

RAPORT

realizat în urma evaluării datelor expuse în documentele
”Raport cu privire la procesul de pilotare a ghidului metodologic de predare a fizicii”
și
”Raport privind analiza și interpretarea rezultatelor la teste”

ca urmare a finalizării fazei din 2014 a proiectului ”Reforma Predării Fizicii în Învățământul Preuniversitar” (RPFIP)

Preambul

Acest raport este realizat la cererea responsabililor proiectului ”Reforma Predării Fizicii în Învățământul Preuniversitar” (RPFIP), ca evaluare externă, bazată pe datele indicate de aceștia și neverificate din surse externe.

În consecință, raportul reprezintă o evaluare colegială, în care autorii documentului original sunt priviți ca cercetători *bona fide*. Raportul actual se concentrează în principal asupra aspectelor care țin de testarea de cunoștințe care a fost una din activitățile proiectului, precum și asupra corectitudinii studiului, a designului de cercetare și a analizelor întreprinse.

Acest raport este structurat în mai multe secțiuni:

1. Aspecte legate de testare
 - Dezvoltarea itemilor
 - Dezvoltarea testelor
 - Administrarea testelor
 - Securitatea testelor
2. Aspecte legate de designul cercetării
3. Concluzii

1. ASPECTE LEGATE DE TESTARE

Dezvoltarea itemilor

Itemii au fost dezvoltați de echipe antrenate pentru această activitate, urmărind pașii considerați în genere importanți pentru acest proces. Experți români în problematica pentru care erau dezvoltați itemii (profesori de fizică), au lucrat împreună cu membrii echipei de proiect, sub coordonarea și după un antrenament din partea unui expert internațional a cărui competență este recunoscută (prof. univ. dr. Eduardo Cascallar). Modelul de dezvoltare a itemilor este unul din cele mai celebre și consacrate programe de testare educațională internațională, TIMSS (Trends in International Mathematics and Science Study). Astfel, itemii sunt compuși dintr-un enunț-rădăcină și un număr de 4 răspunsuri precodificate, dintre care unul este corect și restul sunt distractori.

Itemii au fost dezvoltați special pentru testările din cadrul proiectului și nu au fost aplicați anterior decât în cadrul pretestărilor incluse în proiect. Au fost dezvoltați itemi diferiți pentru cele trei faze de testare, pentru a combate efectele memoriei și învățării itemilor asupra

rezultatelor testelor. Nici un participant la testare nu a primit vreun item pe care îl mai primisese anterior, în afara proiectului sau într-o fază anterioară a acestuia. Itemii fazei 1 (intermediare) au fost aplicați și în cadrul fazei 3 (finale), însă doar claselor noi care au intrat în proiect, în pilotare, la acel moment.

Pentru prima testare itemii nu au fost pretestați. Pentru a doua fază de testare au fost pretestați doar itemii dificili (aprox. 25%) dintre itemii testului aplicat. Pentru a treia fază a testării, toți itemii au fost pretestați (februarie 2013). Raportul nu oferă informații despre rezultatele pretestării, despre calitatea itemilor stabilită în urma acestui proces, ori despre posibilele modificări operate în structura itemilor sau testelor în urma procesului de pretestare.

În lumina acestor aspecte, consider că procesul de dezvoltare a itemilor a urmat buna practică în domeniu.

Dezvoltarea testelor

Modelul de măsurare pentru care s-a optat inițial este acela al teoriei răspunsului la item (IRT, Item Response Theory). Acesta este cel mai modern model de măsurare existent în psihometrie și aplicat astăzi în testarea educațională. Deși itemii au fost dezvoltați cu intenția de a fi aplicați într-un test construit pe baza IRT, acest lucru nu a fost posibil, dar utilizarea unor itemi dezvoltați pe baza IRT într-un test dezvoltat pe baza teoriei clasice a testării (CTT, Classical Test Theory) este perfect acceptabilă.

Așadar, testele care au fost administrate în cele din urmă în cele trei faze de evaluare ale proiectului au fost tributare teoriei clasice a testării, însumând numărul de răspunsuri corecte date de fiecare participant pentru a obține scorul participantului, fără a pondera în vreun fel răspunsurile date la diferiți itemi în funcție de dificultatea sau de alți parametri ai respectivilor itemi. Testul inițial a fost compus din 31 de itemi, iar testele intermediar și final din 33 de itemi, ceea ce consider a fi suficient pentru asigurarea unei varianțe bune a scorurilor. Aproximativ 25% din itemii fiecărui test au fost de dificultate redusă, aproximativ 50% au fost de dificultate medie și aproximativ 25% au fost de dificultate ridicată. Dificultățile nu au fost stabilite prin pilotare, ci la momentul redactării, prin prescripțiile date experților care au redactat itemii.

Testele au o fidelitate bună, iar itemii sunt acceptabil de omogeni. Pentru clasa a VII-a/ a VIII-a, consistența internă (indicele Cronbach alpha) a celor trei forme ale testului este plasată la .77, .74 și .79. Omogenitatea itemilor (coeficientul de corelație punct-biserial cu restul testului) este cuprinsă pentru cele trei forme ale testului între .12-.52, .13-.47 și .14-.41. Pentru clasa a IX-a/ a X-a, consistența internă (indicele Cronbach alpha) a celor trei forme ale testului este plasată la .77, .78 și .83. Omogenitatea itemilor (coeficientul de corelație punct-biserial cu restul testului) este cuprinsă pentru cele trei forme ale testului între .11-.51, .12-.47 și .15-.47.

În lumina acestor aspecte, consider că procesul de dezvoltare a testelor a urmat buna practică în domeniu.

Administrarea testelor

Administrarea testelor a fost realizată în trei faze diferite: o fază inițială (martie 2012), una intermediară (iunie 2012) și una finală (martie 2013).

Administrarea testelor a fost făcută în fiecare fază de operatori profesioniști din cadrul rețelei de operatori a institutului Gallup România. Procesul de administrare a fost supervizat prin supervizori de teren provenind din aceeași rețea de cercetare.

Implicarea unor persoane străine de proiect în procesul de administrare a fost atent controlată: profesorii elevilor testați, deși au putut participa la testare, nu au avut dreptul de a oferi indicații elevilor.

În lumina acestor aspecte, consider că procesul de administrare a testelor a urmat buna practică în domeniu.

Securitatea testelor

Testele au fost create special pentru acest proiect, toți itemii fiind scriși și utilizați doar în acest proiect. Redactarea itemilor a fost realizată în condiții de siguranță, iar păstrarea lor a fost de asemenea realizată corect. Testele au fost pilotate (parțial în faza 2 și complet în faza 3) cu un control bun al securității. Înainte de administrarea lor în proiectul propriu-zis, testele nu au fost cunoscute de operatorii care le-au aplicat sau de supervizorii acestora. Testele au fost tipărite, multiplicare și primite de operatori doar în prezența administrării. Toate testele au fost administrate în toate punctele de administrare în aceeași zi. Implicarea persoanelor străine de proiect a fost controlată cu grijă și atunci când li s-a permis accesul în sala de testare (cum li s-a permis de exemplu profesorilor elevilor testați), aceste persoane nu au avut voie să dea indicații.

Ca parte a efortului de asigurare de oportunități egale și pentru familiarizarea cu formatul itemilor testului, itemi similari în format cu cei ce urmau să fie administrați în proiect au fost făcuți cunoscuți profesorilor elevilor ce urmau să fie testați, cu un număr de zile înaintea testării. Nici un item propriu-zis ce urma să fie utilizat în teste nu a făcut parte din acești item de exercițiu.

În lumina acestor aspecte, consider că procesul de securitate a testelor a urmat buna practică în domeniu.

2. ASPECTE LEGATE DE DESIGNUL CERCETĂRII

Designul de cercetare utilizat de proiect este un design tipic pentru dezvoltarea unor intervenții bazate pe dovezi empirice ("evidence-based interventions") ca parte a unor politici bazate pe dovezi empirice ("evidence-based policies"). Designul a fost dezvoltat pentru a pune în evidență efectul avut de implementarea Ghidului Metodologic de Predare a Fizicii asupra performanței elevilor de liceu.

Metoda utilizată a fost aceea a experimentului, însă designul de cercetare este unul cvasi-experimental, fiind absentă distribuția randomizată a participanților în grupurile experimentale/de control. Considerăm că o distribuție aleatorie ar fi fost imposibilă cu eșantioane rezonabile: de vreme ce fizica se învață în clasă, ar fi fost imposibil să fie distribuți

elevii în grupuri diverse și ar fi fost posibil cel mult să fie distribuite clase întregi ca aparținând unuia sau altuia dintre grupuri. Aceasta ar fi extins însă foarte mult numărul total de participanți. Așadar, consider designul de cercetare a fi unul rezonabil și probabil cel mai bun la care se putea ajunge în condițiile constrângerilor de buget existente.

Designul cercetării a fost dezvoltat cu grijă astfel încât să fie controlate o serie din efectele care ar fi putut influența rezultatele: au fost incluși în design elevi aflați în două niveluri de școlarizare diferite, anume clasele 7-8 și apoi clasele 9-10, clasele elevilor au fost distribuite atât unui grup experimental (de intervenție) cât și unui grup de control etc. Pentru clasele 7-8 grupul experimental a conținut 143 de participanți, iar grupul de control 125. Pentru clasele 9-10 grupul experimental a conținut 74 de participanți, iar grupul de control 86.

Datele au fost analizate pe baza procedurilor ANOVA, în mod specific fiind aplicat modelul factorial complet cu măsurători repetate. Rezultatele extrase din analiză arată faptul că pentru clasele 7-8, atunci când se iau în considerare diferențele între prima testare și testarea finală, grupul de control arată un increment de 4%, pe când grupul experimental arată un increment de 18%. Diferența este de 16%, rezultatul grupului experimental fiind de 4.5 ori mai bun decât cel al grupului de control. Rezultatul este semnificativ la un prag de probabilitate de $p < .01$.

Pentru clasele 9-10, atunci când se iau în considerare diferențele între prima testare și testarea finală, grupul de control arată un increment de 4.2%, pe când grupul experimental arată un increment de 8.2%. Diferența este de 4%, rezultatul grupului experimental fiind de 2 ori mai bun decât cel al grupului de control. Rezultatul este semnificativ la un prag de probabilitate de $p < .05$ ($p = .013$).

În concluzie, consider ca designul de cercetare utilizat este potrivit pentru a pune în evidență o relație cauzală între manipularea realizată (introducerea ghidului de predare a fizicii) și efectul obținut (îmbunătățirea performanțelor academice la fizică). Consider că procedura statistică aplicată este cea indicată pentru acest tip de design. Consider că analiza statistică a fost realizată corect și că rezultatul a fost interpretat corect. În consecință, consider că datele analizate indică o relație cauzală semnificativă statistică și cu un efect notabil între introducerea ghidului de predare a fizicii și incrementul în performanțele academice la această materie.

3. CONCLUZII

Date fiind aspectele menționate mai sus, consider că aspectele legate de dezvoltarea itemilor, dezvoltarea testelor, administrarea testelor și securitatea testelor, designul de cercetare, culegerea, prelucrarea și interpretarea datelor, așa cum au fost aceste aspecte descrise în acele documente ale proiectului menționate în titlul acestui raport, arată urmarea bunelor practici internaționale pentru realizarea acestor activități.

Prof. Univ. Dr. Dragoș Iliescu
12 Septembrie 2014