

## TESTO INFORMACIJA IR JOS TAIKYMAI

### Gediminas Trakas

Viena iš populiariausių ir dažniausiai naudojamų priemonių žinioms bei gebėjimams vertinti yra testas. Apie testą ir testavimą šiuo metu girdime gana dažnai. Tai moksleivių egzaminų užduotys, profesinio pasirengimo testai priimant į darbą, intelekto testai ir pan. Testui, kaip matavimo instrumentui, yra keliami dideli reikalavimai, jo parengimas yra ilgas ir pakankamai sudėtingas. Geras testas turi patenkinti daugybę sąlygų, todėl jį rengiant reikia įvertinti nemažai testo kokybę nusakančių charakteristikų. Viena iš tokių charakteristikų, naudojama testų teorijoje ir leidžianti įvertinti net keletą parametrų, yra užduoties ar viso testo informacija. Šiame straipsnyje, apibrėžę, kas tai yra, pabandysime šiai charakteristikai padedant atsakyti į keletą klausimų, išskylančių testus rengiantiems ir juos taikantiems specialistams.

Klasikinė testų teorija, pradėta vystyti XX a. pradžioje ir dar pakankamai plačiai naudojama šiais laikais, susidūrė su keletu apribojimų, kurių pati negalėjo išspręsti (Hambelton ir kt., 1991). Išėitį pasiūlė užduoties sprendimo teorija (*angl. Item Response Theory, IRT*), kuri priklauso moderniajai testų teorijai, o jos pradininku laikomas danų mokslininkas Georgas Raschas, 1960 metais išleidęs knygą "Probabilistic Models for Some Intelligence and Attainment Tests". 1993 metais MESSA Press išleido trečiąją šios knygos leidimą su testų teorijos specialisto Benjamino Wrighto įžanga ir komentarais. Įžangoje šią G. Rascho knygą jis įvardija svarbiausiu psichometrijos darbu nuo Thurstone'o straipsnių, skelbtų 1925–1929 metais (Rasch, 1993).

IRT teorija siūlo testo uždaviniui priskirti funkciją, nusakančią sąryšį tarp testuojamųjų žmonių charakteristikų (gebėjimų, žinių lygio) ir tikimybės išspręsti tą uždavinį. Ši funkcija turi priklausyti nuo uždavinio parametrų. Pagal šios funkcijos išraišką išskiriami atskiri IRT modeliai. Pati funkcija vadinama charakteristine uždavinio funkcija, o jos kreivė – charakteristine uždavinio kreive (Hambelton ir kt., 1991).

Šiame darbe remiamasi vienu iš IRT modelių, kuris pagal uždavinio parametrų skaičių vadinamas dviparametriniu IRT modeliu. Šio modelio uždavinio charakteristinės funkcijos matematinė išraiška yra:

$$P(\theta) = \frac{e^{a(\theta-b)}}{1 + e^{a(\theta-b)}}, \quad (1)$$

kur  $P(\theta)$  yra tikimybė, kad atsitiktinai pasirinktas egzaminuojamasis, kurio lygis gebėjimų skalėje yra  $\theta$ , teisingai išspręs uždavinį;  $b$  yra parametras, nusakantis uždavinio sunkumą, o  $a$  – parametras, nusakantis uždavinio skiriamąją gebą, t. y. kaip stipriai uždavinys atskiria geresnius nuo blogesniųjų tikrinamo dalyko atžvilgiu.

Iš pradinių testo atlikimo duomenų, t. y. žinant, kuriuos uždavinius kiekvienas testuojamasis išsprendė teisingai ar neteisingai, vienu metu yra įvertinami ir uždavinių parametrai, ir testuojamųjų gebėjimai. Gebėjimų įverčiai  $\theta$  dažniausiai pateikiami normuotoje  $z$  taškų skalėje, kurios vidurkis yra lygus 0, o standartinis nuokrypis – 1.

Testo informacijos funkcija  $I = I(\theta)$  yra apibrėžta, kaip atvirkštinė funkcija gebėjimų įverčio pasikliautinojo intervalo ilgiui kiekviename gebėjimų skalės taške  $\theta$  (Lord, 1980). Todėl testo standartinės matavimo paklaidos *SEM* (*standard error of measurement*) išraiška yra

$$SEM(\theta) = \frac{1}{\sqrt{I(\theta)}}.$$

Šitaip apibrėžta testo informacija nusako, kaip tiksliai testas matuoja tikrinamo dalyko atžvilgiu skirtingų gebėjimų testuojamuosius. Kuo informacija atitinkamame gebėjimų lygyje yra didesnė, tuo mažesnis yra to gebėjimų lygio įverčio pasikliauties intervalas ir paklaida. Kita vertus, kuo tiksliau testas įvertina, tuo juo gaunama informacija (informacija bendrąja prasme) apie testuojamojo gebėjimų lygį yra tikslesnė ir kokybiškesnė. Testavimo specialistas, sakydamas, kad testu surenkama informacija yra didesnė ar mažesnė, tai supranta kaip tvirtinimą apie informacijos kokybę. Atskirai paėmus kokio nors testo informacijos kiekį kokiame nors gebėjimų lygyje, dar nieko pasakyti apie testą ar testuojamąjį negalima. Informacijos kiekis yra prasmingas

tik lyginant atskirų testų informacinės funkcijas ar to paties testo informacijos kiekį skirtinguose gebėjimų skalės taškuose.

Testo informacijos funkcija gali būti išreikšta formule (Hambelton ir kt., 1990, p. 98):

$$I(\theta) = \sum_{i=1}^n \frac{(P_i'(\theta))^2}{P_i(\theta)(1 - P_i(\theta))}, \quad (2)$$

kur  $P_i(\theta)$  yra  $i$ -tojo uždavinio charakteristinė funkcija,  $P_i'(\theta)$  – jos išvestinė, o  $n$  – testo užduočių skaičius. Atskiras  $i$ -tasis šios sumos narys yra vadinamas  $i$ -tojo uždavinio informacijos funkcija. Vietoj  $P_i(\theta)$  įstatę konkrečią išraišką (1), gauname testo informacijos funkciją dviparametriniam IRT modeliui.

Testo informacinės funkcijos savybės plačiai išnagrinėtos F. M. Lordo knygoje (1980), apie tai informacijos galima rasti ir kituose šaltiniuose (Hambelton ir kt., 1990; Hambelton ir kt., 1991). Pirmiausia, ką galima pastebėti iš (2) išraiškos, kad kiekvienas uždavinys į testo informacijos kiekį “įneša” savo informacijos dalį, nepriklausomą nuo kitų užduočių, t. y. testo informacijos funkcija yra tiesinė uždavinių informacijos funkcijų suma. Naudojantis tuo galima apskaičiuoti bet kurios testo dalies informaciją. Be to, žinant atskirų uždavinių informacijos funkcijas galima sudaryti būsimo testo informacijos funkciją bei žinoti testo matavimo paklaidą dar prieš patį testavimą.

Platesnis testo informacijos panaudojimas tapo įmanomas sparčiai vystantis kompiuterinei technikai. Turint pradinius duomenis, testo modeliui įvertinti ir informacijos funkcijai apskaičiuoti pakanka personalinio kompiuterio bei specialios programinės įrangos (dažnai naudojami programiniai paketai ITEMAN, XCALIBRE, QUEST, BIGSCALE, BILOG ir kiti).

Vieni pirmųjų bandymų taikyti IRT ir testo informacijos metodus Lietuvoje buvo analizuojant Vilniaus universiteto stojamąjį matematikos egzaminą į matematikos fakultetą bei amerikietišką matematikos egzaminą AHSME. Apskaičiuota testo informacija parodė, kad daugiausia jos yra surenkama apie stojančiuosius, kurie už egzaminą gavo pažymius 4 ir 5, t. y. ribinius pažymius, kai 4 reiškė neišlaikyta, o 5 – išlaikyta. Konkursas nebuvo labai didelis, tad didžiausia konkurencija ir buvo tarp šių stojančiųjų. Pagal informacijos funkcijos prasmę, testas tiksliausiai įvertino ir geriausiai atskyrė būtent tuos kandida-

tus. AHSME egzaminas buvo vykdomas kaip matematinis konkursas. Informacijos funkcijos kreivė parodė, kad daugiausia informacijos buvo surinkta apie geriausius moksleivius. Taigi tiksliausiai įvertinti ir geriausiai surikiuoti pagal pasiekimus buvo geriausieji moksleiviai. Tai patvirtino testo paskirtį.

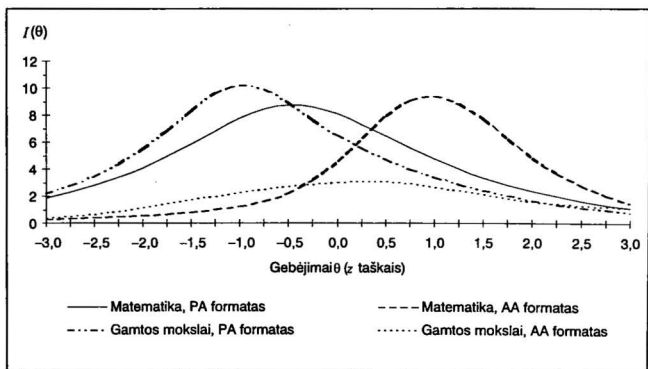
Kartais testo užduotys sudaromos iš uždavinių su pasirenkamaisiais atsakymais. Daug kas į juos žiūri gana skeptiškai. Kokių nors specialių tyrimų, kaip tokio formato uždaviniai tinka moksleivių vertinimui, Lietuvoje nėra, nors pasaulyje ši tema nagrinėjama jau seniai. Galima būtų paminėti Roberto Lukhele'o ir jo kolegų tyrimus (Lukhele ir kt., 1994), atliktus naudojantis IRT metodais. Jų tyrimo objektas buvo Amerikos koledžų tarybos (*US College Board*) istorijos egzamino testas, kuriame buvo klausimų su pasirenkamaisiais atsakymais ir klausimų, į kuriuos reikėjo atsakyti pilnai. Informacijos įvertinimas parodė, kad norint surinkti tokį pat kiekį informacijos apie vidutinį moksleivį vien atviro tipo užduotys užimtų 20 kartų daugiau egzaminavimo laiko nei testas, sudarytas tik iš klausimų su pasirenkamaisiais atsakymais. Tačiau to paties lygio moksleiviui įvertinti viena atviro tipo užduotimi surenkama tiek informacijos, kiek būtų surenkama 4 klausimais su pasirenkamaisiais atsakymais. Šie skaičiai priklauso nuo moksleivio gebėjimų lygio. Pavyzdžiui, apie blogiausius moksleivius surenkamos informacijos kiekiu viena atviro tipo užduotis atitiktų 8 klausimus su pasirenkamaisiais atsakymais. Tai gali būti dėl to, kad tokie moksleiviai, atsakinėdami į klausimus, kurių pateikti galimi atsakymai, dažniau spėlioja, arba tokius testus atlieka labai greitai, nesigilindami į klausimus. Taip nukenčia tokiais testais gaunamų blogesnių moksleivių žinių įvertinimo tikslumas.

1995 metais Lietuvoje vyko Tarptautinės švietimo pasiekimų asociacijos (IEA) matematikos ir gamtos mokslų tyrimas vidurinėje mokykloje – Trečioji tarptautinė matematikos ir gamtos mokslų studija – TIMSS (Čekanavičius, Trakas, Zabulionis, 1997). Lietuvos 7–8 klasių moksleiviams pateikti testai buvo sudaryti iš uždavinių su pasirenkamaisiais atsakymais (žymėkime juos PA) ir iš atvirų uždavinių, t. y. kai reikėjo rašyti sprendimą ir atsakymą (žymėkime juos AA). Sukaupta duomenų bazė leidžia taikyti IRT metodus ir informacijos įverčius.

Svarbus klausimas: kaip PA uždaviniai, su kuriais dauguma Lietuvos septyntokų ir aštuntokų susidūrė pirmą kartą, tinka mūsų moksleivių žinioms

įvertinti? Ar tai nebuvo ta priežastis, kuri lėmė labai žemus Lietuvos rezultatus lyginant su kitomis šalimis, kurių moksleiviams jos yra įprastos? Taigi pažiūrėkime į TIMSS matematikos testo informacijos kreives atskirioms testo dalims: PA ir AA uždaviniai (1 pav.).

Kaip matome, matematikos testo abiejų formų uždavinių grupės surinko panašų informacijos kiekį. Tikrai PA uždaviniai daugiausia informacijos surinko apie vidutinius ir šiek tiek blogesnius (moksleivių gebėjimų skalėje  $\theta$  kinta nuo  $-1$  iki  $0$  z taškų), o AA uždaviniai panašų kiekį informacijos surinko apie šiek tiek geresnius nei vidutiniai moksleivius ( $0,5$ – $1,5$  z taškai). Tai reiškia, kad testas gana vienodu tikslumu įvertino tiek blogesnius, tiek ir geresnius moksleivius. Blogesniųjų įvertinimo tikslumą užtikrino PA uždaviniai, o geresniųjų – AA.

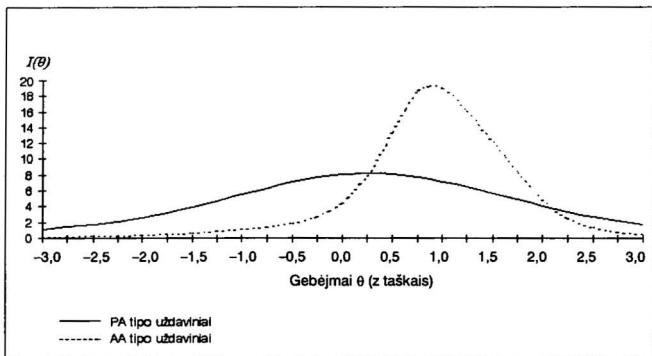


1 pav. TIMSS matematikos ir gamtos mokslų testų informacijos kreivės. 8 klasė

Kitaip yra gamtos mokslų testo atveju. Čia PA formato uždaviniais surinktos informacijos kiekis ir informacijos maksimumo vieta yra panašūs į matematikos testo to paties formato uždavinių, bet AA formato uždaviniais buvo surinkta daug mažiau informacijos nei PA gamtos mokslų uždaviniais, taip pat mažiau nei AA matematikos uždaviniais. Taigi gamtos mokslų testas, sudarytas iš AA uždavinių, būtų daug mažiau tikslus, negu sudarytas iš pana-

šių PA uždavinių. Be to, bendrame testo informacijos kiekyje AA uždavinių informacijos dalis yra labai maža.

Pažiūrėkime į to paties matematikos testo informacijos kreivę 7 klasės moksleiviams (2 pav.). Abiejų tipų uždaviniais surenkamos informacijos maksimumų vietos moksleivių gebėjimų skalėje beveik nesiskiria nuo 8 klasės atveju. Labai stipriai skiriasi AA tipo uždaviniais surenkamos informacijos kiekis. Jis beveik dvigubai didesnis už to paties tipo uždavinių surinktos informacijos kiekį 8 klasėje. Tai rodo, kad testas, sudarytas iš tokių uždavinių, yra daug tikslesnis vertinant gablesnius (0,5–1,5 z taškai) 7 klasės moksleivius, negu testas, sudarytas iš PA tipo uždavinių, bei to paties tipo testas, vertinant gablesnius 8 klasės moksleivius.



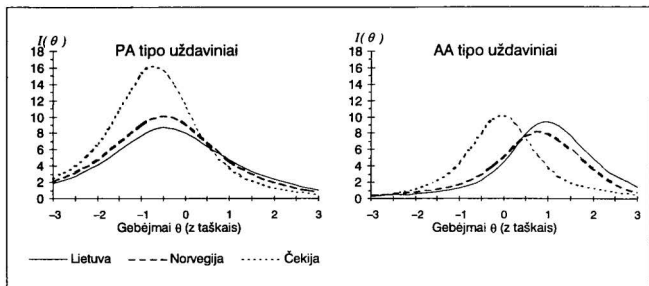
2 pav. TIMSS matematikos testo informacijos kreivės. 7 klasė

Analizuodami IRT informacijos kreives matome, kad tie tariamai neįprasti mūsų moksleiviams PA tipo uždaviniai palyginti tiksliai įvertina septintokų ir aštuntokų matematikos bei aštuntokų gamtos mokslų žinias. Surenkamos informacijos kiekis šiais atvejais yra beveik toks pat. Be to, testu, sudarytu iš PA tipo uždavinių, surenkama beveik vienodas kiekis informacijos apie mokinį gana plačioje jų gebėjimų skalėje.

Panašiai buvo palygintos TIMSS testų PA ir AA dalių informacijos funkcijos atskirai berniukams ir mergaitėms bei skirtingų vietovių moksleiviams

(miesto ir kaimo). Tarp šių grupių nebuvo reikšmingų skirtumų (PA ir AA testų dalimis surenkamos informacijos atžvilgiu).

TIMSS tyrimo metu sukaupta duomenų bazė leidžia palyginti testų rezultatus ne tik Lietuvoje, bet ir su kitomis šalimis (3 pav.). Visos tyrime dalyvusios šalys naudojo tuos pačius testus.



3 pav. TIMSS matematikos testo informacijos kreivės

Palyginkime PA ir AA testo dalių informacijos kreives, nubrėžtas atskirioms šalims: Lietuvai, Norvegijai ir Čekijai. Kaip matome, abiejų matematikos testo dalių PA ir AA informacijos kreivės Lietuvos ir Norvegijos atvejais yra labai panašios. Tai rodo, kad tuo testu surinkta vienoda informacija ir Lietuvoje, ir Norvegijoje, testas vienodai tiksliai įvertino abiejų šalių moksleivius. Čekijos atveju yra kitaip. Čia PA tipo uždaviniais surinkta daugiausia informacijos apie panašaus gebėjimo lygio (atskiroje šalyje) moksleivius kaip Lietuvoje ir Norvegijoje, bet informacijos kiekis yra dvigubai didesnis. AA tipo uždavinių atveju informacijos kiekis nesiskiria, bet Čekijoje šiais uždaviniais daugiau informacijos buvo surinkta apie vidutinį moksleivį (nuo  $-1$  iki  $1$  z taškų), o Lietuvoje ir Norvegijoje – apie geresnį nei vidutinį ( $0,5$ – $1,5$  z taškų). Čekijoje apie vidutinį moksleivį beveik tiek pat informacijos surenkama tiek PA, tiek ir AA tipo matematikos uždaviniais. Lietuvoje ir Norvegijoje AA tipas labiau tinka geresnio nei vidutinio moksleivio matematikos žinioms įvertinti.

Toks informacijos kreivių palyginimas leidžia teigti, kad Lietuvos ir Norvegijos moksleiviai mažiau susidūrę su PA tipo uždaviniais nei jų bendraamžiai iš Čekijos. Gali būti, kad Čekijoje PA tipo uždavinių testai yra dažniau naudojami moksleivių žinių vertinimui.

Kaip matome, naudojant IRT informacijos funkcijas galima įvertinti ne tik bendrus testo parametrus, bet ir tai, kokių gebėjimų lygio testuojamiesiems tas testas tinka. Žinant kokie uždaviniai daugiau „įneš“ informacijos, galima iš anksto sudaryti testą, kuris tiksliausiai įvertins norimą respondentų grupę.

Testus rengiantys specialistai plačiau savo darbe naudodami IRT informacijos metodus galėtų lengviau diagnozuoti silpnąsias testų vietas, jas taisyti. Taip pat šių metodų panaudojimas prisidėtų objektyviai vertinant naudojamus testus – tiek parengtus užsienio specialistų ir atsivežtus bei adaptuotus Lietuvoje, tiek ir naujai sukurtus.

Už paramą rengiant šį straipsnį dėkoju mokslinio darbo vadovui docentui Algirdui Zabulioniui.

## LITERATŪRA

1. Hambleton R. K., Swaminathan H., Rogers H. J. *Fundamentals of Item Response Theory*. SAGE Publications, 1991.
2. Hambleton R. K., Swaminathan H. *Item Response Theory. Principles and Applications*. Kluwer-Nijhoff Publishing, 1990.
3. Lord F. M. *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum Assoc., 1980.
4. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. MESSA Press, 1993.
5. Lukhele R., Thissen D., Wainer H. On the Relative Value of Multiple-Choice, Constructed Response, and Examinee-Selected Items on Two Achievement Tests // *Journal of Educational Measurement*. 1994. Vol. 31. No. 3. P. 234–250.
6. Čekanavičius V., Trakas G., Zabulionis A. Trečioji tarptautinė matematikos ir gamtos mokslų studija. 7–8 kl. moksleivių tyrimo statistinė ataskaita. Vilnius, 1997.

Įteikta 1997 m. birželio mėn.

## TEST INFORMATION AND ITS APPLICATIONS

### Summary

The item and test information functions play key roles in the item response theory. They allow to estimate the standard errors of measurement for each item at a given level of examinee's ability  $q$ . Another important aspect of the application of the item information functions is the determination of the contribution of an individual item to the precision of the total test. The article introduces the concepts of the item and test information functions and suggests some methods of the application of them in the test preparation and interpretation of test results. Some analysis of an impact of the item format (multiple-choice and open-ended) on the test results is presented in this article. This study is based on the data collected through the Third Mathematics and Science Study TIMSS.