

Metodologia de analiză a rezultatelor evaluărilor naționale standardizate în învățământul general

I. Dispoziții generale

1. Metodologia de analiză a rezultatelor evaluărilor naționale standardizate în învățământul general (în continuare - *Metodologie*) definește cadrul conceptual, principiile, procedurile și instrumentele utilizate pentru analiza rezultatelor obținute la evaluările naționale standardizate din învățământul general susținute în Republica Moldova.

2. Metodologia este elaborată și pusă în aplicare de Agenția Națională pentru Curriculum și Evaluare (în continuare – ANCE).

3. Obiectivele specifice ale Metodologiei sunt:

- 1) evaluarea calității psihometrice a testelor utilizate în evaluările naționale standardizate;
- 2) evaluarea gradului de realizare a curriculumului, domeniilor de conținut și a competențelor specifice;
- 3) evaluarea echității rezultatelor pe subgrupuri de candidați și a tendințelor multianuale de evoluție a performanțelor;
- 4) fundamentarea deciziilor referitoare la structura testelor naționale standardizate, revizuirea curriculumului și ajustarea politicilor în educație.

4. Obiectul Metodologiei îl constituie ansamblul datelor obținute în urma administrării evaluărilor naționale standardizate: rezultatele individuale ale candidaților pe itemi și pe test, caracteristicile itemilor (nivel cognitiv, domeniu de conținut, competențe specifice vizate, tip de item, punctaj maxim) și caracteristicile populației de candidați (gen, mediu de reședință, limba de instruire, statut socioeconomic, regiune, tip de instituție de învățământ, statutul de candidat cu cerințe educaționale speciale).

5. Metodologia se aplică la analiza tuturor evaluărilor naționale standardizate din învățământul general, organizate de către ANCE, indiferent de disciplină sau de ciclul de studii.

6. În funcție de scopul principal, de modul de administrare și de consecințele pe care le produc rezultatele, evaluările naționale standardizate organizate de ANCE se clasifică în:

- 1) evaluări certificative ale căror rezultate produc consecințe individuale pentru candidați;
- 2) evaluări de monitorizare a sistemului, aplicate întregii populații eligibile;
- 3) evaluări de monitorizare aplicate unui eșantion reprezentativ;
- 4) evaluări tematice.

7. În sensul prezentei Metodologii, principalele noțiuni utilizate au următoarele semnificații:

- 1) *Item de test* — unitate elementară a testului evaluată pe o scală de punctaj predefinită;
- 2) *Indice de dificultate al itemului (p -value)* — raportul dintre scorul mediu obținut de candidați la un item și punctajul maxim posibil al itemului respectiv;
- 3) *Indice de discriminare (D)* — măsură a capacității unui item de a diferenția între candidații cu performanțe înalte și cei cu performanțe scăzute la test;
- 4) *Coeficient Alpha Cronbach (α)* — indicator al consistenței interne a testului, exprimând gradul de intercorelație dintre itemi. Nu constituie, prin el însuși, o dovadă a unidimensionalității sau a validității testului;
- 5) *Corelația punct-biserială (r_{pbis})* — corelația dintre răspunsul la un item dihotomic și scorul la test. Reprezintă un caz particular al corelației item-test aplicabil itemilor cu punctaj dihotomic;

- 6) *Eroarea standard de măsurare (SEM)* — cuantificarea preciziei cu care testul măsoară abilitatea candidatului, exprimată în unități de scor;
 - 7) *Distractor* — variantă incorectă de răspuns la un item cu alegere multiplă;
 - 8) *Matricea Q* — matrice de specificații care asociază fiecărui item nivelul cognitiv, domeniul de conținut, competențele specifice măsurate.
 - 9) *Grad de realizare* — procentul din punctajul maxim obținut, în medie, de candidați la nivelul unei unități de analiză (item, test, nivel cognitiv, domeniu de conținut, competență specifică).
 - 10) *Nivel cognitiv* — categorie a taxonomiei utilizate pentru clasificarea itemilor în funcție de complexitatea operațiilor mentale solicitate;
 - 11) *Domeniu de conținut* — arie tematică distinctă a curriculumului disciplinei, în care se grupează itemii pentru analiză;
 - 12) *Competență specifică* — capacitate dezvoltată, definită în curriculumul disciplinei, evaluată prin unul sau mai mulți itemi ai testului;
 - 13) *Coeficientul de corelație intraclasă (ICC)* — indicator al proporției din variația totală a unei măsuri care se manifestă între unități de grupare (instituții, clase).
8. Aplicarea prezentei Metodologii se realizează cu respectarea următoarelor principii:
- 1) *Principiul rigorii științifice* — toate analizele se efectuează utilizând metode și formule validate în literatura psihometrică, cu raportarea explicită a pragurilor de interpretare și a limitelor de aplicabilitate;
 - 2) *Principiul transparenței metodologice și al reproductibilității* — toate procedurile, formulele și pragurile de interpretare utilizate sunt prezentate explicit în rapoartele de analiză, iar analizele sunt documentate astfel încât să poată fi reproduse pe baza acelorași date, reguli și proceduri de calcul;
 - 3) *Principiul echității* — analizele verifică în mod sistematic dacă instrumentul de evaluare măsoară același construct pentru toate subgrupurile de candidați, dacă nu introduce obstacole irelevante în raport cu competența vizată și dacă permite interpretări comparabile ale scorurilor. Diferențele constatate pot reflecta inegalități reale de oportunități educaționale;
 - 4) *Principiul comparabilității temporale* — analizele se efectuează în formă consecventă de la o sesiune la alta, pentru a permite analiza tendințelor multianuale;
 - 5) *Principiul protecției datelor personale* — analizele utilizează date individuale exclusiv anonimizate, în conformitate cu legislația privind protecția datelor cu caracter personal;
 - 6) *Principiul utilității pentru factorii de decizie* — rezultatele analizelor sunt formulate într-un mod accesibil, cu accent pe constatări și recomandări operaționale;
 - 7) *Principiul îmbunătățirii continue* — Metodologia este supusă revizuirii periodice, pentru a încorpora progresele științifice din domeniul evaluării educaționale și pentru a integra noi instrumente de analiză;
 - 8) *Principiul validității interpretărilor* — fiecare concluzie se raportează la scopul evaluării și la utilizarea prevăzută a scorurilor și se susține prin dovezi privind conținutul testului, structura sa internă, precizia măsurării și echitatea funcționării;
 - 9) *Principiul delimitării dintre asociere și cauzalitate* — diferențele și relațiile observate în date nu se interpretează ca efecte cauzale decât atunci când evaluarea a fost realizată printr-un design adecvat unei asemenea analize. În celelalte situații, explicațiile posibile se formulează ca ipoteze supuse investigației suplimentare.

II. Date generale pentru analiză

9. Analiza rezultatelor evaluărilor naționale standardizate se realizează pe baza următoarelor categorii de date:

- 1) datele privind itemii testului: identificator unic al itemului, tip de item, nivelul cognitiv vizat, domeniul de conținut, competențele specifice măsurate, punctajul maxim posibil;
- 2) datele privind rezultatele candidaților: scorul obținut la fiecare item de către fiecare candidat, scorul total la test, varianta de răspuns selectată (pentru itemii cu alegere multiplă);
- 3) datele demografice generale ale candidaților: gen, mediu de reședință (urban/rural), limba de instruire, statut socioeconomic, regiunea/raionul, tipul instituției de învățământ, statutul de candidat cu cerințe educaționale speciale;
- 4) datele privind administrarea testului: durata efectivă, adaptările și condițiile speciale de administrare acordate candidaților cu cerințe educaționale speciale, eventualele incidente sau abateri procedurale.

10. Datele necesare analizei sunt furnizate de către structurile responsabile cu administrarea evaluărilor naționale standardizate, în format electronic structurat, în conformitate cu specificațiile tehnice stabilite de către ANCE.

11. Datele recepționate sunt supuse unei verificări preliminare de calitate, prin care se urmărește:

- 1) completitudinea datelor;
- 2) consistența internă a datelor;
- 3) validitatea variabilelor demografice;
- 4) identificarea și tratarea cazurilor atipice.

12. Verificarea calității datelor se realizează secvențial: analiza psihometrică și interpretativă a rezultatelor nu se inițiază înainte de finalizarea verificărilor prevăzute în Anexa 2 și de validarea bazei de date.

13. Eventualele neconformități identificate în etapa de verificare a calității datelor sunt documentate explicit în raportul de analiză.

14. Analiza rezultatelor evaluărilor naționale standardizate, evaluărilor tematice, evaluărilor pe eșantion reprezentativ se efectuează după fiecare sesiune a examenelor, sau a evaluărilor desfășurate pe parcursul unui interval de cel mult 6 luni de la încheierea acestora.

15. Suplimentar față de analiza per sesiune, la fiecare 3–5 ani se poate elabora un raport sintetic multianual, care urmărește tendințele de evoluție a rezultatelor și fundamentează deciziile de revizuire curriculară.

III. Analiza descriptivă a rezultatelor

Secțiunea 1. Indicatori de tendință centrală și distribuție

16. Analiza descriptivă a rezultatelor are ca obiectiv prezentarea performanței globale a candidaților la evaluările naționale standardizate prin indicatori statistici care caracterizează tendința centrală, dispersia, forma distribuției scorurilor, ratele de promovare și diferențele între subgrupuri demografice și instituționale. Aceasta reprezintă prima etapă a oricărui raport de analiză și constituie baza factuală pentru interpretarea ulterioară a calității instrumentului de evaluare și a gradului de realizare a curriculumului.

17. Media aritmetică a scorurilor (sau a notelor) este indicatorul principal al performanței globale. Se calculează la nivel național și separat pe fiecare variabilă de dezagregare. Media trebuie raportată întotdeauna împreună cu numărul de candidați care au stat la baza calculului și cu un indicator de dispersie (deviație standard sau intervale interpercentilice).

18. Mediana scorurilor reprezintă valoarea care împarte distribuția în două jumătăți. Mediana se raportează în paralel cu media. O diferență semnificativă între medie și mediană indică o distribuție asimetrică. Mediana este mai puțin dependentă de valori extreme decât media și oferă o imagine mai stabilă a performanței tipice.

19. Deviația standard a scorurilor exprimă gradul de distribuție a rezultatelor în jurul mediei, în aceleași unități de măsură ca scorul. O deviație standard mare indică o populație de candidați eterogenă din punct de

vedere al nivelului de pregătire. O deviație standard mică indică omogenitate. Coeficientul de variație (raportul dintre deviația standard și medie, exprimat în procente) permite comparații între distribuții cu medii diferite — de exemplu, între discipline sau între sesiuni.

20. Percentilele oferă o descriere mai detaliată a distribuției scorurilor. Se raportează percentilele 10, 25, 50 (mediana), 75 și 90. Aceste praguri de valori reprezintă următoarele:

- 1) P10 — pragul sub care se află 10% dintre candidați (limita superioară a candidaților cu performanțe foarte scăzute);
- 2) P25 — primul cuartil (limita superioară a primului sfert al candidaților);
- 3) P50 — mediana;
- 4) P75 — al treilea cuartil (limita inferioară a sfertului superior);
- 5) P90 — pragul peste care se află 10% dintre candidați (limita inferioară a candidaților cu performanțe foarte ridicate).

Diferența $P75 - P25$ (intervalul intercuartilic, IQR) descrie dispersia centrală a distribuției și este indicatorul recomandat pentru distribuții asimetrice.

21. Coeficientul de asimetrie (skewness) descrie gradul și direcția în care distribuția scorurilor se abate de la simetrie. Valorile coeficientului de asimetrie se interpretează după cum urmează:

- 1) $|\text{skewness}| < 0,5$ — distribuție aproximativ simetrică;
- 2) $0,5 \leq |\text{skewness}| < 1,0$ — asimetrie moderată;
- 3) $|\text{skewness}| \geq 1,0$ — asimetrie puternică.

Valorile pozitive ale coeficientului de asimetrie corespund unei distribuții alungite spre dreapta (concentrare a scorurilor în zona joasă), iar valorile negative unei distribuții alungite spre stânga (concentrare a scorurilor în zona înaltă). Interpretarea acestor valori depinde de tipul și scopul testului: asimetria poate reflecta o nepotrivire între dificultatea testului și distribuția performanțelor candidaților, dar poate rezulta din caracteristicile populației evaluate. În cazul evaluărilor de tip criterial, o distribuție asimetrică negativă, cu majoritatea candidaților concentrată în zona înaltă, poate indica faptul că majoritatea au atins standardul vizat și nu constituie, prin ea însăși, un defect al testului. Simetria distribuției nu reprezintă un obiectiv în sine. Interpretarea asimetriei se formulează ca ipoteză, care se verifică prin examinarea itemilor, a scopului testului și a populației evaluate.

22. Coeficientul de applatizare (kurtosis) descrie cât de „ascuțită” sau „aplatizată” este distribuția față de distribuția normală. Valorile coeficientului de applatizare se interpretează după cum urmează:

- 1) coeficientul de applatizare ≈ 0 - distribuție similară celei normale (mezokurtică);
- 2) coeficientul de applatizare > 0 - distribuție leptokurtică (valori extreme mai frecvente);
- 3) coeficientul de applatizare < 0 - distribuție platokurtică (valori extreme mai rare).

Coeficientul de asimetrie și coeficientul de applatizare oferă indicii utile, dar parțiale, asupra formei distribuției. Forma distribuției reflectă atât proprietățile testului, cât și distribuția abilității în populația evaluată. Valori atipice ale acestor indicatori constituie un semnal pentru examinarea suplimentară a distribuției scorurilor și a itemilor.

23. Distribuția notelor pe scala oficială se raportează la nivel național și separat pentru fiecare variabilă de dezagregare relevantă. Pentru fiecare notă se raportează: numărul de candidați (frecvența absolută), procentul din totalul candidaților (frecvența relativă), procentul cumulat (proporția de candidați cu nota respectivă sau mai mică). Forma distribuției notelor oferă o imagine intuitivă a performanței globale, accesibilă factorilor de decizie și publicului larg.

24. Rata de promovare reprezintă procentul de candidați care au obținut o notă egală sau mai mare decât pragul de promovare stabilit. Rata de promovare se raportează la nivel național și se dezagregă pe variabilele

de interes. La calcularea ratei de promovare trebuie distinsă în mod explicit baza de calcul (candidați prezentați, candidați înscriși, sau toți candidații în drept), pentru a evita comparațiile eronate între sesiuni sau între variabile de interes.

25. Rata de neprezentare reprezintă procentul candidaților înscriși care nu s-au prezentat la proba de evaluare. Variațiile semnificative ale ratei de neprezentare între sesiuni sau între subgrupuri necesită investigare separată — pot reflecta probleme de organizare, lipsă de pregătire percepută, sau dificultăți socio-economice ale anumitor subgrupuri.

Secțiunea 2. Dezagregarea rezultatelor pe variabile de interes

26. Toți indicatorii de tendință centrală și distribuție descriși în secțiunea 1 se calculează nu doar la nivel național, ci și dezagregat pe variabile de interes.

27. Variabilele de dezagregare obligatorii sunt următoarele:

- 1) genul candidatului;
- 2) mediul de reședință;
- 3) limba de instruire;
- 4) regiunea/raionul;
- 5) statut socioeconomic (atunci când datele sunt disponibile și fiabile);
- 6) tipul instituției de învățământ;
- 7) statutul de candidat cu cerințe educaționale speciale (atunci când datele sunt disponibile și fiabile).

28. Dezagregarea se efectuează pentru toți indicatorii principali: media, mediana, deviația standard, percentilele, rata de promovare, distribuția notelor. Rezultatele se prezintă în formă tabelară comparativă, însoțită de reprezentări grafice (diagrame de bare, box-ploturi) care permit identificarea vizuală a diferențelor.

29. Diferențele observate între subgrupuri nu trebuie interpretate doar pe baza diferenței numerice de medii, ci trebuie evaluate prin teste de semnificație statistică, care indică, în cazul rezultatelor provenite dintr-un eșantion, probabilitatea ca diferența observată să fie o consecință a hazardului eșantionării. Pentru fiecare comparație se raportează: valoarea testului, pragul de semnificație p , intervalul de încredere de 95% și mărimea efectului (effect size). Interpretarea pragului de semnificație p ca probabilitate a diferenței datorate hazardului de eșantionare este validă numai atunci când rezultatele provin dintr-un eșantion.

30. Evaluarea unei diferențe între subgrupuri se realizează prin examinarea cumulativă a: diferenței absolute, mărimii efectului, intervalului de încredere, erorii de măsurare asociate, stabilității diferenței în analizele de sensibilitate, relevanței educaționale a diferenței, condițiilor de comparabilitate a măsurării între grupurile comparate, inclusiv a rezultatelor analizei funcționării diferențiate a itemilor, atunci când aceasta este disponibilă.

31. Pentru variabile continue (scoruri, medii), pot fi utilizate următoarele teste statistice: testul t Student, analiza varianței (ANOVA). Aceste teste presupun observații independente provenite dintr-o eșantionare aleatorie simplă. Atunci când evaluarea se realizează pe un eșantion complex (cu ponderi, stratificare sau grupare a candidaților pe unități), se utilizează proceduri de estimare care iau în calcul designul de eșantionare. Pentru variabile categoricale (rate de promovare, rate de neprezentare), se utilizează testul chi-pătrat al independenței.

32. Rezultatele nu se raportează separat pentru subgrupurile la care numărul de candidați este insuficient pentru asigurarea confidențialității datelor sau pentru obținerea unei precizii acceptabile a estimării. Atunci când precizia este insuficientă, rezultatul se prezintă agregat cu alte categorii, se marchează explicit ca instabil sau nu se raportează. Numărul minim de candidați necesar raportării se stabilește ținând cont de precizia estimării și de structura datelor.

IV. Analiza calității testului din perspectiva psihometrică

33. Analiza calității testelor are ca obiectiv principal evaluarea caracteristicilor psihometrice ale instrumentelor de evaluare utilizate în evaluările naționale standardizate, în vederea asigurării că acestea măsoară în mod fidel, valid și echitabil competențele vizate de curriculum.

Secțiunea 1. Indicatori statistici de bază în analiza calității testelor

34. Indicele de dificultate al itemului (p -value) reprezintă raportul dintre scorul mediu obținut de candidați la item și punctajul maxim posibil al itemului respectiv. Indicele de dificultate exprimă gradul în care un item este ușor sau dificil pentru populația de candidați.

35. Valorile indicelui se interpretează după cum urmează:

- 1) item foarte dificil: $p < 0,20$;
- 2) item dificil: $0,20 \leq p < 0,40$;
- 3) item de dificultate medie: $0,40 \leq p \leq 0,70$;
- 4) item ușor: $0,70 < p \leq 0,85$;
- 5) item foarte ușor: $p > 0,85$.

Un test bine construit ar trebui să aibă o distribuție variată a indicilor de dificultate, cu accentul, de regulă, pe itemii cu dificultate medie (0,40–0,70).

36. Indicele de discriminare al itemului (D) reprezintă în ce măsură un item reușește să diferențieze între candidații cu performanțe ridicate și cei cu performanțe scăzute la test. Se calculează ca diferența între scorul mediu la item al grupului superior (27% dintre candidații cu scorurile totale cele mai mari) și scorul mediu la item al grupului inferior (27% dintre candidații cu scorurile totale cele mai mici), raportată la punctajul maxim al itemului.

37. Valorile indicelui de discriminare se interpretează după cum urmează:

- 1) discriminarea foarte bună (item potrivit): $D_i \geq 0,40$;
- 2) discriminare bună (item potrivit): $0,30 \leq D_i < 0,40$;
- 3) discriminare acceptabilă (se recomandă revizuirea itemului): $0,20 \leq D_i < 0,30$;
- 4) discriminare slabă (itemul se supune examinării prevăzute la pct. 49): $0 \leq D_i < 0,20$;
- 5) discriminare inversă (itemul este problematic și se supune cu prioritate examinării prevăzute la pct. 49): $D_i < 0$.

Valorile pozitive înalte semnifică că itemul diferențiază bine. Valorile apropiate de zero semnifică că itemul este fie foarte ușor, fie foarte dificil pentru toți elevii și nu diferențiază între elevi. Valori negative semnifică că itemul este problematic deoarece elevii din grupul cu scoruri joase obțin rezultate mai bune decât elevii din grupul cu scoruri înalte.

38. Coeficientul Alpha Cronbach (α) reprezintă în ce măsură itemii dintr-un test corelează între ei. Valorile se interpretează după cum urmează:

- 1) consistență excelentă (cu atenție la redundanță) — $\alpha \geq 0,90$;
- 2) consistență foarte bună (testul este consistent, iar itemii sunt bine corelați): $0,80 \leq \alpha < 0,90$;
- 3) consistență acceptabilă: $0,70 \leq \alpha < 0,80$;
- 4) consistență discutabilă (testul necesită revizuire): $0,60 \leq \alpha < 0,70$;
- 5) consistență foarte slabă (testul necesită revizuire majoră): $\alpha < 0,60$.

Coeficientul Alpha Cronbach se interpretează cu următoarele precauții: valoarea sa crește odată cu numărul de itemi și poate fi ridicată în prezența unor itemi redundanți. Un coeficient mai scăzut poate apărea în mod legitim la teste care acoperă mai multe componente. Interpretarea se face în raport cu scopul testului, numărul

și structura itemilor. Pentru testele cu structură multidimensională sau mixtă se utilizează, în măsura capacității tehnice disponibile, indicatori alternativi (coeficientul Omega, Alpha stratificat, teoria generalizabilității) sau modele bazate pe Teoria Răspunsului la Item.

Secțiunea 2. Indicatori statistici suplimentari în analiza calității testelor

39. Corelația item-rest (corelația item-test corectată) constituie indicatorul principal, bazat pe corelație, al discriminării itemului, aplicabil atât itemilor dihotomici, cât și celor politomici. Pentru itemii dihotomici, aceasta ia forma corelației punct-biserial corectate. Pentru itemii politomici se utilizează corelația Pearson item-rest. Corelația se calculează în raport cu scorul restului testului (fără itemul analizat), pentru a elimina auto-corelația introdusă de includerea itemului în scorul total.

40. Valorile indicatorului corelației punct-biserial se interpretează după cum urmează:

- 1) itemul discriminează bine și este aliniat cu testul: $r_{pbis} \geq 0,30$;
- 2) discriminare marginală, se recomandă revizuirea itemului: $0,20 \leq r_{pbis} < 0,30$;
- 3) discriminare slabă: $r_{pbis} < 0,20$ (itemul se supune examinării obligatorii prevăzute la pct. 49).

41. Eroarea standard de măsurare (SEM) indică intervalul de eroare în jurul scorului obținut, în cadrul căruia s-ar situa, cu o probabilitate predefinită, scorul real al candidatului. SEM permite construirea intervalelor de încredere în jurul scorului observat. Intervalul de încredere de 95% pentru orice scor se calculează utilizând formula: $\text{scor} \pm 1,96 \times \text{SEM}$.

42. Eroarea standard de măsurare (SEM) este o valoare globală, medie la nivelul întregului test. Precizia de măsurare nu este însă, de regulă, uniformă pe întreaga scală a scorurilor. La examenele naționale, pe lângă SEM globală se estimează și eroarea standard condiționată de măsurare (CSEM), care exprimă precizia la diferite niveluri ale scorului.

43. Analiza distractorilor (pentru itemii cu alegere multiplă) examinează, pentru fiecare item cu alegere multiplă, frecvența de selecție a fiecărei variante de răspuns, inclusiv a variantelor incorecte. Analiza distractorilor permite identificarea: distractorilor nefuncționali (aleși de mai puțin de 5% din candidați), care nu contribuie la măsurare și pot fi înlocuiți; distractorilor „capcană” (aleși preponderent de candidații cu performanțe înalte), care semnalează probleme de formulare a itemului sau ambiguitate în cheia de răspuns; concepțiilor greșite (misconceptions) sistematice — atunci când un distractor specific este selectat masiv, indicând o eroare conceptuală răspândită, care necesită atenție didactică.

44. Analiza item-test examinează modul în care fiecare item se raportează la performanța globală la test, calculându-se coeficientul Alpha Cronbach „dacă itemul este eliminat”. În cazul în care eliminarea unui item conduce la creșterea coeficientului Alpha Cronbach al testului, itemul respectiv reduce consistența internă și necesită revizuire. Această analiză permite identificarea sistematică a itemilor „problematici” care influențează calitatea psihometrică a testului în ansamblu.

45. Analiza item vs. domeniu cognitiv și item vs. competențe (matricea Q) verifică coerența psihometrică intra-grup (itemii care vizează aceeași competență sau același nivel cognitiv). Analiza are la bază matricea de specificații (matricea Q) în care fiecărui item i se asociază explicit nivelul cognitiv vizat, domeniul de conținut și competența/competențele specifice măsurate. Atunci când un item nu corelează cu ceilalți itemi care măsoară aceeași competență, există riscul că itemul nu măsoară ceea ce a fost prevăzut sau că clasificarea sa în matricea Q este eronată. Această analiză contribuie la fundamentarea validității de construct și la verificarea alinierii dintre instrumentul de evaluare și obiectivele curriculare. Valoarea concluziilor depinde de calitatea matricei de specificații, care se validează conform pct. 46.

46. Matricea de specificații (matricea Q) se elaborează și se validează pe baza următoarelor cerințe:

- 1) existența unor reguli explicite de clasificare a itemilor pe niveluri cognitive, domenii de conținut și competențe specifice;

- 2) verificarea reprezentativității itemilor pentru fiecare categorie de analiză, precum și a numărului de itemi care revin fiecărei categorii.

Atunci când un item măsoară mai multe competențe specifice, iar în analiză i se atribuie competența dominantă, această simplificare se menționează explicit în raport, ca limitare a interpretării rezultatelor pe competențe.

47. Analiza speededness (presiunea timpului) evaluează dacă timpul alocat testului este suficient pentru ca toți candidații să ajungă la ultimii itemi. Se examinează procentul de candidați care nu au răspuns la ultimii 5–10% dintre itemi. Un test bine calibrat din punct de vedere al timpului ar trebui să permită cel puțin 95% dintre candidați să răspundă la toți itemii. Atunci când procentul de candidați care nu finalizează testul depășește 5–10%, scorurile la itemii finali nu mai măsoară competența, ci viteza de lucru, fapt ce afectează validitatea concluziilor. Pragul dat aplicat are caracter orientativ, iar orice alt prag aplicat se justifică prin scopul testării, structura și formatul evaluării.

48. Atunci când designul evaluării permite, analiza presiunii timpului se completează cu: timpul de răspuns pe item (la evaluările administrate computerizat), analiza efectului de poziție a itemului și compararea performanței la itemi echivalenți plasați în poziții diferite (la evaluările care utilizează forme cu ordine diferită a itemilor).

49. Niciun indice statistic considerat izolat (dificultatea, discriminarea, corelația item-rest, comportamentul distractorilor) nu constituie temei suficient pentru eliminarea sau reformularea unui item. Decizia asupra unui item se ia prin coroborarea indicatorilor psihometrice cu analiza de conținut și cu poziția itemului în matricea de specificații (matricea Q).

V. Analiza gradului de realizare a curriculumului pe baza performanței la itemii asociați nivelurilor cognitive, domeniilor de conținut și competențelor specifice

50. Analiza gradului de realizare a curriculumului urmărește să stabilească în ce măsură rezultatele obținute de candidați la evaluările naționale standardizate reflectă nivelul de dobândire a competențelor curriculare. Această analiză are rolul de a identifica atât domeniile de conținut și competențele la care candidații demonstrează performanțe satisfăcătoare, cât și ariile problematice care necesită intervenții la nivel de politici educaționale, practici didactice sau dezvoltare curriculară.

51. Gradul de performanță reprezintă procentul mediu din punctajul maxim obținut de candidați la setul de itemi atribuit unei categorii de analiză (nivel cognitiv, domeniu de conținut sau competență specifică). El constituie un indicator descriptiv al performanței la itemii respectivi și, luat izolat, nu demonstrează realizarea integrală a curriculumului, stăpânirea unei competențe sau eficiența predării. Interpretarea gradului de performanță ca grad de realizare a curriculumului sau ca nivel de stăpânire a unei competențe se face numai cu respectarea condițiilor prevăzute la pct. 52 și cu atenția cerută de factorii care influențează valoarea acestui indicator, între care: dificultatea și numărul itemilor atribuiți categoriei, punctajul acordat fiecărui item, reprezentativitatea itemilor pentru conținutul vizat, precizia de măsurare a subscalei și timpul disponibil pentru rezolvare.

52. Un subscor calculat pe un domeniu de conținut, un nivel cognitiv sau o competență specifică poate fi interpretat ca indicator al categoriei respective numai dacă sunt îndeplinite cumulativ următoarele condiții:

- 1) categoria este definită clar și fără ambiguitate în matricea de specificații;
- 2) conținutul categoriei este reprezentat suficient prin itemii testului;
- 3) itemii acoperă mai mult decât o singură sarcină sau o manifestare îngustă a categoriei;
- 4) rezultatul nu este determinat preponderent de unul sau doi itemi cu punctaj ridicat în raport cu punctajul maxim al categoriei;
- 5) rezultatul nu este influențat semnificativ de presiunea timpului;
- 6) itemii categoriei funcționează comparabil între subgrupurile de candidați.

53. Formulările care afirmă realizarea curriculumului, stăpânirea unei competențe sau eficiența predării se utilizează în rapoartele de analiză numai atunci când gradul de performanță este coroborat cu:

- 1) un standard de performanță definit în prealabil pentru categoria analizată;
- 2) dovezi privind reprezentativitatea itemilor pentru conținutul și competențele vizate;
- 3) o precizie de măsurare suficientă a subscalei corespunzătoare;
- 4) coroborarea cu alte surse de date relevante.

54. Analiza performanței la itemii asociați nivelurilor cognitive oferă o imagine a performanței candidaților, identificând punctele forte și slabe pe diverse niveluri de complexitate intelectuală. În prezenta Metodologie se disting nivelurile cognitive conform taxonomiei revizuite a lui Bloom, care clasifică procesele cognitive pe următoarele niveluri:

- 1) Reținerea (Remembering);
- 2) Înțelegerea (Understanding);
- 3) Aplicarea (Applying);
- 4) Analiza (Analyzing);
- 5) Evaluarea (Evaluating);
- 6) Crearea (Creating).

55. Analiza performanței la itemii asociați domeniilor de conținut urmărește să determine în ce măsură candidații au însușit competențele și cunoștințele prevăzute în arii tematice distincte ale disciplinei. Această analiză permite identificarea domeniilor unde curriculumul este realizat corespunzător și a celor unde apar lacune semnificative. Valorile procentuale se interpretează astfel:

- 1) realizare înaltă (cel puțin 70%) — curriculumul este însușit corespunzător pe domeniul respectiv;
- 2) realizare medie (50–69%) — curriculumul este parțial realizat. Se recomandă consolidarea predării și revizuirea materialelor de instruire;
- 3) realizare scăzută (sub 50%) — curriculumul nu este realizat satisfăcător și se necesită măsuri sistemice (formare suplimentară, revizuire curriculară, sprijin ținut).

56. Aceste praguri sunt convenționale și servesc la organizarea și compararea consecventă a rezultatelor între categorii și sesiuni. Pragurile se stabilesc separat pentru fiecare evaluare și disciplină, pe baza descriptorilor de competență din curriculum și printr-o procedură documentată de stabilire a standardelor de performanță.

57. Analiza performanței la itemii asociați competențelor specifice reprezintă un pas esențial în evaluarea aprofundată a curriculumului. Spre deosebire de analiza pe domenii de conținut (care se axează pe tematici), această abordare se concentrează pe abilitățile și capacitățile practice pe care candidații ar trebui să le dezvolte pe parcursul unui ciclu de studiu.

58. Analiza tendințelor multianuale urmărește evoluția rezultatelor în timp, comparativ între sesiuni consecutive ale aceleiași evaluări, pe aceleași niveluri cognitive, domenii de conținut și competențe specifice. Această analiză permite distincția între fluctuațiile aleatorii și tendințele sistemice (creștere, descreștere, stagnare) și oferă o bază longitudinală pentru evaluarea impactului intervențiilor curriculare, schimbărilor de manuale sau programelor de formare a cadrelor didactice.

59. Pentru analiza tendințelor multianuale este necesară existența unei proporții suficiente de itemi comuni securizați între sesiuni, aplicarea unei proceduri de echivalare între formele evaluate, utilizarea unei scale comune construite prin Teoria Răspunsului la Item, demonstrarea stabilității constructului și a itemilor de legătură. În acest regim se pot formula concluzii privind evoluția nivelului de abilitate al candidaților. În absența unor proceduri de echivalare, se raportează evoluția scorurilor brute, a ratelor și a gradului de performanță, cu mențiunea explicită că diferențele constatate nu pot fi atribuite univoc evoluției nivelului de competență al candidaților.

VI. Analize avansate opționale

60. Prezentul capitol descrie un set de analize psihometrice avansate, recomandate pentru aplicare graduală, în măsura în care se dispune de capacitățile tehnice și de software-ul specializat necesar. Aceste analize completează metodologia de bază cu instrumente derivate, în principal, din Teoria Răspunsului la Item (IRT), care oferă o măsură mai precisă a calității itemilor și a echității testului.

61. Analiza funcționării diferențiate a itemilor (DIF) investighează dacă un item funcționează diferit pentru subgrupuri de candidați cu scoruri totale egale la test (control prin variabila scor total). Analiza funcționării diferențiate identifică itemii pentru care, la același nivel de abilitate, candidații din subgrupuri diferite au probabilități diferite de a răspunde corect — ceea ce semnalează o posibilă biasare a itemului (formulări ambigue, contexte culturale specifice unui subgrup, vocabular nefamiliar). Itemii semnalati se supun unei examinări de conținut de către specialiști în disciplină și în evaluare, care stabilesc dacă diferența de funcționare decurge dintr-un factor irelevant pentru competența vizată sau reflectă o diferență relevantă pentru construct.

62. Funcția de informație a testului (TIF) indică precizia cu care testul măsoară abilitatea pe diferite niveluri ale scalei. Spre deosebire de SEM, care este o valoare medie pe tot testul, TIF arată pe care segment al scalei (candidați cu abilitate scăzută, medie sau ridicată) testul măsoară cel mai exact.

63. Modelele liniare ierarhice (multinivel) estimează proporția variației rezultatelor candidaților care se manifestă între instituțiile de învățământ, respectiv în interiorul acestora. Această proporție descrie structura variației observate și nu reprezintă, luată separat, efectul causal al instituției asupra rezultatelor. Ea include și efectele compoziției populației de candidați ai fiecărei instituții, precum și ale contextului socioeconomic în care aceasta funcționează.

64. Descompunerea varianței între școli vs. în interiorul școlilor utilizează modele liniare ierarhice (multinivel) pentru a determina ce procent din variația rezultatelor candidaților este atribuibil școlii pe care o frecventează, respectiv diferențelor individuale între candidați din aceeași școală.

VII. Structura raportului de analiză

65. Raportul de analiză a rezultatelor evaluărilor naționale standardizate prezintă, în formă structurată, rezultatele aplicării prezentei Metodologii pentru fiecare sesiune și disciplină. Raportul are o structură unitară, organizată pe capitole, care reflectă analiza descriptivă a rezultatelor, calitatea instrumentului de evaluare și gradul de realizare a curriculumului, urmate de o secțiune integrată de concluzii și recomandări.

66. Capitolul I al raportului prezintă rezultatele analizei descriptive a performanței candidaților la evaluările naționale standardizate. Pentru fiecare indicator descris în Capitolul III al prezentei Metodologii, se prezintă:

- 1) datele sistematizate în formă tabelară și grafică;
- 2) interpretarea datelor;
- 3) constatările principale;
- 4) comparațiile cu sesiunile anterioare (după caz).

67. Capitolul I al raportului include următoarele secțiuni:

- 1) indicatorii de tendință centrală (media, mediana) la nivel național și dezagregat;
- 2) indicatorii de distribuție (deviația standard, percentilele P10, P25, P50, P75, P90, intervalul intercuartilic);
- 3) indicatorii de formă a distribuției (asimetria, kurtosis);
- 4) distribuția notelor pe scala oficială;
- 5) ratele diferențiate;
- 6) dezagregarea rezultatelor pe variabilele de interes;
- 7) analiza diferențelor statistice între subgrupuri.

68. Capitolul II al raportului prezintă rezultatele analizei calității psihometrice a testului. Pentru fiecare indicator descris în Capitolul IV al prezentei Metodologii, se prezintă:

- 1) datele sistematizate în formă tabelară sau grafică;
- 2) interpretarea datelor;
- 3) constatări;
- 4) recomandări operaționale.

69. Capitolul II al raportului include următoarele secțiuni:

- 1) evaluarea consistenței interne a testului;
- 2) analiza indicelui de dificultate al itemilor;
- 3) analiza indicelui de discriminare al itemilor;
- 4) analiza corelației punct-biserială (după caz);
- 5) analiza distractorilor (pentru itemii cu alegere multiplă);
- 6) analiza item vs. test, item vs. domeniu cognitiv și item vs. competențe;
- 7) analiza speededness;
- 8) rezultatele analizelor avansate descrise în Capitolul VI al Metodologiei (după caz).

70. Capitolul III al raportului prezintă rezultatele analizei gradului de realizare a curriculumului, structurat pe niveluri cognitive, domenii de conținut și competențe specifice. Pentru fiecare unitate de analiză se prezintă:

- 1) datele sistematizate;
- 2) interpretarea datelor;
- 3) ipoteze explicative privind performanța constatată;
- 4) recomandări pentru elaboratorii de curriculum și pentru cadrele didactice.

71. Capitolul III al raportului include următoarele secțiuni:

- 1) analiza gradului de realizare a itemilor asociați nivelurilor cognitive;
- 2) analiza gradului de realizare a itemilor asociați domeniilor de conținut;
- 3) analiza gradului de realizare a itemilor asociați competențelor specifice;
- 4) analiza echității rezultatelor între subgrupuri;
- 5) analiza tendințelor multianuale (în rapoartele sintetice pe 3–5 ani).

72. În secțiunea finală a raportului se prezintă constatările cheie ale analizei psihometrice și curriculare, însoțite de recomandări clare pentru:

- 1) grupurile de lucru pentru elaborarea testelor;
- 2) elaboratorii de curriculum;
- 3) factorii de decizie;
- 4) instituțiile de formare a cadrelor didactice.

Anexa 1. Formule și proceduri de calcul

A.1) Indicele de dificultate al itemului

$$p_i = \frac{m_i}{S_{max,i}},$$

unde: m_i = scorul mediu obținut de candidați la itemul i ; $S_{max,i}$ = punctajul maxim posibil pentru itemul i .

A.2) Indicele de discriminare al itemului

$$D_i = \frac{m_{sup,i} - m_{inf,i}}{S_{max,i}},$$

unde: $m_{sup,i}$ = scorul mediu la itemul i pentru grupul superior (27% din numărul de candidați cu scorurile totale cele mai mari); $m_{inf,i}$ = scorul mediu la itemul i pentru grupul inferior (27% din numărul de candidați cu scorurile totale cele mai mici); $S_{max,i}$ = punctajul maxim posibil pentru itemul i .

A.3) Coeficientul de consistență internă Alpha Cronbach

$$\alpha = \frac{k}{k-1} \cdot \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right),$$

unde: k = numărul de itemi al testului; σ_i^2 = variația scorurilor la itemul i ; σ_x^2 = variația scorului total la test.

A.4) Corelația item-test (Pearson item-total)

Corelația item-test este indicatorul fundamental de discriminare a itemului, aplicabil atât itemilor dihotomici, cât și celor politomici. Forma generală este coeficientul de corelație Pearson între scorul candidatului la item și scorul total la test:

$$r_{it} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \cdot \sum_{i=1}^N (Y_i - \bar{Y})^2}},$$

unde: X_i = scorul candidatului i la itemului analizat; $\bar{X}$ = scorul mediu la item în populația de candidați; Y_i = scorul candidatului i la test; $\bar{Y}$ = scorul total mediu la test; N = numărul total de candidați. Pragurile de interpretare: $r_{it} \geq 0,30$ – discriminare bună; $0,20 \leq r_{it} < 0,30$ – discriminare marginală; se recomandă revizuirea; $r_{it} < 0,20$ – discriminare slabă; itemul se reformulează sau se exclude; valori negative – item defectuos.

A.5) Corelația punct-biserial (caz particular pentru itemi dihotomici)

Atunci când itemul este dihotomic, formula generală Pearson item-test se reduce la corelația punct-biserială:

$$r_{pbis} = \frac{m_1 - m_0}{\sigma_x} \cdot \sqrt{p \cdot q},$$

unde: m_1 = scorul mediu la test al candidaților care au răspuns corect la item; m_0 = scorul mediu la test al candidaților care au răspuns greșit la item; σ_x = devierea standard a scorurilor total la test; p = proporția de candidați care au răspuns corect; $q = 1 - p$.

A.6) Corelația item-rest (corelația item-test corectată)

Corelația item-rest (corelația item-test corectată) elimină auto-corelația introdusă de faptul că scorul total Y_i include și scorul X_i al itemului analizat (aceasta fiind o problemă semnificativă pentru testele cu până la 20 de itemi sau pentru itemi cu punctaj mare în raport cu testul total).

$$r_{i(t-1)} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i^{rest} - \bar{Y}^{rest})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \cdot \sum_{i=1}^N (Y_i^{rest} - \bar{Y}^{rest})^2}},$$

unde: $Y_i^{rest} = Y_i - X_i$, adică scorul total al candidatului i din care se scade scorul la itemul analizat.

A.7) Eroarea standard de măsurare

$$SEM = \sigma_x \cdot \sqrt{1 - \alpha},$$

unde: σ_x = abaterea standard a scorurilor totale la test; α = coeficientul Alpha Cronbach.

A.8) Indicele de selecție a distractorilor

$$p_d = \frac{N_d}{N},$$

unde: N_d = numărul de elevi care au selectat distractorul d ; N = numărul total de candidați care au răspuns la item.

A.9) Diferența între subgrupuri

$$\Delta_g = Realizare_{grup A} - Realizare_{grup B},$$

unde: $Realizare_{grup A}$ = gradul de realizare în grupul A ; $Realizare_{grup B}$ = gradul de realizare în grupul B . Diferența se calculează pe perechi de subgrupuri la nivel general și pe fiecare nivel cognitiv, domeniu de conținut și competență specifică. Praguri de interpretare: $|\Delta_g| < 5$ puncte procentuale – nesemnificativ; 5 – 10 puncte procentuale – moderat; $|\Delta_g| > 10$ puncte procentuale – semnificativ.

A.10) Procedura de analiză a itemilor asociați nivelurilor cognitive

- 1) clasificarea itemilor în funcție de nivelul cognitiv vizat;
- 2) calculul punctajului total per nivel: pentru fiecare candidat, se însumează punctele obținute la toți itemii care aparțin unui anumit nivel cognitiv;
- 3) calculul punctajului maxim per nivel: se însumează punctajele maxime posibile pentru toți itemii dintr-un nivel cognitiv;
- 4) calculul gradului de realizare per nivel: se calculează media gradului de realizare pentru fiecare nivel cognitiv folosind formula

$$GR_{nivel} = \frac{m_{nivel}}{S_{max,nivel}} \cdot 100\%,$$

unde: m_{nivel} = scorul mediu la itemii de nivelul cognitiv analizat; $S_{max,nivel}$ = punctajul maxim posibil pentru nivelul cognitiv analizat.

Valorile procentuale se interpretează conform următorului tabel:

Nivel cognitiv	≥ 70% – Ridicat (performanță satisfăcătoare spre excelentă)	50–69% – Mediu (performanță parțial satisfăcătoare)	< 50% – Scăzut (performanță nesatisfăcătoare)
Reținerea (Remembering)	Elevii regăsesc și recunosc cu acuratețe fapte, termeni, definiții și proceduri din memoria de lungă durată. Rezultatul este compatibil cu o bază factuală solid însușită.	Elevii rețin parțial informațiile factuale, cu omisiuni și confuzii la termeni sau definiții. Necesită exersare sistematică a memorării structurate.	Rezultatul indică lipsuri majore la fapte, termeni și definiții esențiale.
Înțelegerea (Understanding)	Elevii construiesc sensul materialului: interpretează, exemplifică, clasifică, compară și explică cu proprietate. Conceptele sunt înțelese, nu doar memorate.	Elevii înțeleg parțial materialul, explică și exemplifică cu dificultate, iar comparațiile și clasificările conțin erori. Necesită clarificarea conceptelor-cheie.	Rezultatul indică dificultăți majore în construirea sensului conceptelor și în explicare sau exemplificare.
Aplicarea (Applying)	Elevii utilizează corect procedurile, algoritmi și regulile în situații familiare și noi. Curriculumul este realizat la nivel procedural și operațional.	Elevii aplică parțial procedurile, cu erori frecvente în situații noi. Deprinderile de lucru sunt fragile. Necesită exersare pe sarcini variate.	Elevii întâmpină dificultăți majore în aplicarea regulilor și algoritmilor. Curriculumul nu este transpus în competențe funcționale.
Analiza (Analyzing)	Elevii descompun materialul în componente, diferențiază esențialul de accesoriu și identifică relațiile structurale.	Elevii realizează parțial descompunerea și organizarea informației, dar întâmpină dificultăți la relațiile complexe.	Rezultatul indică dificultăți în descompunerea problemelor complexe și în identificarea relațiilor dintre componente.
Evaluarea (Evaluating)	Elevii formulează judecăți întemeiate pe criterii și standarde; verifică, critică și argumentează coerent.	Elevii formulează parțial judecăți, dar criteriile sunt neclare sau aplicate inconsecvent. Argumentarea este superficială	Rezultatul indică dificultăți în formularea de judecăți pe baza unor criterii și în argumentare.
Crearea (Creating)	Elevii reunesc elementele într-un întreg coerent și original: generează, planifică și produc soluții proprii	Elevii produc parțial soluții originale, dar cu structură incompletă sau lipsă de coerență.	Rezultatul indică dificultăți în generarea de soluții originale. Se interpretează cu prudență, dat fiind numărul redus de itemi care vizează, de regulă, acest nivel.

A.11) Procedura de analiză a itemilor asociați domeniilor de conținut

- 1) clasificarea itemilor în funcție de domeniul de conținut atribuit;
- 2) calculul punctajului total per domeniu: pentru fiecare candidat, se însumează punctele obținute la toți itemii care aparțin unui domeniu de conținut;
- 3) calculul punctajului maxim per domeniu: se însumează punctajele maxime posibile pentru toți itemii dintr-un domeniu de conținut;
- 4) calculul gradului de realizare per domeniu: se calculează media gradului de realizare pentru fiecare domeniu de conținut folosind formula

$$GR_{domeniu} = \frac{m_{domeniu}}{S_{max,domeniu}} \cdot 100\% ,$$

unde: $m_{domeniu}$ = scorul mediu la itemii din domeniul analizat; $S_{max,domeniu}$ = punctajul maxim posibil pentru domeniul analizat.

A.12) Procedura de analiză a itemilor asociați competențelor specifice

- 1) Clasificarea itemilor: se grupează itemii în funcție de competența specifică pe care o evaluează. Un item poate contribui la evaluarea mai multor competențe; pentru simplitate, se poate atribui competența dominantă;
- 2) Calculul punctajului total per competență: pentru fiecare candidat, se însumează punctele obținute la toți itemii care evaluează o anumită competență;
- 3) Calculul punctajului maxim per competență: se însumează punctajele maxime posibile pentru toți itemii care evaluează o anumită competență;
- 4) Calculul gradului de realizare per competență

$$GR_{comp} = \frac{m_{comp}}{S_{max,comp}} \cdot 100\% ,$$

unde: m_{comp} = scorul mediu la itemii din competența specifică analizată; $S_{max,comp}$ = punctajul maxim posibil pentru competența specifică analizată.

Valorile procentuale se interpretează conform următorului tabel:

Procent de realizare	Interpretare	Implicații pentru procesul educațional
$\geq 70\%$ (realizare ridicată)	Performanța la itemii asociați competenței este ridicată, rezultatul este compatibil cu o stăpânire solidă a cunoștințelor și deprinderilor vizate.	Menținerea și rafinarea strategiilor existente. Confirmarea concluziei necesită coroborare cu alte surse de date.
50%–69% (realizare medie)	Performanța este intermediară; rezultatul sugerează dificultăți în aplicare și transfer.	Consolidare prin exersare suplimentară și predare diferențiată; ajustarea materialelor și activităților de instruire.

Procent de realizare	Interpretare	Implicații pentru procesul educațional
< 50% (realizare scăzută)	Performanța la itemii asociați competenței este scăzută; rezultatul semnalează posibile lacune în formarea deprinderilor și aplicarea cunoștințelor.	Examinarea prealabilă a calității itemilor și a acoperirii curriculare; dacă rezultatul se confirmă, măsuri de sprijin și analiza concordanței dintre test și curriculum.

PROIECT

Anexa 2. Protocol de validare a datelor

1. Validarea datelor se realizează secvențial, în etape. Fiecare etapă se încheie printr-o listă de control completată. Trecerea la etapa următoare nu se realizează înainte de încheierea etapei precedente, iar analiza psihometrică și interpretativă nu se inițiază înainte de finalizarea tuturor etapelor.

2. Etapele procesului de validare sunt:

- 1) recepția și inventarierea datelor;
- 2) aplicarea regulilor automate de validare;
- 3) efectuarea verificărilor manuale;
- 4) codificarea și tratarea datelor lipsă și a cazurilor speciale;
- 5) validarea bazei de date.

3. Baza de analiză cuprinde, ca structură minimală, următoarele componente:

Componenta	Conținut minim
Fișierul candidaților	Identificator pseudonimizat, statut de prezență, instituția (cod), raionul, mediul de reședință, genul, limba de instruire, statutul socioeconomic (dacă este disponibil), statutul de candidat cu cerințe educaționale speciale și adaptările acordate, forma/varianta de test administrată.
Fișierul răspunsurilor	Identificator pseudonimizat al candidatului, identificator al itemului, punctajul acordat, varianta de răspuns selectată (pentru itemii cu alegere multiplă), codul de date lipsă (dacă este cazul)
Fișierul itemilor (matricea Q)	identificator unic al itemului, tipul itemului, punctajul maxim, nivelul cognitiv, domeniul de conținut, competențele specifice măsurate, poziția itemului în forma administrată, statutul itemului (valid/anulat)
Fișierul de administrare	Durata efectivă, incidentele și abaterile procedurale, condițiile speciale de administrare

4. Fiecare bază de analiză este însoțită de un dicționar al variabilelor, care cuprinde: denumirea variabilei, tipul de date, valorile admisibile, semnificația codurilor și modul de tratare a valorilor lipsă.

5. Înaintea oricărei analize se aplică următorul set minim de verificări automate:

Nr.	Regula de validare	Acțiune la neconformitate
1.	Identificatorul candidatului este unic în fișierul candidaților.	Identificarea și eliminarea duplicatelor; documentarea cazurilor.
2.	Fiecare înregistrare de răspuns corespunde unui candidat existent în fișierul candidaților.	Clarificarea cu furnizorul de date.
3.	Fiecare item din fișierul răspunsurilor există în fișierul itemilor.	Clarificarea cu furnizorul de date.
4.	Punctajul acordat la fiecare item se încadrează în intervalul [0; punctaj maxim al itemului]	Corectare pe baza documentelor primare sau marcarea ca dată lipsă.
5.	Punctajul acordat corespunde valorilor admisibile prevăzute de barem pentru itemul respectiv.	Verificarea baremului și a introducerii datelor.
6.	Scorul total recalculat prin însumarea punctajelor pe itemi coincide cu scorul total din fișier.	Identificarea sursei diferenței; recalcularea scorului total.
7.	Numărul de itemi din fișierul răspunsurilor corespunde structurii formei administrate.	Verificarea completitudinii transferului de date.
8.	Codurile variabilelor demografice se regăsesc în nomenclatoarele stabilite.	Recodificare sau marcarea ca valoare lipsă.

9.	Fiecărui candidat îi este atribuită o formă/variantă de test validă.	Clarificarea cu furnizorul de date.
10.	Statutul de prezență este consistent cu existența răspunsurilor (candidații absenți nu au punctaje; candidații prezenți au cel puțin un răspuns înregistrat).	Verificarea listelor de prezență.
11.	Suma punctajelor maxime ale itemilor corespunde punctajului maxim al testului.	Verificarea matricei de specificații și a baremului.
12.	Fiecare item are atribuite, în matricea Q, nivelul cognitiv, domeniul de conținut și competența specifică	Completarea matricei de specificații.

6. Rezultatele verificărilor automate se consemnează într-un raport de validare care indică, pentru fiecare regulă, numărul de neconformități identificate și modul de soluționare.

7. După aplicarea regulilor automate se realizează următoarele verificări manuale:

- 1) verificarea cheii de răspuns pentru fiecare item cu alegere multiplă, cu prioritate pentru itemii care prezintă discriminare negativă sau indice de dificultate atipic. Eroarea de cheie constituie cea mai frecventă cauză a discriminării negative și se corectează înaintea oricărei decizii privind itemul;
- 2) verificarea itemilor anulați: se consemnează itemii anulați, motivul anulării și modul de tratare a acestora în calculul scorului total și al indicatorilor psihometrici. Tratatamentul aplicat se menține consecvent pentru toți candidații;
- 3) verificarea cazurilor de absență parțială: candidații care au participat la o parte a probelor sau au întrerupt proba se identifică și se marchează distinct;
- 4) verificarea candidaților excluși: se consemnează candidații ale căror rezultate au fost anulate sau excluse, motivul și baza legală a excluderii, precum și tratamentul lor în calculul ratelor;
- 5) verificarea adaptărilor acordate candidaților cu cerințe educaționale speciale, inclusiv a consistenței dintre adaptarea acordată și condițiile efective de administrare.

8. Datele lipsă se codifică distinct, pe categorii, întrucât semnificația lor pentru analiză este diferită:

Cod	Categorie	Semnificație
OM	Omisie internă	Item lăsat fără răspuns, deși candidatul a răspuns la itemi situați ulterior în test.
NA	Item neatins	Item de la finalul testului la care candidatul nu a ajuns (împreună cu toți itemii următori).
AB	Absență	Candidat înscris care nu s-a prezentat la probă.
AN	Item anulat	Item eliminat din evaluare prin decizie documentată.

9. Modul de tratare a fiecărei categorii în calculul indicatorilor se stabilește explicit, se documentează și se menține consecvent între sesiuni, potrivit următoarelor reguli:

- 1) omisiunile interne se tratează ca punctaj zero la calculul scorului total și al indicatorilor de performanță, întrucât candidatul a avut posibilitatea de a răspunde;
- 2) itemii neatinși se tratează ca punctaj zero la calculul scorului total, dar se exclud din calculul indicelui de dificultate și al indicilor de discriminare ai itemului, întrucât includerea lor denaturează estimarea dificultății itemilor finali. Proportia itemilor neatinși se raportează separat;
- 3) candidații absenți se exclud din calculul indicatorilor de performanță și se raportează exclusiv în cadrul ratei de neprezentare;
- 4) itemii anulați se exclud din calculul tuturor indicatorilor psihometrici și din punctajul maxim al testului, al domeniilor și al nivelurilor cognitive afectate.

10. Baza de date rezultată în urma etapelor descrise mai sus se supune validării finale. Momentul validării constituie etapa finală în întregul proces de validare. După acest moment, baza de date nu se modifică. Orice corecție necesară ulterior se realizează prin crearea unei versiuni noi.

PROIECT