
Generalizabilitatea indicilor de validitate în contextul adaptării testelor

Balazsi Robert ¹

Universitatea Babeş-Bolyai, Cluj-Napoca

Stabilirea unor criterii de utilizare a testelor psihologice de către Comisia de Metodologie a Colegiului Psihologilor din România marchează o etapă nouă în procesul de evaluare psihologică. Aceste criterii impun utilizatorului achiziția și folosirea unor probe a căror valoare empirică a fost dovedită. Pe de altă parte, experții care elaborează sau adaptează astfel de instrumente, sunt obligați să întreprindă demersuri în vederea culegerii unor dovezi empirice care să ateste calitățile psihometrice ale testului. Câteodată chiar psihologii practicieni sunt cei care dezvoltă (sau ar trebui să dezvolte) o serie de cercetări de validare și standardizare, a probelor de evaluare, pe populațiile specifice contextului în care își desfășoară activitatea.

Întrebările esențiale adresate de către utilizator, dar și de către comisia de specialitate, ofertanților de teste se referă la: 1) problema validității testului, adică la faptul că testul trebuie să măsoare ceea ce își propune să măsoare, astfel să garanteze consistența concluziilor bazate pe rezultatele oferite de acesta; 2) problema fidelității testului, adică rezultatele obținute să fie stabile, atât pe parcursul testării (consistență internă), cât și în timp (fidelitate test-retest); și, 3) problema existenței unor etaloane standard care să reprezinte repere (în populația de referință) care să permită o interpretare comparativă a rezultatelor individuale.

Lucrarea de față aduce în discuție problema validității testelor de selecție de personal în contextul adaptării acestora la diferite populații. În ce măsură este sau nu necesar un demers de validare a unor probe a căror validitate a fost deja demonstrată în alte contexte? De exemplu, în ce măsură bateriile de teste utilizate în selecția și evaluarea periodică a personalului din transporturi, pot fi aplicate la fel de bine și altor profesii (ex. electricieni, agenți comerciali, militari).

Reformulând în termeni tehnici, lucrarea de față pune problema generalității vs. specificității indicilor de validitate predictivă (Pearlman, Schmidt, & Hunter, 1980).

În anii '60-'70 majoritatea psihologilor din domeniul selecției de personal considerau că indicele de validitate a testelor de selecție are un caracter strict local. Există două argumente majore care susțin această concluzie, și anume: 1) valoarea relativ redusă a indicilor de corelație calculați între teste și performanțe în ocupațiile pentru care inițial aceste teste păreau a fi relevante (Braun, 1988) și 2) variabilitatea crescută a valorilor indicilor de validitate, chiar și în contextul în care testul de selecție și specificul ocupației erau similare (Pearlman, Schmidt & Hunter, 1980). Concluzia firească a acestor observații a fost că testele de selecție sunt specifice contextului utilizării. Nu există posibilitatea de a prezice care dintre teste va funcționa ca și predictor într-o selecție ulterioară. În practica selecției de personal, această concluzie echivala cu necesitatea de a valida testele psihologice pentru orice context pentru care nu au fost încă validate.

Aportul deosebit a lui Schmidt & Hunter (1977) constă în faptul că au observat că această concluzie nu este singura ce poate explica valoarea redusă a corelațiilor între criteriu și predictor, respectiv variabilitatea indicilor de validitate. În esență, orice indice de validitate este exprimat printr-un coeficient de corelație, calculat pe valorile variabilelor predictor și criteriu. Lipsa unei corelații între două variabile poate fi interpretată în două feluri: fie cele două variabile în realitate nu sunt corelate, fie rezultatul calculat este un artefact, datorat unor factori care intervin în procesul de măsurare a performanțelor (ex. fidelitatea testului) și atenuază valoarea calculată a coeficientului de validitate. De asemenea, fidelitatea testului, alături de volumul redus al eșantionului și restricția domeniului de variabilitate a scorurilor (o caracteristică implicită a demersului de validare a probelor de selecție) contribuie și la creșterea variabilității coeficienților de corelație.

Raționamentul specificității indicelui de validitate nu ia în considerare posibilul efect al prezenței artefactelor asupra coeficientului de validitate calculat. Lipsa corelației este

interpretată doar ca un indicator al lipsei oricărei asocieri reale a două variabile.

Raționamentul demersului de verificare a generalității indicelui de validitate se bazează pe estimarea varianței datorate variabilelor artefactuale și raportarea acestuia la varianța totală. În principiu, demersul de verificare a generalității indicelui de validitate se bazează pe sinteza psihometriei și a meta-analizei. Conform psihometriei clasice scorul măsurat se compune din nivelul aptitudinii (scor real) și eroarea de măsurare. Varianța scorului măsurat reprezintă suma dintre varianța scorului real și varianța erorii de măsurare.

$$\text{Var}(x) = \text{Var}(\tau) + \text{Var}(\epsilon)$$

Calculul nivelului aptitudinii se face prin stabilirea unui interval de încredere utilizând estimatori ai erorii de măsurare.

Această formulă și raționamentul aferent a fost preluat de Schmidt & Hunter (1977) și adaptat la situația de generalizare a indicelui de validitate, astfel realizându-se o sinteză a elementelor psihometrice și de meta-analiză (Braun, 1988).

$$\text{Var}(r) = \text{Var}(\rho) + \text{Var}(\epsilon)$$

Varianța indicilor de validitate este egală cu varianța reală a indicelui de validitate și varianța artefactuală. Această ultimă componentă include cele trei surse de eroare: fidelitatea testelor de evaluare a variabilei predictor și criteriu, eroarea standard a indicilor de validitate și eroarea datorată diferențelor de restricție domeniului de variabilitate a scorurilor a variabilei criteriu sau predictor.

Un coeficient de corelație, calculat pe un eșantion, reprezintă gradul real de asociere a două variabile, doar în contextul unor măsurători perfecte (valoarea indicelui de fidelitate este 1). Din moment ce măsurătorile sunt afectate de erori de măsurare, în calculul indicilor de validitate trebuie să ținem cont de aceste erori. Metoda uzuală este de a calcula indicele de validitate utilizând corecția pentru atenuare (vezi Nunnally & Bernstein, 1994).

$$r_{xy,cor} = r_{xy} \sqrt{(r_{xx} * r_{yy})}$$

Rezultatul ($r_{xy,cor}$) astfel obținut reprezintă valoarea pe care o ia corelația (r_{xy}) ținând cont de faptul că măsurarea nu este una perfectă ($r_{xx}=r_{yy}=1$).

Mărimea eșantionului afectează mărimea erorii standard. Cu cât eșantionul este mai mic, cu atât vor fluctua mai mult valorile coeficienților de corelație calculați pe eșantioanele selectate aleator (Schmidt, Hunter, Pearlman & Hirsh, 1985). Dacă urmărim tabelul valorilor critice ale lui r vom observa că pentru un alfa setat la 0.05, pentru un eșantion de $N=22$, r critic are valoarea de 0.42, în timp ce pentru $N=92$, această valoare este de 0.2. Cu cât eșantionul este mai mic, cu atât mai mare este fluctuația expectată a valorilor calculate (Braun, 1988).

Restricția domeniului de variabilitate a scorurilor are ca rezultat reducerea variabilității valorilor măsurate. Această reducere artificială a variabilității atrage după sine distorsionarea estimării valorii coeficientului de corelație calculat. Acest fenomen apare de obicei în contextul demersului de validare relativă la criteriu, corelația între predictor și criteriu fiind calculată doar pentru subiecții care au fost selectați (ex. validitatea relativă la criteriu a testelor de admitere de tip GRE, predictor fiind scorul la test și criteriul performanțele înregistrate pe parcursul anilor de studiu a subiecților declarați admiși). Din punctul de vedere al generalizării indicelui de validitate este mai important faptul că restricția domeniului de variabilitate are valori diferite în studii diferite. Cauza acesteia este că punctul de cut-off variază de la studiu la studiu, în funcție de scopul selecției.

Logica demersului de verificare a generalității indicelui de validitate este următorul: din moment ce eroarea de eșantionare și erorile de măsurare (fidelitate redusă și restricția domeniului de variabilitate) afectează calculul indicelui de validitate, este firesc să observăm o anumită variabilitate a acestuia. Estimarea varianței artefactuale permite inferențe asupra generalității sau specificității indicelui de validitate (Braun, 1988).

Demersul include următorii pași:

- se calculează valoarea medie a indicilor de validitate ținând cont de factorii de corecție pentru fidelitatea redusă și restricția domeniului de variabilitate a scorurilor. De asemenea se estimează varianța observată a indicilor de validitate.
- se estimează valoarea varianței artefactuale.
- se calculează valoarea $\text{Var}(\rho) = \text{Var}(r) - \text{Var}(\epsilon)$. Dacă varianța reală a coeficienților de validitate este zero (sau valoare apropiată de zero, raportat la o valoare

critică a unei distribuții care respectă o distribuție χ^2), atunci se poate afirma că varianța observată se poate explica în întregime prin varianța artefactuală. Adică coeficientul de validitate se poate generaliza, ponderea factorilor locali fiind neglijabilă. O valoare diferită de zero este interpretată ca un argument al existenței unei fluctuații în scorurile reale ale indicilor de validitate. Această fluctuație însă nu se poate explica prin artefacte ale eșantionării sau măsurării. Ca urmare, este atribuită factorilor specifici contextului. În acest caz, cercetarea va încerca să identifice variabilele moderatoare care explică această fluctuație a valorilor reale ale indicelui de validitate.

O altă modalitate de a exprima prezența varianței indicilor de validitate este calculul raportului $V(\epsilon)/V(r)*100$. Valoarea minimă pentru care se verifică ipoteza generalizării indicelui de validitate este 75%. Sub această valoare se consideră că ponderea factorilor contextuali este semnificativă (Schmidt, Gast-Rosenberg, & Hunter, 1980).

Metoda descrisă reprezintă demersul clasic elaborat de Schmidt & Hunter (1977). Componenta esențială a acestui model, estimarea varianței artefactuale (implicit a celei reale), a fost mult criticată pe parcursul ultimelor decenii. Majoritatea modelelor alternative au fost elaborate prin modificarea acestei componente critice a demersului inițial (Raju & Burke, 1983; James, Demaree, & Mulaik, 1986).

Varianta bayesiană a demersului de verificare a generalității indicelui de validitate critică ipoteza $Var(p)=0$ a demersului inițial. Din punct de vedere practic este greu de acceptat ca elementele contextuale ale selecției să nu aibă nici un efect asupra indicelui de validitate. Această ipoteză este înlocuită cu ipoteza apriori (care permite o estimare a indicelui de validitate în populație) și o funcție a verosimilității (indicele de validitate calculat în studiul actual, care este un indicator independent al valorii indicelui de validitate în populație), pe baza cărora se estimează probabilitatea ipotezelor posteriori (o nouă valoare estimată a indicelui de validitate la nivelul populației) (Brannick, 2001).

Întrebarea finală este ce ne spune demersul de generalizare a validității despre generalitatea, respectiv specificitatea, indicelui de validitate. În acest context întrebarea utilizatorilor și a celor care oferă astfel de teste

este: mai e sau nu necesar să validăm testele psihologice utilizate pentru selecție de personal pentru toate contextele în care sunt folosite? La această întrebare nu există însă, deocamdată, un răspuns ferm. La nivel metodologic, modelele actuale ale verificării generalității sunt din ce în ce mai complexe, astfel sporind precizia estimărilor. La nivel instituțional, din 1999 problema generalității validității a fost inclusă în Principiile elaborate de SIOP (Society for Industrial and Organizational Psychology). Acest document apreciază că rezultatele demersului de verificare a generalității reprezintă argumente suficiente pentru selecția instrumentelor de evaluare psihologică. De asemenea, se consideră că existența unei convergențe între rezultatele oferite de aceste studii și studiile de validare locală, sporesc credibilitatea celor din urmă (Braun, 1988).

Bibliografie

- Brannick, M., T. (2001) Implications of Empirical Bayes Meta-Analysis for Test Validation. *Journal of Applied Psychology* 2001. Vol. 86, No. 3,468-480.
- Braun, H., W., H., (1988) Test Validity. Editura Lavrence Erlbaum Associates, London.
- James, L. R., Demaree, R. G., & Mulaik, S. A. (1986). A note on validity generalization procedures. *Journal of Applied Psychology*, 71, 440-450.
- Nunnally, J., & Bernstein, I. (1994) Psychometric test theory. Editura McGraw Hill, New York.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65, 373-406.
- Raju, N. S., & Burke M. J. (1983). Two new procedures for studying validity generalization. *Journal of Applied Psychology*, 68,382-395.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540.
- Schmidt, F. L., Gast-Rosenberg, I. F., & Hunter, J. E. (1980). Validity generalization results for computer programmers. *Journal of Applied Psychology*, 65, 463-661.
- Schmidt, F. L., Hunter, J. E., Pearlman, K., & Hirsh, H. R. (1985) Forty questions about validity generalization and meta-analysis. *Personnel Psychology*, 38, 697-798.