

DOI /10.5281/zenodo.18933413

TEST NATIJALARINING INFIT VA OUTFIT STATISTIK TADQIQOTI

Q.A. Amonov¹, Y.Z. Karimov²

¹O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi Bilim va malakalarni baholash agentligi huzuridagi Ilmiy – o'quv amaliy markazi, 100084, Toshkent sh., Bog'ishamol k., 12

²O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi Bilim va malakalarni baholash agentligi, 100084, Toshkent sh., Bog'ishamol k., 12

Qisqacha mazmuni. Ushbu maqolada tarix fanidan o'tkazilgan test sinovlari natijalari asosida test savollarining javob namunalari sinaluvchilar qobiliyat darajasiga va aksincha qobiliyat darajasi javob namunasiga qanchalik sezgir ekanligi, shuningdek, test topshiriqlarining qiyinlik darajalari va sinaluvchilarning qobiliyat darajalari o'rtasidagi mutanosiblik o'rganildi. Tadqiqot natijalari ayrim test topshiriqlariga sinaluvchilar tomonidan berilgan javoblar mutanosib emasligi va topshiriqlariga berilgan javoblar ko'proq tasodifiy va ehtiyotsizlik tufayli bo'lganligini ko'rsatdi.

Kalit so'zlar: tarix, infit va outfit statistikasi, Rash modeli, dispersiya, qiyinlik darajasi, qobiliyat darajasi.

1. Kirish

Ta'lim tizimida amalga oshirilayotgan islohotlar, eng avvalo, zamonaviy bilim va kompetensiyalarni egallagan, fan hamda texnika taraqqiyotining hozirgi bosqichiga mos yuqori malakali kadrlarni tayyorlash zaruratini belgilab bermoqda. Mazkur jarayonda ta'lim oluvchining individual xususiyatlarini inobatga oluvchi, shaxsga yo'naltirilgan yondashuv asosida kompetensiyaga asoslangan bilimlarni shakllantirish ustuvor ahamiyat kasb etadi. Shu bilan birga, ta'lim oluvchilarda shakllangan bilim, ko'nikma va kompetensiyalarni baholashning ilmiy asoslangan, ratsional, obyektiv hamda shaffof mexanizmlarini ishlab chiqish va ularni

ta'lim jarayoniga joriy etish ta'lim tizimining dolzarb vazifalaridan biri hisoblanadi. Ushbu nuqtai nazardan, bilimlarni baholash jarayoni faqatgina ta'lim oluvchilarning o'zlashtirish darajasini aniqlash bilan cheklanmay, balki ularning intellektual salohiyati, kognitiv kompetensiyalari hamda mustaqil va tanqidiy fikrlash qobiliyatini kompleks baholashning muhim pedagogik vositasi sifatida e'tirof etiladi [1-4].

An'anaviy baholash shakllari ko'pincha faqat bilimlarning reproduktiv jihatini, ya'ni yodlangan ma'lumotni qayta ifodalash qobiliyatini aniqlashga qaratilgan bo'ladi. Biroq zamonaviy ta'lim jarayoni shaxsning

ijodkorligini, muammoli vaziyatlarda to'g'ri qaror qabul qila olish qobiliyatini ham baholashni taqozo etadi [5]. Shunday ekan, ta'limda qo'llanilayotgan baholash mexanizmlarini global tajriba asosida yangilash, ularning obyektivligi, shaffofligi va haqqoniyligini ta'minlash zamonaviy islohotlarning eng dolzarb yo'nalishlaridan biriga aylanmoqda. Shuningdek, baholash jarayonining natijalari butun ta'lim tizimining rivojlanish yo'nalishlariga ham bevosita ta'sir ko'rsatadi. Chunki o'lchov natijalariga asoslangan tahlillar ta'lim dasturlarini takomillashtirish, o'quv-uslubiy materiallarni boyitish va o'qitish metodikasini yangilash imkonini beradi [6-8].

Test sinovlarini samarali tashkil etishda sinaluvchilarning turli xil qobiliyat darajalarini inobatga olish muhim ahamiyat kasb etadi. Shu bois test topshiriqlari bir xil darajada emas, balki turli qiyinlik darajalarida tuzilishi maqsadga muvofiqdir. Bunday yondashuv o'quvchilarning bilim va ko'nikmalarini yanada aniqroq o'lchash, ularning individual tayyorgarlik darajasini to'liqroq baholash imkonini beradi [9].

Test topshiriqlari qiyinlik darajasining tabaqalashtirilishi sinaluvchilarning imkoniyatlarini to'liq ochib berishga, baholash natijalarining xolis va ishonchli bo'lishiga xizmat qiladi. Ayniqsa, yuqori qobiliyatli o'quvchilar uchun murakkabroq savollar, pastroq tayyorgarlikka ega bo'lganlar uchun esa nisbatan sodda savollarni qo'llash

baholash tizimining samaradorligini oshiradi [10].

Ushbu maqolada tarix fanidan o'tkazilgan test sinovi natijalarining Rash modeli asosidagi tahlillari yoritilgan bo'lib test topshiriqlarining ichki (infit) va tashqi (outfit) moslik statistikasi o'rganilgan. Tadqiqotda qo'llanilgan test topshiriqlari to'plami jami 45 ta topshiriqdan iborat bo'lib, ulardan 10 tasi (36-45-topshiriqlar) ochiq turdagi test topshiriqlari bo'lib, har biri A va B qismlarga ajratilgan. Shu sababli, umumiy topshiriqlar soni 55 taga yetdi. Testni bajarish uchun ajratilgan vaqt javoblar varaqasini to'ldirish jarayoni bilan birgalikda 80 daqiqani tashkil etdi. Test sinovida jami 3483 nafar sinaluvchi ishtirok etdi.

Test topshiriqlari mazmun sohasi va qamrovi tarix fani bo'yicha Davlat ta'lim standarti asosida ishlab chiqilib, quyidagi yo'nalishlarni o'z ichiga oldi:

Qadimgi dunyo tarixi – insoniyatning ilk davrlari, qadimgi davlatlar va ularning sivilizatsiyaga qo'shgan hissasi;

O'rta asrlar tarixi – Sharq va G'arb davlatlarining siyosiy, ijtimoiy va madaniy taraqqiyoti;

Yangi davr tarixi – Yevropa va Osiyodagi siyosiy jarayonlar, islohotlar va inqiloblar;

Eng yangi davr tarixi – XIX–XX asrlarning ijtimoiy-siyosiy voqealari, urushlar va xalqaro munosabatlar;

O'zbekiston tarixi - qadimgi davrdan mustaqillikkacha bo'lgan davrlar, milliy davlatchilik taraqqiyoti, mustaqillik yillarida amalga oshirilgan islohotlar. Mazkur topshiriqlar o'quvchilarning tarixiy bilimlarini baholash

bilan birga, ularning tahlil qilish, solishtirish va tarixiy voqealar orasidagi sabab-oqibat bog'liqliklarini aniqlash qobiliyatlarini ham sinovdan o'tkazishga yo'naltirilgan.

2. Ichki (infit) va tashqi (outfit) moslik statistikasi

Test topshiriqlarining ichki (infit) moslik statistikasida javoblar namunasining qobiliyat darajasiga va, aksincha, qobiliyat darajalarining javoblar namunasiga qay darajada sezgir ekanligini ko'rsatadi. Ya'ni, u o'quvchi o'z darajasiga mos vazifalarda modeldan qanchalik chetlanishini aniqlaydi. Ichki (infit) moslik haddan tashqari qiyin yoki oson vazifalarga unchalik sezgir emas, chunki uning e'tibori mos darajadagi topshiriqlarga qaratiladi [11]. Test topshiriqlarining ichki mosligi aniqlangan mezonlardan katta bo'lsa, Gutman namunasi [12] bilan bog'liq bo'lgan, mezonlardan kichik bo'lsa, muqobil o'quv dasturi bilan bog'liq bo'lgan ma'lumotlarni beradi.

Tashqi(outfit) moslik statistikasi an'anaviy χ^2 usuliga asoslangan bo'lib, u test topshiriqlarining qiyinlik darajalari va sinaluvchilarning qobiliyat darajalari o'rtasidagi mutanosiblikni ko'rsatadi. Test topshiriqlari uchun esa tashqi moslik uning uchun aniqlangan mezonlardan katta bo'lsa, maqsadga muvofiq bo'lmagan javoblarni, kichik bo'lsa, tasodifiy yoki

ehtiyotsizlik bilan berilgan javoblarni bildiradi. Tashqi va ichki moslik statistikasi quyidagi ifodalar bilan aniqlaniladi [13].

$$\text{outfit}_i = \frac{\sum_s (X_{si} - E(X_{si}))^2}{n_i}, \quad (1)$$

$$\text{infit}_i = \frac{\sum_s (X_{si} - E(X_{si}))^2}{\sum_s \text{Var}(X_{si})} \quad (2)$$

bu yerda $E(X_{si})$ - X_{si} ning matematik kutilishi, $\text{Var}(X_{si})$ - dispersiya, n_i - test topshiriqlari soni, n_s - sinaluvchilar soni.

Ko'plab ilmiy tadqiqotlarda tashqi va ichki moslik ko'rsatkichlarining pastki va yuqori chegaralari mos ravishda 0,7 va 1,7 oraliqda belgilangan. Ayrim tadqiqot natijalariga ko'ra ushbu chegaralarni 0,5 va 1,5 oralig'ida qabul qilish ham maqsadga muvofiqligi ta'kidlangan [14]. Bundan tashqari, ichki va tashqi moslik statistikalarning test topshiriqlari soni hamda ularning qiyinlik darajasi bilan bevosita bog'liq ekanligi aniqlangan [15].

1-jadvaldan test topshiriqlari ichki moslik 0,7-1,3 oralig'idan tashqariga chiqmaganligini, bu esa ushbu test topshiriqlarida Gutman

namunasi va o'quv dastur bilan bog'liq muammolar mavjud emasligini ko'rsatadi [12].

1-jadval

Rash modeli asosida aniqlangan test topshiriqlari qiyinlik darajalari (β) hamda (1)-ifoda yordamida hisoblangan ichki(infit) va tashqi(outfit) moslik statistikalarining qiymatlari berilgan.

| No | ID | β | Infit | Outfit |
|----|-----------|---------|-------|--------|
| 1 | X110216 | -3,56 | 0,95 | 0,74 |
| 2 | X110215 | -3,52 | 0,92 | 0,59 |
| 3 | X110200 | -3,3 | 0,91 | 0,55 |
| 4 | X110182 | -2,9 | 0,91 | 0,65 |
| 5 | X110218 | -2,9 | 0,89 | 0,57 |
| 6 | X110176 | -2,6 | 0,99 | 0,89 |
| 7 | X110198 | -2,5 | 0,92 | 0,74 |
| 8 | X111326 | -2,5 | 0,1 | 0,98 |
| 9 | X110192 | -2,4 | 0,86 | 0,7 |
| 10 | X110202 | -2,38 | 0,86 | 0,6 |
| 11 | X110194 | -2,3 | 0,96 | 0,81 |
| 12 | X110196 | -2,3 | 0,9 | 0,66 |
| 13 | X111656 | -2,05 | 0,8 | 0,56 |
| 14 | X110184 | -1,93 | 1,02 | 1,04 |
| 15 | X110190 | -1,9 | 0,89 | 0,8 |
| 16 | X11165635 | -1,75 | 0,87 | 0,74 |
| 17 | X110210 | -1,58 | 0,93 | 0,85 |
| 18 | X110229 | -1,25 | 1,03 | 1 |
| 19 | X110236 | -1,13 | 0,95 | 0,84 |
| 20 | X110231 | -1,1 | 1,07 | 1,05 |
| 21 | X110259 | -1,08 | 0,95 | 0,87 |
| 22 | X110222 | -1,05 | 1,02 | 0,96 |
| 23 | X1117382 | -1,02 | 0,83 | 0,77 |
| 24 | X1117282 | -1,01 | 0,83 | 0,73 |

| | | | | |
|----|-----------|-------|-------|------|
| 25 | X110230 | -1 | 0,9 | 0,82 |
| 26 | X1102522 | -0,97 | 0,85 | 0,79 |
| 27 | X110238 | -0,95 | 1,01 | 0,94 |
| 28 | X110186 | -0,92 | 1,12 | 1,11 |
| 29 | X110208 | -0,85 | 0,88 | 0,82 |
| 30 | X110206 | -0,8 | 1,06 | 1,08 |
| 31 | X111738 | -0,8 | 0,92 | 0,9 |
| 32 | X110178 | -0,8 | 0,87 | 0,8 |
| 33 | X110188 | -0,74 | 1,05 | 1,01 |
| 34 | X110180 | -0,74 | 0,84 | 0,77 |
| 35 | X11165634 | -0,6 | 1,08 | 1,09 |
| 36 | X111718 | -0,53 | 0,9 | 0,86 |
| 37 | X110204 | -0,48 | 0,94 | 0,9 |
| 38 | X110233 | -0,35 | 1,45 | 1,58 |
| 39 | X1102532 | -0,35 | 0,81 | 0,76 |
| 40 | X110243 | -0,17 | 0,85 | 0,79 |
| 41 | X110252 | -0,13 | 0,83 | 0,77 |
| 42 | X110256 | 0,12 | 0,99 | 0,96 |
| 43 | X1117182 | 0,35 | 0,92 | 0,87 |
| 44 | X1102592 | 0,35 | 0,85 | 0,8 |
| 45 | X 110212 | 0,37 | 0,86 | 0,84 |
| 46 | X110263 | 0,42 | 0,92 | 0,88 |
| 47 | X1102562 | 0,46 | 1,03 | 1,06 |
| 48 | X110240 | 0,69 | 1,22 | 1,6 |
| 49 | X1102632 | 0,9 | 0,96 | 0,95 |
| 50 | X110253 | 1 | 0,92 | 0,85 |
| 51 | X110221 | 1,44 | 1,11 | 1,43 |
| 52 | X1102442 | 1,6 | 1 | 1,05 |
| 53 | X110244 | 2,18 | 0,92 | 0,72 |
| 54 | X111728 | 3,03 | 1,024 | 0,93 |
| 55 | X1102432 | 3,05 | 0,95 | 0,75 |

Tashqi moslik (Outfit) statistikasida ID raqamlari X110215, X110200, X110202, X110196 va X111656 bo'lgan test topshiriqlarining tashqi mosligi 0,7 dan kichik, ammo X110233, X110240 va X110221 bo'lgan test topshirig'ining tashqi mosligi 1,3 dan katta ekanligini ko'rish mumkin. Tashqi(outfit) moslik darajasi 1,3 dan katta bo'lgan test topshiriqlariga sinaluvchilar tomonidan

berilgan javoblar mutanosib emasligi aniqlandi. Bu esa o'z navbatida test topshiriqlariga javoblar ko'proq tasodifiy va ehtiyotsizlik tufayli bo'lganligini anglatadi [16]. Test topshiriqlarining ichki va tashqi mosligiga o'xshash jarayonni amalga oshirib, qobiliyatlarning ichki va tashqi moslik statistikasini tahlil qilish mumkin.

Xulosa

Natijalar asosida shuni aytish mumkinki, aksariyat topshiriqlarda infit va outfit moslik darajasi maqbul bo'lib, ular ishonchli va barqaror baholash vositasi sifatida qo'llanilishi mumkin. Jadvaldagi tahlil natijalari mazkur test topshiriqlarining psixometrik xususiyatlari qoniqarli ekan-

ligini, ularni o'quvchilarning bilim va ko'nikmalarini baholashda samarali tarzda qo'llash mumkinligini ko'rsatadi. Shu bilan birga, ayrim topshiriqlarda outfit qiymatlariga asoslanib, sifatni yanada oshirish bo'yicha kichik tuzatishlar kiritish maqsadga muvofiqdir.

Adabiyotlar

1. Bloom, B. S. Taxonomy of Educational Objectives: The Classification of Educational Goals. New York, 1956, Longmans, Green and Co.
2. Anderson, L. W., & Krathwohl, D. R. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. 2001, New York: Longman.
3. Compass O. O. L. 2030: A Series of Concept Notes. 2019.
4. Spencer L. M., Spencer S. M. Competence at Work: Models for Superior Performance. 1993, New York: John Wiley & Sons.
5. Andrich D. Rasch Models for Measurement. Sage Publications. 1988.
6. Bond T. G., Fox C. M., Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Routledge. Psychology Press, 2015.
7. A.R. Sattiyev, M.Dj. Ermamatov. Rayt xaritasi, ichki va tashqi moslik statistikalari: kimyo fanidan milliy sertifikat uchun o'tkazilgan test sinovi natijalari. "Axborotnoma" ilmiyuslubiy jurnal. 2024-yil, 1-son, 18-96 betlar.
8. A.B. Normurodov, I.A. Boyxonov. Rayt xaritasi, ichki va tashqi moslik statistikalari: biologiya fanidan o'tkazilgan test sinovi natijalari. "Axborotnoma" ilmiyuslubiy jurnal. 2023-yil, 2-son, 56-83 betlar.
9. Kreiner, S., Christensen, K. B. Validity and objectivity in health-related scales: Analysis by graphical loglinear Rasch models. Statistical Methods in Medical Research, 2011, 20(1), pp. 25-44.
10. Wright, B. D., Masters, G. N. Rating Scale Analysis: Rasch Measurement. MESA Press. 1982.
11. M.Dj. Ermamatov, A.R. Sattiyev, A.B. Normurodov, Z.O. Olimbekov, A. A. Baratov. Fizika fanidan o'tkazilgan test sinovi natijalari: Rayt xaritasi, ichki va tashqi moslik statistikalari, Rash modeli bilan moslik. "Axborotnoma" ilmiyuslubiy jurnal. 2023-yil, 1-son, 4-62 betlar.
12. Guttman, L. The basis for scalogram analysis. In Stouffer et al. Measurement and Prediction. The American Soldier V. IV. New York: Wiley. 1950.
13. B.D. Wright and M.H. Stone. Best Test Design, MESA Press, Chicago, 1979.
14. Linacre, J. M. What do Infit and Outfit, Mean-square and Standardized mean? Rasch Measurement Transactions. 2002, 16(2), 878 p.
15. Smith, R. M. The distributional properties of Rasch item fit statistics. Educational and Psychological Measurement. 1991, 51(3), pp. 541-565.
16. A.R. Sattiyev, M.Dj. Ermamatov. Matematika fanidan milliy sertifikat uchun o'tkazilgan test sinovlari natijalari tahlili, "Axborotnoma" ilmiyuslubiy jurnal. 2023-yil, 2-son. 36-55-betlar.

INFIT AND OUTFIT STATISTICAL RESEARCH OF TEST RESULTS

K.A. Amonov¹, Y.Z. Karimov²

¹Scientific-Study Practical Center under the Agency for Assessment Knowledge and Competences, Tashkent 100084, Bogishamol 12

²Agency for Assessment Knowledge and Competences, Tashkent 100084, Bogishamol 12

Abstract. In this article, based on the results of history subject tests, the sensitivity of test item response patterns to the ability levels of test takers and conversely, the sensitivity of ability levels to response patterns was examined. In addition, the alignment between the difficulty levels of test items and the ability levels of test takers was analyzed. The findings of the study showed that for some test items, the responses given by test takers were not consistent, and that many of the responses appeared to be the result of randomness or carelessness.

Keywords: history, infit and outfit statistics, Rasch model, variance, difficulty level, ability level.