

Vilniaus universitetas
Matematikos ir informatikos fakultetas

Aistė Elijo

TESTŲ TEORIJOS PAGRINDAI

Paskaitų konspektas

2012

Ivadas

Šis leidinys skirtas studentams, pasirinkusiems kursą „Statistika švietimo tyrimuose“, kurio metu nagrinėjamos įvairios švietimo tyrimuose vartojamos su statistika susijusios teorijos ir metodikos. Jis taip pat bus naudingas klausantiems „Baigtinių populiacijų statistikos“ kurso, kur, be kita ko, aptariami ir testų teorijos, kaip sudėtingų tyrimų instrumentų kūrimo metodikos, pagrindai. Abiem atvejais testų teorijos pagrindai apima tik dalį šių kursų metu nagrinėjamų temų. Kadangi literatūros šia tematika lietuvių kalba nėra daug, leidinys „Testų teorijos pagrindai“ skirtas trumpai apžvelgti pagrindines testų teorijos sąvokas, testo kūrimo etapus, veiksnius, įtakojančius testo užduoties kokybę.

Leidinyje iš pradžių aptariamos bendrosios testo charakteristikos – validumas, patikimumas, atitiktis testo / tyrimo programai ir matricai.

Antroje leidinio dalyje nagrinėjami įvairūs testo „ląstelių“ – klausimų tipai, jų privalumai ir trūkumai, o taip pat pateikiami patarimai, kurie gali būti naudingi sudarant įvairių tipų klausimus.

Trečioje dalyje pateikiamos bendros testų sudarymo taisyklės, įvairių tipų klausimų derinimo aspektai, kontroliniai klausimai testo sudarytojui.

Ketvirtoje dalyje kalbama apie vertinimo įtaką testo užduoties kokybei, jos turinio pagrįstumui.

Galiausiai penktoje dalyje trumpai aptariama, kaip kiekybiniai rodikliai (statistinė testo analizė) gali pateikti svarbios informacijos apie testo užduoties turinį ir jos kokybę.

Turinys

Įvadas

I. Bendrosios testo charakteristikos

Testo patikimumas ir validumas

Atitiktis programai ir matricai

II. Testo užduoties „ląstelė“ - klausimas / punktas

Pasirenkamo atsakymo klausimai

Taip / Ne klausimai

Susiejimo / įrašymo iš duoto sąrašo klausimai

Trumpo atsakymo klausimai

Trumpo sprendimo / pagrindimo klausimai

Struktūruoti klausimai

III. Kai kurios bendrosios testų sudarymo taisyklės

Užduoties rengimo etapai

Sąlygos korektiškumas

Lygių galimybių principas

Sudėtingumas

Pagrindinės taisyklės

Kontroliniai klausimai testo užduoties rengėjui

IV. Truputis apie vertinimą

Sprendimo / atsakymo aiškumas

Taškų paskirstymo racionalumas

V. Statistinė testo analizė - kiekybinė testo kokybės patikra

Kam reikalinga statistinė analizė?

Testo klausimų statistinė analizė

Surinktų taškų analizė

Patikimumas

Literatūros sąrašas

I. Bendrosios testo charakteristikos

Testo patikimumas ir validumas

Kalbant apie testo charakteristikas, paprastai labiausiai išskiriami du bendrieji testo kokybės rodikliai – testo validumas ir testo patikimumas. Trumpai juos aptarsime.

Testo **validumas** apibrėžia, kaip tiksliai jis matuoja tai, ką ketina matuoti. Gali pasirodyti, kad tai savaime suprantamas dalykas, tačiau iš tiesų tai yra svarbiausias testo kokybės aspektas. Deja, neretai testų kūrėjų įsivaizdavimas apie tai, ką jie nori pamatuoti, ir tai, ką realiai testas matuoja, ženkliai skiriasi. Be abejo, testo validumas priklauso nuo to, koks klausimas / problema / gebėjimas yra tiriami: vieno klausimo atžvilgiu testas gali būti validus, kito – ne. Jeigu testui trūksta validumo, net jei jis turi daug kitų gerų savybių, jo pateikiama informacija yra beprasmė ir klaidinanti.

Priklausomai nuo testo paskirties, kalbama apie validumą

- turinio atžvilgiu
- kriterijaus atžvilgiu
- konstrukto atžvilgiu

Egzamino arba švietimo tyrimo užduoties kontekste mus labiausiai domina *validumas turinio atžvilgiu*. Paprastai kalbant, jis parodo, kiek testas reikalauja iš testuojamojo pademonstruoti pasiekimus, kurių tikimasi mokant šio dalyko arba kurie yra apibrėžti šio dalyko standartuose, egzamino ar tyrimo programoje.

Vertinant testo validumą, pirmiausiai reikėtų atsakyti į šiuos klausimus:

- Ar užduoties turinys atitinka programos uždavinius dalyko turinio ir gebėjimų atžvilgiu?
- Ar testas nereikalauja ypatingų gebėjimų, kurie nėra susiję su matuojamu dalyku? (pavyzdžiui, reikalingas aukštas skaitymo gebėjimų lygis, atliekant matematikos ar gamtos mokslų užduotis; rašymo lygis teksto suvokimo teste ir pan.)

Testo **patikimumas** apibrėžia, kaip nuosekliai (pastoviai) jis matuoja tai, ką matuoja. Kitaip tariant, jei tą testą tam pačiam testuojamajam duotume atlikti po kažkurio laiko, ar jis taip pat jį atliktų? Galime sakyti, kad testo patikimumas tam tikra prasme parodo, ar smarkiai testo atlikimą įtakoja kažkokie atsitiktiniai dalykai, ar vis dėlto jis yra savyje suderintas.

Jei testo patikimumą matuoja tam tikri statistiniai kriterijai (apie juos plačiau kalbama paskutinėje dalyje), tai testo validumo kiekybiniais kriterijais negalime tiesiogiai pamatuoti, todėl turime būti itin atidūs. Svarbu atsiminti, kad testas gali būti **patikimas**, tačiau tuo pačiu **ne validus**, o **validus** testas tuo pačiu yra ir **patikimas** (jei nėra patikimumo, negalime pasitikėti ir validumu). Kitaip tariant, patikimumas yra būtinas, tačiau ne pakankamas validumui užtikrinti. Kad geriau pailiustruotume šį testo patikimumo ir validumo ryšį, pateiksime kelis pavyzdžius.

Įsivaizduokime, kad didelei kinų kalbos nemokančių žmonių grupei duotume atlikti kinų kalbos testą, kuriame visi klausimai yra pasirenkamo atsakymo tipo, tai yra atsakantieji kiekvienam klausimui turi pasirinkti vieną iš 4 pasirenkamų atsakymų. Visiškai atsitiktinai rinkdamiesi atsakymus, vidutiniškai testuojamieji teisingai atsakytų apie 25 proc. klausimų. Panašus rezultatas būtų ir kitą kartą pamatavus. Taigi galima būtų sakyti, kad testas yra patikimas, tačiau ar jis yra validus – ar išties jis matuoja testuojamųjų kinų kalbos žinias?

Kitas pavyzdys – norime pamatuoti žmonių sąžiningumą. Tam sudarome anketą su kruopščiai parinktais įvairiausiais klausimais – situacijomis, kuriose atitinkami pasirinkimai iliustruotų žmogaus sąžiningumą, ir paprašome testuojamųjų atsakyti į tuos klausimus. Pavyzdžiui, klausiamo, kaip dažnai jie meluoja, ar pasisavina ne jiems priklausančius daiktus ir t.t. Greičiausiai žmonių atsakymai būtų gana suderinti tarpusavyje, tačiau ar galėtume pasitikėti tokios anketos rezultatais? Ar tikrai patys sąžiningiausi atsakytų, kad jie niekada nemeluoja? O gal būtent vagys ir melagiai teigiamai atsakytų į visus klausimus? Taigi, labai abejotina, kad toks testas būtų validus - tai yra tikrai matuotų testuojamųjų sąžiningumą.

Galbūt šie pavyzdžiai atrodo gana nutolę nuo realių egzaminų ar švietimo tyrimų situacijų, bet galime įsivaizduoti ir artimesnę bei subtilesnę situaciją – tarkime, pateikiame penktos klasės matematikos testą dvyliktokams. Jei testas

gerai sudarytas, jis matuos gana patikimai ir, aišku, matuos testuojamųjų matematikos žinias, tačiau ar tas, kurias norime matuoti, ar tas, kurias skelbiamės matuojantys? Turbūt ir šiuo atveju testas stokos validumo.

Atitiktis egzamino arba tyrimo programai ir matricai

Vienas iš svarbių testo validumo aspektų yra susijęs su tuo, kaip testo užduotis atitinka egzamino arba tyrimo programą ir matricą, nes būtent jose apsibrėžiame, ką užduotimi norėsime tikrinti.

Atitiktis programai vertinama vienareikšmiškai: ar šiuo klausimu tikrinamos žinios ir gebėjimai yra apibrėžti programoje, ar nėra. Kartais šis vertinimas nėra lengvas dėl to, kad programos reikalavimai nėra aiškiai suformuluoti, skamba dviprasmiškai ir pan.

Egzamino arba tyrimo programa paprastai apibrėžia šiuos užduoties aspektus:

- Turinio sritys ir temos
- Veiklos sritys, gebėjimai, kompetencijos, etc.

Testo matrica apibrėžia turinio sričių ir veiklos sričių / gebėjimų, kompetencijų proporcijas užduotyje arba visumoje tyrimo užduočių.

Testo matricos struktūra

Gebėjimai	Pirmas	Antras	Trečias	Viso
Turinio sritys, temos				
Pirma				
Antra				
Trečia				
Ketvirta				
Viso				100 %

Egzamino matrica yra susitarimas tarp egzamino rengėjų ir egzamino kandidatų (kokios bus konkrečios egzamino užduoties skirtingų dalių proporcijos). Ji taip pat leidžia egzamino rengėjams nepamiršti validumo turinio atžvilgiu, užtikrinti, kad egzamino užduotis / testas tinkamai „padengia“ mokomo dalyko turinį, atspindi jo programą. Paprastai konkrečios egzamino užduoties matricoje leidžiamos paklaidos tarp dalių gali būti ± 5 proc.

Egzamino atitiktis matricai gali būti atliekama remiantis keliais principais:

1. klausimo taškų priskyrimo matricos langeliui aspektu:

visas uždavinys / klausimas talpinamas į vieną matricos langelį

ARBA

uždavinys / klausimas taškais skaidomas į skirtingus matricos langelius.

2. klausimu matuojamų gebėjimų priskyrimo aspektu:

laikoma, kad į aukštesnius gebėjimus (veiklos sritis) automatiškai įeina ir žemesni gebėjimai (veiklos sritys), pvz., į taikymus ar problemų sprendimą automatiškai įeina žinių reprodukovimas

ARBA

laikoma, kad aukštesni gebėjimai (veiklos sritys) ir žemesni gebėjimai (veiklos sritys) gali būti matuojami atskirai.

3. turinio / tematikos sričių aspektu:

laikoma, kad tam tikros tematikos sritys automatiškai apima kitas tematikos sritis (pavyzdžiui, į geometrijos sritį automatiškai įeina skaičiavimų sritis)

ARBA

laikoma, kad visos tematikos sritys matuojamos atskirai.

Tyrimo matrica apibrėžia tyrimo testų uždavinių / klausimų *visumos proporcines dalis* (neturi susitarimo savybių, greičiau pačių rengėjų nuostata, kokiomis proporcijomis, pavyzdžiui, „padengiamas“ mokymo turinys)

II. Testo užduoties „ląstelė“ - klausimas / punktas

Testo užduoties kokybė neįmanoma be jos pagrindinių „ląstelių“ (angliškai vadinamų *items*, lietuviškai – *uždavinių, klausimų, punktų*) kokybės. Šioje dalyje aptarsime kai kuriuos dažnai pasitaikančius klausimų tipus, jų sudarymo ypatumus ir iššūkius, pateiksime kai kuriuos bendruosius patarimus jų naudojimui ir tarpusavio derinimui.

Paprastai išskiriami šie pagrindiniai klausimų tipai:

- Fiksuoto atsakymo
 - Pasirenkamo atsakymo klausimai
 - Taip/Ne klausimai
 - Susiejimo klausimai
- Trumpo atsakymo
- Trumpo sprendimo / pagrindimo
- Struktūruoti
- Probleminiai
- Esė, projektai, etc.

Čia trumpai aptarsime visus, išskyrus paskutinius du, klausimų tipus, kadangi pastarieji kol kas retai naudojami egzaminų ar statistinių tyrimų metu.

Pasirenkamo atsakymo klausimai

Šio tipo klausimuose iš testuojamojo nereikalaujama pateikti sprendimo, tik pasirinkti iš kelių pateiktų jo(s) manymu teisingą atsakymą. Paprastai vienas variantas yra teisingas (dažnai vadinamas **raktu**), o kiti – klaidingi (vadinami **distraktoriais**).

Šio tipo klausimai yra gana populiarūs įvairiuose testuose (kartais net pats žodis „testas“ suvokiamas kaip pasirenkamo atsakymo klausimų rinkinys). Šis populiarumas sietinas su visa eile tokio tipo klausimų suteikiamų privalumų:

- Šie klausimai garantuoja lengvą ir visiškai objektyvų vertinimą (dažnai galima naudoti šiuolaikines skanavimo sistemas, todėl subjektyvus žmogiškas faktorius praktiškai visiškai eliminuojamas).
- Per trumpą laiką galima patikrinti labai įvairias žinias – surinkti daug informacijos, „padengti“ plačią temų skalę.
- Jei šie klausimai yra tinkamai sudaryti, jie gerai koreliuoja su rezultatais, gautais atsakant į kitų tipų klausimus.
- Jie leidžia lengvai identifikuoti tipines klaidas ir ypač tinka, kai galimų atsakymų aibė aiškiai apibrėžta.

2.1. pavyzdys

Skritulio skersmuo yra lygus 6 cm.	
Kam lygus skritulio plotas kvadratiniais centimetrais?	
A	6π
B	9π
C	12π
D	36π

Šiuo atveju visi pateikti pasirenkami atsakymai atitinka tikėtinus galimus skaičiavimus (žr. lentelę).

Pasirenkamas atsakymas	Atsakymo teisingumas	Skaičiavimo variantas, tipinė klaida
A 6π	Distraktorius	$= 2\pi r$: skaičiuojamas apskritimo ilgis, o ne plotas
B 9π	Teisingas atsakymas (raktas)	$=\pi r^2$: teisingai skaičiuojamas plotas, atsižvelgus į tai, kad sąlygoje pateiktas skersmuo, o ne spindulys

C 12 π	Distraktorius	$=2\pi d$: skaičiuojamas apskritimo ilgis, o ne plotas, taip pat ne atsižvelgiama, kad pateiktas skersmuo, o ne spindulys (pamirštama padalinti iš 2)
D 36 π	Distraktorius	$=\pi d^2$: taikoma ploto skaičiavimo formulė, tačiau neatsižvelgta į tai, kad sąlygoje nurodytas skersmuo, o ne spindulys (pamirštama padalinti iš 2, prieš keliant kvadratu)

Norint išvengti cm ir cm^2 rašymo atsakymų variantuose, kas galėtų tapti teisingo ar klaidingo atsakymo užuomina, matavimo vienetai paminėti klausimo formuluotėje. Visi atsakymai surašyti didėjimo tvarka, nesiskiria savo formatu, todėl tikėtina, kad teisingas atsakymas bus pasirinktas tų, kurie išties teisingai atliks skaičiavimus.

Nors ir patrauklūs daugeliu aspektų, pasirenkamo atsakymo klausimai taip pat pateikia ir nemažai iššūkių, turi savo trūkumų, kuriuos verta turėti minty:

- Šio tipo klausimus sudėtinga suformuluoti, ypač parinkti prasmingus distraktorių.
- Pasirenkamo atsakymo klausimais dažnai sunku patikrinti kai kuriuos aukštesnio lygio gebėjimus (pavyzdžiui, problemų kėlimo, argumentavimo, etc.).
- Šio tipo klausimų atsakymus gana lengva „nusirašyti“, taip pat yra spėjimo tikimybė.

Imčių tyrimuose ypatingai svarbūs pliusai:

- lengvas vertinimas (išlaidos)
- platus turinio „padengimas“
- tipinių klaidų identifikavimas

Imčių tyrimuose nebaisūs trūkumai:

- „nusirašymas“ (daug sąsiuvinių, nusirašyti sunkiau)
- spėjimo tikimybė (svarbu ne kuo objektyviau įvertinti konkretų moksleivį, o surinkti informaciją apie visumą – vidutiniškai į spėjimo tikimybę galima atsižvelgti)
- netgi galime išmokti parinkinėti distraktorius

Kai kurie patarimai, sudarant pasirenkamo atsakymo klausimus

- Klausimo ir atsakymo struktūra turėtų būti suderintos
- „Raktas“ turi būti be jokių abejonių vienintelis teisingas
- Distraktoriai turi būti be jokių abejonių klaidingi
- Klausimas turi būti kaip galima trumpesnis (taupomas testavimo laikas)
- Distraktoriai turėtų matuoti tipines klaidas
- Pasirinkimai turi būti tikėtini, paraleliški, panašaus ilgio ir formuluotės ir patrauklūs (kitaip nėra prasminga juos pateikti)
- Jei vartojami skirtingų formatų atsakymai, tai jų turėtų būti apylygiai (pvz., trys atsakymai yra sudaryti iš vieno žodžio, trys – iš dviejų žodžių junginių ir pan.)
- Pasirinkimus, jei įmanoma, verta sudėti „iš eilės“
- Klausimo formuluotė neturėtų reikalauti didelių skaitymo ir lingvistinių gebėjimų (nebent jie būtų tikrinami)
- Pasirinkimai negali sutapti
- Paprastai distraktoriai neturėtų persidengti ar būti sinonimiški vienas kitam (pavyzdžiui, vanduo ir H₂O)
- Paprastai teste teisingas atsakymas pateikiamas įvairiose pozicijose (A, B ir t. t.), tačiau nebūtinai kiekvienoje pozicijoje teisingas atsakymas turi pasikartoti lygiai tiek pat kartų. Teisingo

atsakymo pozicijų eiliškumas turėtų būti atsitiktinis, neturėtų būti galima įžiūrėti prasmingų „figūrų“ ar žodžių (pvz., ABBA, CACA, ABCDEFEDCBA...).

- Reikia vengti stiprių užuominų klausimo formuluotėje
- Svarbu nenaudoti gramatinių užuominų atsakymų variantuose
- Geriausia nenaudoti neigiamų sakinių klausimo formuluotėje
- Reikėtų vengti tokių formuluočių kaip „visi aukščiau išvardinti“, „nė vienas iš aukščiau išvardintų“
- Galvojant apie visą testą, svarbu vengti klausimų, kurie „sufleruoja“ atsakymą į kitą testo klausimą
- Neprasmingų ar visai neįtikėtinų atsakymų neturėtų būti

Taip / Ne klausimai

Šie klausimai iš dalies yra panašūs į pasirenkamo atsakymo klausimus, kur pasirinkimų yra labai mažai – tik du, dažnai klausimas suformuluojamas kaip teiginys ir reikia pasirinkti vieną iš dviejų pateikiamų atsakymų – *Taip* arba *Ne*; *Teisinga* arba *Klaidinga* ir pan. Vis dėlto šis klausimų tipas turi savo specifiką, todėl juos aptarsime atskirai.

Šio tipo klausimų teikiami privalumai:

- garantuoja lengvą ir visiškai objektyvų vertinimą
- per trumpą laiką galima patikrinti labai įvairias žinias, surinkti daug informacijos, „padengti“ plačią temų skalę
- reikalauja dar mažiau laiko nei pasirenkamo atsakymo klausimai, dėl to testo klausimų skaičius gali būti apie 50 procentų didesnis nei pasirenkamo atsakymo klausimų atveju (tuo labiau lyginant su kitų tipų klausimais)
- ypatingai tinka, kai prasmingų atsakymų į klausimą gali būti tik du (viena alternatyva teisinga, kita – ne)

Tačiau šio tipo klausimai turi ir labai daug ženklių trūkumų:

- mažiau patikimi ir validūs nei pasirenkamo atsakymo klausimai
- tinkamai / prasmingai juos suformuluoti yra sunkiau nei gali pasirodyti iš pradžių
- sunku patikrinti daugumą aukštesnio lygio gebėjimų
- lengva „nusirašyti“, labai didelė spėjimo tikimybė
- nepateikia praktiškai jokios informacijos apie tipines klaidas
- galimas neigiamas efektas (jei suformuluoti kaip teiginiai, kurie nėra teisingi)

2.2. pavyzdys

Ar Kolumbas atrado Ameriką 1492 metais? (Taip / Ne)

Turbūt klausimo kokybė tik padidėtų, jei jį pakeistume į tokį:

2.3. pavyzdys

Kelintais metais Kolumbas atrado Ameriką?

- A 492 m.
- B 1482 m.
- C 1492 m.
- D 1692 m.
- E 1942 m.

Arba į tokį:

2.4. pavyzdys

_____ atrado Ameriką _____ metais

- | | |
|--------------------|---------|
| a) Kolumbas | A) 492 |
| b) Kortežas | B) 1482 |
| c) Francis Dreikas | C) 1492 |
| d) Magelanas | D) 1692 |
| e) Napoleonas | E) 1942 |

Kai kurie patarimai, sudarant Taip/Ne klausimus

- Venkite gramatinių užuominų klausimų formuluotėje
- Atminkite, kad apibendrinantys žodžiai „visi“, „visada“, „niekada“, „niekas“ dažniausiai nurodo neteisingą tvirtinimą
- Venkite neproporcingai daug „Taip“ arba „Ne“ atsakymų
- Venkite tikslių formuluočių iš vadovėlio (taip tikrinate tik atmintį, ne supratimą)
- Kiekvienu klausimu tikrinkite vieną žinių „gabaliuką“
- Venkite sunkių formuluočių

Susiejimo / įrašymo iš duoto sąrašo klausimai

Šio tipo klausimuose reikalaujama atrasti atitinkamą ryšį tarp pateiktų informacijos „gabaliukų“. Paprastai pateikiami du stulpeliai ir kiekvieną pirmo stulpelio dalį reikia susieti su antrojo stulpelio atitinkama dalimi.

Tokio tipo klausimai logiškai tinka tada, kai reikia gebėti susieti, pavyzdžiui:

- įvykius ir datas,
- įvykius ir žmones,
- įvykius ir vietas,
- apibrėžimus ir savybes,
- užsienio kalbų žodžius ir jų vertimus,
- įstatymus ir jų iliustracijas,
- taisykles ir jas atitinkančius pavyzdžius,
- autorius ir jų darbus,
- įrankius ir jų panaudojimą ir t.t.

Vis dėlto šio tipo klausimai turi ir labai daug trūkumų:

- Jie labiausiai tinkami matuoti tik žinias, atsiminimą – nėra tinkami matuoti supratimą ir kitus aukštesnius gebėjimus (labai sunku sukurti susiejimo klausimą, kuris matuotų aukštesnius taksonomijos gebėjimus)
- Galbūt tik Taip/Ne klausimų tipas labiau nei susiejimo klausimai „pasakinėja“ galimą atsakymą
- Egzistuoja didžiulė priklausomybė tarp pasirinkimų
- Padarius vieną klaidą, yra neišvengiama kita
- Jei nėra tikrai tinkamai sudaryti, susiejimo klausimai gali reikalauti daug laiko ir ta prasme būti labai neefektyvūs (reikalauja daug laiko, o patikrina nedaug žinių ar gebėjimų)

Kai kurie patarimai, sudarant susiejimo klausimus

- Kiekviename susiejimo klausime reikia naudoti tik tos pačios srities medžiagą (nemaišyti žmonių, vietų, datų, įrankių, įvykių ir t.t. viename stulpelyje)
- Verta atidžiai patikrinti, ar nėra kokių nors užuominų dėl galimo porų susiejimo (Ką mažiausiai turi žinoti asmuo, kad teisingai susietų poras?)
- Abiejų stulpelių variantus reikėtų išvardinti logine tvarka, „iš eilės“
- Jei tik įmanoma, visas klausimas turėtų būti viename puslapyje (kitaip komplikuoja atsakymus, veltui gaišamas laikas)
- Susiejimo klausimų teste neturėtų būti daug

Trumpo atsakymo klausimai

Tam tikra prasme šio tipo klausimai yra panašūs į pasirenkamo atsakymo klausimus – kaip ir juose, čia reikia pateikti tik atsakymą, o sprendimas, paaiškinimas, pagrindimas nėra būtini ir nėra tikrinami. Tačiau, skirtingai nei pasirenkamo atsakymo klausimų atveju, šiuose klausimuose galimi atsakymai nepateikiami – atsakymą respondentas turi pats pateikti, o ne pasirinkti iš pasiūlytų.

Šio tipo klausimų pagrindiniai išskirtini privalumai yra šie:

- Tokiais klausimais galima nesunkiai „padengti“ plačią temų skalę
- Jie atrodo įprastai ir „natūraliai“
- Šiuo atveju nereikia galvoti apie atspėjimo tikimybę
- Kuriantiems klausimą nereikia parinkinėti tinkamų distraktorių, kas, kaip minėjome anksčiau, tikrai nėra lengva užduotis

Vis dėlto tokio tipo klausimai taip pat turi ir trūkumų:

- Jų vertinimas ne toks objektyvus kaip fiksuoto atsakymo klausimų atveju, nors naujų skanavimo technologijų šviesoje šis trūkumas vis mažėja
- Jie paprastai neleidžia identifikuoti tipinių klaidų (nebent vertinant atsakymus naudojami atitinkami kodai)
- Tokio tipo klausimais sunku patikrinti aukštesnio lygio gebėjimus (išskyrus kai kurias išimtis)
- atsakymus į tokius klausimus gana lengva „nusirašyti“

Kai kurie patarimai, sudarant trumpo atsakymo klausimus

- Jei tik įmanoma, šias užduotis geriau formuluoti kaip klausimus, o ne kaip tvirtinimus
- Klausimų formuluotės turėtų būti aiškios, vienareikšmiškos, užkertančios kelią įvairioms interpretacijoms
- Klausimai turėtų būti suformuluoti taip, kad laukiamas atsakymas būtų trumpas ir aiškus (taip lengviau užtikrinti vertinimo objektyvumą)
- Reiktų vengti formuluočių iš vadovėlio
- Jei užduotis formuluojama kaip tvirtinimas (su praleistais žodžiais, kuriuos reikia įrašyti), vertėtų praleisti esminius dalykus, o ne nereikšmingas detales
- Svarbu vengti gramatinių ar kitokių užuominų

Kadangi pasirenkamo atsakymo ir trumpo atsakymo klausimai yra gana panašūs, žemiau pateikiame jų palyginimą tam tikrais testų sudarymo aspektais.

Problema	Pasirenkamo atsakymo klausimų atveju	Trumpo atsakymo klausimų atveju
Turinio „padengimas”	Labai tinkamas (tinkamiausias)	Labai tinkamas
Tipinių klaidų identifikavimas	Paprastas ir pigus	Įmanomas kodavimo procese, tačiau daug brangesnis
„Nusirašymas”	Problematiškas egzaminuose, testuose, kontr. darbuose, tačiau imčių tyrimuose praktiškai nereikšmingas	Problematiškas egzaminuose, testuose, kontr. darbuose, tačiau imčių tyrimuose praktiškai nereikšmingas
Distraktorių parinkimas	Sunku, bet dažnai įmanoma	Nereikia sukti galvos. Be to, kartais uždavinys / klausimas netenka prasmės, jei tampa pasirenkamo atsakymo.

Trumpo sprendimo / pagrindimo klausimai

Dažniausiai, norint patikrinti aukštesnius gebėjimus, anksčiau minėtų klausimų tipų nepakanka. Todėl naudojami klausimai, reikalaujantys pateikti sprendimą, pagrindimą, paaiškinimą, parodyti, kaip buvo gautas atsakymas arba jį pagrįsti. Šiuo atveju tikrinamas ne tik atsakymo, bet ir sprendimo ar pagrindimo teisingumas. Paprastai šių klausimų sprendimas, pagrindimas yra sudėtingesnis nei pasirenkamo atsakymo ar trumpo atsakymo klausimų atveju.

Tokie klausimai turi svarbių privalumų:

- jie yra įprasti, atrodo „natūraliai“
- jais galime tikrinti mąstymo, argumentavimo ir kitus aukštesnio lygio gebėjimus
- pačių klausimų formuluotės yra labiau atviros, nereikia parinkinėti prasmingų distraktorių
- Šiuo atveju nereikia galvoti apie atspėjimo tikimybę

Vis dėlto šie klausimai turi ir savo trūkumų, pagrindiniai jų yra šie:

- Iš egzaminuojamojo tokie klausimai reikalauja daug daugiau laiko, todėl jais galima „padengti“ ženkliai mažiau temų
- Jie paprastai neleidžia identifikuoti tipinių klaidų (nebent vertinant atsakymus naudojami atitinkami kodai)
- Jų vertinimas yra ne toks objektyvus, sudėtingesnis ir dėl to brangesnis

Struktūruoti klausimai

Šiuo atveju kalbame ne apie vieną klausimą, o dažniausiai logiškai susietą klausimų seką. Paprastai tokie klausimai yra sudaryti taip, kad neišsprendus arba neteisingai išsprendus ankstesnį klausimą, nesudaroma kliūtis gauti taškus už vėlesnį klausimą, pavyzdžiui, jei ankstesniame klausime padaroma skaičiavimo klaida, o gautu atsakymu reikia remtis vėliau, tai tolimesniuose klausimuose vertinama atsižvelgiant į ankstesnįjį mokinio gautą atsakymą (t. y., vertinimo instrukcijos priderinamos mokinio gautam ankstesniam atsakymui); arba pirmuose

klausimuose neprašoma ko nors apskaičiuoti, o pateikiamas atsakymas ir prašoma įrodyti, kad jis teisingas, – tokiu atveju respondentas vis tiek turi pademonstruoti, ar moka tai apskaičiuoti, ir jam neužkertamas kelias atsakyti į tolimesnius klausimus.

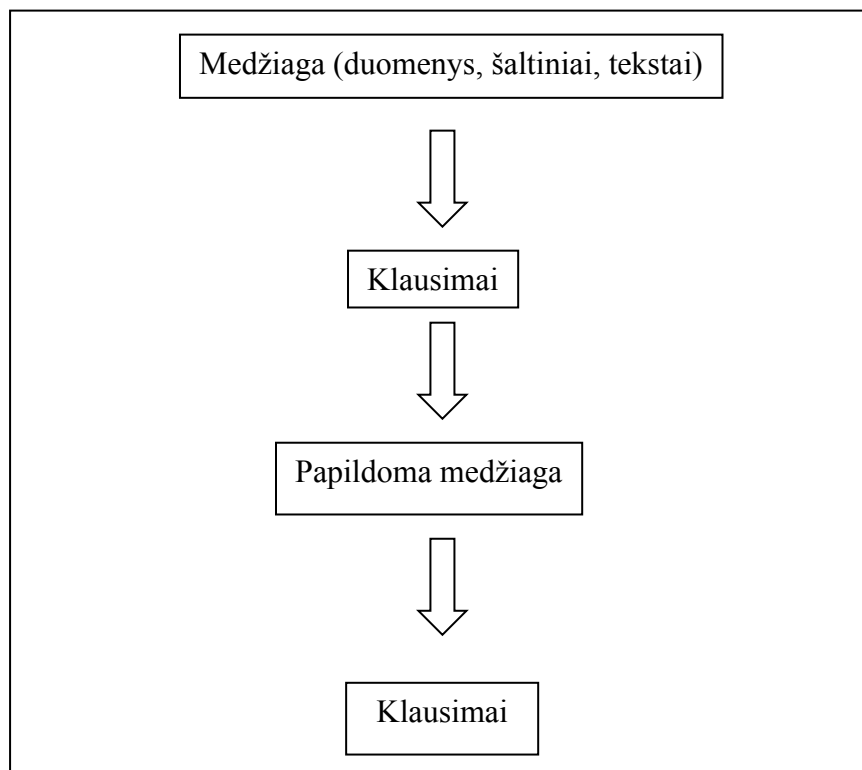
Šie klausimai yra patrauklūs, nes:

- Suteikia galimybę įsitraukti į kontekstinę ar probleminę situaciją, pateikiant eilę palaipsniui sunkėjančių klausimų
- Jie tinka tikrinti įvairių lygių gebėjimus

Tuo pačiu šis klausimų tipas turi ir trūkumų:

- Parinkti prasmingą susietų klausimų grupę nėra lengva
- Tokių klausimų atsakymas reikalauja daug laiko, galime „padengti“ ne tiek daug temų
- Jų vertinimas yra sunkesnis ir brangesnis

Paprastai struktūruotų klausimų forma yra tokia:



Kai kurie patarimai, sudarant struktūruotus klausimus

- **Medžiaga** turėtų būti:
 - realistiška, autentiška, teisinga
 - tinkamo lygio
 - leidžianti užduoti eilę tinkamų klausimų
- **Klausimai** gali būti:
 - fiksuoto atsakymo
 - trumpo atsakymo
 - reikalaujantys sprendimo ar pagrindimo
- **Klausimai** turėtų būti:
 - įvairaus lygio, leidžiantys moksleiviui pademonstruoti įvairaus lygio gebėjimus
 - susiję tarpusavy (kitaip nėra prasminga formuluoti struktūruotą klausimą; geriau kelis kitų tipų klausimus)
 - pradiniai klausimai formuluotini taip, kad net į juos neatsakius, galima būtų atsakinėti į po to einančius (tiesa, tyrimuose kodavimo instrukcija gali numatyti galimas klaidas ir atsižvelgti į tolimesnį sprendimą).

Struktūruotuose klausimuose paprastai didelis dėmesys skiriamas kontekstui, kuris galėtų būti:

- Autentiškas ir realus
- Pseudorealus

Svarbu - nuotolis nuo moksleivio:

- asmeninis gyvenimas,
- mokyklos gyvenimas,
- darbas, sportas ir laisvalaikis apskritai

- bendruomenės ir visuomenės gyvenimas (su kuriuo susiduriama kasdien)
- mokslinis kontekstas

Kontekstas jokių būdu neturėtų būti dirbtinis. Jei uždavinys (klausimas) „aplipdomas“ realaus gyvenimo atributais tik tam, kad „gražiau“ atrodytų, būtų „su kontekstu“, tai yra neprasminga ir nereikalinga. Tokiu būdu, kompromituojama pati konteksto idėja ir sukuriamas papildomas, nereikalingas balastas (pavyzdžiui, jei nupiešiamas laikrodis ir klausiama, koks kampas tarp minutinės ir valandinės rodyklės, nors visiškai aišku, kad laikrodis čia niekuo dėtas). Tokių uždavinių reikėtų vengti.

2.5. pavyzdys:

Vieną dieną pardavėja nuvažiavo **300 km** per **x^2-4** valandų.
 Kitą dieną ji nuvažiavo **325 km** per **$x+2$** valandų.
 Parašykite ir suprastinkite santykį tarp jos vidutinio greičio pirmą ir antrą dienomis.

[Jan de Lange “Real problems with real world mathematics”]

III. Kai kurios bendrosios testų sudarymo taisyklės

Ankstesnėje dalyje aptarėme kai kurių klausimų tipų savybes, privalumus ir trūkumus, o taip pat specifinius patarimus, formuluojant vieno ar kito tipo klausimus. Šioje dalyje aptarsime bendrąsias testų sudarymo taisykles – tai, ką reikia turėti minty sudarant bet kokio tipo klausimus, o taip pat visos egzamino ar tyrimo užduoties kokybiškumą.

Užduoties rengimo etapai

Kuriant testo užduotį, paprastai reikia pereiti tokius etapus:

- Parengti (apsibrėžti) testo programą, pagal kurią bus sudaromas testas – tuo būdu siekiama užtikrinti validumą turinio atžvilgiu (žinome, ką norime tikrinti, ir stengiamės sudaryti užduotį, kuri būtent tai ir tikrintų)
- Parengti / surinkti daugiau testo klausimų nei reikės galutiniame testo variante
- Peržiūrėti visus testo klausimus, peržvelgiant, ar jie atitinka testo sudarymo taisykles ir ar testas validus turinio atžvilgiu (ar atitinka testo programą, matricą)
- Duoti testą peržiūrėti kitam (kitiems) tos srities specialistui (-ams)
- Kai įmanoma, išbandyti klausimus ar net visą testą
- Atrinkti tinkamus klausimus, išmesti, keisti ar redaguoti tuos, kurie nėra tinkami

Sąlygos korektiškumas

Peržiūrint testo užduotį, visų pirma aiškinamasi, ar korektiškai suformuluotos atskirų klausimų sąlygos – tikrinama, ar:

- sąlyga yra aiški ir logiška (aišku, ko prašoma, nėra galimybės suprasti dviprasmiškai);
- sąlyga neturi dalykinių klaidų (įeina ir korektiškas simbolių vartojimas);

- sąlygoje nereikalaujama ne vertinamo dalyko žinių (pvz., kokio nors žaidimo taisyklių žinojimo);
- nepažeidžiami religiniai įsitikinimai, kultūrinis identitetas, kokios nors žmonių grupės interesai;
- naudojami grafikai ir paveikslai yra tinkamos kokybės;
- pateiktos naudojamų šaltinių nuorodos;
- klausime aiškiai nurodyta, kokio atsakymo tipo / formato tikimasi.

Kalbant apie sąlygos korektiškumą verta pastebėti, kad egzaminu arba tyrimu paprastai tikrinamos mokinių žinios ir gebėjimai, o ne pastabumas, reakcija ir pan. Todėl uždavinių / klausimų formuluotėse vengiama painiojančių išsireiškimų. Jei naudojama retesnė ar painesnė klausimų forma, pavyzdžiui, vietoje „kuris teisingas“ vartojama „kuris neteisingas“, siūloma žodį „neteisingas“ rašyti pakreiptu ar paryškintu šriftu. Norint išvengti nesusipratimų suprantant sąlygas, yra patariama paryškinti esminę klausimo vietą, pavyzdžiui, „išvardykite **tris** priežastis“, „apskaičiuokite **reikšmių** sritį“ (o ne „apibrėžimo sritį“, su kuo dažnai nebūtinai dėl nemokėjimo painiojama). Norint, kad egzaminas būtų **patikimas**, reikia mažinti atsitiktinio nepastabumo klaidas.

Jei nuo to priklausys vertinimas, sąlygos klausime būtina nurodyti, kokio atsakymo tipo / formato tikimasi, pavyzdžiui, jei norima, kad atsakymas būtų suapvalintas arba pagrįstas, arba pateikti du atsakymai ir pan., reikia tą paminėti. Jei tai nebus vertinama arba atsakymo forma nesvarbi, arba yra vertinamas būtent atsakymo formos žinojimas, tai sąlygoje nurodyti atsakymo formos nereikia.

Lygių galimybių principas

Peržiūrint užduotį, reikia įvertinti, ar klausimas yra vienodo sunkumo vaikinams ir merginoms, skirtingų religijų, kultūrų, įsitikinimų ir pomėgių žmonėms. Pasitelkiant realaus gyvenimo situacijas, kai kurių klausimų / uždavinių kontekstas gali būti šiek tiek labiau priimtinas vaikinams arba merginoms, tačiau tuomet visame teste turėtų būti vienodas skaičius klausimų, priimtinesnių vaikinams, ir klausimų, priimtinesnių merginoms; tokie klausimai turėtų būti panašaus sunkumo.

Sudėtingumas

Klausimas laikomas:

- *sudėtingu*, jei numanomas jo sunkumas yra 33 ir mažiau procentų.
- *lengvu*, jei numanomas jo sunkumas yra 67 ir daugiau procentų.
- *vidutinio* sudėtingumo, jei jo numanomas sunkumas yra tarp 34 ir 66 procentų.

Sprendžiant apie klausimo sudėtingumą galvojama apie tuos mokinius, kurie ateis laikyti konkretaus egzamino ar dalyvaus tyrime.

Kiekviename egzamine paprastai turi būti **įvairaus sudėtingumo** klausimų. Lengvų taškų (klausimų arba jų dalių) turi būti tiek, kad jais būtų galima pamatuoti vos egzaminą išlaikiusiųjų žinias ir gebėjimus. Egzaminas neturi būti taip sudarytas, kad surinkti žemiausią egzamino išlaikymo ribą būtų galima iki galo neišsprendus nė vieno uždavinio, neatsakius pilnai į joki klausimą, o taškus surinkus iš įvairių klausimų atsitiktinai arba net atspėjus. Mokiniai, kurie pasiekė tik egzamino išlaikymo ribą, taip pat turi būti pademonstravę tam tikras žinias ir gebėjimus.

Dažnai teste klausimai išdėliojami sunkumo tvarka, tačiau tai nėra absoliučiai būtina. Kiekviename uždavinių tipe (tiek atvirojo atsakymo formato, tiek uždarojo) turi būti skirtingo sunkumo klausimų. Svarbu, kad testas neprasidėtų sunkiais klausimais. Daugiausia mąstymo ir laiko reikalaujantys klausimai paprastai pateikiami testo pabaigoje. Jei įmanoma, skirtingų turinio sričių, skirtingo sąlygos formato (pavyzdžiui, pratimai ir tekstiniai uždaviniai), skirtingų gebėjimų reikalaujantys klausimai išdėliojami pakaitomis.

Pagrindinės taisyklės

Apibendrinant galima pasakyti, kad testo užduotis turi būti taip suformuluota, kad atitiktų bent šias pagrindines taisykles:

- Klausimai turi būti suformuluoti taip, kad turinys, o ne forma sąlygotų atsakymą
- Klausimų sunkumo lygis turi būti tinkamas tiems, kurie atlieka testą

- Neįmanoma tiksliai nustatyti klausimų sunkumą prieš testavimą, vis dėlto bent apytikslis sunkumo įvertinimas yra svarbus – verta klausimus teste sudėti nors iš dalies „sunkėjimo tvarka“
- Klausimai turi būti įvairaus sunkumo ir tikrinti įvairius taksonomijos lygius / gebėjimus (ne tik žinias)
- Dažniausiai verta naudoti įvairių tipų klausimus, nes vienos rūšies arba daug vienodo tipo klausimų retai geba pamatuoti įvairialypį gebėjimą
- Geriausia vieno tipo klausimus sudėti vienoje vietoje (palengvinamas vertinimas; taip pat daug lengviau egzaminuojamiesiems, nes „šokinėjimas“ nuo vieno tipo klausimų prie kito dažnai blaško)
- Nurodymai, kaip atsakinėti į klausimus, turi būti kaip galima pilnesni, glaustesni ir aiškesni
- Testo klausimai turi atitikti mokymo / egzamino programą ir užtikrinti testo validumą

Kontroliniai klausimai testo užduoties rengėjui

Parengus testą, verta dar ne kartą jį peržiūrėti ir pabandyti atsakyti į tokius klausimus „pamąstymui“:

- Kas ketinama (norima) tikrinti konkrečiu klausimu (kokios žinios, kokie gebėjimai)?
- Kas tuo klausimu iš tiesų tikrinama (kokios žinios, kokie gebėjimai)?
- Kokio (kokių) atsakymų tikėtės? Ar aiški takoskyra tarp teisingų ir klaidingų atsakymų?
- Ar galimi daliniai atsakymai į tą klausimą? Jei taip – kokie? Kaip būtų vertinami?
- Ar tikslingai pasirinktas klausimo tipas?
- Ar prasmingi pasirenkamo atsakymo klausimų distraktoriai? Kokios klaidos jais matuojamos?
- Ar skirtingi testo klausimai padengia įvairias tos srities žinias ir gebėjimus?

IV. Truputis apie vertinimą

Kai kurie reikalavimai testų, jų klausimų sudarymui yra sąlygoti vertinimo proceso. Kaip ir už ką priskirti klausimo taškus? Kaip vertinti pasirenkamo atsakymo užduotį, jei galimi keli teisingi atsakymai? Ką daryti, jei atsakymai „persidengia“? O taip pat daugybė kitų klausimų iškyla, kai pradedame vertinti užduoties atlikimą. Ugdymo proceso kontekste, klasėje, mokytojas gali priimti sprendimus, atsižvelgdamas į konkrečią situaciją, pakeisdamas savo išankstines vertinimo nuostatas, palygindamas skirtingų mokinių atsakymus, skirdamas dalinius taškus, papildomai įvertindamas netikėtą kūrybingą sprendimą ir t.t. Tačiau egzamino ar tyrimo atveju to dažniausiai negalime daryti. Todėl ir laisvės formuluojant užduotis ir galvojant apie jų vertinimą turime kur kas mažiau. Egzamino ar tyrimo užduoties atveju, net jei klausimas kitais atžvilgiais yra puikus, tačiau jo neįmanoma teisingai įvertinti arba vertinimas yra labai komplikotas, dažniausiai jo reikia atsisakyti arba kitaip formuluoti. Todėl, rengiant egzamino arba tyrimo užduotį, formuluojant jos klausimus, verta iš karto galvoti, kaip jis bus vertinamas. Tuo pačiu, vertinant egzamino ar tyrimo užduoties turinį, būtina žiūrėti ir į vertinimo instrukciją – neatsiejamą užduoties dalį. Vertinimo instrukcijose paprastai yra pateikiamas galimas teisingas sprendimas, atsakymas ir jo vertinimo normos – kiek taškų ir už ką yra skiriama. Trumpai aptarsime šiuos aspektus.

Sprendimo / atsakymo aiškumas

Būtina peržvelgti, ar sprendime nėra per didelių „šuočių“, kurie apsunkina sprendimo eigos supratimą. Atsakymuose / sprendimuose neturi būti nei dalykinių (bet kokio dalyko), nei faktinių klaidų, nei subjektyvių teiginių. Sprendimas / atsakymas aprašomas taip, kad jį vertinant kiek įmanoma būtų galima išvengti subjektyvumo.

Taškų paskirstymo racionalumas

Paprastai taškas skiriamas už „žingsnį“, t. y., pademonstruotą žinojimą arba gebėjimą. Dažniausiai „žingsnis“ vertinamas vienu tašku, nepriklausomai nuo to, kiek jis sudėtingas. Jei numatoma, kad gali pasitaikyti sprendimo atvejai, kuriems būtų sudėtinga pritaikyti pateiktą sprendimo ir taškų paskirstymo schemą, tokiu atveju pateikiami keli skirtingi būdai su jų taškų paskirstymu. Jei įvairiems sprendimo būdams taškų paskirstymas yra analogiškas arba lengvai priderinamas, pateikti kelių sprendimo būdų nereikia. Numatomus keblius atvejus būtina komentuoti pastabose.

V. Statistinė testo analizė - kiekybinė testo kokybės patikra

Kam reikalinga statistinė analizė?

Kuriant standartizuotus testus, dažniausiai neapsieinama be atskirų jų klausimų ir visų testų pilotavimo. Šių testų bandymų tikslas paprastai yra ištirti:

- kaip veikia testo klausimai,
- ar nėra nekorektiškų sąlygų,
- ar jie nėra per lengvi arba per sunkūs,
- ar tinkamai parinkti pasirenkamo atsakymo klausimų distraktoriai,
- ar tinkamai atskiria gerai mokančius nuo silpnųjų,
- ar visas testas matuoja patikimai ir t.t.

Jei testuojamųjų imtis pakankamai didelė, į šiuos klausimus padeda atsakyti statistinė testo analizė. Be abejo, dažnai ji gali tik parodyti, kur gali slypėti pavojai ar problemos, o galutinį sprendimą galima priimti tik matant konkrečią užduotį to dalyko specialisto akimis. Vis dėlto ji pateikia labai svarbią informaciją, kuri gali padėti užtikrinti testo kokybę.

Egzamino užduoties atveju paprastai nebūna galimybės iš anksto išbandyti užduoties klausimus su pakankamai didele imtimi, todėl po egzamino atliekama statistinė testo analizė gali atrodyti nelabai reikalinga, nes, atvirai kalbant, „šaukštai jau po pietų“. Vis dėlto kuriant tęstinę užduočių kokybės sistemą, nagrinėjant egzamino užduoties turinio pagrįstumą, statistinės analizės teikiama informacija yra labai svarbi.

Tuo tarpu visuose rimtuose statistiniuose švietimo tyrimuose, ypač tarptautiniuose, bandomasis tyrimas yra esminė kiekvieno tyrimo ciklo dalimi. Paprastai išbandoma apie dukart daugiau klausimų, kurių dalis po bandomojo tyrimo, remiantis statistinės testų analizės duomenimis, yra keičiama, adaptuojama arba išvis atsisakoma.

Testo klausimų statistinė analizė

Kiekvienam testo klausimui paprastai nustatoma visa eilė statistikų, kurių kiekviena pateikia tam tikrą informaciją apie to klausimo kokybę.

1. kuri dalis (procentais) respondentų pasirinko atitinkamą atsakymą (A, B, C ir t.t., jei klausimas buvo su pasirenkamaisiais atsakymais) ar surinko atitinkamą skaičių taškų (0, 1, 2 ir t. t.).

Šia informacija ypač naudinga remtis nagrinėjant pasirenkamo atsakymo klausimų kokybę. Kokia dalis atsakinėjusiųjų pasirinko vieną ar kitą pateiktą atsakymą? Kokie distraktoriai buvo ypač patrauklūs, o kurių niekas nesirinko? Kodėl?

2. klausimo sunkumas

Šio parametro skaitinė reikšmė yra santykis:

$$\frac{\text{(visų respondentų už šį klausimą surinktų taškų suma)}}{\text{(visų už šį klausimą teoriškai galimų surinkti taškų suma)}}$$

Jei klausimas buvo vertinamas vienu tašku, tai jo sunkumas tiesiogiai parodo, kuri dalis kandidatų į tą klausimą atsakė teisingai. Kitu atveju jis demonstruoja, kokią dalį visų galimų už tą klausimą surinkti taškų iš tiesų surinko į jį atsakinėjusieji. Kuo didesnė ta dalis – tuo iš tiesų lengvesnis buvo klausimas, kuo mažesnė – tuo jis buvo sunkesnis. Be abejo, realioje egzamino užduotyje visi klausimai negali būti vidutinio („optimalaus“) sunkumo. Vieni būtinai turi būti lengvesni, kiti – sunkesni. Todėl nagrinėjant klausimų sunkumą, verta atkreipti dėmesį į lengvų, vidutinių ir sunkių klausimų proporcijas, o taip pat į tuos klausimus, kurie buvo netikėtai ženkliai lengvesni ar sunkesni nei buvo tikėtasi. Pavyzdžiui, jei buvo manoma, kad kažkoks klausimas bus gana lengvas, o jis pasirodė esąs sunkus – ar tai tik todėl, kad mokinių žinios yra gerokai mažesnės nei tikėtasi? O gal klausimo formuluotėje buvo netikslumų, gal gremėzdiška klausimo forma sutrikdė atsakinėjusiuosius? Arba atsakymai buvo ne visai tikslūs? Iš kitos pusės, jei į klausimą, kuris atrodė gana sunkus, teisingai atsakė daug mokinių, vėlgi reikėtų ne iš karto pulti džiaugtis, kad mokinių žinios tokios plačios, o atidžiai pasižiūrėti, ar klausimo formuluotėje nebuvo užuominų? O gal distraktoriai buvo visiškai neįtikėtini, todėl bent kiek maštantis mokinys turėjo pasirinkti teisingą atsakymą?..

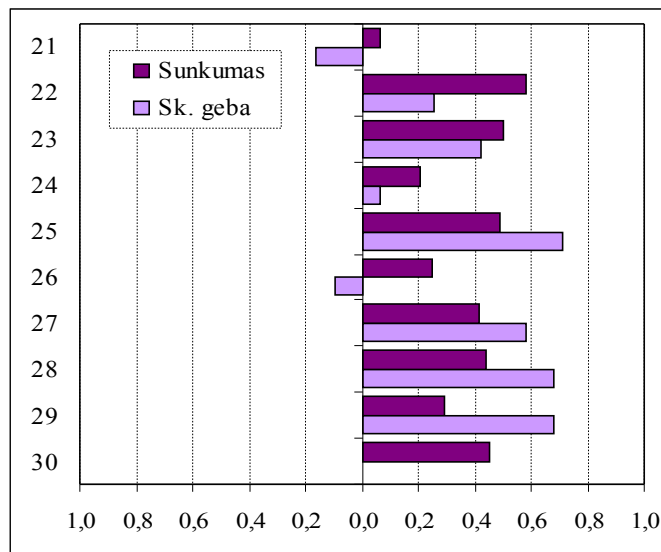
3. klausimo skiriamoji geba

Šis parametras rodo, kaip atskiras testo klausimas išskiria stipresnius ir silpnesnius respondentus. Ji paprastai skaičiuojama imant geriausiai ir blogiausiai pasirodžiusių viso testo atžvilgiu respondentų dalių (pavyzdžiui, 27 proc.) to klausimo sunkumus ir skaičiuojant jų skirtumus. Jeigu gerai atsakinėjusieji labai gerai atsakė į tą klausimą, o blogai atsakinėjusieji – labai blogai, tai šis skirtumas bus didelis. Jeigu ir „stiprieji“, ir „silpnieji“ panašiai į jį atsakinėjo, tai skirtumas tarp tų dviejų sunkumų, aišku, bus nedidelis. Be abejo, labai lengvų ir labai sunkių klausimų skiriamoji geba yra maža, nes į juos arba beveik visi atsako, arba beveik visi neatsako, jie negali labai gerai atskirti vieno nuo kito. Tačiau jei kalbame apie vidutinio sunkumo klausimus, tai paprastai gerai atskiriančiais laikomi tie, kurių skiriamoji geba yra 0,4-0,5, labai gerai – 0,6 ir daugiau. Neigiama skiriamosios gebos reikšmė rodo, kad silpnesnieji (sprendžiant pagal visą testą) už tą klausimą surinko daugiau taškų nei stipresnieji. Tai labai stiprus indikatorius, kad klausimas greičiausiai nėra korektiškas ir reikia ieškoti klaidos klausimo, pateiktų pasirenkamų atsakymų formuluotėse arba vertinimo instrukcijoje. Pavyzdžiui, gali būti, kad tiesiog ne tas atsakymas buvo koduojamas kaip teisingas, o gal pats klausimas buvo suformuluotas taip, kad jį klaidingai suprato stipresnieji mokiniai. Bet koku atveju, akivaizdu, kad toks klausimas turi kažkokių problemų.

4. klausimo koreliacija su visu testu

Tai to klausimo ir viso testo taškų koreliacijos koeficientas. Šis parametras rodo, kuria dalimi atskiras klausimas matuoja taip, kaip ir visa užduotis. Natūraliai, daugiataškio klausimo koreliacija su visu testu yra didesnė nei vienataškio. Tačiau apskritai jei kurio nors klausimo koreliacija yra žema, tai rodo, kad tas klausimas matavo skirtingai nei visas testas, todėl vėlgi verta pasitikrinti, ar jame nėra kokių nors nekorektiškumų, arba galbūt jis išties yra visai kitoks ir matuoja iš esmės kitokius gebėjimus nei likę testo klausimai (pavyzdžiui, ne matematinius, o skaitymo ar rašymo gebėjimus; arba ne istorines žinias, o gebėjimą suskaičiuoti ir pan.).

Aptartas sąvokas iliustruosime grafiškai pateikta informacija apie vieno nedidelio testo sunkumą ir skiriamą gebą.

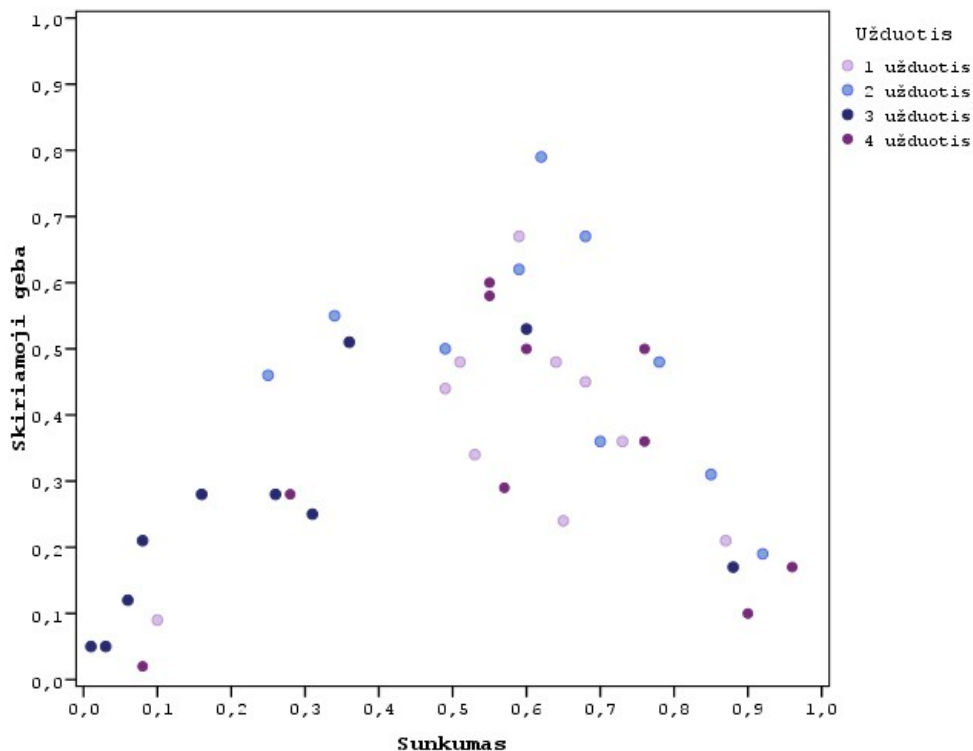


5.1. pav.

Paveikslėlyje 5.1. pavaizduotame pavyzdyje matome, kad 4 iš 10 testo dalies klausimų buvo gana sunkūs, o likę buvo vidutinio sunkumo – svyravo apie 40 proc. Labai lengvų klausimų šioje testo dalyje nebuvo. Taip pat matome, kad, nors 23 ir 25 klausimo sunkumas buvo praktiškai vienodas, tačiau jų skiriamoji geba ženkliai skiriasi – 25 klausimas kur kas ryškiau atskiria stipresnius mokinius nuo silpnesniųjų. Kai kurių klausimų (pavyzdžiui, 24-ojo) skiriamoji geba labai maža, o 30-ojo – lygi 0. Tai reiškia, kad šis klausimas praktiškai nesuteikia jokios informacijos, nes niekaip neatskiria geriau mokančiųjų nuo blogai mokančiųjų, toms abiem grupėms klausimo sunkumas yra toks pat, galima įtarti, kad teisingai į jį atsakinėjo daugmaž atsitiktinai. Dar rimtesnė situacija su 21 ir 26 klausimais, nes jų skiriamoji geba yra neigiama. Tai reiškia, kad turime atidžiai peržvelgti tuos klausimus – kažkodėl juos geriau atsakinėjo prasčiau visą testą išsprendusieji.

5. Testo klausimų sunkumo ir skiriamosios gebos priklausomybė

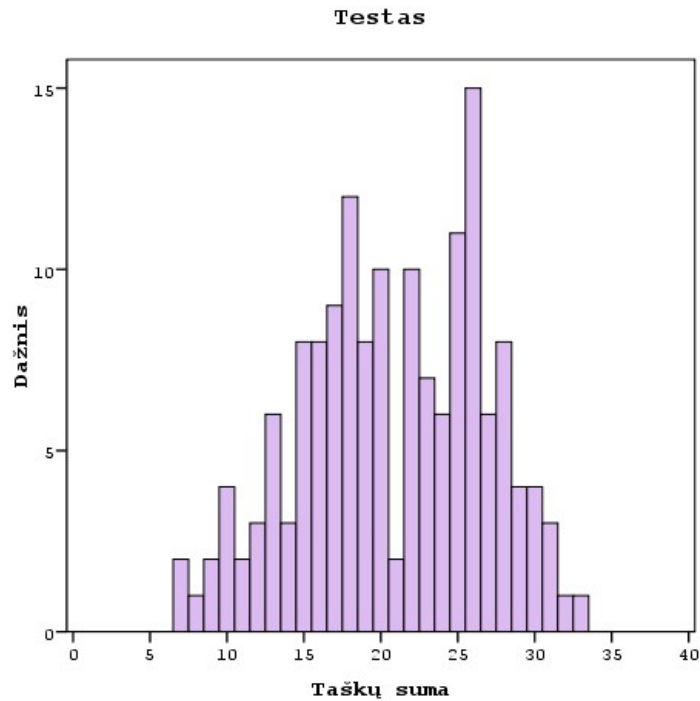
Pateiktame 5.2. grafike matome ryšį tarp klausimų sunkumo ir skiriamosios gebos. Akivaizdu, kad trečiosios testo dalies klausimai buvo pernelyg sunkūs respondentų grupei – jų ir sunkumas, ir skiriamoji geba yra žemi.



5.2. pav.

Surinktų taškų analizė

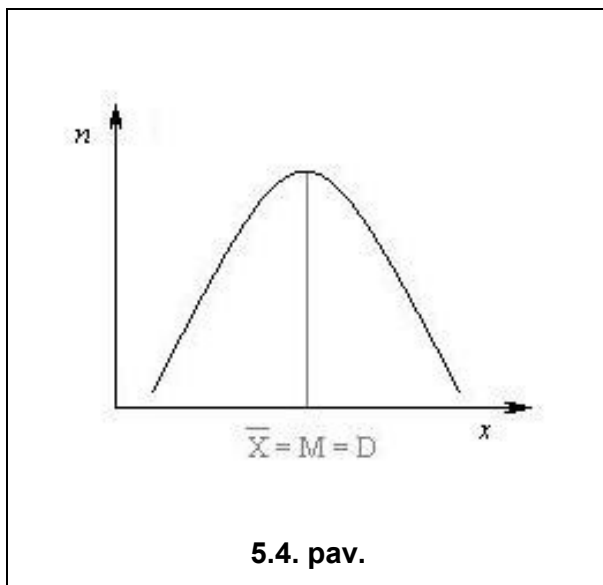
Nagrinėjant egzamino užduoties ar bet kokio kito testo kokybę, remiantis statistiniais kriterijais, paprastai būna svarbūs ne tik atskirų klausimų rodikliai, bet ir visos užduoties suderinamumas, jos visos kaip matavimo vieneto rodikliai. Paprasčiausia informacija, kuri labai greitai po egzaminų būna pateikiama visuomenei, yra susijusi su už egzaminą surinktų taškų pasiskirstymu. Žemiau pateikiamas vienas tokios informacijos pavyzdys (žr. 5.3. pav.).



5.3. pav.

Paprastai besidomintys egzamino rezultatais ieško „varpo formos“ grafiko. Jis demonstruoja panašų į normalųjį rezultatų pasiskirstymą. Kaip interpretuoti skirtingos formos grafikus? Ką reiškia, jei varpo formos grafikas pasislinkęs į dešinę ar kairę puses? Trumpai aptarsime šių pasiskirstymo diagramų formas.

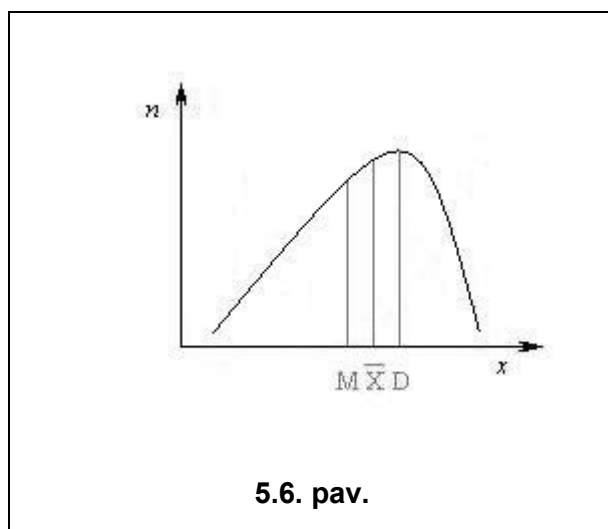
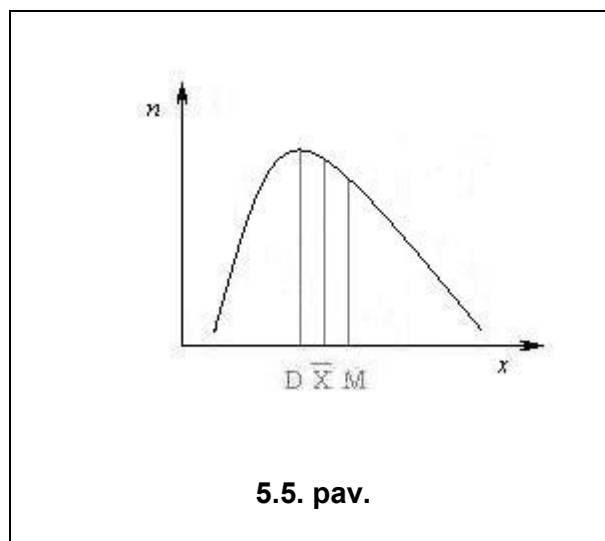
Visuose pateiktuose paveikslėliuose x ašyje yra atidėti galimi surinkti taškai, o y ašyje - skaičius (n) surinkusių atitinkamą taškų kiekį.



5.4. pav.

5.4. paveikslėlyje turime „visiškai normalų“, idealiai vidutinį atvejį – realiai surinktų taškų vidurkis (\bar{X}), taškų mediana (arba, galime sakyti, galimų surinkti taškų vidutinė reikšmė, M), ir moda – taškų suma, kurią gavo daugiausiai respondentų (D), sutampa. Tai reiškia, kad testas idealiai atitiko populiaciją (aišku, jei testo tikslas buvo vidutiniškai gerai pamatuoti ir išdėlioti visus respondentus).

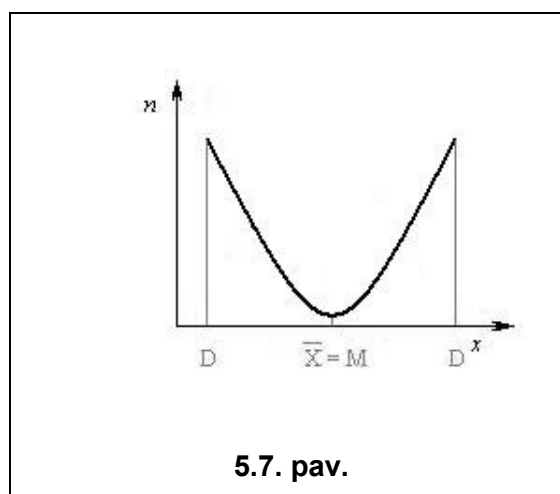
5.5. paveikslėlyje, nors varpo forma tarsi išlaikyta, tačiau visa diagrama (grafikas) tarsi pasislinkę į kairę. Šiuo atveju taškų suma, kurią surinko daugiausiai respondentų, yra kairiau nuo medianos (vidutinių galimų surinkti taškų). Realiai surinktų taškų vidurkis šiuo atveju irgi yra pasislinkęs į kairę, yra mažesnis nei mediana. Tai reiškia, kad egzamino užduotis buvo sunkesnė nei vidutinė. Tokia egzamino užduotis yra tikslinga, kai reikia tiksliau išskirstyti geresnius mokinius, tinka, kai reikia atrinkti nedidelę dalį geriausiųjų.



5.6. paveikslėlyje matome lengvesnio nei vidutinis egzamino rezultatus. Šiuo atveju moda (taškų suma, kurią surinko daugiausiai respondentų), o taip pat realiai surinktų taškų vidurkis yra pasislinkę į dešinę. Tokia egzamino užduotis geriau išskiria, pamatuoja žemesnių gebėjimų respondentus, yra ypač tinkama, kai reikia patvirtinti kažkokį

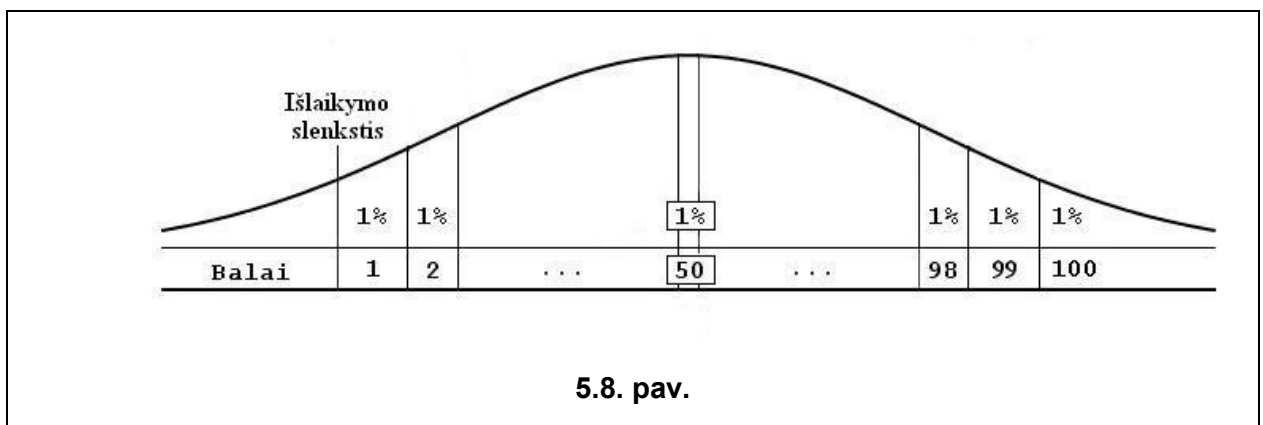
minimalų išlaikymo lygį, tačiau ne taip gerai diferencijuoja didesnių gebėjimų respondentus.

5.7. paveikslėlyje pavaizduota taškų pasiskirstymo diagramos forma nėra tipinė egzamino užduotims, tačiau išskirtiniais atvejais galėtų pasitaikyti. Šiuo atveju mediana (visų taškų vidutinė reikšmė) ir realiai surinktų taškų vidurkis idealiai sutampa, tačiau pats varpas yra tarsi



apsivertęs aukštyn kojom. Modos šiuo atveju yra dvi ir priešingose pusėse. Tai reiškia, kad daugiausiai respondentų pasiskirstė tarp dviejų kraštutinumų – arba surinko labai mažai taškų, arba surinko labai daug taškų. Vidutinių rezultatų šiuo atveju beveik nebuvo. Čia daugiausiai informacijos (didžiausias atskyrimas, diferencijavimas) yra būtent apie vidutinius taškus. Toks testas būtų prasmingas tuo atveju, jei, pavyzdžiui, reikėtų atskirti lygiai pusę respondentų. Tada turėtų būti labai aiški riba, kurią arba išlaikytų (ir tada surinktų daug taškų), arba negalėtų jos pasiekti (ir tada surinktų labai mažai taškų).

Galiausiai su statistika ir taškų pasiskirstymu yra susijęs ir norminis valstybinių egzaminų vertinimas. Šiuo atveju visi respondentai išdėliojami surinktų taškų mažėjimo tvarka. Fiksuojama minimali taškų suma, kurią surinkus egzaminas laikomas išlaikytu („Išlaikymo slenkstis“). Tada tarp išlaikiusiųjų egzaminą pirmasis procentas surinkusiųjų daugiausiai taškų gauna 100 balų įvertinimą, antrasis – 99 balų įvertinimą ir t.t. (žr. 5.8. pav.). Priklausomai nuo egzamino tikslų išlaikymo slenkstis gali būti mažesnis ar didesnis, o pati užduotis kuriama taip, kad vidurkis būtų pasislinkęs į kairę arba dešinę.



Patikimumas

Jau pačioje leidinio pradžioje minėjome, kad viena iš svarbių testo užduoties savybių yra jo patikimumas. Patikimumas rodo, kiek testo rezultatai yra pastovūs, tai yra nepriklauso nuo kažkokių atsitiktinių aplinkybių. Jei tokio pastovumo nėra, tada negalime kalbėti ir apie testo validumą. Paprastai pagal atsakymus į atskirus testo klausimus apskaičiuojama:

- testo atskirų užduočių patikimumo statistinis indikatorius - Kronbacho alfa,
- Kronbacho alfa be klausimo, kuri parodo užduoties patikimumą, išmetus iš jos atitinkamą klausimą.
- Be abejo, jeigu užduoties patikimumas „išmetus“ klausimą padidėja, tai reiškia, kad tas klausimas išsiskiria iš kitų atitinkamos užduoties klausimų. Tai gali būti indikatorius, kad jį reikia peržiūrėti.
- Užduotis laikoma patikima, jei jos bendras patikimumo rodiklis ne mažesnis nei 0,5.

Pateiktoje lentelėje (žr. 5.9. pav.) matome, kad 3-iosios dalies patikimumas yra gana žemas, tai vėlgi yra indikatorius, kad tos dalies klausimuose gali būti problemų.

Dalis	Kronbacho alfa
1 dalis	0,61
2 dalis	0,65
3 dalis	0,41
4 dalis	0,77

5.9. pav.

Klausimas	Kronbacho alfa be klausimo
1	0,62
2	0,58
3	0,57
4	0,60
5	0,51
6	0,52
7	0,60
8	0,47
9	0,72
10	0,59

5.10. pav.

Antroje lentelėje (5.10 pav.) matome, kad 1 ir 9 klausimai visos testo I dalies Kronbacho alfa sumažina, tai yra Kronbacho alfa išmetus tuos klausimus tik padidėtų lyginant su visos testo dalies Kronbacho alfa (0,61). Tai rodo, kad šie klausimai nėra gerai suderinti su likusiais ir gali būti, kad juose yra problemų. Todėl tokiu atveju būtų gerai juos peržiūrėti.

Literatūra

1. Aiken L.R., Tests and Examinations: Measuring Abilities and Performance. John Wiley & Sons, 1998.
2. Crocker L., Algina J., Introduction to Classical and Modern Test Theory, Wadsworth Pub Co, 2006.
3. Downing S. M., Haladyna T. M. (ed.), Handbook of Test Development, Lawrence Erlbaum Assoc., 2006.
4. Foy P., Kennedy A. M., PIRLS 2006 User Guide for the International Database. Boston College, 2008.
5. Haladyna T. M., Developing and Validating Multiple-Choice Test Items, Lawrence Erlbaum Assoc., 1999.
6. Haladyna T. M., Writing Test Items to Evaluate Higher Order Thinking, Allyn&Bacon, Boston, 1997.
7. Martin M.O., Mullis I.V.S., Chrostowski S. J., TIMSS 2003 Technical Report, Boston College, 2004.
8. Weiss C.H., Vertinimas: programų ir veiklos krypčių tyrimo metodai. Homo Liber, Vilnius, 2006.