

# Evaluarea utilizabilității unui scenariu de învățare a biologiei implementat pe o platformă de realitate îmbogățită

Costin Pribeanu, Dragoș Daniel Iordache și Alexandru Balog

Institutul Național de Cercetare-Dezvoltare în informatică – ICI

Bd. Mareșal Averescu, Nr. 8-10, 011455, București

E-mail: [pribeanu@ici.ro](mailto:pribeanu@ici.ro), [iordache@ici.ro](mailto:iordache@ici.ro), [alex@ici.ro](mailto:alex@ici.ro)

**Rezumat.** Combinația realului cu virtualul din sistemele de realitate îmbogățită necesită tehnici de interacțiune adecvate care trebuie testate cu utilizatori pentru a evita problemele de utilizabilitate. Evaluarea formativă are ca scop identificarea problemelor de utilizabilitate cât mai devreme și este utilă în dezvoltarea sistemelor interactive caracterizate prin tehnici de interacțiune inovative. Acest articol prezintă o abordare în evaluarea formativă a unui scenariu de biologie dezvoltat pe o platformă educațională bazată pe realitate îmbogățită. Evaluarea a fost efectuată în timpul unei școli de vară și după terminarea acesteia. Au fost utilizate două metode : inspecția de utilizabilitate și testarea cu utilizatori. Ideea de bază a acestei abordări a fost de a efectua testarea cu utilizatori de două ori. În acest sens, scenariul a fost testat cu un număr mic de utilizatori în timpul școlii de vară cu scopul de a obține un feedback rapid de la elevi având cunoștințe bune de biologie. Apoi, testarea a fost repetată în condiții diferite și cu un număr mai mare de utilizatori reprezentativi. Pe parcursul ambelor experimente, au fost colectate măsuri cantitative și calitative administrând un chestionar de utilizabilitate și analizând fișierelor de log. În acest articol se descriu experimentele, se prezintă metodele utilizate și se analizează comparativ rezultatele.

**Cuvinte cheie:** proiectare centrată pe utilizator, utilizabilitate, evaluare formativă, realitate îmbogățită.

## 1. Introducere

Dezvoltarea sistemelor de realitate îmbogățită – AR (Augmented Reality) provoacă proiectanții să găsească noi paradigme de interacțiune care să fructifice posibilitățile de combinare a realului cu virtualul. Obiectele reale devin parte a spațiului de interacțiune, fiind astfel utilizate ca obiecte de interacțiune versatile, având diferite roluri. În pofida proliferării aplicațiilor de realitate îmbogățită, se constată o lipsă de metode de proiectare centrată pe utilizator (Bach & Scapin, 2004 ; Swann & Gabbard, 2005). Pe de o

parte, sistemele AR sunt scumpe și necesită efort de cercetare și proiectare pentru a dezvolta software de vizualizare în 3D. Pe de altă parte, conform cu Gabard et al. (2004), componentele de interacțiune sunt adesea slab proiectate și reduc utilizabilitatea sistemului.

Evaluarea formativă a utilizabilității este efectuată iterativ, pe parcursul ciclului de dezvoltare cu scopul identificării și eliminării problemelor de utilizabilitate cât mai devreme (Teofanos & Quesenbery, 2005). Cu cât aceste probleme sunt identificate mai devreme, cu atât mai puțin costisitor este efortul de rezolvare. Acest tip de evaluare este denumită formativă, pentru a o distinge de evaluarea sumativă, care este efectuată, de regulă, după ce un sistem sau o componentă a fost dezvoltată (Scriven, 1991). Evaluarea sumativă este efectuată prin testarea cu un număr relativ mare de utilizatori reprezentativi și are ca scop identificarea punctelor tari / slabe precum și compararea unor soluții de proiectare alternative sau a unor sisteme similare.

Evaluarea formativă poate fi efectuată prin testarea sistemului de către 2-5 experți (denumită și evaluare euristică) și/sau prin testarea cu un număr relativ mic de utilizatori (denumită și evaluare centrată pe utilizator).

În această lucrare se prezintă o abordare în evaluarea formativă a unui scenariu de învățare a biologiei implementat pe o platformă AR, în cadrul proiectului de cercetare ARiSE (Augmented Reality in School Environments). Proiectul este realizat într-un consorțiu de șapte parteneri, în care fiecare partener are un rol specific. Platforma hardware a fost proiectată de Franunhofer IAIS (Germania) iar programele de aplicație de Czech Technical University in Prague (Republica Cehă). Selectarea lecțiilor relevante și evaluarea pedagogică a scenariilor este făcută de Universitatea din Siauliai (Lithania) iar evaluarea utilizabilității de către ICI București (România). În consorțiu participă și două școli: Siauliai City Juventa Basic-School (Lituania) și Rabanus-Maurus Gymnasiums Mainz (Germania).

Obiectivul principal al proiectului este testarea eficacității pedagogice a introducerii tehnologiei AR în școli și crearea unei colaborări la distanță între elevi. ARiSE va crea o nouă tehnologie pentru învățare, platforma educațională de realitate îmbogățită – ARTP (Augmented Reality Teaching Platform) în trei etape, rezultând trei prototipuri de cercetare. Fiecare prototip va fi caracterizat de un nou scenariu de interacțiune. Primul prototip are ca disciplină țintă biologia pentru clasa a 7-a. Sarcina generică este vizualizarea și are ca scop creșterea înțelegerii și motivației elevilor pentru

învățarea sistemului digestiv uman.

Pentru a avea un feedback rapid atât de la elevi cât și de la profesori, fiecare prototip este testat cu utilizatori în timpul școlii de vară ARiSE care este organizată anual. La școala de vară participă cele două școli partenere în proiect, din Germania și Lituania, precum și 2-3 școli invitate. Din fiecare școală, au fost selectați de către profesori câte 4 elevi (2 fete și 2 băieți) pe baza cunoștințelor la biologie și a abilităților de comunicare, incluzând vorbirea limbii engleze. Având în vedere aceste criterii, grupul nu este reprezentativ pentru populația utilizatorilor.

Prima versiune a scenariului de biologie a fost implementată în 2006 și testată cu utilizatori la prima școală de vară ARiSE care a fost organizată în Hamrun, Malta. Intrucât rezultatele evaluării utilizabilității nu au fost satisfăcătoare (Lamauskas et al., 2007; Pribeanu et al. 2007), tehnicile de interacțiune au fost reproiectate și testate din nou, la a doua școală de vară care a avut loc în București în octombrie 2007.



Figura 1. Două eleve testând aplicația de biologie în 2007

Idea de bază a modului de abordare a evaluării formative a fost să testăm cu un număr mic de utilizatori în timpul școlii de vară, cu scopul de a obține un feedback rapid de la elevi având cunoștințe bune în biologie și de a repeta apoi testarea în condiții diferite și cu un număr mai mare de utilizatori reprezentativi. Aceasta înseamnă de fapt a efectua evaluarea în două etape și

de a analiza și compara rezultatele.

Evaluarea formativă a fost completată cu o inspecție de utilizabilitate, care a furnizat măsuri cantitative (număr de probleme de utilizabilitate identificate pe grade de severitate) și calitative (descrieri detaliate ale fiecărei probleme de utilizabilitate).

Restul articolului este organizat după cum urmează. În secțiunea următoare vom descrie cele două experimente. În secțiunea 3 vom prezenta metoda și rezultatele evaluării euristice. În secțiunea 4 vom prezenta rezultatele testării cu utilizatori în cele două experimente, vom compara și discuta similaritățile și diferențele dintre acestea. În aceeași secțiune, vom face și o analiză a rezultatelor obținute cu ajutorul celor două metode de evaluare formativă. Lucrarea se încheie cu concluzii și direcții de continuare.

## **2. Testarea cu utilizatori**

### **2.1 Contextul evaluării**

#### **In timpul școlii de vară**

Cea de a doua școală de vară a fost organizată în ICI București, în perioada 24-28 Octombrie 2007. Au participat două grupuri de câte patru elevi însoțiți de profesori de la școlile partenere în proiect din Germania și Lituania și trei școli generale din București.

Scenariul de biologie a fost testat în prima zi iar cel de chimie a doua zi. Testarea și discuțiile cu utilizatorii au avut loc dimineața iar discuțiile între parteneri după-amiaza.

#### **După școala de vară**

Al doilea experiment a fost organizat în ICI București, în perioada 1-15 noiembrie 2007. La testare au participat elevii din două clase a 8-a, fiecare de la altă școală. Elevii au venit în grupuri de 6-8, acompaniați de un profesor, astfel încât testarea a fost organizată în două sesiuni.

De regulă, testarea a avut loc după-amiaza (elevii au ore dimineața). Ambele scenarii au fost testate în aceeași zi, fapt care a condus la unele restricții de timp. Ordinea testării celor două scenarii a fost schimbată în

fiecare zi, astfel încât jumătate dintre elevi au testat mai întâi scenariul de biologie iar restul scenariul de chimie.

## 2.2 Platforma hardware-software

ARTP este un mediu AR de tip „seated” (Wind, Riege & Bogen, 2007): utilizatorii au în față un ecran „see-through”, pe care sunt suprapuse imagini ale obiectelor virtuale (imagini generate de calculator) peste imaginea observată a unui obiect real.

În cazul nostru, obiectul real este un mulaj aplatizat al sistemului digestiv uman. Platforma cuprinde patru module independente, organizate în cadrul unei mese pe care este amplasat obiectul real. Testul a fost realizat pe platforma din ICI, echipată cu 4 module. Așa cum se observă în Figura 1, un mulaj este partajat de doi elevi care stau față în față.

Ca instrument de interacțiune a fost utilizat un dispozitiv de poziționare și selectare alcătuit dintr-o telecomandă Wii Nintendo pe care este montată o baghetă metalică la capătul căreia este o bilă colorată din plastic. Telecomanda este utilizată și pentru selectarea unui item din meniu.

## 2.3 Participanți și sarcini

### Școala de vară

În total, 20 de elevi dintre care 10 băieți și 10 fete au participat la școala de vară. Nici unul dintre elevi nu a fost familiar cu tehnologia AR. 12 elevi au fost din clasa a 8-a (cu vârste cuprinse între 13 și 14 ani), 4 din clasa a 9-a (cu vârste cuprinse între 14 și 15 ani) și 4 din clasa a 10-a (cu vârste cuprinse între 15 și 16 ani). Vârsta diferită a elevilor este datorată diferențelor legate de curriculumul la disciplina chimie în fiecare țară.

În cadrul testării, elevii au avut de îndeplinit patru sarcini de lucru: urmărirea unui program demo și efectuarea a trei exerciții. Sarcinile au fost prezentate prin intermediul unei interfețe vocale în limba națională a elevilor.

Programul demo a prezentat lecția: structura și funcțiile sistemului digestiv uman.

Exercițiul 1 a avut ca scop identificarea organelor din sistemul digestiv uman, solicitând elevului să selecteze organul al cărui nume este afișat pe ecran. Exercițiul 2 a avut ca scop testarea cunoștințelor privind modul în

care alimentele sunt digerate la nivelul fiecărui organ, solicitând elevului să selecteze elementele nutritive care sunt transformate la nivelul unui anumit organ. Exercițiul 3 a avut, de asemenea, ca scop testarea cunoștințelor privind modul în care alimentele sunt digerate la nivelul fiecărui organ, solicitând elevului să selecteze toate organele la nivelul cărora este transformat un anumit element nutritiv.

### **După școala de vară**

În total, 42 de elevi (19 băieți și 23 de fete) au participat la testare. Toți elevii au fost din clasa a 8-a, având vârsta de 13-14 ani.

Elevii au primit 3 sarcini: lecția demo, primul exercițiu și unul dintre exercițiile 2 și 3. Numărul de sarcini a fost micșorat datorită restricțiilor de timp. După terminarea sarcinilor repartizate, elevii au fost liberi să efectueze al treilea exercițiu sau să repete unul dintre cele efectuate.

## **2.4 Metodă**

### **Măsurarea utilizabilității și calității în utilizare**

Standardul ISO 9241-11 definește utilizabilitatea ca măsură în care un anumit produs poate fi utilizat de către utilizatori specificați, pentru a îndeplini obiective specificate, într-un context de lucru specificat, cu eficacitate, eficiență și satisfacție.

Standardul ISO 9126, definește utilizabilitatea prin capabilitatea produsului software de a fi ușor de înțeles, învățat, utilizat și considerat atractiv de către utilizator, atunci când este folosit în condiții specificate. În acest standard, utilizabilitatea definită în context devine calitate în utilizare, având patru componente: eficacitate, productivitate, eficiență și satisfacție.

Această definiție sugerează o perspectivă mai largă asupra utilizabilității, care să includă utilitatea și atitudinea utilizatorului față de sistem.

### **Chestionarul de utilizabilitate**

Pentru a răspunde obiectivelor evaluării în ARiSE a fost adoptată o perspectivă mai largă asupra utilizabilității. Un model cunoscut capabil să anticipeze acceptanța tehnologiei, după ce utilizatorii au avut ocazia să o testeze, este modelul de acceptanță a tehnologiei – TAM (Technology

Acceptance Model), elaborat de Davies (1989). Teoria TAM susține că intenția de utilizare a unui sistem este influențată de atitudinea utilizatorilor față de sistem, care este influențată, la rândul său, de utilitatea percepută și ușurința în utilizare a sistemului.

*Tabelul 1. Chestionarul de utilizabilitate*

1	Ajustarea ecranului "see-through" este ușoară
2	Ajustarea ochelarilor stereo este ușoară
3	Ajustarea căștilor este ușoară
4	Postul de lucru este confortabil
5	Observarea obiectului real prin ecran este clară
6	Înțelegerea modului de operare cu aplicația realitate îmbogățită este ușoară
7	Suprapunerea dintre proiecție și obiectul real este clară
8	Învățarea modului de lucru cu aplicația de realitate îmbogățită este ușoară
9	Reamintirea modului de lucru cu aplicația de realitate îmbogățită este ușoară
10	Înțelegerea explicațiilor vocale este ușoară
11	Citirea informației pe ecran este ușoară
12	Selectarea unui item din meniu este ușoară
13	Corectarea erorilor este ușoară
14	Colaborarea cu colegii este ușoară
15	Utilizarea aplicației mă ajută să înțeleg mai rapid lecția
16	Utilizând aplicația voi obține rezultate mai bune la teste
17	După utilizarea aplicației voi ști mai multe despre acest subiect
18	Sistemul face învățarea mai interesantă
19	Lucrul în grup cu colegii este stimulat
20	Îmi place să interacționez cu obiecte reale
21	Efectuarea exercițiilor este captivantă
22	Aș dori să dispun de acest sistem în școală
23	Intenționez să utilizez acest sistem pentru învățare
24	Voi recomanda altor colegi să utilizeze acest sistem
25	În general, apreciez că sistemul este ușor de utilizat
26	În general, consider că sistemul este util pentru învățare
27	În general, îmi place să învăț cu acest sistem
28	În general, apreciez că sistemul este incitant

Așa cum arată Dillon și Morris (1998), TAM furnizează din timp un feedback util privind măsura în care utilizatorii vor accepta sau nu o nouă tehnologie. Modelul TAM este utilizat în prezent pentru evaluarea unei game largi de produse și sisteme informatice (Venkatesh et al, 2006; Sun & Zhang, 2006). Pe baza acestui model, a analizei altor chestionare de

utilizabilitate și a rezultatelor obținute la Școala de vară ARiSE din 2006, a fost elaborat un nou chestionar, prezentat în Tabelul 1.

Chestionarul de utilizabilitate cuprinde 28 de itemi. Primii 24 itemi ținesc mai multe dimensiuni: ergonomia platformei AR, utilizabilitatea aplicației, utilitatea percepută, atitudinea și intenția de utilizare). Ultimii 4 itemi sunt generali și măsoară ușurința în utilizare, utilitatea pentru învățare, caracterul atractiv al învățării și caracterul incitant al utilizării.

De asemenea, chestionarul cuprinde 2 itemi deschiși, prin care elevii au fost invitați să descrie câte 3 dintre cele mai pozitive și cele mai negative aspecte. În acest fel, măsurile cantitative au fost completate cu măsuri calitative.

Participanții au fost rugați să evalueze itemii pe o scală Likert cu 5 trepte (de la 1 – dezacord total, la 5 – acord total). Gradul de fidelitate a scalei (Cronbach's Alpha) a fost 0.931 la școala de vară și 0.948 după școala de vară, ceea ce este satisfăcător.

### **3 Evaluarea euristică a utilizabilității**

#### **3.1 Metodă și procedură**

Evaluarea euristică se face prin testarea interfeței de către 2-5 experți care evaluează utilizabilitatea în raport cu principii larg recunoscute (euristici) și recomandări. Problemele de utilizabilitate – UP (Usability Problems) sunt ierarhizate în funcție de impact în severe, moderate și minore.

Evaluarea euristică furnizează două categorii de măsuri:

- Măsuri cantitative: numărul de probleme de utilizabilitate din fiecare categorie.
- Măsuri calitative: descrieri detaliate ale problemelor individuale de utilizabilitate.

Problemele de utilizabilitate (măsurile calitative) au fost documentate pe baza unei scheme simplificate, care are la bază clasificarea problemelor de utilizabilitate elaborată de către Hvannberg & Law (2004). În Tabelul 2 este prezentată structura descrierii care a fost utilizată pentru proiectul ARiSE.

Pe parcursul dezvoltării prototipului, am observat că multe dintre problemele de utilizabilitate sunt datorate unei ghidări neadecvate a utilizatorului. Această lipsă de ghidare este datorată caracterului inovativ al



tehnicilor de interacțiune, care presupun manipularea de obiecte reale, precum și lipsei inerente de recomandări specifice de utilizabilitate. Din acest motiv, evaluarea euristică a fost făcută în raport cu criteriile ergonomice elaborate de Bastien și Scapin (1993) și adaptate apoi de Bach & Scapin (2003) pentru sisteme bazate pe realitate mixtă.

Inspecția de utilizabilitate este o evaluare de tip formativ, care este efectuată, de regulă, înaintea testării cu utilizatori. Datorită faptului că prima versiune îmbunătățită a aplicației a fost disponibilă abia la școala de vară, evaluarea euristică a fost făcută în paralel cu testarea ARTP de către elevii de la școlile din București, în perioada 1-15 Noiembrie 2007, de către 2 experți din ICI.

*Tabelul 2. Descrierea unei probleme de utilizabilitate*

	Identificator UP (de ex. UP1, UP2...)
Sarcină	Sarcina afectată de UP (de ex. Ex1)
Context	Contextul în care a apărut problema: ce acțiune a utilizatorului a declanșat-o, locația în interfață și alte elemente de context
Descriere	O descriere succintă a UP
Impact	Severitatea problemei (severă, moderată sau minoră)
Sugestii	Recomandări / sugestii către dezvoltatori pentru eliminarea UP
Evaluator	Evaluatorul care a identificat și documentat UP
Data	Data identificării / verificării
Status	Starea UP (rezolvată / nerezolvată)

### 3.2 Rezultate

Au fost identificate și documentate 19 probleme de utilizabilitate, între care 3 severe, 12 moderate și 4 minore.

În Tabelul 3 sunt prezentate rezultatele evaluării euristice a scenariului de biologie, grupate pe categorii de probleme de utilizabilitate și grad de severitate (1=minoră, 2=moderată, 3=severă).

Cele mai multe probleme au fost cele de selectare a unui organ, apoi cele de vizualizare și cele de suprapunere dintre imaginea observată a obiectului real și imaginea generată de calculator. Tot în această categorie intră și problema calibrării camerei video, realizată cu un program special.

O problemă importantă de ergonomie a ARTP este legată de utilizarea ochelarilor stereo 3D fără fir, care clipesc datorită interferenței dintre

emițătoarele de infraroșii. In ceea ce privește explicațiile vocale, problemele sunt legate de consistența explicațiilor cu lecția / exercițiile, care a fost afectată de modificări ale aplicației.

*Tabelul 3* Principalele categorii de probleme de utilizabilitate

Categorie	1	2	3	Total
Selectie	2	3	2	7
Suprapunere	1	4		5
Visualizare	1	2	1	4
Calibrare		1		1
Ergonomie		1		1
Explicații vocale		1		1

Din punctul de vedere al sarcinilor afectate, 10 dintre probleme afectează toate sarcinile de lucru, 5 probleme afectează efectuarea celui de al doilea exercițiu, 3 probleme afectează urmărirea programului demo și o problemă este asociată cu exercițiile 1 și 3.

Din punctul de vedere al criteriilor ergonomice, 9 probleme se referă la incitare (prompting), 1 la feedback, 3 la lizibilitate, 1 la acțiuni minimale (selectarea ultimului articol din meniu), 1 la densitatea informației (informație inutilă afișată pe ecran), 3 la compatibilitate (două legate de mărimea obiectului real și una de ochelari), iar una este un bug software. Așa cum se observă, 13 dintre problemele de utilizabilitate (68.4%) sunt probleme de ghidare a utilizatorului.

## 4 Rezultate ale testării cu utilizatori

### 4.1 Măsurile cantitative

In Figura 2 sunt prezentate comparativ rezultatele de la școala de vară (Oct) și după școala de vară (Nov). In general, elevii de la școlile din București au apreciat scenariul de biologie mai mult decât cei care au participat la școala de vară (3.96 vs. 3.85).

Paternal general al răspunsurilor este similar, cu excepția itemilor 14, 21 și 25 la care au fost înregistrate medii mai scăzute la școala de vară, ceea ce arată că participanții au considerat că este dificilă colaborarea cu colegii, că exercițiile nu sunt prea captivante iar ARTP este în general greu de utilizat.

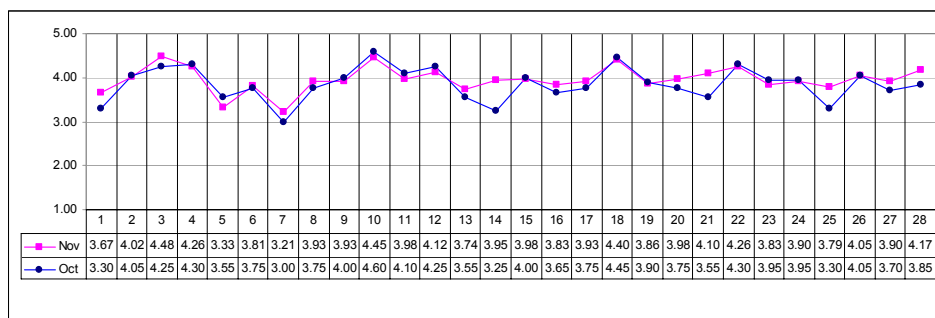


Figura 2. Comparație cu rezultatele de la școala de vară (valori medii)

Testul t pentru eșantioane independente arată că diferențele sunt semnificative, din punct de vedere statistic ( $\alpha=0.05$ ,  $df=60$ ), numai pentru itemii: 14 ( $t=-2.164$ ,  $p=0.034$ ) și 21 ( $t=-2.231$ ,  $p=0.029$ ).

În general, rezultatele pot fi considerate acceptabile, întrucât mediile sunt peste 3.00 (neutru). Itemii 5 și 7 au fost evaluați în ambele experimente sub 3.60. Acești itemi sunt tipici pentru sistemele AR iar mediile scăzute arată probleme legate de claritatea percepției vizuale.

Alți 12 itemi au fost valori cuprinse între 3.50 și 4.00 în ambele eșantioane. 8 itemi au valori medii peste 4.00 în ambele eșantioane. Valorile medii pentru cinci dintre aceștia sunt peste 4.25:

- Item 3: ajustarea căștilor este ușoară
- Item 4: postul de lucru este confortabil
- Item 10: utilitatea interacțiunii multimodale în mediul AR.
- Item 18: valoarea motivațională a ARTP.
- Item 22: intenția de utilizare, care denotă o acceptanță generală a ARTP.

În general, atitudinea față de ARTP a elevilor de la școlile din București a fost mai favorabilă decât cea a participanților la școala de vară, fapt susținut de mediile mai ridicate la itemii 18-21 și la itemii generali 27 și 28.

## 4.2 Măsuri calitative

Pentru a analiza răspunsurile elevilor la întrebările deschise, au fost identificate cuvintele cheie (atribute), care au fost apoi grupate pe categorii. Unii elevi au menționat numai unul sau două aspecte în timp ce alții s-au referit la mai multe aspecte în cadrul unei propoziții sau au descris ce le-a

plăcut / displăcut mai mult în mai multe propoziții. În consecință, au rezultat un total de 82 de aspecte pozitive la școala de vară și 85 de aspecte pozitive după școala de vară, respectiv 69 de aspecte negative la școala de vară și 79 de aspecte negative după școala de vară.

### Aspecte pozitive menționate de elevi

În Tabelul 4, sunt prezentate comparativ principalele categorii de aspecte pozitive menționate de elevi.

Tabelul 4. Categoriile de aspecte pozitive

Categorie	Oct	Nov	Total
Suport educațional	27	33	60
AR și vizualizare 3D	13	15	28
Interesant și motivant	8	8	16
Explicații vocale	7	8	15
Distractiv, provocativ	7	7	14
Post de lucru confortabil	-	11	11
Inovativ, experiență plăcută	4	-	4
Ușor de utilizat și altele	3	3	6
Total	82	85	167

Principala diferență o constituie cele 11 aspecte pozitive referitoare la confortul oferit de postul de lucru. O explicație este faptul că elevii de la școlile din București au venit la testare după ore, când erau deja obosiți. Din același motiv, numărul de aspecte pozitive menționate este de numai 85, deși au fost de peste două ori mai mulți copii care au participat la testare.

Suportul educațional include aspecte ca: mai bună înțelegere („Înțeleg mai bine poziția reală a organelor”), folositor pentru învățare („Pot să învăț ușor locul fiecărui organ”), ușurința în memorare („Imi amintesc mai ușor conținutul lecției”), învățarea mai rapidă („Este bun să înveți rapid”).

Elevilor le-a plăcut tehnologia AR care oferă posibilitatea interacțiunii în 3D (Inveți subiectul în 3D”) și explicațiile vocale care ghidază interacțiunea cu sistemul („Explicațiile sunt bune și descriptive”). De asemenea, ei au apreciat sistemul ca fiind amuzant („asemănător jocurilor”) și motivant („Sistemul te motivează să înveți acest subiect”, „Sistemul face învățarea mult mai interesantă”).

Analizând coloana de total, constatăm că suportul educațional reprezintă 35.9% din totalul aspectelor pozitive menționate de ambele eșantioane de

elevi. Interacțiunea multimodală și vizualizarea 3D însumează 43 de aspecte pozitive, adică 25.7%, iar cele 34 de aspecte referitoare la motivație reprezintă 20.3% din totalul de 167.

### Aspecte negative menționate de elevi

În Tabelul 5, sunt prezentate comparativ, principalele categorii de aspecte negative menționate de elevi în cele două experimente. Aceste aspecte negative reflectă probleme de utilizabilitate identificate și menționate informal de către utilizatori.

Tabelul 5. Categorii de aspecte negative

Categorie	Oct	Nov	Total
Probleme de selecție	25	25	50
Dureri ale ochilor și ochelarii	13	18	31
Obiectul real prea mare sau greu de manevrat	10	15	25
Probleme cu sunetul și ajutorul câștilor	10	1	11
Suprapunere și vizualizare	4	7	11
Greu de utilizat	3	4	7
Alte probleme	4	9	13
Total	69	79	148

Cel mai frecvent a fost menționată dificultatea de a selecta un organ cu ajutorul unelei de interacțiune („Adeseori este dificil să poziționezi pe organul cerut”, „Chiar dacă știi care este organul, este greu să îl selectezi”). Câțiva elevi au menționat și lipsa de acuratețe în suprapunerea imaginii generate pe imaginea obiectului real observată pe ecran. În total au fost menționate 50 de probleme de selecție, ceea ce reprezintă 33.8% din total.

A doua categorie de probleme de utilizabilitate a fost disconfortul provocat de ochelarii stereo 3D fără fir. Majoritatea elevilor au acuzat dureri de ochi la sfârșitul sesiunii. Rezultatele testării arată că pentru contextul de utilizare dat (mai mult de un singur modul într-o cameră, distanță mică între module) este recomandabilă utilizarea de ochelari stereo cu fir, pentru a evita interferențele între transmițătoarele de infraroșii. În total, problemele cauzate de ochelarii fără fir reprezintă 20.9% din problemele de utilizabilitate menționate de elevi.

Primele două categorii de probleme de utilizabilitate, cauzate mai ales de lipsa de acuratețe a percepției vizuale, sunt consistente cu media scăzută înregistrată la itemii 1, 5, 7, 13 (dificultatea de a corecta o eroare) și la

itemul general 25. De asemenea, aceste probleme sunt confirmate și de rezultatele evaluării euristice.

Mulți dintre elevi au menționat că mulajul este prea mare și greu de manipulat („nu mi-a plăcut faptul că mulajul trebuie mutat”, „fiecare elev ar trebui să aibă propriul mulaj”). Cauza este dimensiunea obiectului real în raport cu aria de selecție, fapt care necesită deplasarea acestuia în funcție de sarcina de lucru. În același timp, un mulaj este partajat de doi elevi care stau față în față și sunt frecvente situațiile în care ambii au nevoie să interacționeze cu aceeași parte a mulajului. Această categorie de probleme de utilizabilitate este a treia ca pondere, reprezentând 16.9% din total.

Participanții la școala de vară s-au plâns de probleme de sunet, fapt care se reflectă și în media mai scăzută înregistrată la itemul 3.

### 4.3 Măsurile ale eficacității și eficienței

#### Măsurile colectate la școala de vară

Primul exercițiu a fost mai ușor de rezolvat (doar indicarea organelor) dar dificil de utilizat. Erorile (Min=0, Max=13, SD=3.9) sunt datorate în primul rând dificultăților în selectarea organului dorit, care au fost mai frecvente în cazul organelor de mici dimensiuni (esofag, duoden, pancreas). Cu toate acestea, toți elevii au reușit să termine exercițiul cu succes. Timpul de execuție a variat între 116 sec. (cu 2 erori) și 852 sec. (cu 10 erori), cu o medie de 381.8 sec (SD=218.1).

În Tabelul 6 este arătată dificultatea de a selecta organele de mici dimensiuni în primul exercițiu.

Ultimele două exerciții au fost mai greu de rezolvat, având în vedere că între organe și nutrienți este o relație N:M (un nutrient poate fi digerat în mai multe organe iar în același organ pot fi digerați mai mulți nutrienți). Al doilea exercițiu a fost o sarcină mai ușor de utilizat, întrucât nutrienții sunt selectați apăsând butonul de pe telecomandă. Se poate considera că erorile sunt datorate în primul rând lipsei de cunoștințe, ceea ce este un argument în favoarea eficacității pedagogice a scenariului.

Doi elevi nu au reușit să rezolve al doilea exercițiu. Un singur elev nu a făcut erori iar 3 elevi au făcut 10, 11 respectiv 19 erori. Restul elevilor au comis între 1 și 7 erori. Timpul de execuție a variat între 83 sec. (1 eroare) și 673 sec. (19 erori) cu o medie de 254.9 sec (SD=186.1).

Tabelul 6. Erori în selectarea organelor de mici dimensiuni

Denumire organ	Număr erori	Număr elevi	Număr mediu erori
Cavitatea bucală	13	9	1.44
Duoden	28	12	2.33
Pancreas	21	10	2.10
Esôfag	14	8	1.75
Intestin gros	6	6	1.00
Intestin subțire	2	2	1.00
Altele	5	3	1.67
<b>Total</b>	<b>89</b>		

Patru elevi au eșuat în rezolvarea celui de al treilea exercițiu. Toți au comis erori: 7 elevi au făcut 1-10 erori, 5 au făcut 11-20 erori și 4 elevi peste 20 de erori. La acest exercițiu, erorile au fost datorate atât lipsei de cunoștințe cât și dificultății în selectarea organelor. Timpul de execuție a variat între 95 sec. (1 eroare) și 727 sec (39 erori) cu o medie de 381.6 sec. (SD=178).

În total, 14 elevi au reușit să efectueze toate exercițiile în scenariul de Biologie. Timpul total de execuție a variat între 309 sec. (7 erori) și 1964 sec. (28 erori). Numărul total de erori a variat între 6 și 56, cu o medie de 23.3 erori. Timpul total mediu a fost de 1060 sec, adică 17.67 min. și a fost calculat pentru cei 14 elevi care au îndeplinit toate sarcinile.

### Măsuri colectate după școala de vară și comparație

În Tabelul 7, sunt prezentate comparativ măsurile eficacității (rata de îndeplinire a sarcinii și numărul mediu de erori) și eficienței (timpul mediu de execuție) pentru fiecare exercițiu.

Tabelul 7. Măsuri ale eficacității și eficienței

Ex.	La școala de vară			După școala de vară		
	Rată	Erori	Timp	Rată	Erori	Timp
1	100%	4.45	381.8	80%	6.88	455.8
2	90%	4.94	254.9	91%	6.28	318.4
3	80%	13.69	381.6	94%	15.90	401.4

Un elev din 17 nu a reușit să rezolve cel de al treilea exercițiu. Toți au comis erori: 7 elevi au făcut peste 20 de erori. Timpul de execuție a variat între 174 sec. (3 erori) și 917 sec (21 erori) cu o medie de 401.8 sec.

(SD=226.8).

În total, 32 de elevi (78%) au reușit să efectueze toate exercițiile repartizate, dintre care 11 elevi au efectuat și al treilea exercițiu. 6 elevi au rezolvat un singur exercițiu în timp ce 3 elevi nu au reușit să rezolve nici unul. Timpul total de execuție pentru cei 11 elevi care au executat toate exercițiile repartizate a variat între 705 sec. (22 erori) și 1972 sec. (10 erori). Numărul total de erori a variat între 8 și 50, cu o medie de 20.73 erori (SD=12.73). Timpul total mediu a fost de 1207.8 sec, adică 20.1 (SD=8.75).

Există diferențe între ratele de succes la exercițiile 1 și 3. Participanții la școala de vară au făcut mai puține greșeli. Cu toate acestea, în ambele experimente al treilea exercițiu a fost terminat cu multe erori. De asemenea, există diferențe între numărul de erori și timpul de execuție. În timp ce elevii de la școlile din București au venit la testare după ore, participanții la școala de vară nu au avut altceva de făcut iar evenimentul în sine a fost motivant și a creat un spirit competitiv.

## 5. Concluzii și direcții de continuare

Evaluarea măsurilor subiective ale satisfacției utilizatorului bazată pe măsurile cantitative și calitative colectate cu ajutorul chestionarului de utilizabilitate relevă mai multe aspecte pozitive.

ARTP are valoare educațională: sistemul facilitează înțelegerea, învățarea mai rapidă, ușurează memorarea și este util pentru testarea cunoștințelor. Sistemul are o valoare motivațională ridicată: este atractiv, stimulat și incitant, face învățarea mai interesantă. Elevii au apreciat interacțiunea 3D utilizând tehnici specifice AR precum și explicațiile vocale care asigură ghidarea pe parcursul procesului de învățare.

În general, acceptanța ARTP este bună: elevii au apreciat utilitatea sistemului pentru învățare și au exprimat interesul pentru a îl utiliza în viitor precum și dorința de a îl avea în școală.

Mai multe probleme de utilizabilitate au fost identificate, atât prin evaluare euristică cât și cu ajutorul chestionarului. Claritatea percepției vizuale ca și ușurința în utilizare trebuie îmbunătățite. Ochelarii fără fir trebuie înlocuiți cu ochelari cu fir, care nu obosesc ochii.

Repetarea testării cu utilizatori a confirmat rezultatele preliminare obținute la școala de vară. Aspectele negative menționate de elevi au fost utile pentru completarea problemelor de utilizabilitate identificate de



evaluarea euristică. În acest sens, se cuvine a fi menționat că problemele legate de manipularea obiectului real au fost identificate de către elevi.

Pe ansamblu, se poate concluziona că evaluarea formativă a fost un ajutor prețios pentru proiectanți. O nouă versiune a aplicației de biologie a fost instalată în 2008. Chestionarul de utilizabilitate a fost conceput pentru a efectua și evaluarea sumativă a ARTP. În acest sens, testarea cu utilizatori efectuată în 2007 reprezintă un prim pas în colectarea unui număr suficient de observații pentru a permite o evaluare sumativă a scenariului de biologie. Evaluarea va continua în 2008 cu testarea versiunii îmbunătățite a aplicației.

### Confirmare

Această lucrare este finanțată din proiectul european ARiSE (FP6-027039). Multumim celor trei școli din București pentru sprijinul acordat și d-nelor profesoare care au însoțit elevii la testare : Elena Apostolescu (Sc. Generală Nr. 56 « Jose Marti »), Nicoleta Stan și Georgeta Prodănescu (Sc. Generală nr. 172 « Sfântul Andrei »), Florica Aldea și Gheorghiza Nistoroiu (Sc. Generală nr. 191).

### Referințe

- Bach, C. & Scapin, D., 2004. Obstacles and perspectives for Evaluating mixed Reality Systems Usability. In Mixer workshop, Proceedings of IUI-CADUI Conference 2004, pp. 72-79. ACM Press.
- Bach, C. & Scapin, D.L., 2003. "Adaptation of Ergonomic Criteria to Human-Virtual Environments Interactions". In *Proceedings of Interact'03*. IOS Press, pp. 880-883.
- Bastien, A. & Scapin, D.L., 1993 Ergonomic criteria for the evaluation of human-computer interfaces. Technical report No. 156, INRIA, Roquencourt, France. 1993.
- Bowman, D., Gabbard, J., and Hix, D., 2002. A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. Presence: Teleoperators and Virtual Environments, vol. 11, no. 4, pp. 404-424
- Davis, F.D., 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly*, Vol.13, No.3, pp. 318-340.
- Dillon, A. and Morris, M., 1998. From "can they?" to "will they?": extending usability evaluation to address acceptance. *AIS Conference Paper*, Baltimore, August 1998.
- Gabbard, J., Hix, D., Swan, E., Livingston, M., Herer, T., Julier, S., Baillot, Y. & Brown, D., 2004. A Cost-Effective Usability Evaluation Progression for Novel Interactive Systems. In *Proceedings of Hawaii International Conference on Systems Sciences*, Track 9, p. 90276c, IEEE.

- ISO 9126-4:2001 Software Engineering - Software product quality. Part 4: Quality of use.
- ISO/DIS 9241-11:1994 Information Technology – Ergonomic requirements for office work with visual display terminal (VDTs) - Guidance on usability.
- Lamanauskas, V., Pribeanu, C., Vilkonis, R., Balog, A., Iordache, D.D., Klangauskas, A., 2007. "Evaluating the educational value and usability of an augmented reality platform for school environments: some preliminary results. *Proceedings of 4th WSEAS/IASME International Conference on Engineering Education*. (Agios Nikolaos, Crete Island, Greece, 24-26 July 2007). Mathematics and Computers in Science and Engineering, ISSN 1790-5117. WSEAS Press, pp. 86-91.
- Pribeanu, C., Balog, A., Iordache, D., 2007. „Usability evaluation of an interaction scenario for AR-based educational software systems”. *Proceedings of ICTNSE International Conference - 2007*, Siauliai 3-4 December 2007, pp. 141-146.
- Scriven, M., 1991. *Evaluation thesaurus*. 4th ed. Newbury Park, CA: Sage Publications.
- Sun, H., Zhang, P., 2006. The role of moderating factors in user technology acceptance. *International Journal of Human-Computer Studies*, 64 (2006), Elsevier. Pp. 53-78
- Swann II, J., E., Gabbard, J., 2005. Survey of User-Based Experimentation in Augmented Reality. In. *Proceedings of 1<sup>st</sup> International Conference on Virtual Reality*. July 22-27, Las Vegas, Nevada, 2005.
- Theofanos, M. & Quesenbery, W., 2005. Towards the Design of Effective Formative Test Reports. In *Journal of Usability Studies*, Issue 1, Vol.1. pp. 27-45.
- Venkatesh, V., Davis, F.D., Morris, M.G., 2007. Dead Or Alive? The Development, Trajectory And Future Of Technology Adoption Research. *Journal of the AIS*, Vol. 8, Issue 4, pp. 267-286.
- Wind, J., Riege, K., Bogen M., 2007. Spinnstube®: A Seated Augmented Reality Display System, In *Virtual Environments, Proceedings of IPT-EGVE – EG/ACM Symposium*, pp. 17-23., Eurographics.