatnost  $pr(D|H_0)$  u brojnik, Bayesov faktor označavamo  $B_{01}$ . Kada uspoređujemo rezultate sa standardnom vjerojatnošću odnosa testova (eng. standard likelihood ratio tests), pogodno je staviti nul-hipotezu u nazivnik pa dobijmo Bayesov faktor  $B_{10}$ . Postoje razne interpretacije Bayesovih faktora. Jeffreys prema [KASS1993] je predložio interpretaciju Bayesovog faktora kao skalu od pola jedinice računajući  $\log_{10}(B^{-1})$  što je prikazano u tablici 2.1.

**Tablica 2.1. Interpretacija Bayesovog faktora računajući  $\log_{10}(B^{-1})$**

$\log_{10}(B^{-1})$	dokazi prema $H_1$
0-0,5	nema vrijednosti (eng. not worth more than a bare mention)
0,5-1	značajan (eng. substantial )
1-2	snažan (eng. strong)
>2	Odlučan (eng. decisive )

### 2.1.3. Primjer

Prikazat ćemo pojednostavljenu verziju klasičnog eksperimenta na učinku uzimanja kalcija na krvni tlak (Lyle i ostali, 1987, Moore i McMabe, 1993 [BAYA1997]).U tom eksperimentu, 21 muškarac crne puti je odabran slučajno da uzima veće doze kalcija ili placebo u periodu od 12 tjedana. Deset muškaraca je primalo kalcij, ostali su bili kontrola. Krvni tlak je izmjeren svima na početku i na kraju eksperimenta, i razlika  $\bar{X}$  (početni minus završni tlak) je zabilježena što je prikazano u tablici 2.2. Podaci kažu sljedeće, vrijednosti t-statistike i p-vrijednosti su bile:  $t=1.634$ ,  $p=0.059$ , osiguravajući blagi dokaz protiv nulte hipoteze, ( $\mu_E > \mu_C$ ).

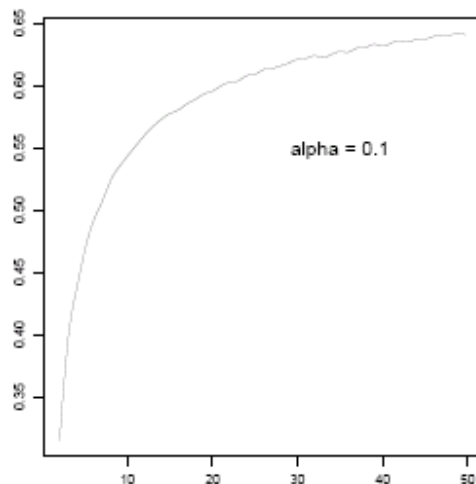
**Tablica 2.2. Originalan eksperiment ( $\epsilon_0$ )**

Grupa	Postupak	n	$\bar{x}$	s
Eksperimentalna	kalcij	10	5.0	8.743
Kontrolna	placebo	11	-0.273	5.901

Ovaj eksperiment je odabran ne samo zbog toga što se često ponavljao u literaturi već i zato što daje mogućnosti za replikacije: izvođači eksperimenata možda žele smanjiti p-vrijednost (cilj [g1]), vrednujući eksperiment (cilj [g2]), studija učinka kalcija kod drugih populacija je uspoređivana sa učinkom kalcija kad su nekih eksperimentalni uvjeti (doza kalcija, osobine individualca,...) promijenjeni (cilj [g3]). U kasnijim replikacijama je pokazano da se učinak kalcija više odrazio na individualce crne puti nego bijele.

Prior koji je korišten u modelu (2.11) ima  $k = E(\tau^2) = 1/2$  dajući nam tako stav da ne očekujemo da će se veličina učinka puno razlikovati, ali još uvijek će biti nekako različita (varianca podataka kada je  $\delta = 0$  je  $Var(t_0|\delta = 0) \approx 1.12$ ). Uzeto je da je  $a = 2$  što je najmanja vrijednost za koju je  $Var(\tau^2)$  definiran. Ovo odgovara 6 stupnju slobode za Student-t distribuciju na  $\delta$ .

Jednostavnije, pretpostavimo da su  $n_E^r$  i  $n_C^r$  jednaki u repliciranom eksperimentu, i označimo poznatu vrijednost jednostavno sa  $n_r$ . U originalnom eksperimentu je  $n_E^o = 10$  i  $n_C^o = 11$  tako da je pretpostavka jednake veličine uzorka u replikaciji sasvim prirodna. U originalnom eksperimentu dobivena vrijednost  $p_0 = 0.059$  što rezultira značajno samo na nivou  $\alpha = 0.1$ . Stoga treba pretpostaviti da "uspjeh" u ovom slučaju znači provođenje  $t_r$  značajno na nivou  $\alpha = 0.1$  (za "uspjeh" na nivou  $\alpha = 0.05$ , i  $p_r$ -vrijednost treba biti znatno manja nego promatrana  $p_0$ , Mayoral, 1998 prema [BAYA1997]). Na slici 2.2 se vidi mogućnost "uspjeha" koji je odbijanje nulte hipoteze u  $\varepsilon_r$  kao funkcija  $n_r$ . Kao što se može vidjeti na slici 2.2, vjerojatnost replikacije originalnog eksperimenta (koji je odbijanje hipoteze "bez učinka" kod nadopune kalcija i smanjivanja krvnog tlaka) se povećava s brojem promatranja  $n_r$  u replikaciji. No, vjerojatnost uspjeha nije velika (za  $n_r = 50$  je samo 0.65).



Slika 2.2. Reprodukcijsko odbacivanje u zatvorenoj replikaciji

U tablici 2.3 je prikazana vjerojatnost uspjeha slučaja D1 koristeći formulu (2.13) za tri različite vrijednosti početka S. Kao i u prethodnim slučajevima uspjeh replikacije se ne očekuje sa malom veličinom uzorka. Kako god, u suprotnosti s istraženim ciljem, vidimo da vjerojatnost uspjeha ne raste brzo.

**Tablica 2.3. Vjerojatnost uspjeha za slučaj D1**

$n_r$	5	10	20	30	50	100
S = 0.3	0.000	0.000	0.371	0.7095	0.9035	0.9885
S = 0.4	0.000	0.4305	0.8495	0.9450	0.9865	1.000
S = 0.5	0.000	0.7450	0.9425	0.9835	0.9985	1.000

Promatrajući slučaj D2 u tablici 2.4 je osigurana vjerojatnost uspjeha za različite vrijednosti B i veličina uzoraka replikacije  $n_r$ .

**Tablica 2.4. Vjerojatnost uspjeha za slučaj D2**

$n_r$	5	10	20	30	50	100
<b>B = 2</b>	0.6485	0.746	0.7925	0.8255	0.8325	0.8415
<b>B = 3</b>	0.387	0.458	0.633	0.664	0.703	0.7185
<b>B = 4</b>	0.3475	0.416	0.488	0.5035	0.516	0.5465

Nije iznenađujuće (dobivena velika količina nepouzdanosti prisutne u modelima) da porast  $n_r$ , samo slabo povećava vjerojatnost uspjeha ako pogledamo na prilično velike ili male vrijednosti Bayesovog faktora. Analiza nam govori da nema dovoljno informacija za predviđanje definitivnih rezultata.

## 2.2. Oblikovanje eksperimenta i replikacije

Kod oblikovanja eksperimenta i njegove replikacije promatrat ćemo model eksperimenta sa paralelnim grupama. To znači da je svaka grupa nosilac svog eksperimentalnog faktora. Primjerice, kada se kontrolna grupa poučava na tradicionalni način, a ekperimentalna grupa koristi neki drugi, pristup poučavanja.

### 2.2.1. Cilj istraživanja

Na početku i u osnovi svakog eksperimenta nalazi se na neki problem, tj. određeno pitanje na koje ne možemo pouzdano odgovoriti. Izvori problema su mnogostruki. Najprije sam život nameće mnoštvo neuravnoteženih situacija, koje treba uravnotežiti da bi se snašli u tim situacijama.

#### **Cilj eksperimenta**

Kod pojedinih istraživanja najčešće se problem istraživanja ne može istražiti u cijelosti. Zato se određivanjem cilja ili ciljeva eksperimenta treba precizno navesti koji će aspekt ili aspekti problema biti obuhvaćeni istraživanjem. To je potrebno da bi se određeni znanstveni problem suzio i da bi se zatim mogle definirati hipoteze koje proizlaze iz ciljeva istraživanja. Osnovni cilj eksperimenta jest utvrditi postoji li, koliki je i kakav je utjecaj određene nezavisne varijable na određenu zavisnu varijablu.



### **Cilj replikacije**

Ciljevi replikacije mogu biti različiti. Neki od temeljnih eksplicitnih ili implicitnih mogućih ciljeva replikacije su:

- [g1] **Redukcija slučajne pogreške** (eng. Reduction of random error).
- [g2] **Potvrda zaključka** (eng. validation of conclusions).
- [g3] **Proširivanje zaključka** (eng. extension of conclusions).
- [g4] **Uočavanje zabluda** (eng. bias detection).

### **2.2.2. Sudionici**

Od vrste i cilja eksperimenta zavisi na kakvim će se ispitanicima izvršiti ispitivanje. U eksperiment po pravilu treba uključiti ispitanike koji su dovoljno homogeni s obzirom na poznate relevantne faktore, tj. takve faktore koje bi mimo nezavisne varijable mogli na značajni način utjecati na pojavu koja se ispituje. Ako je namjera eksperimenta ne samo da provjeri vezu između varijabli na konkretnoj skupini ispitanika nego da i dobivene rezultate generalizira na određenu populaciju tada sudionici treba da predstavljaju reprezentativni uzorak takve populacije. Kontrolu u eksperimentu čine postupci kojima se detektira, pa zatim, eliminira ili/i stabilizira utjecaj relevantnih faktora u eksperimentu. Irelevantne faktore zanemarujemo.

Kada se nastoji održati relevantne faktore pod kontrolom, temeljno je pravilo da ispitanici budu *izjednačeni* po svim relevantnim faktorima izuzev onoga kojeg mjerimo. Na taj način imamo logički temelj za tvrdnju da je promjena u zavisnoj varijabli izazvana isključivo promjenama nezavisnoj. Jedan od načina na koji to možemo obaviti je uvođenje kontrolne i eksperimentalne skupine. Prvu čine ispitanici na koje nije primijenjena nezavisna varijabla i to je kontrolna grupa. Oni čine osnovicu na temelju koje imamo pravo govoriti o veličini utjecaja nezavisne varijable. Druga skupina je eksperimentalna, i na nju se nezavisna varijabla primjenjuje, te se mjeri učinak primjene.

Eksperimentalna i kontrolna skupina trebaju biti izjednačene u svim bitnim obilježjima, a to znači u zavisnoj varijabli i svim onim obilježjima koja bi mogla utjecati na zavisnu varijablu pokraj nezavisne varijable. Treba odabrati sudionike eksperimentalne i kontrolne grupe tako da grupe budu ekvivalentne. Ekvivalentnost ili izjednačavanje grupa vrši se (Backer L.A 2000 prema [GRUB2007]):

- *određivanjem parova ispitanika*; parovi se odrede na način da u paru budu ispitanici koji su najsličniji u zavisnoj varijabli i onim obilježjima koji bi mogli utjecati na zavisnu varijablu; nakon toga ispitanici se raspoređuju u eksperimentalnu i kontrolnu skupinu tako da jedan član para ulazi u eksperimentalnu skupinu, a drugi u kontrolnu skupinu; ispitanici koji previše odskaču iz neke distribucije ili koji se ne mogu spariti, izbacuju se iz eksperimenta
- *izjednačavanjem skupina u nekim poznatim relevantnim faktorima*; već prema vrsti eksperimenta to izjednačavanje vrši se prema nekim vanjskim karakteristikama (spol, dob, razina obrazovanja, itd.) ili/ i na osnovu inicijalnog testa; obično se promatraju mjera aritmetičke sredine kao mjera centralne tendencije i mjera standardne devijacije kao mjera disperzije; statistička značajnost se provjerava uporabom t-testa.

Postoje dva načina izjednačavanja parova ispitanika:

**1. Precizno izjednačavanje** (eng. exact matching)

Nakon što smo sortirali ispitanike prema rezultatima inicijalnog testa određuju se parovi ispitanika s jednakim rezultatima inicijalnog testa. Možemo krenuti od kontrolne grupe i tražiti odgovarajuće ispitanike u eksperimentalnoj grupi, i obratno. Ako postoji više ispitanika koji predstavljaju moguće uparivanje za odabranog ispitanika, onda se primjenjuje metoda slučajnog odabira.

**2. Izjednačavanje u rasponu** (eng. caliper matching)

Nakon što smo sortirali ispitanike prema rezultatima inicijalnog testa treba definirati raspon bodova koji dolaze u obzir kod uparivanja ispitanika prema rezultatima inicijalnog testa. Zatim se traže parovi ispitanika s rezultatima inicijalnog testa koji su u definiranom rasponu bodova  $\pm r$ . Možemo krenuti od kontrolne grupe i tražiti odgovarajuće ispitanike u eksperimentalnoj grupi, i obratno. Prvo se traže ispitanici s jednakom brojem bodova, zatim ako takav ne postoji, traži se ispitanik kojem je rezultat inicijalnog testa u rasponu  $\pm r$ , i tako dalje, povećava se raspon bodova ili dok ne se nađe odgovarajući ispitanik ili dok ne se dostigne definirani raspon. Ukoliko postoji više ispitanika koji predstavljaju moguće uparivanje za odabranog ispitanika, pristupa se metodi slučajnog odabira.

Isti postupak se provodi i kod odabiranja ispitanika za replicirani eksperiment vodeći računa da eksperimentalna i kontrolna grupa bude izjednačena, tj da grupe budu ekvivalentne.

Nakon što se provelo izjednačavanje parova ispitanika prema rezultatima inicijalnog testa statistički se provjerava da li su eksperimentalna i kontrolna grupa ekvivalentna. Najprije treba izračunati aritmetičku sredinu rezultata inicijalnog testa za eksperimentalnu i kontrolnu grupu po formuli [PETZ2004] :

$$\bar{X}_C = \frac{\sum X}{N_C} \qquad \bar{X}_E = \frac{\sum X}{N_E} \qquad (2.18)$$

Nakon toga se izračuna standardna devijacija za svaku grupu po formuli:

$$\sigma_C = \sqrt{\frac{\sum (X - \bar{X}_C)^2}{N_C - 1}} \qquad \sigma_E = \sqrt{\frac{\sum (X - \bar{X}_E)^2}{N_E - 1}} \qquad (2.19)$$

Da bismo usporedili da li je razlika između aritmetičkih sredina značajna treba izračunati i standardnu pogrešku razlika aritmetičkih sredina koja se računa po formuli:

$$S_{\bar{X}_C - \bar{X}_E} = \sqrt{\frac{\sigma_C^2}{N_C} + \frac{\sigma_E^2}{N_E}} \qquad (2.20)$$

Ako kažemo da je neka razlika statistički značajna, onda smo ustvrdili da razlika koja je nađena (bez obzira na veličinu razlike!) nije slučajna, već da razlika vrlo vjerojatno postoji među populacijama. Naprotiv, ako ustvrdimo da neka razlika nije statistički značajna, to drugim riječima znači da razlika koju smo prilikom mjerenja dobili, može biti i slučajna posljedica variranja uzoraka, a da među populacijama, kojima ti uzorci pripadaju, možda i nema nikakve razlike.

Ako je neka razlika između dvije aritmetičke sredine barem dva puta (točnije 1,96 puta) veća od svoje pogreške, onda se može smatrati statistički značajnom, jer je vrlo malo vjerojatno da će se tako velika razlika dogoditi slučajno. Obično se uzima razina značajnosti od 5%, što znači: ako zapravo ne postoji nikakva razlika između dviju aritmetičkih sredina, onda bi se takva konkretna razlika koju smo dobili, mogla slučajno dogoditi samo pet puta u 100 mjerenja, a to je malo vjerojatno, pa zato možemo uzeti da je razlika statistički značajna. Prema tome, razina slučajnosti od 5% znači zapravo šansu od 5% da smo pogriješili. Koliko je puta neka razlika veća od svoje pogreške možemo ustanoviti tako da ćemo razliku podijeliti njezinom pogreškom, tj računanjem *t*- vrijednosti:

$$t = \frac{\bar{X}_C - \bar{X}_E}{s_{x_C - x_E}} \quad (2.21)$$

Ukoliko *t*-vrijednost smatra statistički značajnom, to znači da se prethodno definirana eksperimentalna i kontrolna grupa statistički značajno razlikuje, onda ih je potrebno ponovo presložiti metodom slučajnog odabira i izjednačiti u rasponu, pa opet provjeriti ekvivalentnost grupa.

Značajnost razlike između aritmetičkih sredina malih uzoraka računa se po drugačijem postupku. Princip je jednak kao kod velikih uzoraka, ali se računaska operacija razlikuje utoliko što, pod pretpostavkom da oba uzorka potječu iz iste populacije, izračunavamo zajedničku standardnu devijaciju za oba uzorka.

Ako se te grupe razlikuju u veličini, onda treba ispitati statističku značajnost razlike između aritmetičkih sredina malih uzoraka. Zajedničku standardnu devijaciju smijemo izračunati samo onda ako se standardne devijacije samih grupa zaista značajno ne razlikuju, pa to treba najprije provjeriti. Značajnost razlike među standardnim devijacijama malih uzoraka izračunava se pomoću tzv. F-testa, koji stavlja u omjer veću varijancu prema manjoj varijanci (varijanca =  $s^2$ ) po formuli:

$$F = \frac{\text{veće } \sigma^2}{\text{manje } \sigma^2} \quad (2.22)$$

Prema [PETZ2004], ako su i kontrolna i eksperimentalna grupa jednako velike, nije ni potrebno testirati F-testom da li im se varijance statistički značajno razlikuju ili ne, jer će se, ako se one razlikuju, pogreška u računu biti neznatna. Dakle, ako su grupe jednako velike ili ako im se varijance ne razlikuju značajno, onda smijemo pristupiti izračunavanju zajedničke standardne devijacije koja se izračunava po formuli:

$$\sigma_{\text{zajednicka}} = \sqrt{\frac{(N_C - 1)\sigma_C^2 + (N_E - 1)\sigma_E^2}{N_C + N_E}} \quad (2.23)$$

### 2.2.3 Provedba

Oblikovanja eksperimenta i njegove replikacije se provodi prema određenom planu. Nakon kratkog uvoda u kojem je objašnjena svrha eksperimenta i njegova organizacijska pravila, od ispitanika se prikupljaju osobni podaci i informacije o dosadašnjoj upoznatosti s područnim znanjem putem upitnika. Skupljanje podataka je još jedan bitan faktor koji istraživači moraju uzeti u obzir kada provode eksperiment: ako se podaci ne skupljaju korištenjem preciznih procedura ili ako važne mjere nisu sakupljene, analize podataka i rezultata su izložene kritikama. Postoje razni mjerni instrumenti za prikupljanje podataka, a to su: test, upitnik ličnosti, skala procjene i anketni upitnik.

Nakon provedbe inicijalnog testiranja ispitanici se raspoređuju u jednu od skupina eksperimentalnu ili kontrolnu tako da grupe budu ekvivalentne.

#### **Testovi**

Test je standardizirani mjerni postupak kojim se izazivaju određena ponašanja radi njihovog mjerenja i opažanja [MEJO2003]. Da bi se test koristio u eksperimentu kao mjerni instrument mora imati sljedeće karakteristike: pouzdanost, valjanost, objektivnost, osjetljivost i baždarenost.

*Pouzdanost* je osobina testa da u ponovljenim mjerenjima daje iste ili slične rezultate. Što je pouzdanost veća, veća je nezavisnost testa od nesistemskih izvora pogreške. Pouzdanost i valjanost su međusobno povezane. Pouzdanost je nužan uvjet valjanost testa: nemoguće je znati mjeri li test ono što se tvrdi da se mjeri ako su rezultati dobiveni tim testom nepouzdan. Što je neki test pouzdaniji, to veća može biti njegova valjanost. Međutim, pouzdanost sama po sebi ne može jamčiti valjanost.

*Valjanost* je metrijska karakteristika nekog mjernog postupka (npr. testa, upitnika ličnosti, skale sudova) koja nam pokazuje da li on mjeri i u kojem stupnju mjeri upravo ono što smatramo da mjeri. Može se reći da postoje dva osnovna tipa valjanosti, a to su teorijska valjanost i praktična valjanost.

*Objektivnost* je karakteristika testa koja omogućava nezavisnost rezultata testa od pristranosti ispitivača i njegove subjektivnosti. Određuje se koreliranjem rezultata koje su dobili različiti ispitivači primjenjujući isti mjerni postupak na istim ispitanicima. Iako se obično smatra posebnom metrijskom karakteristikom, objektivnost se može smatrati jednim aspektom pouzdanosti mjerenja.

*Osjetljivost* testa je karakteristika testa da razlikuje ispitanike prema sposobnosti koju mjeri. Što je test osjetljiviji moguće je finije međusobno razlikovanje ispitanika na temelju rezultata koje postižu u testu.

Pod pojmom *baždarenja* u prirodnim znanostima podrazumjeva se provjera ispravnosti mjernog instrumenta. U društvenim i humanističkim znanostima pojam baždarenja se odnosi na utvrđivanje referencijskih standarda pomoću kojih će se vrednovati uradak ispitanika. Ti standardi su aritmetička sredina i standardna devijacija, a nazivaju se norme mjernog instrumenta. Baždarenost testa znači da su rezultati testa usporedivi s normama, odnosno da je moguće odrediti kakav je rezultat, u odnosu na populaciju, postigao pojedinac.

### ***Provedba inicijalnog testa (pred-test)***

Na početku eksperimenta potrebno je provesti inicijalno testiranje (ili pred- test) zbog toga što ispitanici koji sudjeluju u eksperimentu imaju različito znanje i vještine. Zbog toga se na početku mora utvrditi njihovo poznavanje i razumjevanje određenog područja da bi se na kraju eksperimenta mogle odrediti veličine njihovih promjena. Provedbom inicijalnog testa ćemo prikupiti informacije o postojanju statistički značajnih razlika u predznanju ispitanika. Nakon toga se određuju ispitanici za eksperimentalnu i kontrolnu grupu tako da grupe budu ekvivalentne.

### ***Provedba testova provjere i završnog testa (post-test)***

Svi testovi koji se provode u eksperimentu moraju provjeravati poznavanje istog područja, isto tako moraju biti bodovani prema istoj ljestvici da bi se lakše usporedili rezultati sa inicijalnim testom i drugim testovima. Uobičajeno je koristiti ljestvicu od 0- 100 bodova. Pitanja koja se koriste u inicijalnom testu mogu se ponavljati i u drugim testovima, ali to nije uobičajena praksa. Koliko će biti testova za provjeru znanja između inicijalnog i završnog testa ovisi o samom trajanju eksperimenta. Uobičajeno je da trajanje inicijalnog testa kao i ostalih testova bude jednako. Za razliku od inicijalnog test spomenuti testovi nam omogućavaju prikupljanje informacija o postojanju statistički značajnih razlika među grupama (eksperimentalnom i kontrolnom) tijekom postupka.

## **2.3. Analiza podataka**

Cilj analize podataka statističkim i statističko-matematičkim metodama jest sređivanje i osmišljavanje početnih podataka na način da se dobiju poruke koje podaci nose u sebi, a koje nisu vidljive na prvi pogled. Najčešće, broj podataka je toliki da se na temelju pregleda ne mogu izvući zaključci u pogledu hipoteza koje se istraživanjem provjeravaju. Za analizu podataka je bitno da polazni podaci budu sređeni, adekvatno organizirani i upisani. Sljedeće bitno je odabir odgovarajuće metode analize podataka da bude u skladu s problemom koji se istražuje. To znači da treba odabrati onu metodu analize podataka koja će najviše "izvući " iz podataka i omogućiti da se precizno odgovori na hipotezu ili hipoteze koje se istraživanjem provjeravaju. Analiza eksperimentalnih podataka treba biti više nego aplikacija statističkog testa , treba također biti pokušaj za objašnjavanje varijabilnosti subjekata i izdvojenih podataka [DALY1996].

U analizi podataka se provodi testiranje standardne značajnosti (eng. standard significance) za istraživanje učinka eksperimentalnog faktora na zavisnoj varijabli odnosno zavisnim varijablama ako ih je više (za zavisnu varijablu se obično uzima znanje učenika ili razni aspekti znanja).

Snaga statističkog testa ovisi o tri važne komponente (Sawyer i Ball prema [DALY1996]):

- nivo značajnosti  $\alpha$  (eng. significance level  $\alpha$  )
- veličini učinka (eng. effect size), o kojoj će biti govora kasnije
- broju ispitanika

Veći broj ispitanika, manje pogreška i veća preciznost pridonosi većoj snazi testa. Obično je ustaljena praksa za određivanje razine značajnosti  $\alpha = 0.05$ . Kako statistički testovi ne

osiguravaju uvid u snagu veze ili učinak interesa, također je poželjno povezivati testove s indeksima veličine učinka (eng. effect size).

### 2.3.1. Definiranje hipoteza

Hipoteze proizlaze iz problema koji želimo istraživati, odnosno iz hipotetičke teorije koju provjeravamo. "Hipotezu možemo definirati kao provizorni, prema mišljenju onoga koji je čini, jedan od mogućih odgovora na problem, ali koji odgovor tek daljnje istraživanje treba potvrditi ili odbaciti" [BUJA1967]. "Hipoteza nije drugo doli jedno pretpostavljeno objašnjenje koje izražavamo u obliku suda (pozitivnog ili negativnog), a kojeg moramo tek provjeriti" (Supek, 1981 prema [MEJO2003]). Hipoteze moraju biti precizne i jasne, a to se najbolje postiže ako se formuliraju na jednostavan način.

Postoje dvije vrste hipoteza:

1. nul-hipoteza ( $H_0$ )
2. radna ili direktivna hipoteza ( $H_1$ )

Nul- hipoteza formulira se niječno. Npr. ne postoji povezanost između pojava A i B. Radna hipoteza formulira se afirmativno. Npr. postoji pozitivna povezanost između pojava A i B.

Odluke:

- a) ako je p vrijednost manja od  $\alpha$ , prihvaća se  $H_1$
- b) ako je p vrijednost veća od  $\alpha$ , prihvaća se  $H_0$

Prilikom testiranja nul- hipoteze mogu se pojaviti dva tipa greške (Tablica 2.5.):

**Tablica 2.5. Moguće greške kod testiranja nul- hipoteze**

ODLUKA	NEMA razlike između dvije aritmetičke sredine	POSTOJI razlika između dvije aritmetičke sredine
ODBACUJEMO nul-hipotezu	<b>POGREŠKA TIPA 1 (<math>\alpha</math>)</b>	nema pogreške
PRIHVACAMO nul-hipotezu	nema pogreške	<b>POGREŠKA TIPA 2 (<math>\beta</math>)</b>

Tip pogreške 1 je da se odbaci nul-hipoteza kada je ona točna, odnosno istinita, a tip pogreške 2 da se prihvati nul- hipoteza kada ona nije točna i kada bi je trebalo odbaciti. Kada je nul- hipoteza istinita, a rezultati našeg testa značajnosti vode nas do zaključka da je pogrešna, kažemo da se radi o pogrešci tipa 1. Kada je nul- hipoteza pogrešna, a mi zaključimo na temelju testa značajnosti da je ne treba odbaciti, to opisujemo pogreškom tipa 2. (Edwards, 1968 prema [MEJO2003]). Problem je u razini statističke značajnosti koju istraživač odabire ta testiranje nul-hipoteze. Opasnost od prvog tipa pogreške prijeti kada je odabrana razina značajnosti vrlo blaga (npr.  $p \leq 0.10$ ), a opasnost od drugog tipa pogreške kada je odabrana razina statističke značajnosti vrlo stroga ( $p \leq 0.001$ ). U društvenim i humanističkim znanostima standardne razine statističke značajnosti su:  $p \leq 0.01$  i  $p \leq 0.05$ . Kada se testiranje nul- hipoteze vrši na tim razinama statističke značajnosti, opasnost od pogreški u zaključivanju tipa 1 i tipa 2 nije velika.

### **Definiranje nul-hipoteza za eksperiment i replikaciju**

Kod eksperimentata i njihovih replikacija koji prate predloženu metodologiju definira se  $n$  hipoteza za testove provjere znanja i jedna hipoteza za završni test. Prilikom ispitivanja statističke značajnosti odgovarajućeg testa te se nul-hipoteze prihvaćaju ili odbacuju. Za svaki test (ako ga ima) i završni test definira se nul-hipoteza na sljedeći način:

$H_0$ : "Ne postoji statistički značajna razlika između kontrolne i eksperimentalne grupe kod rezultata  $i$ -tog testa."

gdje je  $i$   $n$ -ti test provjere ili završni test ( $i = 1, \dots, n+1$ ).

Za testiranje nul-hipoteze se uobičajeno koristi t-test (jednosmjerni t-test ili dvosmjerni t-test). Preduvjet za korištenje t-testa je pretpostavka da su varijable normalno distribuirane.

### **2.3.2. Veličina učinka**

Za razliku od testova značajnosti koji govore o vjerojatnosti kojom se rezultati na testovima eksperimentalne grupe razlikuju od slučajnih rezultata, veličina učinka, kao numerička vrijednost, otkriva da li je eksperimentalni faktor koji se istražuje učinkovit i kolika je točno ta učinkovitost. Veličina učinka (eng. effect size) je stupanj na kojem je fenomen studije prezentiran u populaciji. Ako su svi ostali faktori konstantni tada je veća veličina učinka, veća vjerojatnost da će učinak biti detektiran i nulta hipoteza odbijena [DALY1996].

Dobivena pozitivna vrijednost pokazuje učinkovitost eksperimentalne grupe, dok negativna vrijednost veličine učinka pokazuje suprotno, odnosno, u provedenom eksperimentu kontrolna grupa je ostvarila bolje rezultate.

Uobičajeno je koristiti veličinu učinka kod usporedbe dva eksperimenta, odnosno eksperimenta i njegove replikacije. Postoje razni načini za izračun veličine učinka. Postoje četiri vrste veličina učinka: razlika aritmetičkih sredina (eng. mean difference), standardna razlika aritmetičkih sredina (eng. standardized mean difference) ili Glass-ova  $\Delta$ , Cohen-ov  $d$ , Hedges-ov  $g$ . Najpoznatija je svakako Glass-ova  $\Delta$ , dok je Cohen-ova veličina učinka gotovo postala standardnom zbog toga što je Cohen je definirao granice vrijednosti veličine učinka, kod istraživanja s dvije grupe pa ćemo njih dvije i pobliže opisati.

#### **Standardna razlika aritmetičkih sredina – Glass $\Delta$**

Da bi se usporedile veličine učinaka iz različitih istraživanja, Glass je 1976. godine predložio uvođenje standardizirane veličine učinka koja se naziva Glass-ova  $\Delta$  (Maxwell, S. E. i Delaney, H. D. 2004 prema [GRUB2007]). Standardna razlika aritmetičkih sredina se računa na način da se razlika između aritmetičkih sredina eksperimentalne i kontrolne grupe podijeli sa standardnom devijacijom kontrolne grupe. Računa se po sljedećoj formuli:

$$\Delta = \frac{\bar{X}_E - \bar{X}_C}{\sigma_C} \quad (2.24)$$

gdje su  $\bar{X}_E$  i  $\bar{X}_C$  aritmetičke sredine redom eksperimentalne i kontrolne grupe te  $\sigma_C$  standardna devijacija kontrolne grupe. Glass, McGaw i Smith [GRUB2007] smatraju da eksperimentalni

faktor može utjecati na varijaciju zavisne varijable kao i na njenu aritmetičku sredinu, pa stoga preporučuju uporabu samo standardne devijacije kontrolne grupe.

### **Cohen-ova veličina učinka ( $d$ )**

Ova veličina učinka se zbog svoje popularnosti smatra standardom. (Thalheimer, W., Cook S. 2002 prema [GRUB2007]). Cohen je definirao granične vrijednosti veličine učinka kod istraživanja s dvije grupe (Tablica 2.6.).

**Tablica 2.6. Granične vrijednosti veličine učinka prema Cohen-u**

0,2 mala	0,5 umjerena	0,8 velika
----------	--------------	------------

To omogućava usporedbu veličine učinka eksperimentalnog faktora s poznatim referentnim vrijednostima. Cohen-ova veličina učinka se računa po formuli:

$$d = \frac{\bar{X}_E - \bar{X}_C}{\sigma_{zajednicka}} \quad (2.25)$$

gdje je  $\sigma_{zajednicka}$  zajednička standardna devijacija populacija eksperimentalne i kontrolne grupe (2.23) koja se računa ako F-vrijednost (omjer varijanci) nije statistički značajna.

*Ukupna veličina učinka* se računa kao zbroj svih veličina učinka (n-testova provjere i završnog testa) podjeljeno sa brojem testova, u ovom slučaju  $n+1$ .

### **2.3.3. Statistička analiza**

U statističkoj analizi provodi testiranje standardne značajnosti (eng. standard significance) za istraživanje učinka postupka (eng. treatments) na zavisnoj varijabli odnosno zavisnim varijablama ako ih je više (za zavisnu varijablu se obično uzima znanje učenika ili razni aspekti znanja).

Treba izračunati aritmetičku sredinu (prema formuli 2.18), medijan (vrijednost koja se u nizu rezultata poredanih po veličini nalazi točno u sredini) i standardnu devijaciju (prema formuli 2.19) za inicijalni test, n testova provjere i završni test za eksperiment i njegovu replikaciju. Nakon toga potrebno je izračunati razliku za aritmetičku sredinu, medijan i standardnu devijaciju za svaki n test provjere i inicijalni test isto tako i za završni test i inicijalni test. Isti postupak se ponovi i za replicirani eksperiment.

Rezultati testiranja statističke nul-hipoteze se prikazuju za svaku nul-hipotezu posebno. Nul-hipoteza se testira koristeći t-test (jednosmjerni ili dvosmjerni) za zavisnu grupu za eksperiment i njegovu replikaciju. Potrebno je izračunati veličinu učinka ( $\Delta$ ), odrediti stupnjeve slobode (df,  $df = N-1$ , gdje je N veličina uzorka), odrediti t-vrijednost, graničnu vrijednost t (koja se uzima iz tablice granične vrijednosti t uz prethodno određenu razinu značajnost  $\alpha$ ) i p- vrijednost.



## 2.4. Rezultati i interpretacija

Interpretacija rezultata izuzetno je važna faza u istraživanju. U interpretaciji rezultata prelazi se od činjenica na razmišljanje, na logički utemeljene dedukcije (Mužić, 1979. prema [MEJO2003]). Da bi se učinila kvalitetna interpretacija, treba dobro poznavati problem koji se istražuje i logiku metode kojom su podaci obrađeni. Također, bitno je i iskustvo u interpretiranju rezultata istraživanja a koje se ne može naučiti iz knjiga. Dobiveni rezultati se uspoređuju s polaznom hipotezom odnosno hipotezama kada ih ima više i zaključuje se da li ih dobiveni rezultati potvrđuju ili ne potvrđuju. Zatim te rezultate treba usporediti sa rezultatima repliciranog eksperimenta i zaključiti da li se slažu s tim rezultatima te u čemu se slažu ili ne slažu. U generalizaciji rezultata treba voditi računa o reprezentativnosti uzorka ispitanika i o kvaliteti prikupljenih podataka.

Kada je jednom eksperimentalni proces završen, važno je ne preuveličavati rezultate eksperimenta, osobito ako se testira određena hipoteza. Glavna poteškoća je sposobnost generaliziranja rezultata eksperimenta na ostale uzorke populacije i procjena pogreške rezultata iz laboratorija u industrijsku praksu. Faktori koji ograničavaju svaku generalizaciju su određene opasnosti vanjskoj valjanosti [DALY1996].

Kod provjere nul-hipoteze originalnog eksperimenta treba izračunati razliku između n-testova provjere i završnog testa te inicijalnog testa za kontrolnu i eksperimentalnu grupu. Isti postupak se ponovi i za replicirani eksperiment. Da bi se nul-hipoteza mogla prihvatiti ili oboriti potrebno je izračunati t-vrijednost i p-vrijednost. Kad smo izračunali t-vrijednost onda je potrebno iz tablice graničnih t-vrijednosti na određenoj razini značajnosti očitati granični t koji ćemo usporediti sa dobivenom t-vrijednošću. Ako t-vrijednost prijeđe graničnu vrijednost na određenoj razini značajnosti ili ako je p-vrijednost manja od razine značajnosti onda se *odbija* nul-hipoteza. Ako je t-vrijednost manja od granične vrijednosti na određenoj razini značajnosti ili p-vrijednost veća od razine značajnosti onda se nul-hipoteza *prihvaća*.

Ako promatramo apsolutnu vrijednost dobivene t-vrijednosti i ako je ona manja od granične vrijednosti to znači da razlika nije statistički značajna, a ako razlika nije statistički značajna to znači da smo prihvatili nul-hipotezu. U tom slučaju pozitivne t-vrijednost znači da je kontrolna grupa statistički neznačajno bolja od eksperimentalne, a negativna t-vrijednost znači da je eksperimentalna grupa statistički neznačajno bolja od kontrolne.

Ako je apsolutna vrijednost dobivene t-vrijednosti veća od granične vrijednosti to znači da je razlika statistički značajna, a ako je razlika statistički značajna znači da smo odbacili nul-hipotezu. U tom slučaju pozitivna t vrijednost znači da je kontrolna grupa statistički značajno bolja od eksperimentalne, a negativna t-vrijednost znači da je eksperimentalna grupa statistički značajno bolja od kontrolne.

Isto tako dobro je izračunati i veličinu učinka koja nam govori da li je eksperimentalni faktor učinkovit i kolika je ta učinkovitost. Veličina  $\Delta \geq 0.5$  se smatra učinkom praktičnog značenja (Cohen prema [PFAH2004]). Ima raznih načina interpretacije veličine učinka, jedan od njih je u terminima standardne devijacije na način da se izjednači vrijednost veličine učinka i standardne devijacije za koliko je eksperimentalna grupa bolja od kontrolne. Primjerice, veličina učinka od 0,5 znači da je eksperimentalna grupa bolja od kontrolne za 0,5 standardne devijacije.

Kod učenja i poučavanja uz pomoć inteligentnih tutorskih sustava ciljana veličina učinka iznosi  $2\sigma$  [BLOO1984]. Bloom je u svom istraživanju utvrdio da je prosječan rezultat studenta u eksperimentalnoj grupi za oko dvije standardne devijacije bolji od prosječnog rezultata studenta u tradicionalnoj grupi. No međutim, očekuje se manji rezultat koji je pozitivan i na razini drugih sličnih istraživanja.

### 2.4.1. Valjanost

Konstrukcija valjanosti je stupanj na kojem korištene varijable u studiji točno mjere koncepte koje treba izmjeriti [PFAH2004]. U eksperimentu je uvijek potrebna kontrola koja će omogućiti da se promjene u zavisnoj varijabli mogu isključivo pripisati djelovanju nezavisne varijable. U eksperimentu treba postojati stroga kontrola svih potencijalnih činitelja smetnje koje bi mogle utjecati na zavisnu varijablu.

Kod istraživanja postoje dva tipa valjanosti o kojoj treba voditi računa prilikom provođenja eksperimenta, a to je unutrašnja i vanjska valjanost. U eksperimentu treba postojati stroga kontrola svih potencijalnih činitelja smetnje koji bi mogli utjecati na zavisnu varijablu. Eksperiment u kojem je to postignuto je eksperiment visoke unutarnje valjanosti [MEJO2003]. Osnovni cilj u svim znanostima uključujući društvene znanosti je dobivanje općenitih spoznaja (znanja). Općenito znanje je znanje koje nije ograničeno određenim vremenom i mjestom. Jednostavnije kazano, možemo reći da smo proizveli općenito znanje kada imamo sigurnost da demonstrirana veza odgovara zadanom vremenu, mjestu i metodologiji. Ako je studija oblikovana da testira teoriju i ako zadovoljava uvjete teorije, onda studija može imati vanjsku vrijednost i može nam reći nešto o ostaloj populaciji i postavkama koji također odgovaraju uvjetima teorije. Ako je studija vođena nejasnoćom ili nekom neodgovarajućom teorijom tada otkrića neće biti generalizirana i studiji će nedostajati vanjska vrijednost [LUCA2003].

#### *Unutarnja valjanost*

Unutarnja valjanost [PFAH2004] je stupanj na kojem zaključci mogu biti temeljeni na veličini učinka nezavisne varijable na zavisnu. Moguće opasnosti uključuju učinke selekcije, ne-slučajni gubitak subjekata, usavršavanje učinka i učinak zrelosti. Moguće ih je izbjeći na sljedeći način:

- Učinak selekcije je izbjegavan slučajnim prijelazom subjekata. Zatim, postojeće razlike u sposobnostima između grupa su dobivene skupljanjem pred-test rezultata i mjerenjem stupnja iskustva subjekta.
- Ne-slučajno odustajanje subjekata je izbjegavano eksperimentalnim nacrtom, npr. prijelazom grupe samo direktno prije tretmana, i ne prije pred-testa na početku eksperimenta.
- Činjenica da su postupci eksperimentalne i kontrolne grupe mogu biti različiti u broju scenarij blokova i u dozvoljenom vremenu za izvedbu svakog scenarij bloka može se izazvati učinak usavršavanja. Nadalje, čak i da se uspije ovo izbjegavati pažljivim oblikovanjem, modeli planiranja korišteni u oba postupka se mogu slabo razlikovati u svrsi i rukovanju.
- Učinak zrelosti može biti uzrokovan ako je subjekt informiran prije ili za vrijeme testa da će na kraju eksperimenta ispuniti post-test sa jednakim pitanjima. Dok prva informacija nije dana subjektima i svi se materijali prikupljaju poslije pred-testa može se pretpostaviti da se ovaj učinak neće pojaviti.

### ***Vanjska valjanost***

*Vanjska valjanost* [PFAH2004] je stupanj na kojem rezultati istraživanja mogu biti generalizirani s populacijom van studije ili s stavkama drugih istraživanja. Da bi se povećala vanjska valjanost eksperimenta, istraživači uvode u eksperiment dodatne nezavisne varijable. U tome slučaju eksperimentalni nacrt postaje složeniji, no imaju veću vanjsku valjanost, jer je mogućnost generalizacije veća. Vanjska valjanost pokazuje kada rezultati istraživanja mogu biti legitimno generalizirani na neku specifičnu populaciju (MctTavish i Loether 2002. prema [LUCAS2003]). Vanjska valjanost se odnosi na stupanj na kojem uzrokovana interferencija može biti generalizirana s drugim vremenima, postavkama ili grupama ljudi (Monette, Sullivan, i DeJong 2002. prema [LUCAS2003]). Thomas Cook i Donald Campbell (1979) su utemeljili stav da vanjska valjanost uključuje generalizaciju određenih ciljnih osoba, postavki i vremena, između tipova osoba, postavki i vremena.

Vanjska valjanost može se postići poštujući sljedeće korake [LUCA2003]:

- (1) **Stvaranje valjanosti (eng. construst validity).** Stvaranje valjanosti se odnosi do kojeg raspona mjere precizno reflektiraju teorijski koncept koji trebaju izmjeriti. Iako se o stvaranju valjanosti i vanjskoj valjanosti obično raspravlja odvojeno, stvaranje valjanosti je potreban preduvjet za vanjsku valjanost. Ukoliko mjerenja ne podupiru precizno teoriju tada otkrića ne mogu nadovezati na teoriju i nije moguća generalizacija otkrića.
- (2) **Relevantnost (eng. relevance).** Relevantnost se odnosi na stupanj na kojem oblikovana situacija za testiranje odgovara teoriji i ciljnim uvjetima (eng. scope conditions) koji su postavljeni. Ciljni uvjeti su stavke koje definiraju raspored okolnosti u kojima je teorijska propozicija prikladna.  
Nažalost, mnoge teorije eksplicitno ne definiraju svoje ciljne uvjete, a to je neophodno da bi otkrića bila generalizirana. Ako je teorija podržana u različitim testovima, tada možemo generalizirati otkrića iz jedne situacije u drugu kada su obje situacije opisane jednako i zadovoljavaju iste uvjete.
- (3) **Sposobnost reprodukcije (eng. reproductibility).** Nijedna studija koja se provodi sama ne može postići općenite spoznaje. Zbog ovog razloga, replikacija otkrića istraživanja je važan korak za vanjsku valjanost. Otkrića koja imaju sposobnost reprodukcije su ona koja se ponavljaju pod istim uvjetima, i na ovo se obično odnosi pojam replikacije istraživanja.
- (4) **Konzistencija (eng. consistency).** Konzistencija se odnosi na stupanj na kojem se promatranja u studiji podudaraju međusobno i s teorijom testa. Jednostavnije, konzistencija ispituje koliko se otkriće podudara sa teorijskom propozicijom.
- (5) **Status potvrđivanja (eng. confirmatory status).** S obzirom na činjenicu da otkrića iz bilo koje studije mogu biti generalizirana samo s teorijom, karakteristike teorije testa su najvažnija stavka za određivanje vanjske valjanosti. Status potvrđivanja teorijske propozicije se odnosi na stupanj na kojem je propozicija podržana u brojnim testovima i različitim uvjetima.

Ako eksperiment ima nisku unutrašnju valjanost, tada je i vanjska valjanost niska a generalizacija rezultata može biti besmislena. No, unutarnja valjanost nije nikakva garancija vanjske valjanosti. Moglo bi se kazati da su unutarnja i vanjska valjanost u obrnuto recipročnom odnosu. S povećavanjem jedne, opada druga i obrnuto. Prema tome potreban je određeni kompromis, kojim bi se uz još uvijek zadovoljavajuću unutarnju valjanost postigla i zadovoljavajuća vanjska valjanost.

Najvažniji zaključak koji proizlazi iz primjera je da empirijski bazirano istraživanje nužno ne znači dobro izvedeno istraživanje. Fenton i ostali su izdvojili pet značajnih pitanja koja se trebaju postavljati oko prigovora kod istraživanja [DALY1996]:

- 1) Da li je temeljeno na empirijskom razvoju ili podacima?
- 2) Što eksperiment točno oblikuje?
- 3) Je li temeljeno na izmišljenoj ili realnoj situaciji?
- 4) Da li su korištene mjere prikladne?
- 5) Da li eksperiment prelazi predviđeno vrijeme?

U budućnosti ova pitanja mogu pomoći istraživačima fokusirati se preciznost empirijskog rada koji provode. Votta i Porter su izdvojili neke savjete kako bi se istraživanja još poboljšala [DALY1996]:

- 1) Prihvatanje potrebe za ponavljanjem eksperimenta i objavljivanjem rezultata repliciranog eksperimenta iako se slažu ili ne slažu s originalom.
- 2) Poticanje na istraživanje objašnjenja kada su rezultati repliciranog eksperimenta različiti
- 3) Teorijski test koji nije moguće empirijski provjeriti ne smije biti dopušten
- 4) Modeli različitosti između studenata i profesionalnih istraživača u industrijskim postavkama moraju napravljene tako da se vrijednost studije studenata može razumjeti
- 5) Trebaju se razvijati jeftinije metode eksperimenata, npr. simulacija, efikasnije skupljanje podataka...
- 6) Zahtijeva se više pristupa realnim podacima projekta.
- 7) Empirijski rad se mora prepoznati kao važan i nužan za uspjeh discipline.

## **2.5.Primjer**

Rad Pfahl-a i njegovih suradnika o vrednovanju učinkovitosti učenja korištenjem simulacija u upravljanju projektima razvoja računalne podrške (eng. software project managemant) nam uvelike može pomoći da shvatimo kako provesti eksperiment i njegovu replikaciju, koji su važni faktori koje moramo uzeti u obzir kod provođenja eksperimenta, kako prikupiti podatke, analizirati ih i na kraju, ono što je najvažnije, pravilno interpretirati rezultate dobivene istraživanjem [PFAH2004].

### **2.5.1 Uvod**

U ovom radu su prikazani rezultati dvaput ponovljenog eksperimenta koji se bavi ocjenjivanjem nastavne učinkovitosti korištenja razvojno simulacijskog modela za poučavanje studenata informatike o upravljanju projektima razvoja računalne podrške (eng. software project managemant). Dok eksperimentalna skupina primjenjuje simulacijski model systemske dinamike, kontrolna skupina koristila je COCOMO model koji je već provjereni alat za projektno planiranje. Rezultati svih iskustvenih istraživanja pokazuju da studenti koji koriste simulacijski model steknu bolji uvid i razumijevanje karakterističnih uzoraka kretanja u upravljanju

projektima razvoja računalne podrške. Kombinacija rezultata početnog eksperimenta i dvije replikacije eksperimenta podržavaju i dokazuju ovu tvrdnju.

### ***Opis eksperimenta***

Da bi ispitali učinkovitost na računalu zasnovanog učenja o području upravljanja projektima razvoja računalne podrške korištenjem SD modela za simulaciju, u kontroliranom eksperimentu je primijenjen model kontroliranja skupine prije testa i poslije testa. Ispitanik je morao pristupiti dvama testovima: pre-test i post-test. Učinkovitost pojedinih modela je tada ispitana i ocijenjena uspoređivanjem rezultata od prije i poslije za svakog studenta ponaosob, i usporedbom rezultata između studenata u eksperimentalnoj skupini tj. onih koji su koristili SD model, i studenata u kontrolnoj skupini koji su koristili konvencionalni, već uvriježeni model planiranja projekata, umjesto SD modela. U istraživanju, kontrolna skupina izvršila je svoj zadatak uz pomoć dobro poznatog COCOMO modela. COCOMO je izabran jer je to lako razumljiv model i smatra se uvriježenom praksom u mnogim poduzećima za izradu industrijskog softvera.

Glavni cilj razvoja i primjene modela učenja zasnovanog na simulaciji bilo je omogućavanje djelotvornog učenja o određenim temama o software-skom upravljanoj projektu za studente informatike. Dodatni cilj bilo je povećati interes studenata informatike za temu software-skog vođenja projekta, i upoznati ih s nekim poteškoćama vezanim uz kontroliranje dinamičke složenosti software-skih projekata.

## **2.5.2. Postupak**

Da bi ocijenili učinkovitost postupka kad je korišten SD simulacijski model, koristili su metodu testiranja prije-testa (eng. pre-test) i poslije-testa (eng. post-test). Ova metoda uključuje nasumično pridruživanje studenata eksperimentalnoj skupini (A) i kontrolnoj skupini (B). Provjera znanja kod pre-testa mjeri izvedbu obaju grupa na početku eksperimenta, a provjera znanja post-test mjeri izvedbu obaju skupina na kraju eksperimenta. Studenti nisu znali da su pitanja kod post-testa bila jednaka kao i pitanja pre-testa.

### ***Sudionici eksperimenta (studenti)***

Početni eksperiment je sproveden na studentima koji su diplomirali informatiku na Sveučilištu u Kaiserslauternu (KL) u Njemačkoj, koji su pohađali napredne nastavu iz programskog inženjerstva. Iz razloga što oni još nisu završili svoj magisterij iz informatike ili srodnih polja pretpostavljeno je da je njihov nivo znanja na razini fakultetskog diplomanata. Dvanaest studenata izrazilo je interes za sudjelovanje ali je naposljetku samo devet njih završilo eksperiment.

Prva replikacija početne studije odvijala se tokom ljetne nastave sa sudjelovanjem dvanaest studenata i postdiplomaca (jedan s magisterijem i jedan s doktoratom) Sveučilišta u Oulu, u Finskoj diplomiranih iz informatičkih tehnologija, računalnih znanosti, informatičkog inženjeringa, mikroelektronike ili matematike.

Druga replikacija je izvršena sa trinaest studenata četvrte godine Sveučilišta u Calgriju, u Kanadi, iz područja računalnih znanosti, elektroničkog inženjeringa i računalnog inženjeringa.

U toku svakog od ova tri istraživanja prikupljeni su dodatni podaci o sudionicima. Studente su pitana pitanja osobnog karaktera (starost, spol), nivo sveučilišne naobrazbe (broj semestara,

pohađani kolegiji), osobno iskustvo u razvoju programske podrške, prijašnje znanje o upravljanju projektima razvoja računalne podrške te sklonosti i uvjerenja o načinu izvođenja nastave. Na ova se pitanja odgovaralo na dobrovoljnoj bazi.

### **Provedba eksperimenta**

Početni eksperiment i njegove dvije replikacije su sprovedene po planu naznačenom u tablici 2.7. Nakon kratkog uvoda u kojem je objašnjena svrha eksperimenta i opća organizacijska pitanja, osobni podaci i informacije o dosadašnjoj upoznatosti s predmetima iz eksperimenta su prikupljeni putem upitnika. Onda je obavljeno pred-testiranje i sakupljene su informacije o svim zavisnim varijablama, ponovno se koristeći upitnicima. Nakon pred-testa iznesen je kratak uvod o organizacijskim pitanjima u svezi postupka. Poslije toga sudionici su nasumično raspoređeni u jednu od skupina eksperimentalnu ili kontrolnu. Od dvanaest studenata koji su pristali sudjelovati u početnom eksperimentu, devet njih je sudjelovalo u pred-testu i post-testu. Pet studenata je nasumično raspoređeno u eksperimentalnu a četiri u kontrolnu skupinu. Od dvanaest studenata koji su sudjelovali u prvom repliciranju eksperimenta, šest ih je nasumično raspoređeno u eksperimentalnu i šest u kontrolnu skupinu. Od trinaest studenata koji su sudjelovali u drugoj replikaciji, sedam ih je nasumično raspoređeno u eksperimentalnu skupinu a šest u kontrolnu skupinu. Svaka grupa je bila podvrgnuta određenom postupku. Obrada svake grupe razlikuje se s obzirom na specifičnu strukturu odgovarajućeg scenarija vježbe i vrsti planiranog modela korištenog u osobnoj primjeni scenarija, i pruženih informacija o korištenom planiranom modelu.

Nakon što su izvršili svoje zadatke, obje grupe pristupile su post-testu na kojem su pitanja bila identična onima iz pre-testa, time po drugi put pružajući podatke za zavisne varijable. Na kraju, studenti su dobili priliku da sami ocijene nastavu popunjavanjem još jednog upitnika.

Tokom cijelog postupka, vrijeme omogućeno za izvršavanje pojedinih planiranih koraka bilo je jednako kod eksperimentalne i kod kontrolne skupine. Međutim kod replikacije eksperimenta slijedio opušteniji raspored nego kod samog početnog eksperimenta.

**Tablica 2.7. Raspored zadataka kod eksperimenta i njegovih replikacija**

	<b>Eksperiment</b>	<b>Replikacija1</b>	<b>Replikacija2</b>
<b>Uvod u eksperiment</b>	5 min	5 min	5 min
<b>Osobni podaci i prijašnje znanje</b>	5 min	5 min	5 min
<b>Pre-test</b>			
<b>Interesi</b>	3 min	5 min	5 min
<b>Predznanje o emirijskim obrascima</b>	5 min	5 min	5 min
<b>Razumjevanje jednostavne projektne dinamike</b>	10 min	10 min	10 min
<b>Razumjevanje složene projektne dinamike</b>	12 min	15 min	15 min
<b>Uvod u postupak</b>	5 min	5 min	5 min
<b>Nasumično razmještanje studenata u skupine</b>	5 min	5 min	5 min
<b>Postupak</b>	45 min	80 min	80 min
<b>Post-test</b>	3 min	5 min	5 min
<b>Interesi</b>			
<b>Predznanje o emirijskim obrascima</b>	5 min	5 min	5 min
<b>Razumjevanje jednostavne projektne dinamike</b>	10 min	10 min	10 min
<b>Razumjevanje složene projektne dinamike</b>	12 min	15 min	15 min
<b>Potrebno vrijeme i subjektivna ocjena nastave</b>	5 min	10 min	10 min
<b>Sveukupno</b>	130 min	180 min	180 min

### **Razlika između početnog eksperimenta i replikacija**

Budući da je većina sudionika početne studije na Sveučilištu u Kaiserslautern naznačila da nisu imali dovoljno raspoloživog vremena da rade na materijalu, više vremena im je dano tokom izvođenja replikacija na Sveučilištu u Oulu. Druga razlika je u cjelokupnom postupku eksperimenta. Dok je početni eksperiment na Sveučilištu u Kaiserslautern vršen u dva dana, s tjednom dana između ta dva dana, replikacije na Sveučilištu u Oulu i na Sveučilištu u Calgary-iju odvijale su se tokom jednog dana u oba slučaja.

### **Prikupljanje podataka**

Za vrijeme eksperimenta, prikupljeni su podaci za dvije vrste varijabli, (Y.1, ..., Y.4) i dvije varijable koje sakupljaju subjektivne dojmove o nastavi (Z.1, Z.2), (Tablica 2.8.). Zavisne varijable Y.1, Y.2, Y.3, i Y.4 su sastavljene da sakupe razne aspekte učenja vršenog na ovoj vrsti nastave. Svaki prikupljeni sastav mjeran je kroz skup od 5 – 7 pitanja s odgovorima koji su dani na ravnomjernoj ljestvici. Vrijednost svake zavisne varijable izračunata je kao prosječni rezultat. Vrijednost varijabli Z.1, i Z.2 je izvučena također iz upitnika. Neobrađeni podaci za zavisne varijable Y.1 do Y.4 sakupljeni su tokom pre-testa i post-test uz pomoć upitnika. Svaki odgovor u upitniku je označen sa vrijednosti  $R = [0,1]$  uzimajući jednake udaljenosti između mogućih odgovora, primjer "potpuno neslaganje" je označeno sa "0", "neslaganje" sa "0.25", "nedefinirano" sa "0.5", "slaganje" sa "0.75" i "potpuno slaganje" sa "1".

Subjektivna percepcija kvalitete postupka je razvijena s obzirom na četiri dimenzije ("korisno" nasuprot "beskorisno", "zanimljivo" nasuprot "dosadno", "lako" nasuprot "teškom" i "razumljivo" nasuprot "nerazumljivo") koristeći pet točaka Likerove ljestvice, npr. "iznimno dosadan", "dosadan", "neodređen", "zanimljiv" i "iznimno zanimljiv".

**Tablica 2.8. Eksperimentalne varijable**

<b>Zavisne varijable</b>	
Y.1	Interes za teme o upravljanju projektom ("Interes")
Y.2	Poznavanje tipičnih primjera ponašanja o upravljanju projektima razvoja računalne podrške ("Znanje")
Y.3	Razumijevanje "jednostavne" dinamike projekta ("Razumijevanje jednostavne")
Y.4	Razumijevanje "složene" dinamike projekta ("Razumijevanje složene")
<b>Osobna zapažanja</b>	
Z.1	Dostupna količina vremena nasuprot realnim vremenskim potrebama ("Pritisak vremena")
Z.2	Ocjena nastave

### **2.5.3. Analiza podataka**

Testiranje standardne značajnosti je korišteno za istraživanje učinka postupka na zavisnim varijablama od Y.1 do Y.4. Nulte hipoteze su izvedene na sljedeći način:

$H_{0,1}$ : Nema razlike između pred-test rezultata i post-test rezultata sa eksperimentalnom grupom A i kontrolnom grupom B.

$H_{0,2a}$ : Nema razlike u učinkovitosti relativnog učenja između eksperimentalne grupe A i kontrolne grupe B

$H_{0,2b}$ : Nema razlike u učinkovitosti apsolutnog učenja između eksperimentalne grupe A i kontrolne grupe B.

Za testiranje hipoteze  $H_{0,1}$  koristio se kombinirani jednostruki (eng. one-way paired) t- test, jer su skupljeni podaci za ovu hipotezu unutar subjekata, npr. rezultati post-testa su uspoređeni sa rezultatima pred-testa na subjektima unutar iste grupe. Za testiranje hipoteza  $H_{0,2a}$  i  $H_{0,2b}$  odgovarajući test je bio jednosmjerni (eng. one-side) t-test za nezavisne uzorke .

Preduvjet za apliciranje t-testa je pretpostavka normalne distribucije varijabli u uzorku testa. Obično je ustaljena praksa postavljanja razine značajnosti  $\alpha=0.05$ . Za kontroliranje tipa I pogreške i tipa II pogreške zahtjeva se veća količina učinka ili veća količina uzorka. To predstavlja određenu polemiku u području software istraživanja jer većina istraživanja u ovom području uključuje relativno umjerene veličine učinka i općenito male veličine uzorka. Zbog toga je postavljena razina značajnosti  $\alpha=0.01$ .

## 2.5.4. Rezultati

Tablica 2.9. sadrži deskriptivnu statistiku za originalan eksperiment i njegove dvije replikacije. Stupac "pre-test bodovi" i "post-test bodovi" prikazuju izračunate vrijednosti za aritmetičku sredinu, medijan i standardnu devijaciju podataka koji su prikupljeni tijekom pre-testa i post-testa za originalni eksperiment (E) i njegove dvije replikacije ( R1 i R2). Stupac razlika (eng. difference scores) prikazuje za izračunate vrijednosti aritmetičku sredinu, medijan i standardnu devijaciju razliku između post-testa i pre-testa za originalan eksperiment (E) i njegove dvije replikacije(R1 i R2). Ako je ta razlika negativna ili nula. to znači da nema pozitivnog učinka relativnog učenja. U tablici 2.9. vidimo da se taj fenomen pojavio dvaput u originalnom eksperimentu (varijabla Y.4 grupa A i varijabla Y.1 grupa B), tri puta u prvoj replikaciji (varijabla Y.4 u grupi A i B i Y.2 grupa B) i tri puta u drugoj replikaciji (varijabla Y.1 grupa A, Y.2 grupa B i Y.4 grupa A).

**Tablica 2.9. Vrijednosti zavisnih varijabli (E, R1 i R2)**

	<i>Pre-test bodovi</i>			<i>Post-test bodovi</i>				<i>Razlika bodovi</i>				
	Y.1	Y.2	Y.3	Y.4	Y.1	Y.2	Y.3	Y.4	Y.1	Y.2	Y.3	Y.4
E:originalan eksperiment(KL)												
Grupa A (5 ispit.)												
aritm. sred.	0.69	0.56	0.31	0.37	0.79	0.84	0.66	0.43	0.10	0.28	0.34	0.07
medijan	0.75	0.60	0.29	0.33	0.85	0.80	0.71	0.33	0.10	0.40	0.43	0.00
st. dev.	0.18	0.30	0.26	0.25	0.19	0.17	0.13	0.32	0.09	0.36	0.28	0.19
Grupa B (4 ispit.)												
aritm. sred	0.81	0.50	0.43	0.33	0.79	0.60	0.82	0.46	-0.03	0.10	0.39	0.13
medijan	0.78	0.50	0.36	0.25	0.80	0.60	0.86	0.50	0.00	0.10	0.50	0.17
st. dev.	0.13	0.26	0.31	0.24	0.19	0.16	0.07	0.37	0.09	0.35	0.38	0.34
R1: prva replikacija (Oulu)												
Grupa A (6 ispit.)												
aritm. sred	0.83	0.57	0.41	0.44	0.85	0.97	0.67	0.44	0.03	0.40	0.26	0.00
medijan	0.88	0.60	0.43	0.42	0.85	1.00	0.57	0.50	0.03	0.40	0.14	0.00
st. dev.	0.14	0.23	0.11	0.23	0.15	0.08	0.27	0.09	0.07	0.28	0.25	0.24
Grupa B (6 ispit.)												
aritm. sred	0.70	0.47	0.33	0.33	0.78	0.43	0.74	0.33	0.08	-0.03	0.41	0.00
medijan	0.73	0.40	0.36	0.25	0.83	0.40	0.79	0.33	0.08	0.00	0.43	0.08
st. dev.	0.18	0.21	0.17	0.30	0.21	0.15	0.17	0.24	0.13	0.32	0.19	0.45
R2:druga replikacija (Calgary)												
Grupa A (7 ispit.)												
aritm. sred	0.84	0.63	0.41	0.62	0.87	0.83	0.61	0.52	0.03	0.20	0.20	-0.10
medijan	0.85	0.60	0.43	0.67	0.85	0.90	0.71	0.67	0.00	0.20	0.14	0.00
st. dev.	0.09	0.15	0.27	0.30	0.10	0.23	0.24	0.20	0.12	0.22	0.35	0.19
Grupa B (6 ispit.)												
aritm. sred	0.82	0.60	0.43	0.44	0.91	0.60	0.55	0.53	0.09	0.00	0.12	0.09
medijan	0.88	0.70	0.43	0.50	0.95	0.70	0.57	0.50	0.05	0.00	0.14	0.07
st. dev.	0.18	0.33	0.13	0.30	0.11	0.42	0.11	0.19	0.11	0.18	0.19	0.21



U sljedećem dijelu će biti prikazan rezultat testiranja statističke hipoteze za svaku hipotezu posebno.

### HIPOTEZA $H_{0,1}$

Nul-hipoteza  $H_{0,1}$  je postavljena kao što slijedi: Nema razlike između rezultata pre - test i rezultata post- test između eksperimentalne grupe A i kontrolne grupe B.

Fokusirajući se na jedino na eksperimentalanu grupu, tablica 2.10 pokazuje za svaku zavisnu varijablu odvojeno rezultat od testiranja nul-hipoteze  $H_{0,1}$ , koristeći jednosmjerni t-test (eng. one-tailed t-test) za zavisni uzorak. Stupac 1. prikazuje varijable i odgovarajuće istraživanje za početni eksperiment (E), prvo ponavljanje eksperimenta (R1) i drugo ponavljanje (R2). Stupac 2 predstavlja vrijednost veličine učinka, stupac 3 stupnjeve slobode, stupac 4 t-vrijednost od istraživanja, stupac 5 graničnu vrijednost (eng. critical value) za  $\alpha = 0.10$  i stupac 6 prikazuje p-vrijednost. Ne ispitivajući stupac 4 i 5 tablice 2.10 može se vidjeti da eksperimentalna grupa može postići statistički i praktično važan rezultat za zavisnu varijablu Y.1 jedino u početnom eksperimentu. Za zavisnu varijablu Y.2 i Y.3 i statistički i praktično važan rezultat dobiven u svim tri istraživanja, dok za varijablu Y.4 nije dobiven statistički važan rezultat (ni statistički ni praktičan) u bilo kojem istraživanju. U oba ponavljanja (replikacije), varijabla Y.4 nije izravana podrška dane nul-hipoteze.

**Tablica 2.10. Rezultati post- testa nasuprot pre- testa za grupu A**

Varijable/Studija	$\Delta$	df	t- vrijednost	granični $t_{0,90}$	p- vrijednost
<b>Varijabla Y.1</b>					
E	1.07	4	2.39	1.53	0.04
R1	0.36	5	0.89	1.48	0.21
R2	0.23	6	0.62	1.44	0.28
<b>Varijabla Y.2</b>					
E	0.77	4	1.72	1.53	0.08
R1	1.41	5	3.46	1.48	0.01
R2	0.91	6	2.24	1.44	0.04
<b>Varijabla Y.3</b>					
E	1.23	4	2.75	1.53	0.03
R1	1.06	5	2.61	1.48	0.02
R2	0.59	6	1.55	1.44	0.09
<b>Varijabla Y.4</b>					
E	0.35	4	0.78	1.53	0.24
R1	0.00	5	0.00	1.48	0.50
R2	-0.50	6	-1.33	1.44	0.88

**Tablica 2.11. Rezultati post- testa nasuprot pre- testa za grupu B**

Varijable/Studija	$\Delta$	df	t- vrijednost	granični $t_{0,90}$	p- vrijednost
<b>Varijabla Y.1</b>					
E	-0.29	3	-0.58	1.64	0.70
R1	0.65	5	1.58	1.48	0.09
R2	0.82	5	2.02	1.48	0.05
<b>Varijabla Y.2</b>					
E	0.29	3	0.58	1.64	0.30
R1	-0.10	5	-0.25	1.48	0.60
R2	0.00	5	0.00	1.48	0.50
<b>Varijabla Y.3</b>					
E	1.05	3	2.09	1.64	0.06
R1	2.13	5	5.22	1.48	0.0017
R2	0.63	5	1.54	1.48	0.09
<b>Varijabla Y.4</b>					
E	0.37	3	0.73	1.64	0.26
R1	0.00	5	0.00	1.48	0.50
R2	0.41	5	1.01	1.48	0.18

Fokusirajući se na kontrolnu grupu (B) tablica 2.11. prikazuje za svaku zavisnu varijablu rezultate testiranja nul-hipoteze  $H_{0,1}$  koristeći jednosmerni t-test (eng. one-tailed t-test) za zavisnu uzorak. Sastav tablice je isti kao u tablici 2.10. Ne ispitivajući stupac 4 i 5 u tablici 2.11, možemo vidjeti da kontrolna grupa (B) postiže statistički i praktično važne rezultate za zavisnu varijablu Y.1 jedino u replikaciji. U početnom eksperimentu, vrijednosti varijable Y.1 nisu izravna podrška dane nul-hipoteze. Za zavisne varijable Y.2 i Y.4 nije pronađen niti jedan značajan rezultat. Podatci u početnom eksperimentu i u drugoj replikaciji je barem izravna podrška očekivanom pozitivnom učinku učenja. Rezultati za zavisnu varijablu Y.3 su otkrili statističku i praktičnu značajnost u svim istraživanjima.

### **HIPOTEZA $H_{0,2a}$**

Nulta hipoteza  $H_{0,2a}$  je postavljena kako sljedi: Nema razlike u učinkovitosti relativnog učenja između eksperimentalne grupe A i kontrolne grupe B, npr. razlika između post i pred-test rezultata grupe A nije značajno veća od onih grupe B.

Tablica 2.12. pokazuje za svaku zavisnu varijablu odvojeno rezultate testiranja hipoteze  $H_{0,2a}$  koristeći jednosmjerni t-test (eng. one-tailed) za nezavisne uzorke. Za varijablu Y.1 nul-hipoteza  $H_{0,2a}$  može biti odbijena samo u početnom eksperimentu, gdje vrijednosti za obje replikacije ne podupiru čak ni smjer očekivanog relativnog učenja učinka. Za zavisne varijable Y.2 statistički značajan nivo je postignut u obje replikacije i rezultat početnog eksperimenta podržava smjer širenja relativnog učinka sa praktičnim značenjem, npr. pokazujući srednju veličinu učinka. U nijednoj studiji vrijednosti varijabli Y.3 i Y.4 nisu smatrane značajnima i samo u drugoj replikaciji rezultat za varijablu Y.3 nije kontradiktoran smjeru hipoteze, pokazujući prilično malu veličinu učinka, npr. nema praktičnog značaja.

**Tablica 2.12. Rezultati testiranja hipoteze  $H_{0,2a}$  (Grupa A nasuprot grupe B)**

Varijable/Studija	$\Delta$	df	t- vrijednost	granični $t_{0,90}$	p- vrijednost
<b>Varijabla Y.1</b>					
E	1.38	7	2.06	1.42	0.04
R1	-0.56	10	-0.98	1.37	0.82
R2	-0.54	11	-0.97	1.36	0.82
<b>Varijabla Y.2</b>					
E	0.51	7	0.75	1.42	0.24
R1	1.43	10	2.48	1.37	0.02
R2	1.00	10	1.73	1.37	0.06
<b>Varijabla Y.3</b>					
E	-0.16	7	-0.23	1.42	0.59
R1	-0.65	10	-1.13	1.37	0.86
R2	0.30	11	0.53	1.36	0.30
<b>Varijabla Y.4</b>					
E	-0.23	7	-0.33	1.42	0.62
R1	0.00	10	0.00	1.37	0.50
R2	-0.92	11	-0.65	1.36	0.94

### **HIPOTEZA $H_{0,2b}$**

Nul-hipoteza  $H_{0,2b}$  je objašnjena ovako: Nema razlike u učinkovitosti apsolutnog učenja između eksperimentalne grupe A i kontrolne grupe B, npr. post-test rezultati grupe A nisu značajno veći od onih grupe B.

Tablica 2.13. pokazuje za svaku zavisnu varijablu odvojeno rezultate testiranja hipoteze  $H_{0,2b}$  koristeći jednosmjerni t-test za nezavisne uzorke. Pokazalo se da varijabla Y.2 pokazuje

statistički značajne rezultate u početku eksperimenta i njegove prve replikacije, i praktični značajne rezultate za drugu replikaciju. Stoga, hipoteza  $H_{0,2b}$  a može biti odbačena za Y.2. Zajedno s ovom prvom replikacijom gdje je praktično značajan rezultat za varijablu Y.4 postignut, za varijable Y.1, Y.3 i Y.4 nema značajnih rezultata primijećenih u nijednoj studiji. Podaci za varijable Y.3 i Y.4 u početku eksperimenta, za Y.3 u prvoj replikaciji, i za Y.1 i Y.4 u drugoj replikaciji ne podržavaju čak ni smjer hipoteze.

**Tablica 2.13. Rezultati testiranja hipoteze  $H_{0,2b}$  ( grupa A nasuprot grupe B)**

Varijable/Studija	$\Delta$	df	t- vrijednost	granični $t_{0,90}$	p- vrijednost
<b>Varijabla Y.1</b>					
E	0.01	7	0.02	1.42	0.49
R1	0.36	10	0.63	1.37	0.27
R2	-0.36	11	-0.64	1.36	0.73
<b>Varijabla Y.2</b>					
E	1.45	7	2.16	1.42	0.03
R1	4.40	10	7.63	1.37	0.000009
R2	0.69	10	1.19	1.37	0.13
<b>Varijabla Y.3</b>					
E	-1.53	7	-2.28	1.42	0.97
R1	-0.32	10	-0.56	1.37	0.71
R2	0.33	11	0.60	1.36	0.28
<b>Varijabla Y.4</b>					
E	-0.07	7	-0.11	1.42	0.54
R1	0.63	10	1.08	1.37	0.15
R2	-0.02	11	-0.04	1.36	0.51

### Zaključci

Ovo poglavlje kao prvo sakuplja rezultate početnog eksperimenta i njegove dvije replikacije s obzirom na nul-hipotezu  $H_{0,1}$  (Tablica 2.14.), i nul-hipotezu  $H_{0,2a}$  i  $H_{0,2b}$  (Tablica 2.15.) odvojeno za svaku zavisnu varijablu.

Statističko značenje (stat.sig.): nul-hipoteza se može odbaciti na nivou značajnost  $\alpha = 0.1$ .

Praktično značenje (pract.sig.): nul-hipoteza se ne može odbaciti ali je veličina učinka  $\Delta \geq 0.5$ . Ako je postignut statistički značaj, praktično značenje se ne spominje.

Pozitivni učinak (+): ne može se primijetiti praktično značenje ali je veličina učinka  $\Delta > 0$ . Broj u intervalu pokazuje koliko je potrebno subjekata za postići statističko značenje sa dobivenom veličinom učinka (samo u tablici 2.15).

Bez učinka ili negativnog učinka (-):  $\Delta \leq 0$ .

Tablica 2.14 pokazuje da nul-hipoteza  $H_{0,1}$  može biti odbačena samo u svim eksperimentima za varijablu Y.3 (eksperimentalna i kontrolna grupa). Nadalje, za eksperimentalnu grupu  $H_{0,1}$  može biti odbačena u svim slučajevima za Y.2 i u jednom slučaju za Y.1. Za kontrolnu grupu,  $H_{0,1}$  može biti odbačena u dva slučaja za Y.1 također.

Tablica 2.15 pokazuje da nul-hipoteza  $H_{0,2a}$  može biti odbačena u svim slučajevima za varijablu Y.2. Nadalje, značajan rezultat je postignutu jednom slučaju za varijablu Y.1. S obzirom na nultu hipotezu  $H_{0,2a}$ , statističko testiranje je dalo rezultate statističkog i praktičnog značenja za varijablu Y.2 također. Praktično, značajni rezultati su postignuti u jednom slučaju za varijablu Y.4.

**Tablica 2.14. Rezultati za hipotezu  $H_{0,1}$**

Varijable	Grupa A			Grupa B		
	Eksperiment	Replikacija1	Replikacija2	Eksperiment	Replikacija1	Replikacija2
Y.1	stat. znač.	+	+	-	stat. znač.	stat.znač.
Y.2	stat. znač.	stat. znač.	stat. znač.	+	-	-
Y.3	stat. znač.	stat. znač.	stat.znač.	stat. znač.	stat. znač.	
Y.4	+	-	-	+	-	+

**Tablica 2.15. Rezultati za hipotezu  $H_{0,2}$**

Varijable	Grupa A nasuprot B					
	$H_{0,2a}$			$H_{0,2b}$		
	Eksperiment	Replikacija1	Replikacija2	Eksperiment	Replikacija1	Replikacija2
Y.1	stat. znač.	-	-	+	+	-
Y.2	prakt. znač.	stat. znač.	stat. znač.	stat. znač.	stat. znač.	Prakt. znač.
Y.3	-	-	+	-	-	+
Y.4	-	-	-	-	prakt. znač.	-

## 2.5.4. Interpretacija

Testirajući pozitivan učinak učenja unutar eksperimentalne grupe potvrđuje se statističko značenje djelovanja na promjeni rezultata od pred-testa do post-testa za zavisne varijable Y.1 prema Y.3. Ovo nam osigurava dokaz za pretpostavku da poučavanje koje uključuje SD model umjesto COCOMO značajno povećavaju interes u vođenju projekta, znanje o empirijskim obrascima u software-skim projektima, i razumijevanje jednostavne dinamike projekta od strane studenata koji su sudjelovali u poučavanju. S druge strane, ne može se pronaći nijedan pozitivni učinak za varijablu Y.4 (razumijevanje složenije dinamike projekta) iako koristeći SD model se pretpostavljalo da će biti izvrstan alat za objašnjivanje i analiziranje unutar zavisnosti između duljine projekta, uloženog napora i kvalitete rezultata projekta, npr software produkt.

Testirajući izvedbu relativnog i apsolutnog učenja učinkovitosti između eksperimentalne i kontrolne grupe (hipoteze  $H_{0,2a}$  i  $H_{0,2b}$ ) pokazalo se da poučavanje koje uključuje SD model daje značajnije rezultate za varijablu Y.2 (znanje o empirijskim obrascima u software projektima) nego korištenje COCOMO-a. S druge strane, ne mogu se primijetiti značajne različitosti između eksperimentalne i kontrolne grupe promatrajući varijable Y.1, Y.3 i Y.4 (interes u vođenju projekta te razumijevanje jednostavne i složene projektne dinamike). SD model nudi potrebnu funkcionalnost dok je COCOMO ograničen na model međusobne ovisnosti između trajanja projekta i uloženog napora (jednostavna dinamika projekta). Što više, čini se da postupak eksperimentalne grupe ne iskorištava potpuno sve mogućnosti dostupnih oblika koji se uče između uporabe SD modela i izgradnje koju SD može ponuditi. Za vrijeme poučavanja samo su subjekti eksperimentalne grupe imali dopuštenje koristiti osiguranu SD simulaciju modela. Studenti su se suočili sa pre-definiranim modelom uključujući jako složenu uzročnu vezu i povratne strukture bez da je dopušteno stvarno istraživanje individualne veze između varijabli modela i njihovih učinaka na ponašanje projekta. Bez mogućnosti aktivnog mijenjanja struktura modela čini se da će biti teško razumjeti SD model u njegovoj potpunoj složenosti u usporedbi sa dobro razvijenim i struktuiranim, znatno jednostavnijim COCOMO. Kombinirajući ograničenja, npr. nedostatak aktivnog uplitanja u izgradnju modela, čini se da su šanse male za postići pozitivno učenje učinka za varijablu Y.4 u eksperimentalnoj grupi. Ovo će, kako god, biti preduvjet za postizanje značajnijeg uspjeha učenja učinka u usporedbi sa kontrolnom grupom.

Napokon, činjenica da je izvedba eksperimentalne grupe uzimajući u obzir varijablu Y.1 bila značajno bolja u početnom eksperimentu, ali se ne može u dvije replikacije dovesti u vezu s razlikama osobnih karakteristika subjekta i/ili pozadinom znanja o upravljanju projektima razvoja računalne podrške. Barem u slučaju Oulu eksperimenta može se primijetiti da su subjekti bili općenito zreliji i iskusniji u usporedbi sa Kaiserslautern eksperimentom. Interpretacija svakako treba više istraživanja u budućim replikacijama.

### 2.5.5 Određivanje valjanosti

Konstrukcija valjanosti je stupanj na kojem korištene varijable u studiji točno mjere koncepte koje trebaju izmjeriti. Sljedeće stavke zajedno s konstrukcijom valjanosti su identificirane:

- Razne aplikacije SD modela ne moraju adekvatno prihvatiti specifične prednosti SD modela preko konvencionalnih modela planiranja, dok je često prihvaćeno da model izgrađuje i ne aplikacije postojećeg modela koja su glavna dobit SD simulacijskog modeliranja.
- Postoje indikacije da je razlika između "jednostavne dinamike" i "složene dinamike", kao što je napravljeno za mjerenje varijabli Y.3 i Y.4 bila preočita.
- Teško je izbjegavati "nepoštenu" usporedbu između korištenja SD modela i COCOMO zato što SD modeli nude oblike koji nisu prihvatljivi za COCOMO (npr. simulacija parametara se mijenja preko (eng. time/on-the-fly) modifikacija pretpostavki modela).

Dvije su moguće opasnosti identificirane s obzirom na proučavanje vanjske valjanosti: reprezentativnost subjekta i materijali.

Subjekti koji su sudjelovali u eksperimentu su svi bili studenti informatike ili sličnih područja na naprednom stupnju. Može se očekivati da su rezultati studije na nekom stupnju reprezentativnosti za ovaj tip subjekata. Svaka generalizacija rezultata s obzirom na obrazovanje studenata početnika (npr, bruceša) ili čak uzimajući u obzir poučavanje nad software profesionalcima treba biti napravljena sa oprezom.

U svakom slučaju, točka koju treba istaknuti je predstavljeno istraživanje u svom trenutnom stadiju koje istražuje prirodu i samo prvi korak serije eksperimenata, koji nakon modifikacije tretmana i laganim uključivanjem subjekata s različitim pozadinom može dobiti konkretnije rezultate u budućnosti.

### 3. Replikacija eksperimenta za vrednovanje učinkovitosti sustava xTEx-Sys

Ovaj rad prikazuje rezultate ponovljenog eksperimenta koji se bavi ocjenjivanjem nastavne učinkovitosti sustava e-učenja. Zainteresirani rezultatima Pfahl-a i njegovih suradnika, vlastitim smo istraživanjem pokušali utvrditi učinkovitost učenja uz pomoć inteligentnog tutorskog sustava (ITS) Extendend Tutor Expert System (xTEx-Sys) [STAN2003].

#### 3.1. E-učenje

Jedan od ustaljenih pojmova u području informacijsko-komunikacijske tehnologije je e-učenje (eng. e-learning). Što je e-učenje? Možda bi prikladnije pitanje bilo što predstavlja slovo "e" u e-učenju? Brojne su interpretacije ovog naziva, ali čini se da ponajmanje "e" ovdje predstavlja elektroničko učenje. E stoji za: iskustveno učenje (eng. experience learning ), učenje "posvuda" (eng. everywhere learning), bolje učenje (eng. enhanced) i prošireno učenje (eng. extended learning) [SRCE2008].

E-učenje predstavlja presjek svijeta informacijske i komunikacijske tehnologije (ICT) i svijeta obrazovanja. Ono uključuje brojne strategije učenja kao i podršku različitih tehnologija koje sudjeluju u interakciji učenik-sadržaj-nastavnik. E-učenje ne predstavlja samo Web utemeljenu nastavu ili pak daljinsko učenje već uključuje podršku raznih tehnologija poput CD-ROM uređaja i medija, zatim nastavu temeljenu na računalu, videokonferencijskog sustava, nastavnih sadržaja za učenje koji su isporučeni uz pomoć satelitske komunikacije. Prema tome e-učenje uključuje mnoge putove u kojima se može vršiti individualna izmjena informacija i stjecanje znanja onih koji sudjeluju u takvom procesu. S obzirom na provedbu, razlikujemo *sinkrono* (istodobna prisutnost učitelja i učenika) i *asinkrono* e-učenje (zasniva na Internetu i Web tehnologijama) [STAN2004].

Današnji oblici e-učenja obuhvaćaju različite aspekte korištenja ICT-a u obrazovanju, pa ovisno o intenzitetu i načinu korištenja ICT-a razlikujemo nekoliko oblika e-učenja [WIKI2008]:

- **klasična nastava** - nastava u učionici;
- **nastava uz pomoć ICT-a** - tehnologija u službi poboljšanja klasične nastave;
- **hibridna ili mješovita nastava** - kombinacija nastave u učionici i nastave uz pomoć tehnologija;
- **online nastava** - nastava je uz pomoć ICT-a u potpunosti organizirana na daljinu.

Posebna klasa inteligentnih sustava e-učenja su inteligentni tutorski sustavi (ITS) [GRUB2007b]. Inteligentni tutorski sustavi predstavljaju specijaliziranu programsku podršku namijenjenu učenju i poučavanju, u odabranoj bazi znanja.

Strukturu inteligentnih tutorskih sustava [BURN1988] čine četiri međusobno povezana programska modula:

- modul stručnjaka koji predstavlja osnovu svakog inteligentnog sustava a sadrži područno znanje koje učenik kroz proces učenja treba usvojiti;
- modul učenika - programska jedinica namijenjena učeniku koja omogućava aktivno i dinamično usvajanje znanja, analizu postignutog i vrednovanje rezultata. Modul učenika sadrži opis mogućeg učeničkog ponašanja što uključuje i pogrešna razumijevanja i nedostatke u znanju. Neslaganje između konkretnog ponašanja učenika i ponašanja kojeg sustav očekuje predstavlja signal da sustav aktivira korektivne mjere. Da bi to mogao učiniti, nužno je da sustav sadrži informaciju kako bi čovjek tutor u takvoj situaciji reagirao;
- modul učitelja (tutor modul) je sastavnica sustava zadužena za procese vođenja stjecanja znanja i vještina učenika kroz izbor scenarija poučavanja a uz upotrebu pedagoških znanja;
- komunikacijski modul, predstavlja korisničko sučelje učenika i inteligentnog tutorskog sustava unutar kojeg se ostvaruje interakcija na relaciji učenik-učitelj-znanje.

### 3.2. xTEx-Sys

Extended Tutor-Expert System [STAN2003] ili xTEx-Sys je Web orijentirana inteligentna hipermedijska autorska ljuška koja je evoluirala od on-site verzije (Tutor-Expert System) [STAN1997], preko distribuirane verzije (Distributed Tutor-Expert System) [ROSI2000] u najnoviju proširenu Web-orijentiranu verziju. Svrha takvog sustava je poboljšati proces učenja i poučavanja učenika, te ujedno povećati i učinkovitost rada samog učitelja.

Web orijentiran inteligentna autorska ljuška xTEx-Sys ima četiri sudionika:

- učenika (radi stjecanja znanja i vještina);
- učitelja (radi didaktičkog oblikovanja nastavnih sadržaja);
- stručnjaka (radi izgradnje baze znanja) i;
- administrator (radi nadzora sustava).

Osnovne funkcije xTEx-Sys-a su:

- oblikovanje baze znanja za različita područna znanja;
- oblikovanje i pristup nastavnim sadržajima koji se temelje na bazi područnog znanja;
- učenje, poučavanje i testiranje;
- ocjenjivanje znanja, te nadzor sustava

Nastavni sadržaj predmeta pripremljenog za izvođenje na računalu (eng. courseware) je strukturiran u više razina na *nastavne cjeline* koje su na prvoj razini, *nastavne teme* na drugoj razini, *nastavne jedinice* na trećoj razini i naposljetku *nastavne pojmove*, na četvrtoj razini te testove koje je moguće pridružiti nastavnoj temi, cjelini ili jedinici. Strukturu nastavnog sadržaja (semantičku mrežu) gradi učitelj kroz dekompoziciju čvorova strukture nastavnog sadržaja na nastavne cjeline, nastavne teme, nastavne jedinice te nastavne pojmove.

Kod sustava xTeX-Sys se razlikuju dva tipa testiranja:

- testiranje pomoću statičkih pitanja koje generira učitelj;
- testiranje pomoću dinamičkih pitanja koja generira računalo slučajnim odabirom temeljem baze područnog znanja.

Kod dinamičkog kviza učeniku se distribuira skup pitanja uz pridružene ponuđene odgovore koji mogu biti točni ili netočni. Pritom za svako pitanje mogu biti svi ponuđeni odgovori točni, može biti nekoliko točnih odgovora, može biti samo jedan točan odgovor, ali mogu i svi ponuđeni odgovori biti netočni. Pitanja su u kvizu grupirana su u tri težinske kategorije, a u svakoj kategoriji su po četiri težinske forme. Znači da ima ukupno dvanaest formi pitanja koje učenik dobiva tijekom testiranja svojeg znanja. Ako učenik odgovori točno na oba pitanja, u sljedećoj seriji dobiti će pitanja iz teže skupine, odnosno ako odgovori netočno na oba pitanja sljedeća će biti iz lakše skupine. Time se sustav prilagođava iskazanom učenikovom znanju. Ukupan broj pitanja tijekom testiranja određuje učitelj.

### **3.3. Opis eksperimenta i replikacije**

Za vrednovanje učinkovitosti sustava xTeX-Sys proveli smo dva eksperimenta: originalni eksperiment je proveden 2005/06 i njegova replikacija 2006/07 godine. Na početku eksperimenta studenti su pristupili inicijalnom testu, zatim su imali parcijalne testove i završni test. Učinkovitost je ispitana uspoređivanjem rezultata inicijalnog testa sa rezultatima parcijalnih testova i završnog testa za svakog studenta posebno te usporedbom rezultata između studenata u eksperimentalnoj grupi i studenata u kontrolnoj grupi.

#### **3.3.1 Cilj eksperimenta i replikacije**

Cilj eksperimenta i replikacije je istražiti da li i u kojoj mjeri uporaba sustava xTeX-Sys uz klasične nastavne metode utječe na konačna postignuća studenata i mogu li inteligentni tutorski sustavi biti kvalitetna podrška profesoru u realizaciji tradicionalne nastave.

#### **3.3.2. Sudionici**

U eksperimentu i njegovoj replikaciji su sudjelovali studenti dvaju fakulteta u Splitu: Kemijsko-tehnološki fakultet (KTF) i Fakultet prirodoslovno-matematičkih znanosti i kineziologije (PMF) iz kolegija Uvod u računarstvo. Kontrolnoj grupi su pripadali studenti KTF-a, dok eksperimentalnoj grupi studenti PMF-a.

Originalan eksperiment je počeo u listopadu 2005. i završio u lipnju 2006. U eksperimentu je sudjelovalo 175 studentata ali je njih 120 završilo sve djelove eksperimenta (68%). Od 175 studenata koliko je sudjelovalo u originalnom eksperimentu, njih 86 je pripadalo kontrolnoj grupi, a 109 studenata je pripadalo eksperimentalnoj grupi (Prilog A).



Replikacija je počela u listopadu 2006. i završila u lipnju 2007. godine. U replikaciji je sudjelovalo 127 studenata, njih 70 je završilo sve korake replikacije (55 %). U replikaciji je sudjelovalo 127 studenata, od toga je 52 studenta su pripadala kontrolnoj grupi dok njih 75 eksperimentalnoj grupi (Prilog B). U toku svakog istraživanja prikupljeni su dodatni podaci o studentima. Studente su pitana pitanja osobnog karaktera (starost, spol), nivo sveučilišne naobrazbe, način učenja. Studenti su dobrovoljno odgovarali na takva pitanja.

### 3.3.3. Provedba eksperimenta i replikacije

Originalan eksperiment i njegova replikacija provedeni su po određenom planu. Nakon kratkog uvoda u kojemu je objašnjena svrha eksperimenta, osobni podaci i informacije o dosadašnjoj upoznatosti sa kolegijem su prikupljene putem upitnika. Onda je obavljeno inicijalno testiranje (inic.) i sakupljene su informacije o zavisnoj varijabli, u ovom slučaju znanju studenta. Nakon završenog prvog koraka obje grupe su obavile prvo parcijalno testiranje (kol1.), nakon završenog drugog koraka obe grupe su pristupile rješavanju drugog parcijalnog testa (kol2.) i na kraju eksperimenta su imali završno testiranje (završni).

Da bi se postigla ekvivalentnost grupa potrebno je analizirati dobivene rezultate, odnosno rezultate inicijalnog testa za eksperiment i njegovu replikaciju.

#### 3.3.3.1. Provjera ekvivalentnosti grupa za originalan eksperiment

U originalnom eksperimentu od 86 studenata kontrolne grupe (KTF) njih 40 je završilo sve dijelove eksperimenta, od 109 studenata eksperimentalne grupe (PMF) njih 80 je završilo sve dijelove eksperimenta. Nakon provedenog izjednačavanja parova studenata prema rezultatima inicijalnog testa korištenjem metode izjednačavanja u rasponu treba statistički provjeriti da li se eksperimentalna i kontrolna grupa razlikuju.

U tablici 3.1. su dane frekvencije rezultata inicijalnog testa kontrolne grupe i proširene eksperimentalne grupe.

**Tablica 3.1. Frekvencije bodova inicijalnog testa**

Bodovi	Kontrolna grupa	Proširena eksperimentalna grupa	Eksperimentalna grupa
0-10	0	0	0
10-20	3	0	0
20-30	3	2	2
30-40	7	14	7
40-50	5	11	5
50-60	10	21	10
60-70	5	15	5
70-80	6	10	6
80-90	1	7	1
90-100	0	0	0

Iz tablice 3.1. se vidi da trebamo odabrati 40 studenata iz proširene eksperimentalne grupe u rasponu od 30-100 bodova. U proširenoj eksperimentalnoj grupi nema studenata koji imaju od 10-20 bodova, a ima samo 2 studenta sa rezultatom od 20-30 bodova dok ih je u kontrolnoj grupi 3. Da bi kontrolna grupa i eksperimentalna bile ekvivalentne potrebno je odabrati još 4 studenta slučajnim odabirom iz proširene eksperimentalne grupe.

Nakon provedenog izjednačavanja, u originalnom eksperimentu smo odabrali po 40 studenata za kontrolnu i 40 studenata za eksperimentalnu grupu i njihove frekvencije su prikazane u tablici 3.2.

**Tablica 3.2. Frekvencije kontrolne i eksperimentalne grupe**

Bodovi	Kontrolna grupa	Eksperimentalna grupa
0-10	0	0
10-20	3	0
20-30	3	2
30-40	7	9
40-50	5	6
50-60	10	11
60-70	5	5
70-80	6	6
80-90	1	1
90-100	0	0
Ukupno	N <sub>c</sub> =40	N <sub>e</sub> =40

Da bi provjerili statistički ekvivalentnost grupa treba izračunati aritmetičku sredinu i standardnu devijaciju rezultata inicijalnog testa posebno za kontrolnu i eksperimentalnu grupu. Nakon toga se računa standardna pogreška razlika aritmetičkih sredina te da bi na kraju odredili da li je ta pogreška statistički značajna računamo odnos razlike aritmetičke sredine (eksperimentalne i kontrolne grupe) i standardne pogreške razlike aritmetičke sredine, odnosno t-vrijednost što je prikazano u tablici 3.3. Zajedničku standardnu devijaciju za eksperimentalnu i kontrolnu grupu možemo računati ako su grupe jednako velike ili im se varijance značajno ne razlikuju. Kako su grupe jednako velike ne moramo računati F.

**Tablica 3.3. Ispitivanje statističke značajnosti razlike**

	KONTROLNA GRUPA	EKSPERIMENTALNA GRUPA
$\bar{X}$	50,00	52,31
$\sigma = \sqrt{\frac{(X - \bar{X})^2}{N-1}}$	18,01	14,75
$\sigma_{zajednicka} = \sqrt{\frac{(N_C - 1)\sigma_C^2 + (N_E - 1)\sigma_E^2}{N_C + N_E}}$	16,25	
$t = \frac{\bar{X}_C - \bar{X}_E}{S_{x_C - x_E}}$	-0,627	

Da bi odredili da li je razlika statistički značajna trebamo iz tablice graničnih t- vrijednosti očitati t. Prije toga potrebno je odrediti stupnjeve slobode koji se računaju po formuli (N<sub>C</sub>-1)+(N<sub>E</sub>-1), a u našem slučaju to je 78. Dakle, granična t-vrijednost na razini značajnosti od 5% je 1,99. S obzirom da je apsolutna vrijednost od dobivene t-vrijednosti manja od granične t-vrijednosti, to znači da se eksperimentalna i kontrolna grupa statistički značajno ne razlikuju.

### 3.3.3.2. Povjera ekvivalentnosti grupa za replikaciju eksperimenta

Kod repliciranog eksperimenta od 52 studenata kontrolne grupe (KTF) njih 19 je završilo sve dijelove eksperimenta, a od 75 studenata eksperimentalne grupe (PMF) njih 51 je rješilo sve dijelove eksperimenta. Nakon provedenog izjednačavanja parova studenata prema rezultatima inicijalnog testa korištenjem metode izjednačavanja u rasponu treba statistički provjeriti da li se eksperimentalna i kontrolna grupa razlikuje. U tablici 3.4. su dane frekvencije bodova inicijalnog testiranja.

**Tablica 3.4. Frekvencije bodova inicijalnog testa**

Bodovi	Kontrolna grupa	Proširena eksperimentalna grupa	Eksperimentalna grupa
0-10	0	0	0
10-20	0	0	0
20-30	3	1	1
30-40	9	9	9
40-50	3	11	3
50-60	2	16	2
60-70	0	9	0
70-80	2	5	2
80-90	0	0	0
90-100	0	0	0

Da bi obe grupe bile ekvivalentne trebamo odabrati 19 studenata eksperimentalne grupe iz proširene eksperimentalne grupe. Iz tablice 3.4. se vidi da u rasponu bodova od 20-30 u kontrolnoj grupi ima 3 studenata dok u proširenoj eksperimentalnoj grupi ima 1 student pa treba metodom slučajnog izbora odabrati još 2 studenata. Nakon provedenog izjednačavanja odabrali smo po 19 studenata za kontrolnu i 19 studenata za eksperimentalnu grupu, njihove frekvencije su prikazane u tablici 3.5.

**Tablica 3.5. Frekvencije kontrolne i eksperimentalne grupe**

Bodovi	Kontrolna grupa	Eksperimentalna grupa
0-10	0	0
10-20	0	0
20-30	3	1
30-40	9	9
40-50	3	4
50-60	2	2
60-70	0	1
70-80	2	2
80-90	0	0
90-100	0	0
Ukupno	$N_C=19$	$N_E=19$

Nakon izjednačavanja studenata po grupama sljedeći statistička provjera ekvivalentnosti grupa na isti način kao za originalan eksperiment što je prikazano u tablici 3.6. Zajedničku standardnu devijaciju za eksperimentalnu i kontrolnu grupu možemo računati ako su grupe jednako velike ili im se varijance značajno ne razlikuju. Kako su grupe jednako velike ne moramo računati F.

Granična t-vrijednost na razini značajnosti od 5% uz 36 stupnjeva slobode iznosi 2,02. S obzirom da je apsolutna vrijednost od dobivene t-vrijednosti manja od granične t-vrijednosti, to znači da se eksperimentalna i kontrolna grupa repliciranog eksperimenta statistički značajno ne razlikuju. Kako nema statistički značajne razlike između kontrolne i eksperimentalne grupe za eksperiment i njegovu replikaciju možemo nastaviti sa daljnom analizom podataka.

Tablica 3.6. Ispitivanje statističke značajnosti razlike

	KONTROLNA GRUPA	EKSPERIMENTALNA GRUPA
$\bar{X}$	41,00	44,21
$\sigma = \sqrt{\frac{(X - \bar{X})^2}{N-1}}$	14,97	14,59
$\sigma_{zajednicka} = \sqrt{\frac{(N_C - 1)\sigma_C^2 + (N_E - 1)\sigma_E^2}{N_C + N_E}}$	14,78	
$t = \frac{\bar{X}_C - \bar{X}_E}{S_{x_C - x_E}}$	-0,69	

### 3.3.4. Analiza podataka i interpretacija

U analizi podataka se provodi testiranje standardne značajnosti (eng. standard significance) za istraživanje učinka postupka na zavisnoj varijabli. Nul- hipoteze su postavljene za svaki parcijalni test provjere (kol1, kol2) i završni test provjere:

**H<sub>0kol1</sub>**: "Ne postoji statistički značajna razlika između kontrolne i eksperimentalne grupe kod rezultata prvog parcijalnog testa."

**H<sub>0kol2</sub>**: "Ne postoji statistički značajna razlika između kontrolne i eksperimentalne grupe kod rezultata drugog parcijalnog testa."

**H<sub>0završni</sub>**: "Ne postoji statistički značajna razlika između kontrolne i eksperimentalne grupe kod rezultata završnog testa."

Za testiranje hipoteza koristimo dvosmjerni t-test (eng. two-tailed) za testiranje razlike. Kod dvosmjernog testiranja razlike za statističku značajnost od 5% potreban je t od najmanje 1,96. Preduvjet za korištenje t-testa je normalna raspodjela varijabli. Najprije treba izračunati aritmetičku sredinu (prema formuli 2.19), medijan i standardnu devijaciju (formula 2.20) za inicijalni test, prvi i drugi parcijalni test te završni test za eksperiment i njegovu replikaciju. Nakon toga potrebno je izračunati razliku za aritmetičku sredinu, medijan i standardnu devijaciju za svaki parcijalni test provjere i inicijalni test isto tako i za završni test i inicijalni test. Isti postupak se ponovi i za replicirani eksperiment. T-vrijednost se računa od te razlike. Ako je ta t-vrijednost statistički značajna za svaki test (oba parcijalna testa i završni test), to znači da nezavisna varijabla (sustav e-učenja) ima pozitivan učinak na zavisnu varijablu (znanje učenika).

### 3.3.4.1. Analiza i interpretacija podataka za originalan eksperiment

U tablici 3.7. je prikazana izračunata aritmetička sredina, medijan i standardna devijacija za inicijalni test, prvi i drugi parcijalni test i završni test za eksperiment i njegovu replikaciju.

**Tablica 3.7. Vrijednosti zavisne varijable**

	inic.	kol1	kol2	završni	razlika kol1-inic.	razlika kol2-inic	razlika završni-inic
kontrolna grupa (40 studenata)							
arit. sred.	50,00	40,70	54,95	37,25	-9,29	4,95	-12,75
medijan	51,49	42,50	58,00	37,00	-7,87	6,78	-13,54
st. dev.	18,01	15,76	17,36	13,79	17,73	21,68	14,56
eksperimentalna grupa (40 studenata)							
arit. sred.	52,31	46,12	46,95	51,22	-6,18	-5,36	-1,08
medijan	52,98	49,37	45,50	51,50	-8,59	-4,23	-2,00
st. dev.	14,75	16,80	12,79	12,29	18,97	17,85	13,66

Promatrajući stupce koji prikazuju razliku između parcijalnih testova i završnog testa te inicijalnog testa uočavamo negativne vrijednosti. Nula ili negativna razlika između aritmetičke sredine prvog parcijalnog testa i inicijalnog testa se javlja dva puta. Negativna razlika ili nula između aritmetičke sredine drugog parcijalnog testa i inicijalnog testa se javila samo jednom, dok kod razlike završnog testa i inicijalnog testa negativna razlika ili nula se javila dvaput.

#### Testiranje nul-hipoteza $H_{0_{kol1}}$ , $H_{0_{kol2}}$ i $H_{0_{završni}}$

Da bi testirali hipoteze koristit ćemo dvosmjerni t-test za zavisnu grupu. Potrebno je izračunati veličinu učinka ( $\Delta$ , formula 2.24), odrediti stupnjeve slobode (df,  $df = N - 1$ , gdje je N veličina uzorka), odrediti t-vrijednost (formula 2.21), graničnu vrijednost t (koja se uzima iz tablice granične vrijednosti t uz prethodno određenu razinu značajnost  $\alpha$ ) i p- vrijednost. Testirati ćemo nul-hipoteze zajedno.

U tablici 3.8 su dani podaci potrebni za testiranje nul-hipoteza  $H_{0_{kol1}}$ ,  $H_{0_{kol2}}$  i  $H_{0_{završni}}$ . U prvom stupcu je izračunata veličina učinka, u drugom stupnjevi slobode, u trećem t- vrijednost, u četvrtom granična vrijednost uz razinu značajnosti  $\alpha = 0,05$  i u petom stupcu p-vrijednost.

**Tablica 3.8. Podaci potrebni za testiranje nul-hipoteza  $H_{0_{kol1}}$ ,  $H_{0_{kol2}}$  i  $H_{0_{završni}}$**

	$\Delta$	df	t-vrijednost	granični t $\alpha = 0,05$	p-vrijednost
$H_{0_{kol1}}$	0,17	78	-0,75	1,99	0,4555
$H_{0_{kol2}}$	-0,47	78	2,32	1,99	0,0230
$H_{0_{završni}}$	0,80	78	-3,69	1,99	0,0004

Kod testiranja nul-hipoteze  $H_{0_{kol1}}$  t-vrijednost iznosi -0,75 i njena apsolutna vrijednost je manja od dobivene granične vrijednosti koja je 1,99, a to znači da razlika nije statistički značajna i hipotezu  $H_{0_{kol1}}$  prihvaćamo. S obzirom da smo dobili negativnu t-vrijednost to znači da je eksperimentalna grupa možda bolja od kontrone grupe ali to nije statistički značajno. Veličina učinka je pozitivna, znači da postoji pozitivan učinak učenja iako nije statistički značajno. U terminima standardne devijacije to znači da je eksperimentalna grupa bolja od kontrolne za 0,17 standardne devijacije.

Što se tiče nul-hipoteze  $H_{0_{kol2}}$  t-vrijednost veća od granične vrijednosti znači da je razlika statistički značajna, tj. odbacili smo nul-hipotezu  $H_{0_{kol2}}$ . Znači da postoji statistički značajna razlika između eksperimentalne i kontrolne grupe u drugom parcijalnom testu provjere. Kako je dobivena t-vrijednost pozitivna to je kontrolna grupa sigurno bolja od eksperimentalne. U originalnom eksperimentu nema pozitivnog učinka učenja, kontrolna grupa bolja od eksperimentalne za 0,47 standardne devijacije.

Testirajući nul-hipotezu  $H_{0_{završni}}$  dobivena apsolutna t-vrijednost je veća od granične znači da je razlika statistički značajna i odbacili smo nul-hipotezu  $H_{0_{završni}}$ . Dakle, postoji statistički značajna razlika između kontrolne i eksperimentalne grupe kod rezultata završnog testa. S obzirom da je dobivena t- vrijednost negativna eksperimentalna grupa je sigurno bolja od kontrolne i to je statistički značajno. Veličina učinka je pozitivna i statistički značajna. Eksperimentalna grupa je bolja od kontrolne za 0,80 standardne devijacije.

U sljedećoj tablici (Tablica 3.9.) su dani interpretirani podaci koji pokazuju da/ne postojanje statističke značajnosti te pozitivan odnosno negativan učinak učenja.

**Tablica 3.9. Sažeti podaci potrebni za testiranje nul-hipoteza  $H_{0_{kol1}}$ ,  $H_{0_{kol2}}$  i  $H_{0_{završni}}$**

eksperimentalna grupa nasuprot kontrolna grupa	zavisna varijabla – znanje studenta	
	statistička znač./praktična znač.	poz. veličina učinka/ neg. veličina učinka
$H_{0_{kol1}}$	ništa	+
$H_{0_{kol2}}$	stat. znač.	-
$H_{0_{završni}}$	stat. znač.	+

### 3.3.4.2. Analiza i interpretacija podataka za replikaciju eksperimenta

U tablici 3.10. je prikazana izračunata aritmetička sredina, medijan i standardna devijacija za inicijalni test, prvi i drugi parcijalni test i završni test.

**Tablica 3.10. Vrijednosti zavisne varijable**

	inic.	kol1	kol2	završni	razlika kol1-inic.	razlika kol2-inic	razlika završni-inic
kontrolna grupa (19 studenata)							
arit. sred.	41,00	54,73	31,89	40,78	13,73	-9,10	-0,21
medijan	35,00	55,00	27,00	37,00	14,00	-9,00	3,00
st. dev.	14,97	19,62	23,29	11,78	19,62	23,29	11,78
eksperimentalna grupa (19 studenata)							
arit. sred.	44,21	53,00	44,52	57,15	8,78	0,31	12,94
medijan	38,00	48,00	48,00	56,00	7,00	-6,00	11,00
st. dev.	14,59	17,68	22,15	11,74	17,68	22,15	11,74

Promatrajući stupce koji prikazuju razliku između parcijalnih testova i završnog testa te inicijalnog testa uočavamo negativne vrijednosti. Nula ili negativna razlika između aritmetičke sredine prvog parcijalnog testa i inicijalnog testa se ne javlja ni jednom. Negativna razlika ili nula između aritmetičke sredine drugog parcijalnog testa i inicijalnog testa se javila samo jednom, dok kod razlike završnog testa i inicijalnog testa negativna razlika ili nula se javila jednom.

### Testiranje nul-hipoteza $H_{0_{kol1}}$ , $H_{0_{kol2}}$ i $H_{0_{zavrnsni}}$

Kao i kod originalnog eksperimenta za testiranje nul-hipoteza koristit ćemo dvosmjerni t-test za zavisnu grupu. Znači isto kao i kod originalnog eksperimenta potrebno je izračunati veličinu učinka ( $\Delta$ , formula 2.24), odrediti stupnjeve slobode (df,  $df = N - 1$ , gdje je N veličina uzorka), odrediti t-vrijednost (formula 2.21), graničnu vrijednost t (koja se uzima iz tablice granične vrijednosti t uz prethodno određenu razinu značajnosti  $\alpha$ ) i p- vrijednost. Testirati ćemo nul-hipoteze zajedno.

U tablici 3.11. su dani podaci potrebni za testirati nul-hipoteze  $H_{0_{kol1}}$ ,  $H_{0_{kol2}}$  i  $H_{0_{zavrnsni}}$ . U prvom stupcu je izračunata veličina učinka, u drugom stupnjevi slobode, u trećem t- vrijednost, u četvrtom granična vrijednost uz razinu značajnosti  $\alpha = 0,05$  i u petom stupcu p-vrijednost.

**Tablica 3.11. Podaci potrebni za testiranje nul-hipoteza  $H_{0_{kol1}}$ ,  $H_{0_{kol2}}$  i  $H_{0_{zavrnsni}}$**

	$\Delta$	df	t-vrijednost	granični t $\alpha = 0,05$	p-vrijednost
$H_{0_{kol1}}$	-0,25	36	0,81	2,02	0,4233
$H_{0_{kol2}}$	0,40	36	-1,27	2,02	0,2122
$H_{0_{zavrnsni}}$	1,11	36	-3,44	2,02	0,0015

Kod testiranja nul-hipoteze  $H_{0_{kol1}}$  t-vrijednost je manja od granične vrijednosti, a to znači da razlika nije statistički značajna i hipotezu  $H_{0_{kol1}}$  prihvaćamo. S obzirom da smo dobili pozitivnu t-vrijednost to znači da je kontrolna možda bolja od eksperimentalne grupe ali to nije statistički značajno. Nema pozitivnog učinka učenja i kontrolna grupa je bolja od eksperimentalne za 0,25 standardne devijacije.

Što se tiče nul-hipoteze  $H_{0_{kol2}}$  dobivena je negativna t-vrijednost čija je apsolutna vrijednost manja od granične vrijednosti pa razlika nije statistički značajna, odnosno prihvaćamo nul-hipotezu  $H_{0_{kol2}}$ . Negativna t-vrijednost upućuje da je eksperimentalna možda bolja od kontrolne grupe ali to nije statistički značajno. Veličina učinka je pozitivna ali nije statistički značajna. Eksperimentalna grupa je bolja od kontrolne za 0,40 standardne devijacije.

Kod nul-hipoteze  $H_{0_{zavrnsni}}$  je opet dobivena negativna t-vrijednost čija apsolutna vrijednost u veća od granične znači da je razlika statistički značajna i odbacili smo nul-hipotezu  $H_{0_{zavrnsni}}$ . Dakle, postoji statistički značajna razlika između kontrolne i eksperimentalne grupe kod rezultata završnog testa. S obzirom da je dobivena t-vrijednost negativna eksperimentalna grupa je sigurno bolja od kontrolne i to je statistički značajno. Veličina učinka je pozitivna i statistički značajna. Eksperimentalna grupa je bolja od kontrolne za 1,11 standardne devijacije.

U tablici 3.12. su dani interpretirani podaci koji pokazuju da/ne postojanje statističke značajnosti te pozitivan odnosno negativan učinak učenja.

**Tablica 3.12. Sažeti podaci potrebni za testiranje nul-hipoteza  $H_{0_{kol1}}$ ,  $H_{0_{kol2}}$  i  $H_{0_{zavrnsni}}$**

eksperimentalna grupa nasuprot kontrolna grupa	<u>zavisna varijabla – znanje studenta</u>	
	statistička znač./praktična znač.	poz. veličina učinka/ neg. veličina učinka
$H_{0_{kol1}}$	ništa	-
$H_{0_{kol2}}$	ništa	+
$H_{0_{zavrnsni}}$	stat. znač.	+

### 3.3.5. Zaključak

Ovo istraživanje je provedeno da da bi utvrdili učinkovitost učenja uz pomoć inteligentnog tutorskog sustava xTeX-Sys. U tablici 3.13. su prikazani rezultati testiranja svih nul-hipoteza i sažeti prikaz koja je grupa bolja kod testiranja pojedine nul-hipoteze.

**Tablica 3.13. Rezultati testiranja hipoteza za eksperiment i replikaciju**

	nul-hipoteza H0	bolja grupa
<b>kol1</b>		
E	prihvaćena	eksperimentalna (stat. ne znač.)
R	prihvaćena	kontrolna (stat. ne znač.)
<b>kol2</b>		
E	odbaćena	kontrolna (stat. znač.)
R	prihvaćena	eksperimentalna (stat. ne znač.)
<b>završni</b>		
E	odbaćena	eksperimentalna (stat. znač.)
R	odbaćena	eksperimentalna (stat. znač.)

Kod prvog parcijalnog testa nul-hipoteza je prihvaćena i za originalni eksperiment i njegovu replikaciju. Kod drugog parcijalnog testa nul-hipoteza je odbaćena kod originalnog eksperimenta dok je kod replikacije prihvaćena. Kod završnog testa nul-hipoteza je odbaćena za oba eksperimenta.

U tablici 3.14. su prikazani rezultati veličine učinka za svaki pojedini parcijalni test provjere i završni test kako za eksperiment tako i za replikaciju te je izračunata ukupna veličina učinka.

**Tablica 3.14. Sažeti prikaz veličine učinka za eksperiment i replikaciju**

	veličina učinka	
	originalan eksperiment	replikacija
kol1	0,17	-0,25
kol2	-0,47	0,40
završni	0,80	1,11
ukupna veličina učinka	0,16	0,42

Veličina učinka u originalnom eksperimentu za prvi parcijalni test iznosi 0,17 što je mala veličina učinka prema [COHE1969], ali dobivena t-vrijednost (-0,75) u originalnom eksperimentu upućuje da to nije statistički značajna razlika između kontrolne i eksperimentalne grupe. Što se tiče drugog parcijalnog testa provjere dobivena je mala veličina učinka prema [COHE1969] od -0,47, a dobivena t-vrijednost (2,32) upućuje na postojanje statistički značajne razlike među grupama i to u korist kontrolne grupe što nas posebno iznenađuje. Kod završnog testa dobivena je velika veličina učinka prema [COHE1969] od 0,80 ali je u ovom slučaju dobivena t-vrijednost (3,69) statistički značajna i to u korist eksperimentalne grupe što je bilo i za očekivati. Dobivena je ukupna veličina učinka za originalan eksperiment od 0.16 sigma, što pokazuje prema [COHE1969] malu učinkovitost sustava xTeX-Sys u procesu učenja i poučavanja prema tradicionalnim metodama.



Kod replikacije veličina učinka za prvi parcijalni test iznosi -0,25 što je mala veličina učinka prema [COHE1969], a razlika među grupama nije statistički značajna prema dobivenoj t-vrijednosti (0,81). U drugom parcijalnom testu veličina učinka iznosi 0,40 što je umjerena veličina učinka prema [COHE1969]. Prema dobivenoj t-vrijednosti (-1,27) grupe se statistički značajno razlikuju i to u korist eksperimentalne grupe. U završnom testu veličina učinka iznosi 1,11 što je velika veličina učinka prema [COHE1969], a dobivena t-vrijednost (-3,44) upućuje na to da postoji statistički značajna razlika među grupama i ovaj put u korist eksperimentalne grupe. Ukupna veličina učinka iznosi 0,42 sigma što pokazuje prema [COHE1969] umjerenu učinkovitost opisanog sustava u procesu učenja i poučavanja.

Postavlja se pitanje: Zašto su dobivene male veličine učinka za oba eksperimenta? Da li je sustav xTeX-Sys kvalitetna podrška učeniku u procesu učenja i poučavanja?

S obzirom na rezultate istraživanja očekivana je veća veličina učinka jer prema Bloom-u ciljana veličina učinka je  $2\sigma$ . Postavlja se pitanje: Kako sustav xTeX-Sys utječe na proces učenja i poučavanja? Da li učenik ima koristi od takvog poučavanja? I ako ima, kolika je ta učinkovitost u odnosu na tradicionalni oblik poučavanja?

## 4. Zaključak

Replikacija je pokušaj reprodukcije empirijske studije kako bi se dobila buduća vrijednost otkrića ili uspješan rezultata takvih pokušaja. Replikacija ne varira samo s obzirom na cilj, već također ovisi i o uvjetima pod kojima je provedena, obično ovisi o pretpostavljenoj vezi između originalnog eksperimenta i njegove replikacije. Osim toga koriste se statističke mjere i jedinice potrebne za analizu eksperimenta (ocjenjivanje, testiranje, predviđanje, analiziranje..., itd) koje se mogu bitno razlikovati. Neke se replikacije mogu provoditi sa slabašnom pomoći zajedno sa svim dostupnim podacima (iz originalnog i repliciranog eksperimenta), ostale će imati izolirane analize replikacija, treće mogu imati mogućnost kombinacije oba eksperimenta.

Iz ovoga vidimo da su mogući različiti scenariji replikacija. Ovo je možda jedan od razloga zašto manjka studija pažljivog oblikovanja replikacija. Teško je razviti univerzalan model po kojem bi se uspješno mogao oblikovati eksperiment i njegova replikacija upravo zbog raznih mogućih slučajeva replikacije.

Opisan je hijerarhijski model po kojem se uspješno oblikuje eksperiment i njegova replikacija, navedeni su ekstremni slučajevi kada je replikacija gotovo identična originalnom eksperimentu i drugi slučaj kada je originalan eksperiment i replikacija totalno nepovezana. Najčešći zapis o tome što je replicirani eksperiment se odnosi na "ispravnu" ili "identičnu" replikaciju. Kako joj ime govori, ispravna replikacija se određuje imitirajući što je moguće više originalni eksperiment. Cilj je obično nadati se jednakim zaključcima. Glavna motivacija je "vrijednost" nalaza kod originalnog eksperimenta. Ovi tipovi replikacija su jako izvođeni u eksperimentalnim znanostima, i ponekad čak u društvenim znanostima.

Kao što je prije navedeno, zbog nedostatka studija koje detaljno oblikuju eksperiment i replikaciju, detaljno smo opisali provođenje eksperimenta i replikacije. Dana je struktura eksperimenta i replikacije, iznesni su ciljevi i sudionici, opisana je metoda izjednačavanja sudionika po grupama kao i statistička provjera ekvivalentnosti grupa kako za eksperiment tako i za njegovu replikaciju. Zatim je pokazano kako postaviti nul-hipoteze, odrediti veličinu učinka koja će se računati te kako se analiziraju i interpretiraju rezultati eksperimenta i replikacije. I na kraju poglavlja, potaknuti radom Pfahla i njegovih suradnika prikazani su rezultati dvaput ponovljenog eksperimenta koji se bavi vrednovanjem učinkovitosti učenja uz pomoć simulacijskog modela.

Današnje doba modernih tehnologija i globalizacije donosi brze promjene u svim aspektima ljudskog života. Svakim danom stvaraju se nove informacije, a opći razvoj kontinuirano zahtijeva nova znanja i vještine. Javlja se potreba za što bržim, pravovremenim obrazovanjem, koje će istovremeno biti otvoreno i široko dostupno. U tom pogledu sustav e-učenja je sve prisutniji u procesu učenja i poučavanja, a pri tom utječe i na znanje učenika.

Proveli smo istraživanje o vrednovanju učinkovitosti inteligentnog tutorskog sustava xTEx-Sys kao posebne klase inteligentnih sustava e-učenja a time i sustava e-učenja. Dobivena je ukupna veličina učinka za originalan eksperiment 0,16 sigma što je prema [COHE1969] mala veličina učinka, dok je za replikaciju eksperimenta dobivena veličina učinka od 0,42 sigma što je prema [COHE1969] umjerena veličina učinka. Iako su dobivene male veličine učinka, u repliciranom

eksperimentu postoji pozitivan napredak u vrednovanju učinkovitosti sustava xTEx-Sys. S obzirom na dobivene male veličine učinka postavlja se pitanje: Da li je sustav xTEx-Sys kvalitetna podrška u procesu učenja i poučavanja?

Ovo istraživanje je provedeno sa studentima prve godine KTF-a i PMF-a kada su se studenti prvi put susreli sa sustavom xTEx-Sys što je isto tako moglo loše utjecati na konačne rezultate. Bilo bi dobro da se studenti kroz duži period poučavaju uz pomoć sustava xTEx-Sys kako bi ga bolje upoznali pa da se onda provedu ovakva istraživanja. Vjerujem da bi tada bila veća učinkovitost danog sustava e-učenja.

Dobiveni rezultati nisu ni približno očekivanim rezultatima, ali mogu biti dodatna motivacija i poticaj za nastavak razvoja sustava xTEx-Sys, kako njegova unapređivanja i daljnjeg razvitka tako i sve većoj implementaciji u procesu učenja i poučavanja. Daljna istraživanja trebala bi biti fokusirana prema smjeru povećanja učinkovitosti sustava xTEx-Sys. Što se tiče budućih istraživanja, dobro bi bilo ponoviti ovakav eksperiment i njegovu replikaciju, ali na većem uzorku kako bi se mogli usporediti rezultati s dobivenim rezultatima.

Sustavi e-učenja ne mogu u cijelosti zamjeniti profesora, ali mogu olakšati rad profesora i ujedno im omogućiti više vremena potrebnog za razgovor s učenicima, razglabanje o određenim problemima te češće provjere znanja koje inače nemoguće provoditi zbog nedostatka vremena u tradicionalnoj nastavi. Naime, u današnje vrijeme profesori zbog sve većeg obujma gradiva za koje je predviđeno određeni broj sati, nemaju vremena za ponavljane određenog znanja kao i raspravu sa učenicima o mogućim neispravnostima. Zbog takvih ograničenja učenici mogu ostvariti loše rezultate na ispitima. Upravo zbog takvih problema sustav e-učenja može pomoći profesorima da lakše organiziraju kvalitetniju nastavu u kojoj bi bilo više vremena za posvetiti se potrebama učenika.

---

## 5. Literatura

- [ALMQ2006] Almqvist J. P. F.(2006) Replication of Controlled Experiments in Empirical Software Engineering- A Survey, Department of Computer Science, Faculty of Science, Lund University.
- [BAYA1997] Bayarri M. J., Mayoral A. M. (1997) Dessining "Successful" Replicatins, Universitat de Valencia.
- [BLOO1984] Bloom, B.S. (1984) The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 4-16.
- [BUJA1967] Bujas Z (1967) Uvod u metode eksperimentalne psihologije, Zagreb, Školska knjiga, 11-79.
- [BURN1988] Burns H. L., Capps C. G. (1988) Foundations of intelligent tutoring systems: an introduction, M.C.Poison, J. J.Richardson (Ed.) Foundations of intelligenttutoring systems, Lawrence Erlbaum, London.
- [COHE1969] Cohen, J.(1969) Statistical Power Analysis for the Behavioral Sciences, NY: Academic Press, 19-33.
- [GRUB2007a] Grubišić A., Stankov S., Žitko B.(2007) Experiment Replication in Evaluation of E-Learning System's Effectiveness, Fakultet prirodoslovno-matematičkih znanosti i kineziologije, Split.
- [GRUB2007b] Grubišić A. (2007) Vrednovanje učinka inteligentnih sustava e-učenja (magistarski rad), Zagreb.
- [KASS1993] Kass R. E., Raftery A. E. (1993) Bayes factors and model uncertainty, University of Washington.
- [JALŠE2008] Jalšenjak B., Filozofija.org, FFDI, Zagreb, URL: <http://www.filozofija.org> (25/04/08)
- [LUCA2003] Lucas J. W.(2003) Theoey-Testing, Generalization, and the Problem of External Validity, The University of Akron.
- [MEJO2003] Mejovšek M. (2003) Uvod u metode znastvenog istraživanja, Edukacijsko-rehabilitacijski fakultet, Naklada Slap, 35-112.
- [PETZ2007] Petz B. (2004) Osnovne statističke metode za nematematičare, Naklada Slap.

- 
- [PFAH2004] Pfahl<sup>a</sup> D., Laitenberger<sup>b</sup> O., Ruhe<sup>c</sup> G., Dorsch<sup>d</sup> J., Krivobokova<sup>e</sup> T. (2004) Evaluating the learning effectiveness of using simulations in software project management education: result from a twice replicated experiment, <sup>a</sup>Fraunhofer IESE, Germany, <sup>b</sup>Droege & Comp., Germany, <sup>c</sup>University of Calgary, Canada, <sup>d</sup>Accenture, Campus kronberg 1, Germany, <sup>e</sup>University of Bielefeld, Germany.
- [ROSI2000] Rosić M. (2000) Zasnivanje sustava obrazovanja na daljinu unutar informacijske infrastrukture (magistarski rad), Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu.
- [SRCE2008] Sveučilišni računski centar – Srce, URL: [http://www.srce.hr/ceu/e\\_ucenje.html](http://www.srce.hr/ceu/e_ucenje.html) (24/05/08)
- [STAN1997] Stankov, S. (1997) Isomorphic Model of the System as the Basis of Teaching Control Principles in an Intelligent Tutoring System. Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Split, Croatia, PhD Thesis.
- [STAN2003] Stankov S. (2003-2005) (glavni istraživač) Tehnološki projekt MZT: Web orijentirana inteligentna hipermedijska autorska ljuška (TP-02/0177-01).
- [STAN2004] Stankov, S., Grubišić, A., Žitko B. (2004) E-learning paradigm & Intelligent tutoring systems. In: Kniewald, Z. (ed.): Annual 2004 of the Croatian Academy of Engineering. Croatian Academy of Engineering, Zagreb.
- [WIKI2008] E-learning from Wikipedia, the free encyclopedia, URL: <http://en.wikipedia.org/wiki/E-learning> (24/05/08)

## 6. Prilozi

### Prilog A: rezultati testova za originalan eksperiment

eksperimentalna	grupa			
Ime i prezime	inicijalni test	kolokvij 1	kolokvij 2	završni test
student4	52	59	31	43
student6	78	73	67	68
student7	72	50	41	54
student9	61	24	21	43
student12	31	11	31	23
student13	54	51	50	65
student14	42	26	50	40
student15	57	28	27	46
student16	55	48	36	53
student25	57	56	34	47
student26	72	38	45	58
student27	75	53	53	60
student28	64	53	39	46
student33	58	49	44	53
student34	73	63	60	68
student36	51	35	26	33
student37	36	34	53	41
student39	31	24	53	47
student40	60	66	57	62
student42	55	53	61	80
student43	43	24	41	53
student44	49	61	66	65
student45	42	66	50	59
student46	64	85	69	71
student52	37	55	39	61
student58	34	43	43	44
student59	27	66	42	38
student64	63	29	42	59
student65	37	59	57	50
student66	39	23	53	43
student67	37	75	71	62
student68	45	34	40	32
student71	75	38	45	49
student72	39	28	54	44
student73	45	41	28	40
student79	81	50	62	59
student89	30	36	37	31
student91	60	50	68	62
student93	51	38	46	56
student101	63	56	46	41

---

<b>kontrolna</b>	<b>grupa</b>			
<b>Ime i prezime</b>	<b>inicijalni test</b>	<b>kolokvij 1</b>	<b>kolokvij 2</b>	<b>završni test</b>
student103	55	57	60	47
student104	58	34	67	49
student105	31	25	53	30
student106	72	50	52	50
student107	60	29	48	34
student108	63	36	64	37
student109	25	64	64	52
student113	45	53	73	31
student114	54	49	63	33
student117	39	14	50	14
student119	25	19	5	20
student120	58	60	58	50
student122	45	10	17	12
student126	45	48	61	50
student128	75	56	64	51
student129	49	45	65	37
student134	52	15	50	21
student139	39	38	84	54
student141	13	15	58	17
student143	58	43	52	39
student144	51	43	42	35
student146	16	50	38	32
student149	43	45	71	44
student150	75	45	72	53
student151	81	50	61	50
student153	33	25	63	25
student154	63	58	74	35
student155	75	43	54	55
student156	37	24	57	13
student157	75	64	47	42
student158	64	49	31	36
student159	18	10	22	21
student160	64	44	66	38
student162	78	35	27	50
student164	37	40	52	24
student168	51	39	66	43
student169	63	61	85	63
student170	31	39	55	22
student173	30	39	34	23
student175	55	71	73	58

## Prilog B: rezultati testova za replikaciju eksperimenta

<b>eksperimentalna</b>	<b>grupa</b>			
<b>Ime i prezime</b>	<b>inicijalni test</b>	<b>kolokvij 1</b>	<b>kolokvij 2</b>	<b>završni test</b>
student2	32	33	67	55
student5	49	53	73	51
student10	38	91	40	63
student11	68	85	52	70
student13	31	61	1	34
student15	35	53	36	53
student16	70	81	64	70
student19	80	82	56	66
student20	50	48	52	76
student23	54	75	48	65
student25	38	71	75	61
student30	51	27	30	56
student37	41	48	80	56
student38	33	24	33	37
student39	34	32	69	62
student50	37	47	25	47
student53	42	31	25	52
student70	23	42	0	47
student73	34	23	20	65

<b>kontrolna</b>	<b>grupa</b>			
<b>Ime i prezime</b>	<b>inicijalni test</b>	<b>kolokvij 1</b>	<b>kolokvij 2</b>	<b>završni test</b>
student78	39	53	37	45
student79	27	56	32	20
student80	34	28	49	37
student82	50	36	2	34
student85	35	77	25	46
student90	34	32	9	36
student91	21	37	17	40
student92	35	55	20	37
student95	35	46	26	45
student96	51	54	42	44
student100	33	37	76	37
student103	54	55	0	40
student109	22	60	53	29
student110	38	67	37	12
student113	47	75	27	32
student116	33	59	10	24
student118	41	83	14	55
student119	71	39	67	76
student127	79	91	63	86