

EVALUAREA FORMATIVĂ A UTILIZABILITĂȚII UNUI SCENARIU DE ÎNVĂȚARE A CHIMIEI IMPLEMENTAT PE O PLATFORMĂ EDUCAȚIONALĂ DE REALITATE ÎMBOGĂȚITĂ

Dragoș Daniel Iordache, Costin Pribeanu și Alexandru Balog

iordache@ici.ro, pribeanu@ici.ro, alexb@ici.ro

Institutul Național de Cercetare - Dezvoltare în Informatică, ICI, București

Rezumat: 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. Combinația realului cu virtualul din sistemele de realitate îmbogățită necesită tehnici de interacțiune specifice, care trebuie testate cu utilizatori pentru a evita probleme de utilizabilitate. Acest articol prezintă o abordare în evaluarea formativă a unui scenariu de învățare a chimiei pentru școala generală, implementat pe o platformă educațională, bazată pe realitate îmbogățită. Abordarea integrează două metode de evaluare: inspecția de utilizabilitate și testarea cu utilizatori. Scenariul a fost testat mai întâi cu un număr mic de utilizatori, în timpul unei școli de vară, cu scopul de a obține un feedback rapid de la elevi având cunoștințe bune de chimie. Apoi, testarea a fost repetată în condiții diferite și cu un număr mai mare de utilizatori reprezentativi. În acest articol, vom descrie ambele experimente și vom compara rezultatele obținute prin cele două metode.

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

Abstract: Formative usability evaluation aims to identify usability problems as early as possible in the development life cycle of interactive systems that are featuring novel interaction techniques. The mix of real and virtual in Augmented Reality (AR) systems requires specific interaction techniques which have to be tested with users in order to avoid usability problems. This paper presents an approach to the formative usability evaluation of a learning scenario for Chemistry in secondary schools that has been implemented onto an AR-based educational platform. This approach integrates two evaluation methods: usability inspection and user testing. Firstly, the scenario has been tested with a small number of users during a summer school, in order to get a fast feedback from students having good Chemistry knowledge. Then the user testing has been repeated in different conditions and with a larger number of representative users. In this paper we will describe both experiments and we will compare the results of the two methods.

Key Words: user-centered design, usability, formative evaluation, augmented reality.

1. Introducere

Evaluarea formativă a utilizabilității este efectuată iterativ, pe parcursul ciclului de dezvoltare cu scopul identificării și eliminării problemelor de utilizabilitate cât mai devreme [15]. 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 [12]. 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 inspecție de utilizabilitate sau evaluare euristică) și/sau prin testarea cu un număr relativ mic de utilizatori (denumită și evaluare centrată pe utilizator). Fiecare metodă are avantaje și dezavantaje dar, așa cum au arătat Law & Hvannberg [11], tendința actuală este a găsi modalități de a profita de complementaritatea dintre acestea.

Sistemele 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 specifice de proiectare centrată pe utilizator [1], [14]. 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, așa cum au arătat Gabard et al. [7], componentele de interacțiune sunt adeseori slab proiectate reducând astfel utilizabilitatea sistemului.

În acest articol, se prezintă o abordare în evaluarea formativă a unui scenariu de învățare a chimiei, implementat pe o platformă AR, în cadrul proiectului de cercetare ARiSE (Augmented Reality in School Environments). 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 a avut ca disciplină țintă biologia pentru clasa a 7-a, sarcina generică de interacțiune fiind vizualizarea 3D a sistemului digestiv uman. Al doilea prototip are ca disciplină țintă chimia pentru clasa a 8-a. Paradigma specifică AR, implementată în acest scenariu, este

construirea cu ghidare și are ca scop facilitarea înțelegerii și motivației elevilor pentru învățarea tabelului periodic al elementelor și a reacțiilor chimice.

Proiectul este realizat într-un consorțiu în care fiecare partener are un rol specific. Platforma hardware a fost proiectată de Fraunhofer 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 (Lituania) 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). Platforma a fost realizată în primul an de către Fraunhofer IAIS și a fost înregistrată sub marca Spinnstube® [11].

Pentru a obține 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 și alte două-trei școli din țara gazdă. Din fiecare școală, au fost selectați de către profesori câte 4 elevi (2 fete și 2 băieți), pe baza cunoștințelor la chimie 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 chimie a fost dezvoltată în 2007 și testată cu utilizatori la a doua școală de vară, care a avut loc în București în octombrie 2007. Datorită timpului scurt, nu a fost posibilă o evaluare euristică prealabilă. În aceste condiții, ideea 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 chimie 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, efectuată în paralel cu a doua testare, ceea ce a permis o a doua comparație, între rezultatele obținute cu cele două metode.

Restul articolului este organizat după cum urmează. În secțiunea următoare, vom descrie cele două experimente. În secțiunea 3, vom prezenta rezultatele evaluării euristice. În secțiunea 4, vom prezenta rezultatele testării cu utilizatori, vom compara și vom discuta similaritățile și diferențele dintre acestea. Lucrarea se încheie cu concluzii și direcții de continuare a cercetărilor în acest proiect.

2. Testarea cu utilizatori

2.1. Contextul evaluării

Cea de a doua școală de vară a fost organizată în ICI București, în perioada 24-28 Octombrie 2007. Toți partenerii din proiect au participat la acest eveniment. În ceea ce privește utilizatorii, 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 cele de chimie a doua zi. Testarea și discuțiile cu utilizatorii au avut loc dimineața, iar discuțiile între parteneri după-amiaza.

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 mai întâi scenariul de chimie.

2.2. Platforma hardware-software

ARTP este un mediu AR de tip „seated” [17]: utilizatorii au în față un ecran „see-through” pe care sunt suprapuse imagini ale obiectelor virtuale peste imaginea observată a unui obiect real. Figura 1 ilustrează două eleve care testează aplicația de chimie la școala de vară.

În cazul scenariului de chimie, obiectele reale sunt tabelul periodic al elementelor și un set de bile colorate simbolizând atomi. Tabelul periodic are două părți: partea A, prezentând simbolurile elementelor chimice, și partea B, prezentând doar numerele grupelor și ale perioadelor. Partea B este utilizată pentru a testa măsura în care elevii au înțeles structura internă a atomilor. Fiecare post de lucru a avut propriul tabel periodic.

Pentru selectarea unui item din meniu a fost utilizată o telecomandă Wii Nintendo.

2.3. Participanți și sarcini

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.

Participanții au avut de îndeplinit toate cele 14 sarcini definite în aplicație: o introducere și 13 exerciții grupate în 3 lecții. Prima lecție este despre structura chimică a atomilor și a cuprins două exerciții. A doua lecție este despre formarea moleculelor și cuprinde 8 exerciții. Cea de-a treia lecție este despre reacții chimice și cuprinde 3 exerciții. Sarcinile au fost prezentate prin intermediul unei interfețe vocale în limba națională a elevilor.



Figura 1. Două eleve testând aplicația de chimie la școala de vară

La testarea care a avut loc după școala de vară au participat 42 de elevi (19 băieți și 23 de fete). 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 și procedură

Măsurarea utilizabilității și calității în utilizare

Standardul ISO 9241-11 definește utilizabilitatea dintr-o perspectivă mai largă, 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. Eficacitatea este definită ca fiind măsura în care utilizatorii îndeplinesc obiectivele specificate cu acuratețe și completitudine.

Noua versiune a standardului ISO 9126, definește utilizabilitatea ca un atribut de calitate, prin capacitatea produsului software de a fi ușor de înțeles, învățat, folosit și considerat atractiv de către utilizator, atunci când este folosit în condiții specificate. În acest standard, utilizabilitatea definită de standardul ISO 9241-11 devine calitate în utilizare, având patru componente: eficacitate, productivitate, eficiență și satisfacție.

Această definiție a calității în utilizare 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 din proiectul ARiSE, a fost adoptată o perspectivă mai largă asupra utilizabilității, prin elaborarea unui chestionar care are la bază un model conceptual de acceptanță a tehnologiei. 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 [5]. 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.

Așa cum arată Dillon și Moris [6], 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 [16], [13]. 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.

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 stimulant
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

Chestionarul de utilizabilitate cuprinde 28 de itemi. Primii 24 itemi ținesc mai multe dimensiuni: ergonomia generală a 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.

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.

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 a fost 0.931 (Cronbach’s Alpha) la școala de vară și 0.948 după școala de vară, ceea ce este satisfăcător.

3. Inspecția de utilizabilitate

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, denumite euristici. Utilizabilitatea poate fi evaluată și în raport cu recomandări de proiectare, caz în care se mai numește și evaluare bazată pe recomandări. Problemele de utilizabilitate – UP (Usability Problems) sunt ierarhizate în funcție de impact în trei categorii: 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, identificate în ARiSE, au fost documentate pe baza unei scheme simplificate, care are la bază clasificarea problemelor de utilizabilitate elaborată de către Hvannberg & Law [8]. În Tabelul 2, este prezentată structura descrierii.

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 și 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)

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 [3] și adaptate apoi de Bach & Scapin [2] pentru sistemele bazate pe realitate mixtă. Inspectia de utilizabilitate este o evaluare de tip formativ, care este efectuată, de regulă, înaintea testării cu utilizatori. Datorită faptului că prima versiune a aplicației de chimie 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.

3.2. Rezultate

În Tabelul 3, sunt prezentate rezultatele evaluării, grupate pe categorii de probleme de utilizabilitate și grad de severitate (1=minoră, 2=moderată, 3=severă). Au fost identificate și documentate 31 de probleme de utilizabilitate, între care 6 severe, 21 moderate și 4 minore.

Cele mai multe probleme au fost cele de selectare a unui element chimic, apoi cele de vizualizare și cele de suprapunere dintre imaginea observată a obiectului real și imaginea generată de calculator.

Tabelul 3. Principalele categorii de probleme de utilizabilitate

Categorie	Severitate			Total
	1	2	3	
Dificultăți de selecție	2	11	6	19
Probleme de vizualizare	2	6		8
Acuratețea suprapunerii		2		2
Ochelarii stereo		1		1
Explicații vocale inconsistente		1		1
Total	4	12	3	31

Problema de ergonomie este legată de utilizarea ochelarilor stereo 3D fără fir, care clipesc datorită interferenței dintre emițătoarele de infraroșii. În ceea ce privește explicațiile vocale, problemele sunt legate de consistența explicațiilor cu lecția / exercițiile, care a fost afectată de modificări ale aplicației.

Din punctul de vedere al sarcinilor afectate, 11 dintre probleme afectează toate sarcinile de lucru, 5 probleme lecția introductivă (demo), 8 probleme efectuarea celui de al doilea exercițiu din prima lecție, 3 probleme efectuarea primului exercițiu din a doua lecție, 3 probleme exercițiile din a doua lecție și o problemă exercițiile din ultima lecție.

Din punctul de vedere al criteriilor ergonomice, 9 probleme se referă la feedback, 6 la incitare (prompting), 5 la lizibilitate, 3 la compatibilitate (2 legate de mărimea tabelului periodic și una legată de ochelari), 2 la acțiuni minimale (selectarea ultimului articol din meniu), câte una la încărcare cognitivă, acțiuni explicite,

corectarea erorilor și consistență, iar una este un bug software. Așa cum se observă, 20 dintre problemele de utilizabilitate (64.5%) sunt probleme de ghidare a utilizatorului (incitare, feedback și lizibilitate).

4. Rezultate ale testării cu utilizatori

4.1. Măsurile cantitative

În Figura 2, sunt prezentate comparativ rezultatele de la școala de vară (octombrie) și după școala de vară (noiembrie). În general, rezultatele pot fi considerate acceptabile, întrucât mediile sunt peste 3.00 (neutru). Așa cum se observă, elevii de la școlile din București au evaluat la un nivel mai scăzut decât participanții la școala de vară (3.78 vs. 3.94).

Diferențe relativ mari (peste 0.30) s-au înregistrat la itemii 5, 6, 7, 9, 12, 14, 18, 22, 23 și la itemul general 25. Testul t pentru eșantioane independente arată că diferențele sunt semnificative, din punct de vedere statistic ($\alpha=0.05$, $df=60$), numai pentru itemii: 5 ($t=2.105$, $p=0.039$), 6 ($t=2.366$, $p=0.021$) și 9 ($t=4.348$, $p<0.001$).

Paternul general al răspunsurilor este similar, în sensul că la aceiași itemi s-au înregistrat valori mai scăzute / ridicate, cu excepția itemului 9, care arată că reamintirea modului de operare a aplicației AR) a fost mai dificilă pentru elevii de la școlile din București. Reamintirea modului de operare este legată de ușurința în învățare (itemul 8), care este la rândul său afectată de ușurința în înțelegere a modului de operare cu aplicația (itemul 6), la care diferența este foarte mare (-0.55). Cauza este vârsta mai mică și lipsa de familiaritate cu tehnologia AR, care nu au fost compensate cu un plus de explicații preliminară testării.

Itemii 1, 5, 7 și 13 au fost evaluați în ambele experimente sub 3.50. Primii trei itemi sunt tipici pentru sistemele AR, iar mediile scăzute arată probleme legate de claritatea percepției vizuale. Itemul 13 se referă la dificultatea de a corecta erorile. Mediile mai scăzute care s-au înregistrat la itemii 1-5 (ergonomia generală a ARTP) și 6-14 (utilizabilitatea aplicației de chimie) sunt reflectate și în valorile mai scăzute înregistrate la itemul general 25.

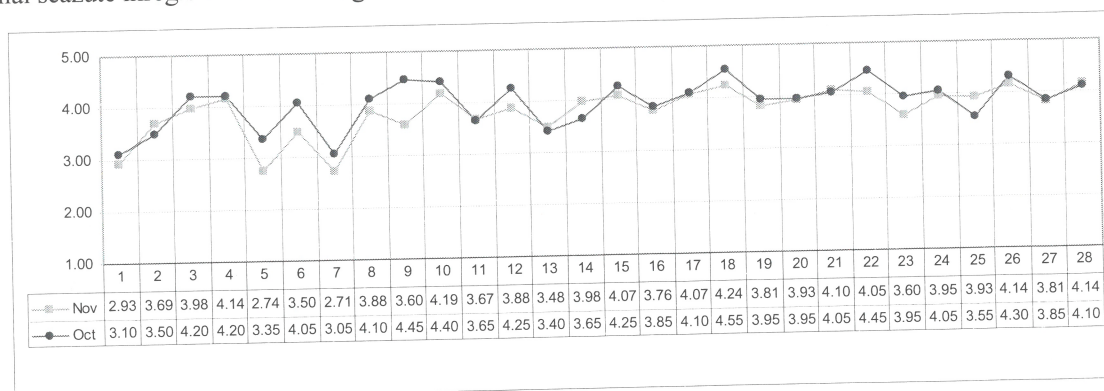


Figura 2. Comparație cu rezultatele de la școala de vară (valori medii)

Alți 9 itemi au fost valori cuprinse între 3.50 și 4.00 în ambele eșantioane. Numai 7 itemi au valori medii peste 4.00 în ambele eșantioane. Cele mai ridicate valori au fost înregistrate la itemul 18 (4.24 vs. 4.55), care arată că ARTP face învățarea mai interesantă.

În general, măsurile cantitative arată că atitudinea față de ARTP a elevilor de la școlile din București a fost mai puțin favorabilă decât cea a participanților la școala de vară, fapt susținut de mediile mai ridicate la itemii 18-20 și la itemul general 27. Același lucru se poate spune și despre utilitatea percepută (itemii 15-17 și itemul general 26) și intenția de utilizare (itemii 22-24).

4.2. Măsurile 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 / plăcut mai mult în mai multe propoziții. În consecință, au rezultat un total de 70 de aspecte pozitive la școala de vară și 95 de aspecte pozitive după școala de vară, respectiv 65 de aspecte negative la școala de vară și 116 de aspecte negative după școala de vară.

Aspecte pozitive menționate de elevi

În Tabelul 4, sunt prezentate principalele categorii de aspecte pozitive, menționate de către participanții la școala de vară octombrie) și elevii de la școlile din București (noiembrie).

Tabelul 4. Categorii de aspecte pozitive

Categorie	Oct	Nov	Total
Suport pentru înțelegere	11	12	23
Suport pentru testare	3	1	4
Suport pentru învățare	6	13	19
Învățare interesantă	10	6	16
Colaborare cu colegii	-	6	6
Interacțiune AR	5	4	9
Captivant, stimulativ	2	5	7
Vizualizare 3D	11	6	17
Ghidare utilizator multimodală	6	4	10
Învățare atractivă	4	3	7
Distractiv, provocativ, inovativ	4	17	21
Post de lucru confortabil	-	7	7
Intenție de utilizare	-	3	3
Ușor de utilizat și altele	8	8	16
Total	70	95	165

Primele trei categorii corespund itemilor 15-17 și 26 din chestionar. După ambele experimente, elevii au apreciat utilitatea pentru înțelegere („Ajută să înțelegem mai bine lecția”, „Înțeleg mult mai bine prin acest sistem”) și învățare („Am învățat mai mult și mai ușor”, „Ne ajută să învățăm mai repede”).

Următoarele patru categorii corespund itemilor 19-21. Participanții la școala de vară au apreciat într-o măsură mai mare că sistemul face învățarea mai interesantă, fapt consistent cu diferența la itemul 18 (+0.31). De asemenea, elevilor le-a plăcut interacțiunea specifică AR („Combinația atomilor este interesantă”) și au găsit că exercițiile sunt captivante („Exercițiile sunt captivante”).

Elevii de la școlile din București au apreciat și colaborarea cu colegii („Lucrul cu colegii este stimulativ”). De asemenea, ei au apreciat confortul oferit de postul de lucru. O explicație este faptul că au venit la testare după ore, când erau deja obosiți.

Elevilor le-a plăcut tehnologia AR, care oferă posibilitatea interacțiunii în 3D și explicațiile vocale care ghidează interacțiunea cu sistemul. De asemenea, ei au menționat caracterul inovativ și incitant al aplicației, care oferă o experiență de învățare nouă și atrăgătoare („Mi-a plăcut sistemul pentru că este atrăgător pentru copii”, „Face ca învățatul să fie mult mai distractiv”, „Tehnologia este fascinantă”, „Face ca învățatul să fie mult mai incitant decât învățatul obișnuit”).

Un sumar al principalelor categorii de aspecte pozitive este prezentat în Tabelul 5.

Tabelul 5. Sumarul aspectelor pozitive

Categorie	Oct	Nov	Total
Suport educațional	20	26	46
Suport motivațional	25	41	66
Vizualizare 3D și interacțiune multimodală	17	10	27
Altele	8	18	26
Total	70	95	165

Analizând coloana de total, constatăm că suportul educațional reprezintă 27.9% din totalul aspectelor pozitive menționate de ambele eșantioane de elevi. Interacțiunea multimodală și vizualizarea 3D însumează 27 de aspecte pozitive, adică 16.4%, iar cele 66 de aspecte referitoare la motivație reprezintă 40% din totalul de 165. Aceasta înseamnă că elevii, în special cei de la școlile din București (43.1% din totalul de 95), au apreciat valoarea motivațională a ARTP.

Numărul total al aspectelor pozitive menționate de către elevii de la școlile din București (95) este relativ mic, având în vedere că numărul participanților a fost de peste două ori mai mare, fapt care confirmă atitudinea mai puțin favorabilă, rezultată din măsurile cantitative. Cu toate acestea, 3 dintre elevi au menționat că ar dori să aibă acest sistem în școală.

Aspecte negative menționate de elevi

În Tabelul 6 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 6. Categorii de aspecte negative

Categorie	Oct	Nov	Total
Suprapunere	19	17	36
Vizualizare	19	11	30
Probleme de selecție	14	14	28
Obiectul real prea mare sau greu de manevrat	10	18	28
Dureri ale ochilor și ochelarii	13	24	27
Probleme cu sunetul și ajustarea căștilor	6	7	13
Greu de utilizat și alte probleme	3	25	28
Total	65	116	183

În ambele experimente, cel mai frecvent a fost menționată lipsa de acuratețe a suprapunerii imaginii proiectate peste obiectul real („Nu se suprapunea imaginea cu tabelul periodic”, „suprapunerea dintre ecran și proiecție nu este foarte clară”), fapt consistent cu mediile cele mai scăzute înregistrate la itemul 7. De asemenea, elevii au menționat numeroase probleme de vizualizare („Nu se vedea proiecția integrală pe ecran”, „Imaginea ar trebui să fie puțin mai clară” și „Imaginea procesată de calculator acoperea bilele reale, fiindu-i practic imposibil să le aranjezi”). A treia categorie de probleme de utilizabilitate este dificultatea în selectarea unui element („Am pus bila pe H și mi se afișa Li”, „Se ajunge cu greutate cu bila la elementele din ultimele perioade”). O altă problemă a fost dificultatea observării bilelor (mai ales a culorii acestora), din cauza suprapunerii informațiilor afișate pe ecran cu fiecare element chimic.

Numărul mare de probleme de utilizabilitate, menționate de elevi la primele trei categorii, este consistent cu rezultatele evaluării euristice, chiar dacă ponderea fiecărei categorii este diferită. Se cuvine a fi menționat faptul că evaluarea euristică identifică problemele de utilizabilitate în mod sistematic, menționându-le o singură dată și precizând sarcinile care sunt afectate.

Mulți dintre elevi au menționat că tabelul periodic este prea mare în raport cu aria de selecție și că este dificil de manipulat bilele (în special, în cazul reacțiilor chimice, când sunt necesare peste 4 bile).

O altă 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 14,8% din problemele de utilizabilitate menționate de elevi.

Numărul total de aspecte negative menționat de elevii de la școlile din București este destul de mare față de numărul de aspecte pozitive menționate, fapt care confirmă și, în același timp, explică atitudinea mai puțin favorabilă comparativ cu participanții la școala de vară, determinată de problemele de utilizabilitate și de un plus de dificultate în înțelegerea subiectului, pe care îl învață în clasa a 8-a.

4.3. Măsuri ale eficacității și eficienței

Tabelul 7 prezintă comparativ măsurile eficacității și eficienței în cele două experimente. În ambele cazuri, exercițiul al doilea din prima lecție a fost greu de efectuat.

Numai 8 elevi din 20 la școala de vară au efectuat ambele exerciții ale primei lecții (înțelegerea tabelului periodic). Timpul de execuție pentru prima lecție a variat între 352 sec. și 781 sec. cu o medie de 481.8 sec. (SD=143.9). Numai 5 elevi din 42 au reușit la al doilea experiment să efectueze exercițiul 2, cu un timp de execuție mediu de 467.8 sec.

La școala de vară, 8 elevi au reușit să efectueze toate exercițiile din lecția 2 (formare molecule și compuși chimici). Timpul de execuție a variat între 281 sec. și 1020 sec., cu o medie de 555 sec. (SD=200.2). Timpul de execuție pentru exercițiile individuale variază între 18 sec. și 327 sec. 10 dintre elevi au efectuat exercițiul 4 în mai puțin de 30 sec. Așa cum se observă în tabel, timpul de execuție a scăzut de la 92 sec. la 41 sec. ceea ce denotă că elevii au învățat foarte repede cum trebuie făcut exercițiul. Apoi timpul de execuție a crescut, ceea ce sugerează că elevii s-au plictisit să repete exercițiul pentru alte molecule / compuși.

La al doilea experiment, 25 de elevi din 42 au reușit să efectueze cel puțin 4 exerciții din lecția 2, dintre care 1 a efectuat 5, iar 3 elevi au efectuat 7 exerciții. 9 elevi au efectuat toate exercițiile, cu un timp variind între 724 sec. și 1630 sec. ($M=1084.4$ sec., $SD=309.6$). Și în acest caz, la exercițiul 4 s-a înregistrat cel mai scurt timp de execuție, dar cu o valoare aproape dublă față de media de la școala de vară. În general, timpii de execuție sunt mai mari, în special, la primul exercițiu, ceea ce arată că elevii au înțeles mult mai greu modul de operare. În acest sens, rezultatele sunt consistente cu cele înregistrate la itemii 6, 8 și 9 din chestionar.

Tabelul 7. Măsuri ale eficacității și eficienței

Sarcină	Școala de vară		După școala de vară	
	Rată succes	Timp	Rată succes	Timp
Le1 Ex 1	90%	261	86%	398
Le1 Ex 2	40%	246	12%	468
Le2 Ex 1	90%	92	71%	215
Le2 Ex 2	100%	72	89%	76
Le2 Ex 3	100%	47	79%	94
Le2 Ex 4	100%	41	90%	71
Le2 Ex 5	100%	72	92%	115
Le2 Ex 6	100%	89	80%	120
Le2 Ex 7	100%	84	81%	195
Le2 Ex 8	100%	76	91%	119
Le3 Ex 1	100%	110	64%	214
Le3 Ex 2	95%	99	50%	148
Le3 Ex 3	80%	149	55%	258

La școala de vară, numai 16 elevi din 20 au reușit să facă toate exercițiile pentru lecția 3 (reacții chimice). Timpul total de execuție variază între 180 sec. și 704 sec. cu o medie de 348 sec ($SD=122.3$). Pe ansamblu, numai 5 elevi (25%) au reușit să efectueze toate exercițiile. Timpul de execuție a variat între 816 sec. și 1468 sec., cu o medie de 1153.8 sec., adică 19.23 min.

La cel de al doilea experiment, 24 de elevi au efectuat cel puțin un exercițiu din lecția 3, dintre care 3 au reușit să efectueze 2 exerciții, iar 2 au reușit să efectueze toate cele trei exerciții cu un timp variind între 384 sec. și 507 sec. ($M=445.2$ sec.). Pe ansamblu, numai 4 elevi (9.5%) au reușit să efectueze toate exercițiile repartizate. Doi dintre ei au efectuat un total de 11 exerciții (4 exerciții suplimentare) iar doi elevi un total de 12 exerciții (5 exerciții suplimentare).

5. Concluzii și direcții de continuare

Pe ansamblu, se poate concluziona că evaluarea formativă a fost un ajutor prețios pentru proiectanți întrucât a condus atât la identificarea avantajelor introducerii tehnologiei AR în școli, dar și la identificarea unor probleme de utilizabilitate.

Evaluarea măsurilor subiective ale satisfacției utilizatorului, bazată pe măsurile cantitative și calitative colectate cu ajutorul chestionarului de utilizabilitate, relevă mai multe aspecte pozitive. ARTP este o platformă utilă pentru învățare: sistemul facilitează înțelegerea, învățarea mai rapidă, ușurează memorarea și este util pentru testarea cunoștințelor. De asemenea, această platformă educațională are o valoare motivațională ridicată. Pe de o parte, aplicația de chimie este atractivă, stimulativă și incitantă, făcând învățarea mai interesantă. Pe de altă parte, elevii au apreciat interacțiunea 3D utilizând tehnici specifice realității îmbogățite, 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-l utiliza în viitor, precum și dorința de a-l 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. Al doilea exercițiu din prima lecție trebuie reproiectat, pentru a fi ușor de utilizat. Trebuie acordată o atenție mai mare ghidării utilizatorului în procesul de selecție a unui element chimic prin furnizarea unui feedback util. De asemenea, vizualizarea trebuie îmbunătățită, pentru a mări lizibilitatea informației afișate pe ecran și a ușura manipularea obiectelor reale. În general, aspectele negative, menționate de elevi, au fost utile pentru completarea problemelor de utilizabilitate identificate de evaluarea euristică.

Repetarea testării cu utilizatori a confirmat rezultatele preliminare obținute la școala de vară, dar a adus

și informații suplimentare, care permit o evaluare mai exactă a scenariului implementat. Numărul mult mai mare de aspecte pozitive referitoare la valoarea motivațională, care au fost menționate de către elevii de la școlile din București, contrastează cu valorile medii mai scăzute decât cele înregistrate pentru participanții la școala de vară. Aceasta înseamnă că problemele de utilizabilitate afectează negativ atitudinea și intenția de utilizare a unei tehnologii.

O nouă versiune a aplicației a fost instalată în 2008. Chestionarul de utilizabilitate a fost conceput pentru a efectua și o evaluare 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 chimie.

Confirmare

Această lucrare este finanțată din proiectul european ARiSE (FP6-027039).

Bibliografie

1. **BACH, C., D. L. SCAPIN:** Obstacles and perspectives for Evaluating mixed Reality Eystems Usability. Mixer workshop, Proc. of IUI-CADUI Conference 2004, ACM Press., pp. 72-79.
2. **BACH, C., D. L. SCAPIN:** Adaptation of Ergonomic Criteria to Human-Virtual Environments Interactions". Proc. of Interact'03, IOS Press, pp. 880-883.
3. **BASTIEN, A., D. L. SCAPIN, D.L.:** Ergonomic criteria for the evaluation of human-computer interfaces. Technical report No. 156, INRIA, Roquencourt, France, 1993.
4. **BOWMAN, D., J. GABBARD, D. HIX:** 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.
5. **DAVIS, F.D.:** Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, MIS Quarterly, Vol. 13, No. 3, 1989, pp. 318-340.
6. **DILLON, A., M. MORRIS:** From „can they?” to „will they?": Extending Usability Evaluation to Address Acceptance. AIS Conference Paper, Baltimore, August 1998.
7. **GABBARD, J., D. HIX, E. SWAN, M. LIVINGSTON, T. HERER, S. JULIER, Y. BAILLOT, D. BROWN:** A Cost-Effective Usability Evaluation Progression for Novel Interactive Systems. Proc. of Hawaii Int. Conf. on Systems Sciences, Track 9, p. 90276c, IEEE.
8. **HVANNBERG, E.T., L – C. LAW:** Classification of usability problems (CUP) scheme. Proc. of Interact 2003, Zurich, Switzerland, 1-5 September.
9. * * *: ISO 9126-4:2001 Software Engineering - Software Product Quality. Part 4: Quality of use.
10. * * *: ISO/DIS 9241-11:1994 Information Technology – Ergonomic Requirements for Office Work with Visual Display Terminal (VDTs) - Guidance on Usability.
11. **LAW, L-C., E. T. HVANNBERG:** Complementarities and Convergence of Heuristic Evaluation and Usability Test: A Case Study of UNIVERSAL Brokerage Platform. Proc. Of NordiCHI Conf. 2002, ACM, pp. 71-79.
12. **SCRIVEN, M.:** Evaluation Thesaurus. 4th ed. Newbury Park, 1991, CA: Sage Publications.
13. **SUN, H., P. ZHANG:** The Role of Moderating Factors in User Technology Acceptance. Int. Journal of Human-Computer Studies, 64 (2006), Elsevier, pp. 53-78.
14. **SWANN II, J., E., J. GABBARD:** Survey of User-Based Experimentation in Augmented Reality. Proc. of 1st Int. Conf. on Virtual Reality. 22-27 July, 2005, Las Vegas, Nevada.
15. **THEOFANOS, M., W. QUESENBERRY:** Towards the Design of Effective Formative Test Reports. Journal of Usability Studies, Issue 1, Vol. 1, 2005, pp. 27-45.
16. **VENKATESH, V., F. D. DAVIS, M. G. MORRIS:** Dead Or Alive? The Development, Trajectory And Future Of Technology Adoption Research. Journal of the AIS, Vol. 8, Issue 4, 2007, pp. 267-286.
17. **WIND, J., K., RIEGE, M. BOGEN:** Spinnstube®: A Seated Augmented Reality Display System. Virtual Environments, Proc. of IPT-EGVE – EG/ACM Symposium, 2007, pp. 17-23., Eurographics.