

ՀՀ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԶԳԱՅԻՆ ԱԿԱԴԵՄԻԱՅԻ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏ

ԱՎԵՏԻՍՅԱՆ ՎԱՐԱԶԴԱՏ ԿՈՐՅՈՒՆԻ

ԳԻՏԵԼԻՔՆԵՐԻ ՍՏՈՒԳՄԱՆ ԹԵՍԵՐԻ ՈՐԱԿԻ ԳՆԱՀԱՏՄԱՆ ՄԵԹՈԴԻ ԵՎ ԾՐԱԳՐԱՅԻՆ ՀԱՄԱԿԱՐԳԻ ՄՇԱԿՈՒՄ

Ե.13.04 – «Հաշվողական մեքենաների, համալիրների, համակարգերի և ցանցերի մաթեմատիկական և ծրագրային ապահովում» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման

ԱՏԵՆԱԽՈՍՈՒԹՅՈՒՆ

Գիտական ղեկավար
Ֆիզմաթ. գիտ. դոկտոր, պրոֆեսոր՝
Մ.Ե. Հարությունյան

ԲՈՎԱՆԴԱԿՈՒԹՅՈՒՆ

ՆԵՐԱԾՈՒԹՅՈՒՆ.....	4
ԳԼՈՒԽ I. ԹԵՍՏԵՐԻ ՈՐԱԿԻ ՎԵՐԼՈՒԾՈՒԹՅԱՆ ՏԵՍՈՒԹՅՈՒՆՆԵՐԸ.....	10
1.1. Թեստերի տեսակները և վերլուծության կազմակերպման սկզբունքները	10
1.2. Թեստերի դասական տեսությունը	17
1.3. Թեստերի մաթեմատիկական տեսությունը.....	29
1.4. Եզրակացություններ.....	44
ԳԼՈՒԽ II. ԱՌԿԱ ՀԱՄԱԿԱՐԳԵՐԻ ՀԵՏԱԶՈՏՈՒԹՅՈՒՆ.....	46
2.1. Թեստավորման արդյունքների վերլուծության ժամանակակից առավել լայն կիրառվող ծրագրերի հետազոտություն	46
2.2. Եզրակացություններ.....	54
ԳԼՈՒԽ III. ԹԵՍՏԵՐԻ ՈՐԱԿԻ ԳՆԱՀԱՏՄԱՆ ՆՈՐ ՄՈԴԵԼԻ ՄՇԱԿՈՒՄ.....	58
3.1. Շենոնի ինֆորմացիայի չափի մեծությունները և դրանց հատկությունները.....	58
3.2. Թեստերի որակի գնահատման նոր մոդելի նկարագրությունը.....	61
3.3. Եզրակացություններ.....	65
ԳԼՈՒԽ IV. ԹԵՍՏԵՐԻ ՈՐԱԿԻ ԳՆԱՀԱՏՄԱՆ ԳՈՐԾԻՔԱՅԻՆ ԾՐԱԳՐԱՅԻՆ ՀԱՄԱԿԱՐԳԻ ՄՇԱԿՈՒՄ	66
4.1. «Թեստերի որակի վերլուծություն» համակարգի հնարավորությունները	67
4.2. Համակարգի կառուցվածքը.....	68
4.3. Օգտվողի ինտերֆեյսը	70
4.4. Որակի վերլուծության և գնահատման մոդուլը.....	79
4.5. Արդյունքների մշակման և հաշվետվության ձևավորման մոդուլը.....	81
4.6. Նոր տեխնոլոգիական լուծումները	86
4.7. Եզրակացություններ	94
ԳԼՈՒԽ V. ՆՈՐ ՄԵԹՈԴԻ ԲՆՈՒԹԱԳՐԻՉՆԵՐԻ ԸՆԴՈՒՆԵԼԻ ԱՐԺԵՔՆԵՐԻ ՈՐՈՇՈՒՄ.....	95

5.1. Թեստերի որակի գնահատման նոր մոդելի փորձարկումների արդյունքները.....	95
5.2. Եզրակացություններ.....	104
ԵԶՐԱԿԱՑՈՒԹՅՈՒՆՆԵՐ.....	106
ՕԳՏԱԳՈՐԾԱԾ ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ	107

ՆԵՐԱԾՈՒԹՅՈՒՆ

Թեմայի արդիականությունը

Համաշխարհային փորձը վկայում է գիտելիքի ստուգման համար կիրառվող քննական թեստերի արդյունավետության մասին: Թեստերը օգտագործվում են (ինքնա)ուսուցման և (ինքնա)ստուգման համար, թույլ են տալիս ստանալ գիտելիքների մակարդակի բավականին օբյեկտիվ գնահատական, բացահայտել ուսուցման բացթողումները: Ուսումնական հաստատություններում մանկավարժական թեստերը օգտագործվում են ընդունելության քննությունների, գիտելիքների ընթացիկ ստուգման համար: Հեռուսուցման ժամանակ լայն կիրառվում են թեստավորման համակարգչային տեխնոլոգիաները:

Համակարգչային թեստավորման ձևով քննությունը թույլ է տալիս ստուգել գիտելիքներն առարկայի նյութի մեծ մասով, ապահովում է գնահատականի օբյեկտիվությունը, ազատում է դասավանդողին գրավոր աշխատանքների աշխատատար ստուգումներից:

Կարևոր է նշել, որ թեստավորման և գիտելիքների գնահատման օբյեկտիվությունը մեծապես կախված է թեստերի բովանդակությունից և որակից: Ներկայումս շատ դեպքերում կիրառվում են թեստեր, որոնք չեն ենթարկվում վերլուծության, որակի գնահատման և փորձաքննության, բնութագրվում են հուսալիության և վավերության ցածր, որոշ դեպքերում՝ անբավարար ցուցանիշներով, որի արդյունքում էլ չի ապահովվում գիտելիքների ստուգման օբյեկտիվությունը, արդյունքների համապատասխանությունը: Այդ իսկ պատճառով թեստերի և թեստային առաջադրանքների որակական բնութագրերի վերլուծության և գնահատման նոր արդյունավետ գործիքների մշակումը արդիական գիտական խնդիր է:

Վերջին տարիներին Հայաստանի Հանրապետությունում ևս լայն կիրառություն են գտնում թեստավորման տեխնոլոգիաները: Դպրոցական ավարտական և ԲՈՒՀ-երի ընդունելության միասնական քննությունները կազմակերպվում են թեստավորման եղանակով՝ ՀՀ կառավարության աշխատակազմի «Գնահատման և թեստավորման

կենտրոն» ՊՈԱԿ-ի կողմից: Համակարգչային թեստավորման եղանակով քննություններ են կազմակերպվում ԲՈՒՀ-երում, ինչպես նաև հետբուհական կրթության ոլորտում, օրինակ ասպիրանտուրայի ընդունելության հնֆորմատիկայի ստուգարքը, անգլերենի «TOEFL» քննությունը, որակավորման հնֆորմատիկայի ստուգարքը:

Թեստի որակի գնահատումը կարող է իրականացվել վիճակագրական եղանակով, փորձնական թեստավորման միջոցով ստացված արդյունքների վերլուծության հիման վրա: Վիճակագրական եղանակով թեստի որակը գնահատելու համար գոյություն ունեն տեսություններ, որոնցից առավել լայն կիրառություն են գտել թեստերի դասական տեսությունը (Classical Test Theory-CTT) և թեստերի ժամանակակից մաթեմատիկական տեսությունը (Item Resonse Theory-IRT): Երկու տեսություններում էլ սահմանվում են թեստի և թեստային առաջադրանքների որակական ցուցանիշները, դրանց հաշվարկման, գնահատման մեթոդները և մաթեմատիկական վիճակագրական ապարատը: Հարկ է նշել, որ վիճակագրական ապարատը բավականին բարդ է տարբեր բնագավառների մանկավարժների համար, իսկ հաշվարկները աշխատատար են նեղ մասնագետի համար: Այդ խնդիրը լուծելու համար մշակվում են համակարգեր, ծրագրային փաթեթներ, որոնց միջոցով իրականացվում է փորձնական թեստավորման արդյունքների վերլուծություն և թեստի որակական բնութագրերի գնահատում:

Ներկայումս գոյություն ունեն թեստերի որակական վերլուծության բազմաթիվ ծրագրային փաթեթներ և համակարգեր, որոնցից առավել լայնորեն կիրառվում են՝ CITAS, Iteman 4, Xcalibre 4, Winsteps, Facets, jMetrik, RUMM 2030, ACER ConQuest, IRTPRO, ConstructMap, Bigsteps, GENIRV, RESCEN: Թվարկվածներից յուրաքանչյուրն ունի իրեն բնորոշ հատկանիշներ, հնարավորություններ, առավելություններ, թերություններ և նախապատվություն՝ կախված կիրառական նշանակությունից, ոլորտից և բնույթից: Հիմնական թերություններից կարելի է առանձնացնել այն, որ նմանատիպ համակարգերը չունեն օգտվողին կողմնորոշված պարզ ինտերֆեյս, երբեմն աշխատելու համար պահանջվում են տեխնիկական խորը գիտելիքներ, հրամանային տողով աշխատելու հմտություններ: IRT –ով վերլուծություն

անցկացնող ծրագրերը բազմագործառության են և կիրառվում են տարբեր բնագավառներում չափումներ գնահատելու համար, իսկ միայն թեստավորման վերաբերյալ վերլուծություն անցկացնելու համար անհրաժեշտ է ծրագրից գտնել, առնաձնացնել և խմբավորել միայն թեստավորման մոդուլները, որն այդքան էլ դյուրին գործ չէ: Թերություն կարելի է համարել նաև այն, որ թեստավորման արդյուքների վերլուծությունը հիմնականում ստացվում է բազմաթիվ աղյուսակների տեսքով, որոնք պահպանվում են TXT ձևաչափով: Գրաֆիկներն էլ իրենց հերթին ստացվում են առանձին ֆայլերի տեսքով՝ JPG կամ PNG ձևաչափով: Այսինքն մեկ միասնական ֆայլի տեսքով հաշվետվություն ստանալու համար անհրաժեշտ է լինում տարբեր ֆայլերից կատարել խմբագրումներ և ստանալ նոր հաշվետվություն, ինչը հարմար և կիրառելի չէ մանկավարժի համար: Բացի այդ, գոյություն ունեցող ծրագրերը հիմնականում անգլերեն լեզվով են, վճարովի են և կապված բազմագործառությանության հետ ունեն բարձր արժեքներ:

Այս իրավիճակում արդիական էր մշակել թեստերի որակի գնահատման ավելի պարզ մեթոդ և հայերեն լեզվով ծրագրային համակարգ:

Հետազոտության նպատակը

Թեստերի որակի ստուգման, գնահատման գործընթացի կազմակերպման համար խնդիր էր դրվել մշակել նոր եղանակ և ծրագրային իրականացում, որը կապահովի բազմակողմանի վերլուծություն հայերեն լեզվի սատարմամբ:

Հետազոտման օբյեկտը

Ուսումնասիրության օբյեկտներն են թեստերի որակի վերլուծության և գնահատման համակարգերն իրականացնող տեխնոլոգիաները, մասնավորապես, Web Server-ի վրա աշխատանքի կազմակերպման սկզբունքները, Java ծրագրավորման լեզվով Microsoft Office փաստաթղթերի ստեղծման և ձևավորման ֆունկցիաները, Java և PHP ծրագրավորման լեզուներով CSV, TXT ֆայլերի հետ աշխատելու միջոցները, PHP ծրագրավորման լեզվով արտաքին ծրագրերի գործարկման հնարավորությունները:

Հետազոտման մեթոդները

Ատենախոսության մեջ կիրառվում են վիճակագրական ու ինֆորմացիայի տեսության մեթոդներ, CTT և IRT տեսությունների սկզբունքներ: Օգտագործվում են Java psychometrics բաց ծրագրային կոդով գրադարանի CTT, IRT ցուցանիշների հաշվարկման մաթեմատիկական և վիճակագրական մեթոդները:

Հետազոտության գիտական նորույթը

1. Մշակվել է թեստերի որակի գնահատման նոր մոտեցում՝ հիմնված ինֆորմացիայի տեսության տարրերի վրա, որն ապահովում է ավելի արագ հաշվարկներ և վավեր վերլուծության իրականացման համար պահանջում է շատ ավելի քիչ թեստավորման արդյունքներ:
2. Մշակված թեստերի որակի գնահատման հայալեզու սատարմամբ ծրագրային համակարգը գործարկվում է սերվերի վրա և ունի web հենքով օգտվողի ինտերֆեյս, ինչն էլ թույլ է տալիս դիմել համակարգին լոկալ ցանցի կամ համացանցի միջոցով՝ գործարկելով այն web զննարկիչի միջավայրում:
3. Համակարգի առանձին բաղադրիչ հանդիսացող վիճակագրական և մաթեմատիկական ցուցանիշները հաշվարկող մոդուլը կարող է կիրառվել նաև առանձին համակարգի վրա որպես desktop հավելված:
4. Մշակված համակարգը կախված չէ սերվերի օպերացիոն համակարգի տեսակից:

Ստացված արդյունքների կիրառական նշանակությունը

Մշակված պարզ և մատչելի web ինտերֆեյսով համակարգն օգտվողներին հնարավորություն է տալիս առանց տիրապետելու թեստերի վիճակագրական տեսության բարդ ապարատին և առանց տեղադրելու լրացուցիչ ծրագրային միջոցներ, կազմակերպել մշակված թեստերի որակի վերլուծություն և գնահատում:

Համակարգի միջոցով գեներացվում է թեստերի որակի վերլուծության և գնահատման հաշվետվություն, որը խմբագրվող rtf (rich text format) ձևաչափի ֆայլ է և բովանդակում է ինչպես հաշվարկվող և գնահատվող ցուցանիշները, աղյուսակները, գրաֆիկները, այնպես էլ նկարագրություններ և ցուցումներ դրանց վերաբերյալ:

Համակարգը կարող է կիրառվել թեստերի մշակամաթերադվող տարբեր մասնագիտությունների մանկավարժների և թեստավորման անցկացմամբ զբաղվող կազմակերպությունների (ուսումնական հաստատություններ, թեստավորման կենտրոններ և այլն) կողմից՝ թեստավորման արդյունքների վերլուծության և թեստերի որակի գնահատման նպատակով:

Ներդրումներ

Մշակված համակարգը ներդրվել է Գավառի պետական համալսարանի կողմից իրականացվող համակարգչային թեստավորումով քննական գործընթացի մեջ: Ներդրված համակարգի միջոցով վերլուծվում են թեստավորման արդյունքները և գնահատվում է թեստերի որակը:

Պաշտպանության ներկայացվող հիմնական դրույթները

- Առաջարկվել է թեստերի որակի գնահատման նոր մեթոդ՝ հիմնված ինֆորմացիայի տեսության տարրերի վրա, որն ապահովում է ավելի արագ հաշվարկներ:
- Որոշվել են առաջարկված մեթոդի ցուցանիշների ընդունելի սահմանները:
- Գնահատվել է վավեր վերլուծության համար անհրաժեշտ թեստավորման արդյունքների քանակը, որը էականորեն ավելի փոքր է մյուս մեթոդների հետ համեմատած:
- Մշակվել է CTT, IRT տեսությունների և առաջարկված նոր մեթոդի հիման վրա թեստերի որակի վերլուծություն և գնահատում իրականացնող երկեզուծրագրային համակարգ, որը գործարկվում է սերվերի վրա և ունի web հենքով օգտվողի ինտերֆեյս, ինչն էլ թույլ է տալիս դիմել համակարգին լոկալ ցանցի կամ համացանցի միջոցով՝ գործարկելով այն web զննարկիչի միջավայրում: Համակարգի առանձին բաղադրիչ հանդիսացող վիճակագրական և մաթեմատիկական ցուցանիշները հաշվարկող մոդուլը կարող է կիրառվել նաև առանձին համակարգչի վրա որպես desktop հավելված: Մշակված համակարգը կախված չէ սերվերի օպերացիոն համակարգի

տեսակից:

Ապրոբացիա

Ատենախոսության արդյունքները զեկուցվել են՝

- ITA 2015 – ITHEA ISS Joint International Events on Informatics Summer Session, Varna, Bulgaria, June 29 - July 12, 2015,
- «Համակարգչային գիտություններ և տեղեկատվական տեխնոլոգիաներ (CSIT 2015)» 10-րդ գիտաժողովում, Երևան, սեպտ.28- հոկտ. 2, 2015,
- «Բնագիտական խնդիրներ հետազոտական էլեկտրոնային ենթակառուցվածքի միջավայրում» գիտաժողովում, Երևան, դեկտեմբերի 18-20, 2015,
- Գնահատման և թեստավորման կենտրոնի համակարգչային և վերլուծական աշխատանքների բաժնում,
- ինչպես նաև ՀՀ ԳԱԱ ԻԱՊԻ ընդհանուր սեմինարում:

ԳԼՈՒԽ I.

ԹԵՍՏԵՐԻ ՈՐԱԿԻ ՎԵՐԼՈՒԾՈՒԹՅԱՆ ՏԵՍՈՒԹՅՈՒՆՆԵՐԸ

Թեստի մշակումը բավականին բարդ, աշխատատար և խտրացիոն գործընթաց է [3], որով մեծապես պայմանավորված է սովորողների ստացած գիտելիքների ստուգման արդյունավետությունը և օբյեկտիվությունը: Կախված դրված նպատակներից, թեստերի մշակման ընթացակարգը կարող է ունենալ առանձնահատկություններ, սակայն թեստերի մշակման փուլերը նույնն են տարբեր առարկայական տիրույթների համար [3]: Տարբեր հեղինակներ նշում և նկարագրում են տարբեր քանակությամբ փուլեր՝ այսպես Վ. Չելիշկովան [22] առանձնացնում է 18 փուլ, Ա. Մայորովը [15]՝ 17, Վ. Պերներզևը [20]՝ 9, Ա. Անաստագին, Ս. Ուրբինյան [9] և Վ. Ավանեսովը [4-8]՝ 3: Թեստի որակի գնահատումը թեստի մշակման գործընթացի կարևոր փուլերից է և իրեն հերթին բաղկացած է երկու մասից.

- թեստի բովանդակության որակի գնահատում փորձագետների կողմից,
- թեստի որակի գնահատում վիճակագրական եղանակով, որը

կազմակերպվում է փորձնական թեստավորման միջոցով ստացված արդյունքների վերլուծության հիման վրա:

Վիճակագրական եղանակով թեստի որակը գնահատելու համար կիրառվում են մի քանի հայտնի թեստերի տեսություններ, որոնցից առավել լայն տարածում են գտել թեստերի դասական տեսությունը (Classical Test Theory - CTT) [9] [13] [15] [22] [45] [74], որը շատ ավելի հին է, և թեստերի ժամանակակից մաթեմատիկական տեսությունը (Item Resonse Theory) [4] [8] [18] [33] [47] [52] [67]:

1.1.Թեստերի տեսակները և վերլուծության կազմակերպման սկզբունքները

Գիտելիքների ստուգումն ուսումնական գործընթացի կարևոր և անբաժանելի բաղադրիչներից է: Գիտելիքների ստուգումը կարելի է իրականացնել մի շարք եղանակներով: Գնահատականների օբյեկտիվության տեսանկյունից մեծ առավելություններ ունի գիտելիքների ստուգման համակարգչային թեստավորման եղանակը: Համակարգչային թեստավորման քննությունը թույլ է տալիս ստուգել

գիտելիքներն առարկայի նյութի մեծ մասով, ապահովում է գնահատականի օբյեկտիվությունը, ազատում է դասավանդողին գրավոր աշխատանքների աշխատատար ստուգումներից: Գիտելիքների ստուգման թեստավորման գործընթացը բաղկացած է երեք փուլից[13].

- թեստի կազմում,
- թեստավորման անցկացում,
- արդյունքների վերլուծություն և թեստի գնահատում:

Թեստավորումն անմիջապես առնչվում է թեստաբանության հետ: Թեստաբանությունը գիտություն է մարդկային գործունեության զանազան բնագավառներում կիրառվող թեստերի մշակման, արդյունավետ օգտագործման, թեստավորման անցկացման, արդյունքների վերլուծության սկզբունքների մասին:

Ըստ Վ. Ավանեսովի [5],[6] սահմանման՝ թեստը հատուկ ձևի, աճող բարդությամբ առաջադրանքների համակարգ է, որը թույլ է տալիս որակյալ և արդյունավետ չափել փորձարկվողների պատրաստվածության մակարդակն ու կառուցվածքը:

Ըստ Ն. Մայորովի [15] սահմանման՝ թեստը գործիք է, որը կազմված է առաջադրանքների որակապես ստուգված համակարգից, թեստավորման անցկացման ստանդարտ գործընթացից և արդյունքների մշակման ու վերլուծության տեխնոլոգիայից և նախատեսված է անհատի որակի ու այնպիսի հատկությունների գնահատման համար, որոնք կարող ենք փոփոխվել ուսուցման ընթացքում:

Գրավոր և բանավոր ստուգարքների և քննությունների ժամանակ օգտագործվող ավանդական հարցերը և խնդիրները, որպես կանոն, չեն կարող առանց փոփոխության կիրառվել թեստերի մեջ: Դրանց համար մշակվում են հատուկ առաջադրանքներ թեստային ձևով: Այդ աշխատանքը ըստ աշխատատարության, դժվարության և կարևորության կարելի է համեմատել ուսումնական նոր առարկայի դասագրքի փուլ առ փուլ մշակման հետ: Փորձը ցույց է տալիս, որ կազմված առաջադրանքների մի զգալի մասը բազմակի վերամշակվում և խմբագրվում է: Սովորողների տարբեր խմբերի վրա ստուգումների արդյունքում հանվում են ամենահասարակ, ամենաբարդ և վատ ձևակերպված առաջադրանքները:

Թեստային առաջադրանքն ուսումնական առարկայի որոշակի ծավալի, բովանդակության և ձևի նյութ է, որը նախատեսված է գիտելիքների ստուգման համար: Առաջադրանքների մշակման նպատակն է ստեղծել գիտելիքների օբյեկտիվ ստուգման և գնահատման թեստեր:

Ըստ առաջադրանքների ներառման ձևի թեստերը լինում են.

- աճող բարդության, երբ թեստում յուրաքանչյուր հաջորդ առաջադրանք բարդ է նախորդից,
- խառը, երբ առաջադրանքները դասավորված են ըստ տվյալ առարկայի թեմաների,
- ադապտիվ, երբ առաջադրանքի ներկայացումը կախված է փորձարկվողի տված պատասխանից, եթե փորձարկվողը պատասխանել է առաջադրանքին ճիշտ, ապա հաջորդիվ ներկայացվում է ավելի բարդ առաջադրանք, իսկ հակառակ դեպքում ավելի հեշտ առաջադրանք:

Ըստ թեստավորման նպատակի առանձնացվում է թեստերի երկու տեսակ [15].

- հաջողությունների թեստեր, երբ գնահատվում է որևէ տիրույթի գիտելիք կամ ընդհանուր զարգացվածության մակարդակ և այլն,
- ունակությունների թեստեր, երբ գնահատվում է ընդհանուր ինտելեկտը:

Ըստ կիրառության ոլորտի առանձնացվում են հետևյալ տեսակները [15].

- նորմատիվ-կարգապահական (norm-referenced test), որը թույլ է տալիս տարբերակել փորձարկվողներին ըստ նրանց պատրաստվածության մակարդակի,
- որակավորման (criterion-referenced test), որը թույլ է տալիս գնահատել տվյալ առարկայական տիրույթի համապատասխան բաժնից փորձարկվողի գիտելիքների մակարդակը:

Ըստ բովանդակության թեստերը բաժանվում են երեք խմբի[5].

- հոմոգեն, երբ ստուգվում են առանձին առարկայից գիտելիքները,
- հետերոգեն, երբ ստուգվում են հարակից առարկաներից գիտելիքները,
- ինտեգրված, երբ ստուգվում են տարբեր առարկաներից գիտելիքներ:

Ըստ անցկացման տեխնոլոգիայի թեստերը լինում են.

- թղթային (բլանկային),
- համակարգչային

Կարգապահական, որակավորման և այլ թեստերի մշակման համար օգտագործվում են թեստային առաջադրանքների տարբեր ձևեր (տեսքեր, տիպեր):

Վ. Ավանեսովը առանձնացնում է թեստային առաջադրանքների չորս հիմնական ձևեր [5].

- փակ,
- համապատասխանության հաստատման,
- ճիշտ հաջորդականության հաստատման,
- բաց:

Հաճախ առաջադրանքների այս ձևերին տալիս են նաև համապատասխանաբար հետևյալ անվանումները.

- պատասխանի ընտրությամբ,
- երկու բազմությունների տարրերի միջև համապատասխանության հաստատման,
- նշված տարրերի շարքում ճիշտ հաջորդականության հաստատման,
- լրացման:

Գործնականում կիրառվում են փակ առաջադրանքների (պատասխանի ընտրությամբ) զանազան ձևափոխություններ՝ կոնկրետ պատասխանով, պատասխանների տարբերակներով (նաև տարբերակների համարներով): Սովորողը առաջարկված տարբերակներից ընտրում է մեկ կամ մի քանի ճիշտ տարբերակները:

Երկրորդ ձևի առաջադրանքներում համապատասխանության մեջ են դրվում երկու բազմությունների տարրերը: Օրինակ, տերմիններ – նրանց սահմանումներ, ցուցանիշներ – նրանց հաշվարկի եղանակները և այլն: Երբեմն այս առաջադրանքները անվանում են համատեղության առաջադրանքներ:

Ճիշտ հաջորդականության որոշման առաջադրանքներում սովորողը գրում է (մուտքագրում է ԷՏՄ) առաջարկված գործողությունների համարներն անհրաժեշտ հերթականությամբ: Առաջադրանքի այս ձևը կարող է օգտագործվել նաև օրենքների, թեորեմների, սահմանումների իմացությունը ստուգելու համար: Որպես առաջադրանքի

տարրեր բերվում են բառեր, որոնք որոշակի դասավորությամբ տալիս են օրենքի (սահմանման) ճիշտ ձևակերպումը: Դրան նման են տարրերը ըստ իրենց բնութագրերի, նշանակության, հատկությունների աճման կամ նվազման կարգով դասավորելու առաջադրանքները (տարրերի ռանժիրովկա): Այս առաջադրանքները երբեմն համարում են որպես առանձին խումբ :

Բաց առաջադրանքները անվանում են նաև «բացթողումները լրացնելու», «նախադասությունը, արտահայտությունը ավարտելու», «ավելացնելու» առաջադրանքներ: Սովորողն ինքն է ձևակերպում պատասխանը: Համակարգչային թեստավորման ժամանակ պատասխանի համար հատկացված է<Մ-ի դաշտում մուտքագրում է բառ, թիվ, մի քանի բառեր և այլն: Կարելի է առանձնացնել նաև առաջադրանքները ըստ պատասխանների ձևավորման տեսակի. կազմվող, տեղադրվող, հաշվվող և հաշվվող-ընտրվող պատասխաններով առաջադրանքներ:

Կազմելու առաջադրանքներում տրվում են տարրեր, որոնցից ստացվում է որոշակի բանաձև, սահմանում և այլն:

Տեղադրման առաջադրանքների տարրերը տրվում են բառ-պատասխանների տարբերակների ձևով, որոնք պետք է անհրաժեշտ հերթականությամբ տեղադրել նախադասության մեջ բաց թողած տեղերում:

Հաշվվող պատասխաններով առաջադրանքները խնդիրներ են: Սովորաբար պահանջվում է կատարել բանավոր հաշվարկ (1-3 գործողություն «կլոր» թվերի հետ) և գրել ճիշտ պատասխանի թիվը:

Բանավոր հաշվարկ պետք է կատարել հաշվվող-ընտրվող պատասխանի առաջադրանքներում, իսկ ապա նշել պատասխանի տարբերակի համարը համապատասխան թվով: Այդպիսի առաջադրանքը պարունակում է անբացահայտ հուշվածք: Սովորողը, որը չգիտի ինչպես լուծել խնդիրը, կսահմանափակվի պատասխանի տարբերակը ընտրելով:

Համակարգչային թեստավորման ժամանակ հնարավոր են գեներացվող փոփոխականներով խնդիրներ, որոնք լուծվում են թեստավորման ծրագրի մեջ ներդրված հաշվիչի օգնությամբ: Այդ խնդիրները պահանջում են որոշակի մաթեմատիկական գործողությունների կատարում և ուսումնական նյութի ավելի խոր

իմացություն:

Այսպիսով, ըստ սովորողի կողմից կատարվող գործողությունների բովանդակության, կարելի է առաջադրանքները բաժանել հետևյալ խմբերի.

- մեկ պատասխանի ընտրության,
- մի քանի պատասխանների ընտրության,
- երկու բազմությունների տարրերի միջև համապատասխանության հաստատման (որոնման),
- առաջարկվող տարրերի շարքում ճիշտ հերթականության հաստատման,
- առաջարկվող տարրերի ռանժիրովկայի (դասավորման),
- բացթողումների լրացման, նախադասության ավարտի,
- տեղադրման,
- պատասխանը կազմելու,
- պատասխանը հաշվելու,
- հաշվելու և պատասխանը ընտրելու:

Թեստային առաջադրանքները կարող են լինել. հարցական, հաստատողական, տեքստային, աղյուսակային, գրաֆիկական:

Գրաֆիկական աշխատանքների պատրաստման աշխատատարությունը զգալիորեն բարձր է տեքստայինից և աղյուսակայինից: Տեխնիկական ուսումնական առարկաների համար առանց գրաֆիկական առաջադրանքների հնարավոր չէ ստեղծել որակյալ թեստեր: Գրաֆիկական առաջադրանքների համար բնութագրական են, օրինակ, հետևյալ ձևակերպումները. «ո՞ր գրաֆիկի վրա է պատկերված...», «ո՞ր գրաֆիկն է ճիշտ ցույց տալիս փոխադարձ կապը... և ...միջև», «ո՞ր գիծն է արտապատկերում...», «ցույց տվե՞ք ...ֆունկցիան» և այլն:

Թեստեր մշակելիս անհրաժեշտ է հաշվի առնել ուսումնական նյութի բովանդակության ընտրության հետևյալ սկզբունքները [5].

- նշանակալիությունը (ընտրել գիտելիքների առավել կարևոր, հիմնական տարրերը),
- գիտական արժանահավատարությունը (խորհուրդ չի տրվում թեստի մեջ մտցնել վիճելի տեսակետներ),

- ներկայացվելիությունը (բովանդակության լրիվությունը և ստուգման համար բավարար լինելը)
- բովանդակության տարբերակայնությունը (թեստի փոփոխման հնարավորությունը կախված ուսումնական առարկայի բովանդակության, ժամերի քանակի և այլ փոփոխությունների հետ),
- աճող բարդությունը (թեթև առաջադրանքներից, որոնք ճիշտ կատարում է սովորողների մեծ մասը, մինչև ամենաբարդերը):

Հայտնի է, որ անհնարին է միևնույն թեստով ճիշտ չափել տարբեր պատրաստվածության աստիճան ունեցող սովորողների գիտելիքները: Դրա համար էլ թեստերը մշակում են պատրաստվածության միջին մակարդակ ունեցող սովորողների (ներառյալ մեծամասնություն են կազմում) գիտելիքները չափելու համար:

Թեստավորման արդյունքների մշակման և թեստի որակի գնահատման նպատակով կիրառվում են CTT և IRT տեսությունները: Այս տեսություններով վերլուծություն իրականացնելու համար դիտարկվում է թեստավորման արդյունքների մատրիցան, որն առաջադրանքներից յուրաքանչյուրին տրված փորձարկվողների պատասխանների աղյուսակն է: Այս աղյուսակի տողերը համապատասխանում են փորձարկվողներին, իսկ սյուները առաջադրանքներին, առաջադրանքին և փորձարկվողին համապատասխանող դաշտում նշվում է տվյալ առաջադրանքից տվյալ փորձարկվողի գրանցած միավորը:

Միաճյուղ բինար գնահատման դեպքում թեստավորման արդյունքներում գրանցված միավորները ներկայացվում են 1 կամ 0 տեսքով՝ համապատասխանաբար ճիշտ և սխալ կատարած առաջադրանքների համար: Թեստավորման մատրիցան այս դեպքում կոչվում է բինար մատրիցա: Նման մատրիցան կիրառվում է հիմնականում փակ տիպի առաջադրանքների համար:

Բազմաճյուղ գնահատման դեպքում թեստավորման արդյունքներում գրանցված միավորները ներկայացվում են որևէ թվային միջակայքով, որը համապատասխանում է տվյալ առաջադրանքից ստացած միավորին: Նման թեստավորման արդյունքների մատրիցան կիրառվում է հիմնականում մի քանի ճիշտ պատասխանների ընտրության, համապատասխանության հաստատման, ճիշտ հաջորդականության հաստատման և

բաց առաջադրանքների համար, որոնց դեպքում տվյալ առաջադրանքի համար կարող է սահմանվել 1-ից տարբեր միավորներ:

1.2. Թեստերի դասական տեսությունը

Թեստերի դասական տեսության հիմնադիրը համարվում է բրիտանացի հայտնի հոգեբան Չ. Սպիրմենը [82] [83] (1863-1945թ.): Թեստերի դասական տեսության զարգացման գործում մեծ ավանդ ունի Լ. Գութմանը [51] (1916-1987): Թեստերի դասական տեսությունը ժամանակակից շարադրանքով ներկայացված է Լ. Քրոկերի, Զ. Ալգինայի [45] գրքում (1986թ.): Թեստերի դասական տեսության ոլորտում կարևոր նշանակություն ունեն Ա. Անաստազի, Ս. Ուրբինայի [9], Զ. Գլասի, Զ. Ստենլիի [10], Վ. Ավանեսովի [4-8], Մ. Չելիշկովայի [22], Վ. Կիմի [13], Ա. Մայորովի [15], Յու. Նեյմանի, Վ. Խլեբնիկովայի [18], Վ. Բեսպալկոյի [11], Պ. Կլայնի [14] գիտական հետազոտությունները, որոնք վերաբերվում են թեստերի կազմման և մշակման, թեստավորման անցկացման, արդյունքների վերլուծության կազմակերպման, թեստերի որակական բնութագրերի ուսումնասիրության, համակարգչային թեստավորման և թեստաբանության այլ հարցերին:

Թեստերի դասական տեսությունը հիմնվում է հետևյալ հիմնադրույթի վրա.

թեստավորման փորձնական ճանապարհով ստացված փորձարկվողի անհատական միավորը (X) իրենից ներկայացնում է փորձարկվողի իրական (ճշգրիտ) միավորի (T) և չափման սխալի (E) գումար [45].

$$X = T + E,$$

ընդ որում T -ն (true score) բնութագրում է փորձարկվողի պատրաստվածության մակարդակը և պատահական մեծություն չէ, իսկ E -ն (error score) պատահական մեծություն է և բնութագրում է փորձարկվողի պատրաստվածության մակարդակի չափման սխալը:

Թեստերի դասական տեսության մաթեմատիկական ապարատի նկարագրման համար կատարված են հետևյալ նշանակումները [1].

N - թեստի փորձարկվողների թիվը,

n - փորձարկվողի համարը,

M - թեստի թեստային առաջադրանքների քանակը,

m - թեստում թեստային առաջադրանքի համարը,

$a = (a_{nm}, n \in [1: N], m \in [1: M])$ – թեստավորման արդյունքների մատրիցը, որտեղ a_{nm} -ը n -րդ փորձարկվողի m -րդ առաջադրանքից ստացած գնահատականն է,

X_n - n -րդ փորձարկվողի անհատական միավորի արժեքը՝

$$X_n = \sum_{m=1}^M a_{nm},$$

$\bar{X}$ - փորձարկվողների միջին անհատական միավորը՝

$$\bar{X} = \frac{\sum_{n=1}^N X_n}{N},$$

X_n^1 - կրկնակի թեստավորման դեպքում n -րդ փորձարկվողի անհատական միավորի արժեքը առաջին թեստավորումից,

X_n^2 - կրկնակի թեստավորման դեպքում n -րդ փորձարկվողի անհատական միավորի արժեքը երկրորդ թեստավորումից,

X_n^e - թեստի կիսման դեպքում n -րդ փորձարկվողի անհատական միավորի արժեքը՝ ըստ թեստի զույգ (even) թվով առաջադրանքների,

X_n^o - թեստի կիսման դեպքում n -րդ փորձարկվողի անհատական միավորի արժեքը՝ ըստ թեստի կենտ (odd) թվով առաջադրանքների,

$\bar{X}_1^m$ - m -րդ առաջադրանքը կատարած փորձարկվողների միջին անհատական միավորը,

$\bar{X}_0^m$ - m -րդ առաջադրանքը չկատարած փորձարկվողների միջին անհատական միավորը,

N_1^m - m -րդ առաջադրանքը կատարած փորձարկվողների թիվը,

N_0^m - m -րդ առաջադրանքը չկատարած փորձարկվողների թիվը,

R_m - m -րդ առաջադրանքին տրված ճիշտ պատասխանների քանակը՝

$$R_m = \sum_{n=1}^N a_{nm},$$

W_m - m -րդ առաջադրանքին տրված սխալ պատասխանների քանակը՝

$$W_m = N - R_m,$$

p_m - ճիշտ պատասխանների մասը՝

$$p_m = \frac{R_m}{N},$$

q_m - սխալ պատասխանների մասը՝

$$q_j = 1 - p_m \quad \text{կամ} \quad q_m = \frac{W_m}{N},$$

p_{m1m2} - $m1$ և $m2$ առաջադրանքներին (երկուսին միասին) տրված ճիշտ պատասխանների մասը,

S_x^2 - N փորձարկվողների անհատական միավորների ցրվածքը՝

$$S_x^2 = \frac{1}{N-1} \left(\sum_{n=1}^N X_n^2 - \frac{(\sum_{n=1}^N X_n)^2}{N} \right), \quad (1)$$

S_m^2 - m -րդ առաջադրանքից ստացված արդյունքների ցրվածքը՝

$$S_m^2 = \frac{1}{N-1} \left(\sum_{n=1}^N a_{nm}^2 - \frac{(\sum_{n=1}^N a_{nm})^2}{N} \right),$$

S_{xi} -կրկնակի թեստավորման դեպքում i -րդ թեստավորումից N փորձարկվողների անհատական միավորների՝ X_n^i միջինից շեղումների քառակուսիների գումարը՝

$$S_{xi} = \sum_{n=1}^N (X_n^i)^2 - \frac{(\sum_{n=1}^N X_n^i)^2}{N},$$

$S_{X1,X2}$ - կրկնակի թեստավորման դեպքում X_n^1 և X_n^2 մեծությունների միջին արժեքներից շեղումների արտադրյալների գումարը՝

$$S_{X1,X2} = \sum_{n=1}^N X_n^1 * X_n^2 - \frac{1}{N} \sum_{n=1}^N X_n^1 * \sum_{n=1}^N X_n^2$$

$r_{X1,X2}$ - X_n^1 և X_n^2 մեծությունների միջև կոռելյացիայի գործակիցը՝

$$r_{X1,X2} = \frac{S_{X1,X2}}{\sqrt{S_{X1} S_{X2}}}, \quad (2)$$

S_e - թեստի կիսման դեպքում N փորձարկվողների՝ զույգ թվով թեստային առաջադրանքներից ստացված անհատական միավորների՝ X_n^e միջինից շեղումների քառակուսիների գումարը՝

$$S_e = \sum_{n=1}^N (X_n^e)^2 - \frac{(\sum_{n=1}^N X_n^e)^2}{N},$$

S_o - թեստի կիսման դեպքում N փորձարկվողների՝ կենտ թվով թեստային առաջադրանքներից ստացված անհատական միավորների՝ X_n^o միջինից շեղումների քառակուսիների գումարը՝

$$S_o = \sum_{n=1}^N (X_n^o)^2 - \frac{(\sum_{n=1}^N X_n^o)^2}{N},$$

S_{eo} - թեստի կիսման դեպքում X_n^e և X_n^o մեծությունների միջին արժեքներից շեղումների արտադրյալների գումարը՝

$$S_{eo} = \sum_{n=1}^N X_n^e * X_n^o - \frac{1}{N} \sum_{n=1}^N X_n^e * \sum_{n=1}^N X_n^o$$

r_{eo} - X_n^e և X_n^o մեծությունների միջև կոռելյացիայի գործակիցը՝

$$r_{eo} = \frac{S_{eo}}{\sqrt{S_e S_o}}, \quad (3)$$

E - փորձարկվողի անհատական միավորի սխալի վեկտորը,

կրկնակի թեստավորման դեպքում E մեծության արժեքը՝

$$E_n^{X1,X2} = X_n^1 - X_n^2,$$

թեստի կիսման դեպքում E մեծության արժեքը՝

$$E_n^{eo} = X_n^e - X_n^o,$$

S_E^2 - N փորձարկվողների անհատական միավորների E սխալների ցրվածքը,

կրկնակի թեստավորման դեպքում S_E^2 մեծության արժեքը՝

$$S_{E(X1,X2)}^2 = \frac{1}{N-1} \left(\sum_{n=1}^N (E_n^{X1,X2})^2 - \frac{(\sum_{n=1}^N E_n^{X1,X2})^2}{N} \right), \quad (4)$$

թեստի կիսման դեպքում S_E^2 մեծության արժեքը՝

$$S_{E(eo)}^2 = \frac{1}{N-1} \left(\sum_{n=1}^N (E_n^{eo})^2 - \frac{(\sum_{n=1}^N E_n^{eo})^2}{N} \right),$$

$\varphi_{m1,m2}$ - $m1$ և $m2$ առաջադրանքների կոռելյացիայի Պիրսոնի գործակիցը,

$\bar{r}_m$ - m -րդ առաջադրանքի կոռելյացիայի գործակցի միջին արժեքը,

$\bar{R}$ - թեստի առաջադրանքների կոռելյացիայի գործակիցների միջին արժեքը՝

$$\bar{R} = \frac{\sum_{j=1}^M \bar{r}_m}{M}, \quad (5)$$

r_t - t թեստի հուսալիության գործակիցը:

Թեստերի դասական տեսության շրջանակներում վերլուծվում են թեստի այնպիսի որակական բնութագրեր [17], ինչպիսիք են՝ թեստի հուսալիությունը, վավերությունը, բարդությունը, տարբերակիչ ունակությունը (discrimination) [22] և այլն:

Թեստի հուսալիություն. Հուսալիությամբ որոշվում են թեստավորման արդյունքների վերարտադրելիությունը և ճշտությունը: Ենթադրենք ունենք

փորձարկվողների հիպոթետիկ խումբ, որը թեստավորումը ավարտելուց հետո անմիջապես մոռանում է թեստի բովանդակությունը: Այդ դեպքում հուսալի թեստի պարագայում բազմաթիվ անգամ կրկնելով թեստավորումը՝ պետք է ստացվեն միևնույն անհատական միավորները: Իսկ փոքր հուսալիության թեստերի դեպքում ամեն անգամ արդյունքները կփոփոխվեն: Թեստի դասական տեսության մեջ թեստի հուսալիությունը որոշվում է հետևյալ բանաձևով [13].

$$r_t = 1 - \frac{S_E^2}{S_x^2},$$

S_E^2 և S_x^2 արժեքները կարելի է որոշել համապատասխանաբար (4) և (1) բանաձևերի միջոցով [13]: Թեստի հուսալիության հաշվարկի համար անհրաժեշտ են երկու փորձարկման արդյունքներ, որոնք կազմակերպվում են հետևյալ մեթոդներով [54].

- 1-ին մեթոդ - թեստավորում երկու զուգահեռ թեստերի միջոցով (հուսալիությունն այս դեպքում կոչվում է parallel-form reliability),
- 2-րդ մեթոդ-կրկնակի թեստավորում նույն թեստի միջոցով (հուսալիությունն այս դեպքում կոչվում է test-retest reliability [36]),
- 3-րդ մեթոդ- թեստի կիսում (split-half method):

Նշված մեթոդներից յուրաքանչյուրն ունի իր առավելություններն ու թերությունները [19] [24]:

Կրկնակի թեստավորման մեթոդով հուսալիությունը կարելի է որոշել կիրառելով երկու՝ X_n^1 և X_n^2 մեծությունների միջև կոռելյացիայի գործակցի որոշման (2) արտահայտությունը [19], որտեղից ստացվում է հետևյալ բանաձևը.

$$r_{t(X1,X2)} = \frac{N * \sum_{n=1}^N X_n^1 X_n^2 - (\sum_{n=1}^N X_n^1) * (\sum_{n=1}^N X_n^2)}{\sqrt{[N * \sum_{n=1}^N (X_n^1)^2 - (\sum_{n=1}^N X_n^1)^2] * [N * \sum_{n=1}^N (X_n^2)^2 - (\sum_{n=1}^N X_n^2)^2]}} \quad \square \quad (6)$$

Թեստը կարելի է օգտագործել, եթե նրա հուսալիության գործակիցը պակաս չէ +0,7-ից [13]:

Գոյություն ունեն հուսալիության մի շարք գործակիցներ [19]:

Թեստի կիսման մեթոդով հուսալիությունը կարելի է որոշել Սպիրման-Բրաունի (Spearman-Brown) կոռելյացիայի գործակցի միջոցով [79]: Այստեղ որոշվում է թեստի

զույգ և կենտ առաջադրանքների արդյունքների միջև կոռելյացիան: Բանաձևն ունի հետևյալ տեսքը [13]՝

$$r_{t(eo)} = \frac{2r_{eo}}{1 + r_{eo}},$$

որտեղ $r_{t(eo)}$ -ը հուսալիության ճշգրտված գործակիցն է, r_{eo} -ն թեստի հուսալիության գործակիցն է ըստ կիսված թեստի մասերի, որը հաշվարկվում է (3) բանաձևով:

Սպիրմեն-Բրաունի բանաձևը [80] թույլ է տալիս գնահատել տրված հուսալիության արժեքի դեպքում թեստի անհրաժեշտ երկարությունը (թեստային առաջադրանքների քանակը):

r_{tk} հուսալիության գործակիցը թեստի երկարության փոփոխման դեպքում [22].

$$r_{tk} = \frac{kr_{t(eo)}}{1 + (k - 1)r_{t(eo)}},$$

որտեղ k -ն թեստի երկարության չափման գործակիցն է:

Օրինակ, ենթադրենք թեստի նախնական հուսալիությունը 0,758 է, և թեստում առաջադրանքների քանակը կրկնապատկվում է: Այդ դեպքում նոր թեստի հուսալիությունը կլինի՝

$$r_{tk} = \frac{2 * 0,758}{1 + (2 - 1) * 0,758} = 0,862:$$

Կարելի է որոշել թեստի անհրաժեշտ երկարության k գործակցի արժեքը՝ ունենալով r_{tk} և $r_{t(eo)}$ արժեքները: Ընդունենք, որ թեստի նախնական հուսալիությունը 0,758 է և պետք է այն հասցնել 0,862-ի: Քանի՞ անգամ է հարկավոր մեծացնել թեստի երկարությունը: Հաշվարկի համար օգտագործվում է հետևյալ արտահայտությունը [22].

$$k = \frac{r_{tk}(1 - r_{t(eo)})}{r_{t(eo)}(1 - r_{tk})} :$$

Մեր օրինակում

$$k = \frac{0,862(1 - 0,758)}{0,758(1 - 0,862)} = 1,994 \approx 2,$$

այսինքն՝ թեստի երկարությունն անհրաժեշտ է մեծացնել երկու անգամ:

Մեկանգամյա թեստավորման եղանակով թեստի հուսալիությունը կարելի է

հաշվարկել նաև Կյուդեր-Ռիչարդսոնի KR-20 բանաձևի միջոցով [63].

$$r_{t(kr20)} = \frac{M}{(M-1)} \left(1 - \frac{\sum_{j=1}^M p_j q_j}{S_x^2} \right) :$$

Թեստի հուսալիությունը գնահատվում է գերազանց, եթե $r_{t(kr20)}$ –ն կազմում է 0,90 ... 0,99, լավ՝ 0,80 ... 0,89, բավարար՝ 0,70 ... 0,79, 0,69-ից փոքր արժեքների դեպքում՝ անբավարար [12]: Գիտելիքների ընթացիկ գնահատման համար խորհուրդ է տրվում օգտագործել 0,80-ից ոչ պակաս հուսալիության գործակից ունեցող թեստ, իսկ վերջնական քննության դեպքում՝ 0,90-ից ավել [13]: Կարևոր է նշել, որ KR-20 բանաձևով է որոշվում այնպիսի հայտնի թեստերի հուսալիությունը, ինչպիսիք SAT և TOEFL-ն են:

Բազմաճյուղ գնահատման սանդղակների դեպքում, այսինքն՝ երբ թեստի միավորը սահմանվում է որևէ թվային միջակայքում (0,1,2,3 և այլն) կիրառվում է Կրոնբախի ալֆա գործակցի բանաձևը, որն ունի հետևյալ տեսքը [46]՝

$$r_{at} = \frac{M}{(M-1)} \left(1 - \frac{\sum_{m=1}^M S_m^2}{S_x^2} \right) :$$

Թեստի հուսալիությունը կարելի է որոշել նաև թեստի առաջադրանքների միջև կոռելյացիայի Պիրսոնի գործակցի միջոցով [13], որը երկճյուղային գնահատման սանդղակների (թեստից ստացված միավորն ընդունում է 0 կամ 1 արժեքը) դեպքում կոչվում է «ֆի» գործակից և որոշվում է հետևյալ բանաձևով [22].

$$\varphi_{m1,m2} = \frac{p_{m1,m2} - p_{m1} * p_{m2}}{\sqrt{p_{m1} * q_{m1} * p_{m2} * q_{m2}}} :$$

Առաջադրանքների միջև կոռելյացիան բնութագրում է թեստի հոմոգենությունը՝ այսինքն՝ առաջադրանքների միջև փոխկապակցվածությունը: Ըստ Չելիշկովայի [22] բնութագրում է «թեստի առարկայական մաքրությունը», այսինքն՝ որքանով են բովանդակային առումով առաջադրանքներն իրար հետ փոխկապակցված: Այս ցուցանիշը արժեքներ է ընդունում [-1,+1] միջակայքից: Առաջադրանքների միջև կոռելյացիան չպետք է լինի շատ բարձր ($\varphi_{m1,m2} \leq 0,3$), հակառակ դեպքում՝ առաջադրանքներն սկսում են իրար կրկնել [6]: Եթե երկու առաջադրանքների միջև կոռելյացիան մոտ է 1-ին, ապա դրանցից մեկն ավելորդ է: Աղյուսակ 1-ում ներկայացված է առաջադրանքների միջև կոռելյացիայի ցուցանիշի արժեքները:

Աղյուսակ 1

Առաջադրանքների միջև կոռելյացիայի ցուցանիշի արժեքները:

Արժեքների միջակայք	Մեկնաբանություն
-1-0	Ներքին փոխկապակցվածության բացակայություն (միմյանցից լրիվ անկախ առաջադրանքներ)
0-0,3	Թույլ ներքին փոխկապակցվածություն
0,3-0,6	Լավ ներքին փոխկապակցվածություն
0,6-0,8	Բարձր ներքին փոխկապակցվածություն
0,8-1	Շատ բարձր ներքին փոխկապակցվածություն

Թեստի հուսալիությունը՝ ըստ թեստի առաջադրանքների միջև կոռելյացիայի գործակցի, հաշվարկվում է հետևյալ բանաձևով [19]՝

$$r_{t(\varphi)} = \frac{M\bar{R}}{1 + (M - 1)\bar{R}} :$$

Հուսալիության արժեքներն ամբողջական ներկայացված են աղյուսակ 2-ում:

Աղյուսակ 2

Հուսալիության արժեքները:

Արժեք	Մեկնաբանություն
0-0,599	Շատ ցածր
0,600 – 0,699	Ցածր: Որոշ քանակությամբ առաջադրանքների փոփոխման անհրաժեշտություն
0,700-0,799	Բավարար
0,800- 0,899	Լավ: Շատ քիչ քանակությամբ առաջադրանքների փոփոխման անհրաժեշտություն:
0,900-0,949	Շատ լավ
0,950 -1	Գերազանց

Թեստի վավերություն. Թեստի կարևոր որակական բնութագրերից է նաև

վավերությունը [13] (անգլերեն validate): Վավերություն նշանակում է թեստային արդյունքների պիտանիություն նախատեսված նպատակով կատարվող թեստի համար [8]:

Առանձնացվում են վավերության երեք տեսակ. **բովանդակային, կոնցեպտուալ և հատկանիշային** [13][15]:

Թեստի բովանդակային վավերությունը բնութագրում է, թե ինչքանով է թեստը համապատասխանում առարկայական տիրույթին: Բովանդակային վավերությունը պարտադիր չէ որ արտացոլի առարկաների լրիվությունը: Օրինակ, նորմատիվ – կողմնորոշված թեստի համար բոլոր թեմաների ընդգրկման լրիվությունը կարող է լինել ավելի փոքր, քան հայտանիշ – կողմնորոշված թեստի համար: Այստեղ կարևոր է առանձին ենթաթեմաների ու հայտերի ուսումնասիրման խորությունը: Դա թույլ կտա մեծ արդյունավետությամբ տարբերակել սովորողներին:

Թեստի կոնցեպտուալ վավերությունը որոշվում է այն դեպքերում, երբ փորձարկվողի չափվող հատկության ներկայացումը տրված է ինչ-որ վերացական կերպարի, մոդելի տեսքով, այսինքն՝ փորձարկվողի անձի որոշակի որակները բացահայտելու համար ստեղծվում է կոնցեպտուալ մոդել, որը թեստի օգնությամբ հաստատվում է կամ հերքվում:

Հատկանիշային վավերությունը ենթադրում է արտաքին հատկանիշի առկայություն, որի հետ կոռելյացիայի մեծությունը որոշում է թեստի վավերությունը: Հատկանիշային վավերությունը լինում է **ընթացիկ և կանխատեսվող**: Ենթադրենք՝ որևէ փորձարկվող ինչ-որ առարկայից թեստով ցույց տվեց գերազանց գիտելիքներ, իսկ այդ նույն առարկայից դասավանդողի նշանակած գնահատականները անբավարար են: Եթե որպես արտաքին հայտանիշ ընտրվում են դպրոցական գնահատականները, թեստի հատկանիշային վավերությունը կլինի ցածր, նույնիսկ եթե թեստն ունի բարձր հուսալիություն: Թեստի կանխատեսվող վավերությունը բնութագրում է թեստավորման արդյունքների կոռելյացիան ապագայում հանդես եկող արտաքին հատկանիշի հետ:

Ընդհանուր դեպքում որպես **վավերության գործակից** ընդունվում է թեստավորման արդյունքների և հայտանիշի հարաբերության արժեքը, վավերության V_t

գործակիցը կարելի է հաշվել հետևյալ բանաձևով [1].

$$V_t = \frac{\frac{\sum_{n=1}^N (X_n^{exp} * X_n)}{N} - \bar{X}_{exp} * \bar{X}}{S_{X(exp)} * S_X} * \frac{N}{N-1},$$

որտեղ $\bar{X}_{exp} = \frac{\sum_{n=1}^N X_n^{exp}}{N}$ - փորձագետի կողմից նշանակած գնահատականների միջին արժեքն է, $S_{X(exp)} = \sqrt{\frac{\sum_{n=1}^N (X_n^{exp} - \bar{X}_{exp})^2}{N-1}}$ - փորձագետի կողմից նշանակած գնահատականներից ստանդարտ շեղումն է: V_t գործակցի արժեքները ներկայացված են աղյուսակ 3-ում:

Աղյուսակ 3

Վավերության գործակցի արժեքները:

Արժեքների միջակայք	Մեկնաբանություն
0-0,09	Վատ վավերություն
0,10-0,39	Անբավարար վավերություն
0,40-0,49	Բավարար վավերություն
0,50-0,74	Լավ վավերություն
0,75-0,79	Շատ լավ վավերություն
0,8 -1	Գերազանց վավերություն

Այս մեթոդի թերություններն են՝ փորձարկվողների խմբի սովորողների ընտրության որակից մեծ կախվածությունը և դասավանդողների սուբյեկտիվ գնահատականների օգտագործումը:

Թեստային առաջադրանքների վավերությունը կարելի է որոշել առաջադրանքի կոռելյացիայի կետային բիսերիալ (point biserial) գործակցի միջոցով: Այս գործակիցը բնութագրում է յուրաքանչյուր առաջադրանքի կոռելյացիան փորձարկվողի թեստային միավորի (անհատական միավորի) հետ՝ r_{pb}^m : r_{pb}^m -ն չափվում է հետևյալ բանաձևով [13]՝

$$r_{pb}^m = \frac{\bar{X}_1^m - \bar{X}_0^m}{S_x} \sqrt{\frac{N_1^m * N_0^m}{N * (N-1)}},$$

որտեղ S_x -ը բոլոր փորձարկվողների անհատական միավորների ստանդարտ շեղումն է՝

$$S_x = \sqrt{S_x^2}:$$

Վ. Ս. Ավանեսովը նշում է, որ թեստային առաջադրանքի վավերության համար պետք է ապահովվի $r_{pb}^m \geq 0,5$ պայմանը [3]:

Թեստի բարդություն. Թեստի բարդությունը թեստի այն բնութագրիչն է, որն արտացոլում է թեստը լուծելու վիճակագրական ցուցանիշը: Թեստի բարդությունը բնութագրվում է թեստի բարդության ինդեքսով և որոշվում է թեստին տրված ճիշտ պատասխանների քանակի (N_1^m) և փորձարկվողների քանակի (N) հարաբերությամբ [15].

$$U_m = 1 - \frac{N_1^m}{N}, \quad (7)$$

որտեղ U_m -ը բարդության ինդեքսն է, իսկ N_1^m/N հարաբերությունը՝ բարդությունը:

Բարդության ինդեքսի միջոցով թեստից հանվում են չափազանց բարդ և չափազանց դյուրին առաջադրանքները: Այն արժեքներ է ընդունում 0-ից 1 միջակայքից, թեստային առաջադրանքի բարդության ինդեքսը պետք է ունենա արժեք 0,3 ... 0,8 միջակայքում: Լավագույն է համարվում 0,5 արժեքը: Թեստի կազմման ժամանակ բարդության ցածր գործակից ունեցող թեստային առաջադրանքները տեղադրվում են թեստի սկզբում, իսկ բարդության բարձր գործակից ունեցողները՝ թեստի վերջում: Բարդության ինդեքսի արժեքները ամբողջական ներկայացված են աղյուսակ 4-ում:

Աղյուսակ 4

Բարդության արժեքները:

Արժեքների միջակայք	Մեկնաբանություն
0-0,19	Շատ բարդ
0,2-0,49	Բարդ
0,5-0,69	Միջին բարդության
0,7-0,89	Հեշտ
0,9-1	Շատ հեշտ

Թեստի տարբերակիչ ունակություն. Թեստի տարբերակիչ ունակությունը բնութագրում է թեստի՝ ուժեղ և թույլ փորձարկվողներին տարբերակելու ունակությունը: Թեստային առաջադրանքի տարբերիչ ունակությունը (D) հաշվելու համար այդ առաջադրանքի կատարումը հանձնարարվում է փորձարկվողների՝ ըստ ունակության երկու հակադիր («բարձր» – «ցածր») խմբերի: Այս գործակիցը որոշվում է հետևյալ

սկզբունքով[15].

- իրականացվում է փորձարկվողների բավականին մեծ խմբի թեստավորում ($N > 100$),
- թեստավորման արդյունքների հիման վրա առանձնացվում է փորձարկվողների մոտ 25 ... 27% բարձր ցուցանիշներով խումբ՝ p_{best} և ցածր ցուցանիշներով խումբ՝ p_{worst} , համապատասխանաբար N_b և N_w փորձարկվողների թվով,
- հաշվարկվում է վերոնշյալ խմբերում ճիշտ պատասխանների քանակները՝ N_b^1 և N_w^1 ,
- հաշվարկվում է տարբերակիչ ունակության D գործակիցը հետևյալ բանաձևով՝ [15]

$$D = \frac{N_b^1}{N_b} - \frac{N_w^1}{N_w}.$$

Եթե $D > 0.2$, ապա դիտարկվող թեստի տարբերակիչ ունակությունը բավարար է [15]:

D ցուցանիշը արժեքներ է ընդունում $[-1, +1]$ հատվածում: Եթե $D = 1$, ուրեմն առաջադրանքն ունի առավելագույն տարբերիչ ունակություն: $D = 0$ -ն նշանակում է, որ առաջադրանքը բոլորովին չի տարբերում ուսումնական նյութին տիրապետող և չտիրապետող փորձարկվողներին: $D = -1$ տարբերակը հանդիպում է շատ հազվադեպ և նշանակում է, որ առաջադրանքը տարբերում է փորձարկվողներին, բայց հակադարձված, այսինքն՝ ճիշտ պատասխանում են ուսումնական նյութին չտիրապետողները, իսկ նյութին տիրապետողները պատասխանում են սխալ:

Տարբերակիչ ունակության արժեքները ամբողջությամբ ներկայացված է աղյուսակ 5-ում:

Աղյուսակ 5

Տարբերակիչ ունակության արժեքները:

Արժեք	Մեկնաբանություն
0-0,2	Անբավարար
0,21-0,40	Բավարար
0,41-0,60	Լավ
0,61-1	Գերազանց

Ստանդարտ շեղում (Std. Dev). Ճիշտ կառուցված թեստը պետք է լավ տարբերակի փորձարկվողներին: Դա նշանակում է, որ անհատական թեստային միավորները պետք է բավարար չափով տարբերվեն մեկը մյուսից: Անհատական միավորների լրիվ համընկման դեպքում ցրվածքը (variance) հավասար է զրոյի: Եթե անհատական միավորները չեն համընկնում, ապա շեղումներ կարող են լինել ինչպես բացասական այնպես էլ դրական:

Ըստ Չելիշկովայի [21] թեստային միավորների ցրվածքը օպտիմալ է, իսկ բաշխումը մոտ է նորմալի, եթե տվյալ առաջադրանքի թեստային միավորի միջին արժեքի համար բավարարվում է հետևյալ պայմանը՝

$$\bar{X} \approx 3S_x:$$

Հատկացվող ժամանակ. Հետազոտությունները ցույց են տալիս, որ առաջադրանքի կատարման ժամանակը կախված է առաջադրանքի տեսակից: Չորս-հինգ տարբերակներից բաղկացած առաջադրանքների դեպքում նախատեսվում է միջինը մինչև 1 րոպե, կարճ պատասխանով առաջադրանքների դեպքում՝ 2 րոպե, իսկ ազատ ոճով գրվող երկար պատասխանների դեպքում՝ 5 րոպե:

1.4. Թեստերի մաթեմատիկական տեսությունը

Բացի թեստերի դասական տեսությունից, ներկայումս մեծ զարգացում ունի թեստավորման IRT տեսությունը (մաթեմատիկական տեսություն) [45][47], որն ուղղված է փորձարկվողի և առաջադրանքների պարամետրերի լատենտային [16] որակների գնահատմանը: Այս տեսության մեջ լայնորեն կիրառություն են գտել մաթեմատիկական մոդելները [18]: IRT մեկ պարամետրով մոդելն [24] առաջարկվել է Գ.Ռաշի [79] կողմից: IRT-ի մեկ պարամետրով մոդելի կատարելագործված տարբերակներն են Բիրնբաումի երկու և երեք պարամետրերով մոդելները [37]: IRT տեսության զարգացման ուղղությամբ մեծ գործունեություն է ծավալվել Դ.Էնդրիչի [26] և Բ.Ռայտի [90] կողմից: IRT տեսությամբ հնարավոր է վերլուծել ինչպես միաճյուղ (0 կամ 1), այնպես էլ բազմաճյուղ (մի քանի միավորներով սահմանվող) գնահատումով թեստերի արդյունքներ:

IRT տեսության մաթեմատիկական ապարատը բաղկացած է մի շարք մոդելներից [67].

միաչափ միաճյուղ մոդելներ.

- նորմալ մոդել (Normal Ogive Model),
- 1 պարամետրով միաճյուղ մոդել (Ռաշի մոդել - 1PLM),
- 2 պարամետրով միաճյուղ մոդել (2PLM),
- 3 պարամետրով միաճյուղ մոդել (3PLM),
- ոչ պարամետրական մոդել,

միաչափ բազմաճյուղ մոդելներ[75].

- մասնակի միավորներով մոդել (PCM),
- ընդհանրացված մասնակի միավորներով մոդել (GPCM),
- միավորների սանդղակավորման մոդել (RSM),
- նոմինալ պատասխանների մոդել (NRM),

բազմաչափ միաճյուղ մոդել:

IRT-ում n -րդ փորձարկվողի պատրաստվածության մակարդակը նշանակվում է θ_n - ով և որոշվում հետևյալ բանաձևով՝

$$\theta_n = \ln(p_n/q_n): \quad (8)$$

m -րդ առաջադրանքի բարդության լոգարիթմը նշանակվում է β_m -ով և որոշվում հետևյալ բանաձևով՝

$$\beta_m = \ln(p_m/q_m): \quad (9)$$

Ենթադրվում է, որ IRT-ում β_m -ի և θ_n -ի միջև գոյություն ունի որոշակի կախվածություն: Գործնականում առաջադրվում է հետևյալ խնդիրը. ըստ թեստին տված փորձարկվողի պատասխանների, գնահատել β_m և θ_n լատենտային պարամետրերի արժեքները [21][13]:

Ըստ Գ. Ռաշի փորձարկվողի պատրաստվածության մակարդակը՝ θ_n և թեստային առաջադրանքի բարդության մակարդակը՝ β_m գտնվում են միևնույն սանդղակում և չափվում են նույն միավորներով: Փորձարկվողի թեստը հաջողությամբ հանձնելու ֆունկցիայի արգումենտը θ_n և β_m մեծությունների տարբերությունն է՝

$|\theta_n - \beta_m|$: Եթե տարբերությունը մեծ է և բացասական, ապա տվյալ m -րդ թեստային առաջադրանքը բարդ է տվյալ n -րդ փորձարկվողի համար: Եթե տարբերությունը մեծ է և դրական, ապա տվյալ m -րդ թեստային առաջադրանքը շատ հեշտ է տվյալ n -րդ փորձարկվողի համար: Ռաշի մոդելում փորձարկվողի հաջողության հասնելու ֆունկցիան ունի մեկ պարամետր՝ $|\theta_n - \beta_m|$, որի պատճառով էլ այն երբեմն կոչվում է IRT մեկ պարամետրով միաճյուղ մոդել:

Մեկ պարամետրով միաճյուղ մոդել (Ռաշի մոդել)

Ռաշի մոդելը ներկայացվում է գիտելիքների θ մակարդակով փորձարկվողի կողմից առաջադրանքը ճիշտ պատասխանելու հավանականության ֆունկցիայի (item response function) տեսքով: i -րդ առաջադրանքի համար Ռաշի մոդելը ներկայացվում է հետևյալ բանաձևով՝

$$P_i(\theta) = \frac{1}{1 + e^{-D(\theta - b_i)}}, \quad (10)$$

որտեղ $P_i(\theta)$ -ն գիտելիքների θ մակարդակով փորձարկվողի կողմից i -րդ առաջադրանքը ճիշտ պատասխանելու հավանականությունն է, b_i -ն i -րդ առաջադրանքի բարդության պարամետրն է (IRT «b» պարամետր), D -ն մասշտաբավորման գործակիցն է, որի արժեքն է 1.7: IRT «b» պարամետրը բնութագրում է առաջադրանքի բարդությունը:

Առաջադրանքը ճիշտ պատասխանելու հավանականությունը՝ $P_i(\theta)$ -ն, մոնոտոն աճում է θ լատենտային բաղադրիչի աճի դեպքում: Եթե առաջադրանքի բարդությունը մեծ է փորձարկվողի պատրաստվածության մակարդակից ($\theta < b$), ապա անհավանական է, որ փորձարկվողը ճիշտ կպատասխանի առաջադրանքին: Եթե փորձարկվողի պատրաստվածության մակարդակը նույնն է ինչ որ առաջադրանքի բարդությունը ($\theta = b$), ապա փորձարկվողը 50% հավանականությամբ ճիշտ կպատասխանի առաջադրանքին: Եթե θ -ն գերազանցում է b -ի արժեքին, ապա ճիշտ պատասխանելու հավանականությունը մեծ է 50 %-ից:

(7)-ից հետևում է, որ

$$\begin{aligned} \lim_{\theta \rightarrow -\infty} P(\theta; b, D) &= 0, \\ \lim_{\theta \rightarrow +\infty} P(\theta; b, D) &= 1, \end{aligned}$$

այսինքն՝ փորձարկվողի պատրաստվածության θ մակարդակի աճի դեպքում մեծանում է առաջադրանքը ճիշտ կատարելու հավանականությունը: «Անվերջ լավ» պատրաստված փորձարկվողը ($\theta \rightarrow \infty$) թեստավորման ժամանակ ցանկացած առաջադրանք ճիշտ կկատարի 1 հավանականությամբ: Փորձարկվողի պատրաստվածության մակարդակի նվազեցման դեպքում առաջադրանքը ճիշտ կատարելու հավանականությունը ևս նվազում է: Թեստավորման ժամանակ «անվերջ վատ» պատրաստված փորձարկվողը ($\theta \rightarrow -\infty$) ցանկացած առաջադրանք ճիշտ կկատարի 0 հավանականությամբ:

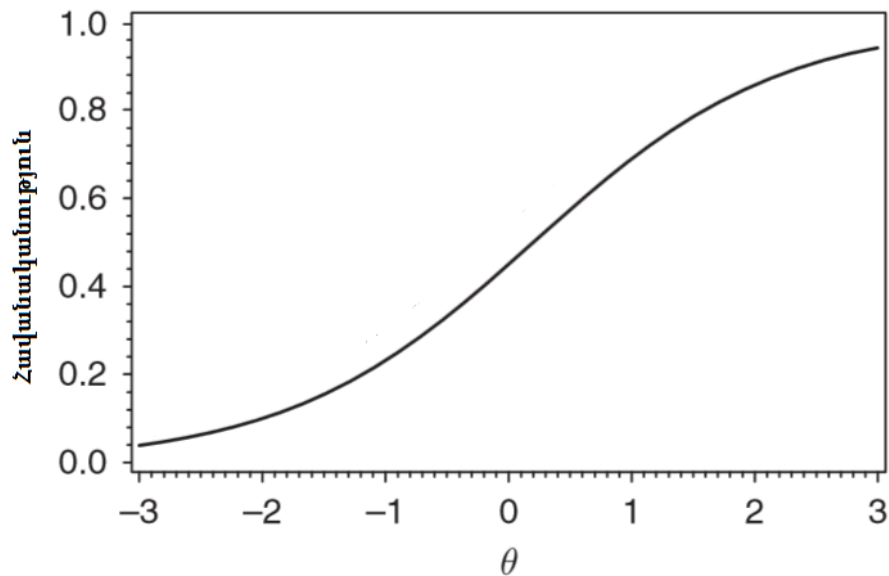
Կարելի է նկատել, որ նույն պատրաստվածության մակարդակով «b» պարամետրի արժեքի մեծացման դեպքում առաջադրանքը ճիշտ կատարելու հավանականությունը փոքրանում է: Այսինքն՝ որքան մեծ է առաջադրանքի «b» պարամետրի արժեքը, այնքան ավելի մեծ պատրաստվածության մակարդակ պետք է ունենա փորձարկվողը՝ տվյալ առաջադրանքը ճիշտ կատարելու համար:

Առաջադրանքի բնութագրական կոր

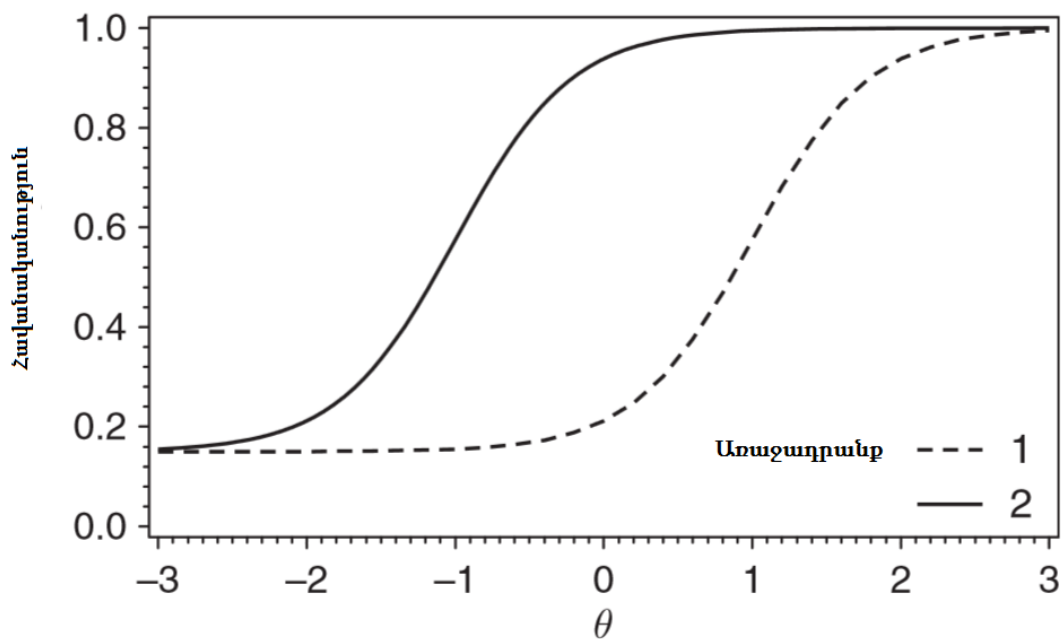
Առաջադրանքի պատասխանների ֆունկցիայի [52] գրաֆիկական ներկայացումը առաջադրանքի բնութագրական կորն է [47] (Item Characteristic Curve-ICC), որը ցուցադրում է գիտելիքների մակարդակի լատենտային բաղադրիչի և ճիշտ պատասխանելու հավանականության կախվածությունը:

Գրաֆիկում արսցիսների առանցքի վրա նշվում է փորձարկվողների պատրաստվածության մակարդակը, իսկ օրդինատների առանցքի վրա ճիշտ պատասխանելու հավանականության արժեքները:

Նկար 1-ում բերված է 1 պարամետրանոց առաջադրանքի բնութագրական կորը:



Նկար 1. Մեկ պարամետրանոց առաջադրանքի բնութագրական կորը



Նկար 2. Տարբեր «b» պարամետրերով երկու առաջադրանքների բնութագրական կորերը: Առաջադրանք 1-ը ավելի բարդ է:

Ինչպես երևում է նկար 2-ի գրաֆիկից, որքան առաջադրանքի բարդությունը մեծանում է, այնքան կորը տեղափոխվում է աբսցիսների առանցքով դեպի աջ: Ըստ Ռաշի մոդելի առաջադրանքի բարդությունը աբսցիսների առանցքի վրա այն կետն է, որին տարված ուղղահայացը հասնում է օրդինատների առանցքի 0,5 արժեքին

համապատասխանող կորի հետ: Այսինքն՝ առաջադրանքի բարդությունը արեւելիսների առանցքի վրա այն կետն է, որի դեպքում ճիշտ պատասխանելու հավանականությունը 0,5 է: Առաջադրանքի բնութագրական կորի ֆունկցիան ունի շրջման կետ, որի արեւելիսը հավասար է «b» պարամետրի արժեքին, իսկ օրդինատը 0,5 արժեքին: Շրջման կետում բնութագրական կորի և շոշափողի հետ կազմած անկյան տանգենսը կախված է «b» պարամետրի արժեքից և հավասար է 0,25-ի:

Թեստի բնութագրական կորը (Test Characteristic Curve) բնութագրում է թեստի միավորների (true score) և փորձարկվողի պատրաստվածության մակարդակի (θ) միջև կախվածությունը: Այն ստացվում է առաջադրանքների բնութագրական կորերի գումարից:

Բիրնբաումի երկու պարամետրով միաճյուղ մոդել

Ռաշի մեկ պարամետրով միաճյուղ մոդելի դեպքում առաջադրանքը բնութագրվում է միայն իր բարդությամբ (IRT «b» պարամետր): CTT-ի նմանությամբ անհրաժեշտություն է դառնում առաջադրանքը բնութագրել նաև իր տարբերակիչ ունակությամբ (discrimination ability) : Ռաշի մոդելի կատարելագործված տարբերակում, որը կոչվում է Բիրնբաումի երկու պարամետրով մոդել, փորձարկվողի թեստը հաջողությամբ հանձնելու ֆունկցիայի պարամետրերի շարքում ներառվում է նաև տվյալ առաջադրանքի տարբերակիչ ունակության «a» պարամետրը՝ item discrimination parameter: Իսկ Բիրնբաումի երեք պարամետրերով մոդելում ներառվում է նաև տվյալ թեստային առաջադրանքի ճիշտ պատասխանը գուշակելու հավանականությունը:

Երկու պարամետրով միաճյուղ մոդելը մաթեմատիկորեն ներկայացվում է հետևյալ բանաձևի տեսքով՝

$$P_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}}, \quad (11)$$

որտեղ a_i -ն i-րդ առաջադրանքի տարբերակիչ ունակությունն է:

Ինչպես երևում է բանաձևից, երբ $a_i = 1$, մոդելը փոխակերպվում է մեկ պարամետրանոց միաճյուղ մոդելի: Ինչպես մեկ պարամետրով մոդելի դեպքում, այնպես էլ այս դեպքում առաջադրանքի բնութագրական կորը մոնոտոն աճում է և ունի

շրջման կետ: Շրջման կետի արքսիսը հավասար է «b» պարամետրի արժեքին, իսկ օրդինատը 0,5 արժեքին:

Առաջադրանքի բնութագրական կորի վրա շրջման կետում կորի և շոշափողի հետ կազմած անկյան տանգենսը հավասար է $a/4$: Կարելի է համոզվել, որ

$$\lim_{\theta \rightarrow -\infty} P(\theta; b, a, D) = 0,$$

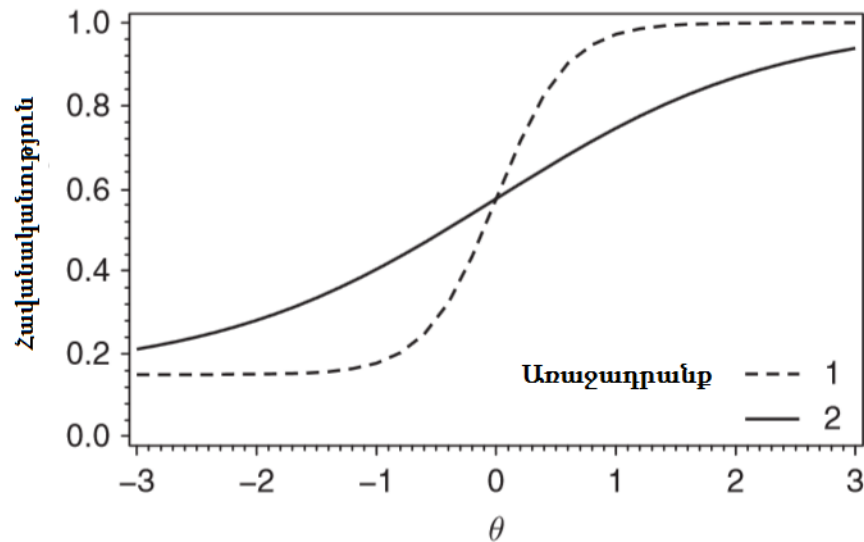
$$\lim_{\theta \rightarrow +\infty} P(\theta; b, a, D) = 1:$$

Շրջման կետում բնութագրական կորի և շոշափողի թեքության անկյունը a պարամետրի արժեքից կախված է: Որքան մեծ է a պարամետրի արժեքը, այնքան մեծ է թեքության անկյունը և հակառակը:

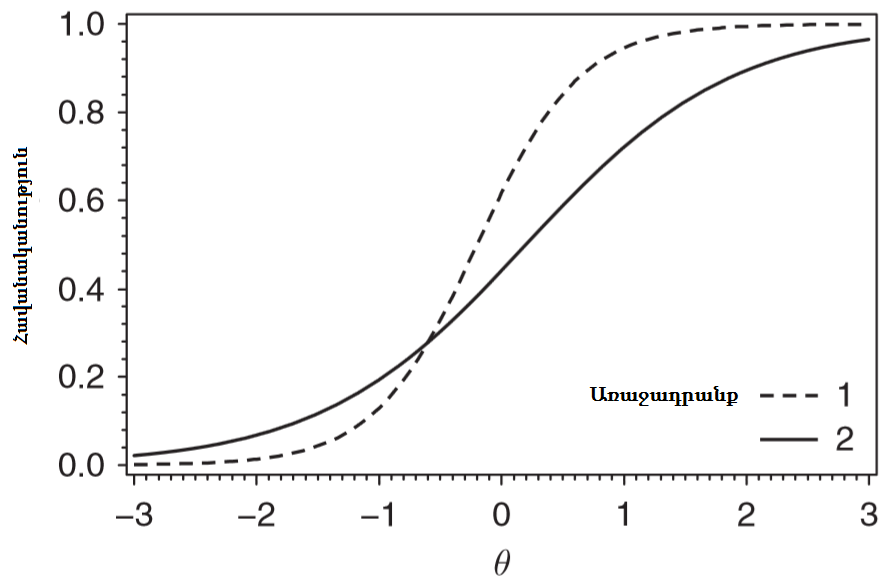
a պարամետրի փոքր արժեքների դեպքում b արժեքից փոքր-ինչ մեծ պատրաստվածության մակարդակով փորձարկվողների կողմից տվյալ առաջադրանքը ճիշտ կատարելու հավանականությունը գործնականորեն հավասար է b արժեքից փոքր-ինչ փոքր պատրաստվածության մակարդակով փորձարկվողների կողմից տվյալ առաջադրանքը ճիշտ կատարելու հավանականությանը: Այսինքն՝ այս դեպքում առաջադրանքը շատ թույլ է տարբերակում պատրաստվածության տարբեր մակարդակով փորձարկվողներին:

a պարամետրի մեծ արժեքների դեպքում b արժեքից փոքր-ինչ մեծ պատրաստվածության մակարդակով փորձարկվողների կողմից տվյալ առաջադրանքը ճիշտ կատարելու հավանականությունը խիստ տարբերվում է b արժեքից փոքր-ինչ փոքր պատրաստվածության մակարդակով փորձարկվողների կողմից տվյալ առաջադրանքը ճիշտ կատարելու հավանականությունից: Իսկ այս դեպքում առաջադրանքը լավ տարբերակում է պատրաստվածության բարձր և ցածր մակարդակով փորձարկվողներին:

Նկար 3-ում ներկայացված են երկու առաջադրանքներ տարբեր « a » պարամետրերով: «Առաջադրանք 1»-ն ունի ավելի մեծ տարբերակիչ ունակություն, քան «Առաջադրանք 2»-ը: Այսինքն «Առաջադրանք 1»-ը կարող է ավելի լավ տարբերակել համեմատաբար բարձր θ մակարդակով և համեմատաբար ցածր θ մակարդակով փորձարկվողներին:



Նկար 3. Երկու առաջադրանքների տարբեր «a» պարամետրերով:



Նկար 4. Երկու առաջադրանքների տարբեր «a» և «b» պարամետրերով:

Բիրնբաումի երեք պարամետրով մոդել

Թեստերում շատ հաճախ կիրառվում են հնարավոր տարբերակներով առաջադրանքներ, հետևաբար առաջ է գալիս պատահականությամբ պատասխանելու գործոնը: Ուստի խնդիր է դրվում առաջադրանքը բնութագրել ոչ միայն երկու պարամետրերով, այլև մի նոր պարամետրով, որը կբնութագրի տվյալ առաջադրանքը պատահականությամբ ճիշտ գուշակելու հավանականությունը: Բիրնբաումի երեք պարամետրով մոդելում ներառվում է նոր պարամետր՝ «C» պարամետր [38] և արտահայտվում է հետևյալ տեսքով՝

$$P_i(\theta) = C_i + (1 - C_i) \frac{1}{1 + e^{-Da_i(\theta - b_i)}}, \quad (12)$$

որտեղ C_i –ն նվազագույն պատրաստվածության մակարդակ ունեցող փորձարկվողների կողմից i -րդ առաջադրանքը ճիշտ պատասխանելու հավանականությունն է: C պարամետրը ընդունում է արժեքներ $[0,1]$ միջակայքից:

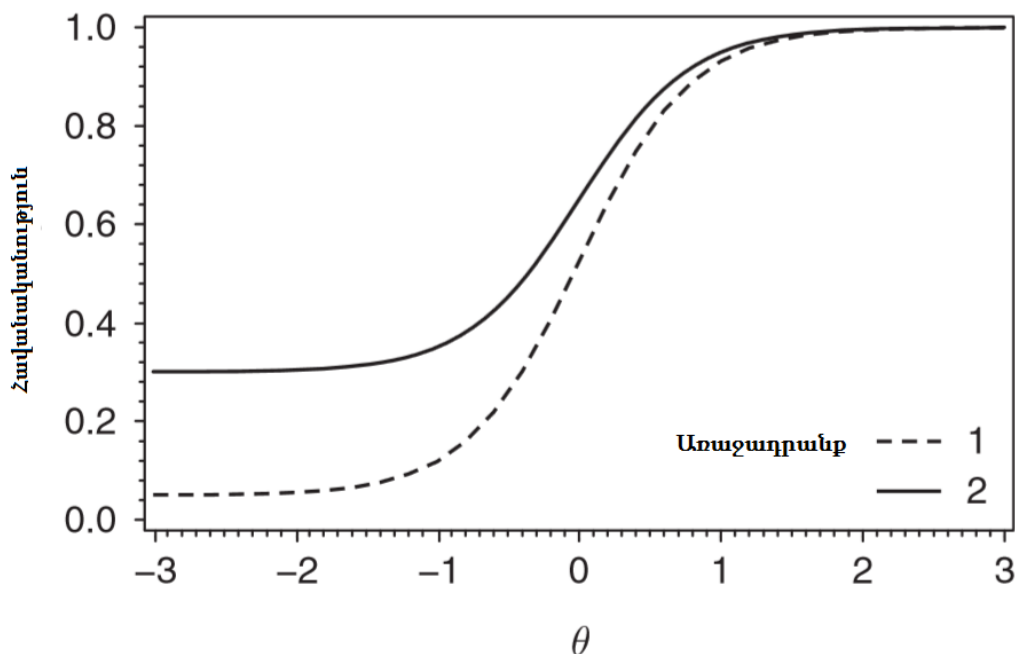
Կարելի է նկատել, որ $C = 0$ դեպքում մոդելը ձևափոխվում է երկու պարամետրով մոդելի, իսկ $C = 0, a = 1$ դեպքում մեկ պարամետրով մոդելի: Երեք պարամետրով մոդելի դեպքում առաջադրանքի բնութագրական կորի արժեքների բազմությունը դառնում է ոչ թե $[0,1]$ միջակայքը, այլ $[C,1]$ միջակայքը: Բնութագրական կորի շրջման կետի աբսցիսը դառնում է b , իսկ դրան համապատասխանող օրդինատի արժեքը՝ $\frac{C+1}{2}$: Առաջադրանքի բնութագրական կորի վրա շրջման կետում շոշափողի հետ կազմած անկյան տանգենսը հավասար է $\frac{a(1-C)}{4}$:

(9)-ից հետևում է, որ

$$\lim_{\theta \rightarrow -\infty} P(\theta; b, a, c, D) = C,$$

$$\lim_{\theta \rightarrow +\infty} P(\theta; b, a, c, D) = 1,$$

այսինքն՝ անվերջ փոքր պատրաստվածությամբ փորձարկվողի կողմից տվյալ առաջադրանքը ճիշտ կատարելու հավանականությունը C -է:



Նկար 5. Առաջադրանքներ տարբեր C պարամետրերով: Առաջադրանք 2 –ն ունի C պարամետրի ավելի մեծ արժեք:

Այս մոդելում ի տարբերություն նախորդների ենթադրվում է, որ նույնիսկ ոչինչ չիմացող փորձարկվողն ունի առաջադրանքը ճիշտ կատարելու հավանականություն, որի արժեքը 0-ից տարբեր է: Նկար 5-ում ներկայացված են երկու առաջադրանքների բնութագրական կորեր:

Բազմաճյուղ մոդելներ

Բազմաճյուղ առաջադրանքների գնահատման համար օգտագործվում են մի շարք մոդելներ [76]:

Մասնակի միավորներով մոդելը (PCM) միաճյուղ մեկ պարամետրով մոդելի ընդլայնված տարբերակն է : (10) -ը կարելի է գրել հետևյալ տեսքով՝

$$P_{ix}(\theta) = \frac{1}{1 + e^{-D(\theta - b_i)}} = \frac{\exp(D(\theta - b_i))}{1 + \exp(D(\theta - b_i))} = \frac{P_{i1}(\theta)}{P_{i0}(\theta) + P_{i1}(\theta)}, \quad (13)$$

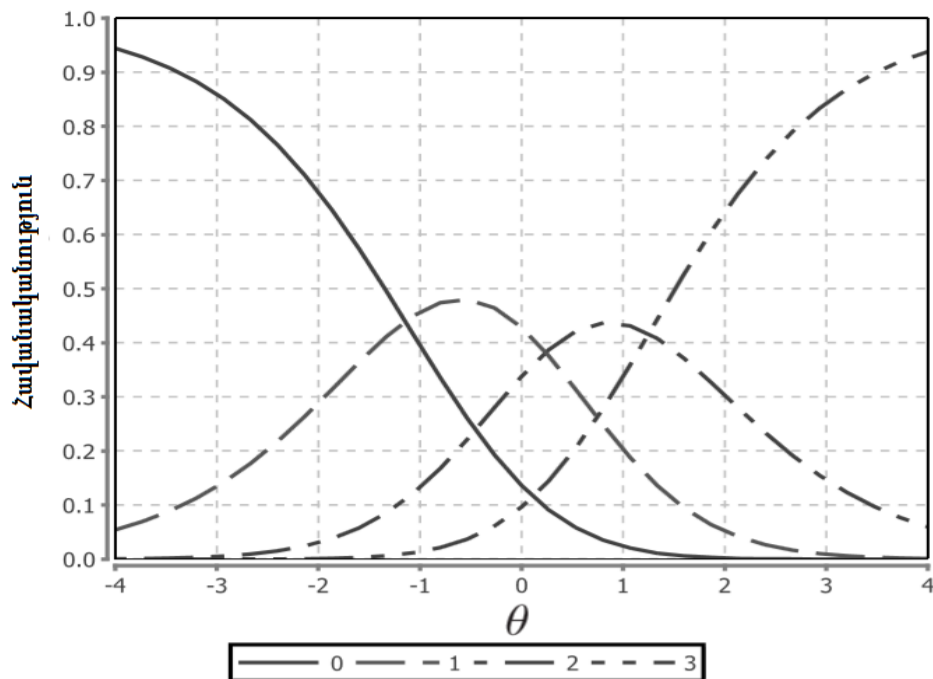
որտեղ $P_{i1}(\theta)$ -ն պատահականությամբ ընտրված θ պատրաստվածության մակարդակ ունեցող փորձարկվողի կողմից 1 միավորով գնահատվող i -րդ առաջադրանքը ճիշտ կատարելու հավանականությունն է, իսկ $P_{i0}(\theta)$ -ն պատահականությամբ ընտրված θ պատրաստվածության մակարդակ ունեցող փորձարկվողի կողմից 0 միավորով գնահատվող i -րդ առաջադրանքի ճիշտ կատարելու հավանականությունն է:

Այսպիսով X միավորանոց և m_i հնարավոր միավորների կարգեր ունեցող i -րդ առաջադրանքի դեպքում մասնակի միավորներով բազմաճյուղ մոդելն արտահայտվում է հետևյալ բանաձևով՝

$$P_{ix}(\theta) = \frac{\exp \sum_{k=0}^x (D(\theta - b_{ik}))}{\sum_h^{m_i} \exp \sum_{k=0}^h (D(\theta - b_{ik}))}, x = 1, 2, \dots, m_i, \quad (14)$$

(14) բանաձևով արտահայտվող ֆունկցիան հաճախ կոչվում է միավորների կատեգորիաների պատասխանների ֆունկցիա (score category response function - SCRF):

Նկար 6-ում պատկերված են չորս կատեգորիաներով բազմաճյուղ մասնակի միավորներով մոդելի առաջադրանքի տարբերակների բնութագրական կորերը (option characteristic curves):



Նկար 6.2 Որս կատեգորիաներով բազմաճյուղ մասնակի միավորներով մոդելի առաջադրանքի տարբերակների բնութագրական կորերը:

Ընդհանրացված մասնակի միավորներով մոդել (GPCM)

Ընդհանրացված մասնակի միավորներով մոդելի և մասնակի միավորներով մոդելի միջև տարբերությունը առաջադրանքի տարբերակիչ ունակության a պարամետրի առկայությունն է: Այս մոդելն ունի հետևյալ տեսքը՝

$$P_{ix}(\theta) = \frac{\exp \sum_{k=0}^x (Z_{ik}(\theta))}{\sum_h^{m_i} \exp \sum_{k=0}^h (Z_{ik}(\theta))}$$

$$Z_{ik}(\theta) = D_{ai}(\theta - b_i + d_{ix}), \quad (15)$$

որտեղ d_{ix} -ը i -րդ առաջադրանքի x -րդ կատեգորիայի հարաբերական բարդությունն է:

Հաշվի առնելով, որ

$$b_{ix} = b_i - d_{ix},$$

(15) արտահայտությունը կարելի է գրել հետևյալ տեսքով՝

$$Z_{ik}(\theta) = D_{ai}(\theta - b_{ix}):$$

Ինֆորմացիոն ֆունկցիա

Առաջադրանքների բնութագրական ֆունկցիաները օգտագործվում են փորձարկվողի պատրաստվածության մակարդակի և առաջադրանքը ճիշտ կատարելու հավանականության միջև կախվածությունը բնութագրելու համար, իսկ այդ չափումների որակը նկարագրելու համար օգտագործվում է ինֆորմացիոն ֆունկցիան: Առաջադրանքի ինֆորմացիոն ֆունկցիան ($I_i(\theta)$) նկարագրում է առաջադրանքի՝ փորձարկվողի պատրաստվածության մակարդակի չափման վրա ազդեցությունը: Որքան մեծ է ինֆորմացիայի քանակը, այնքան մեծ է ազդեցությունը փորձարկվողի պատրաստվածության մակարդակի գնահատման վրա, կամ այլ կերպ ասած այնքան առաջադրանքը որոշիչ է փորձարկվողի պատրաստվածության մակարդակի գնահատման գործում: IRT-ում ինֆորմացիոն ֆունկցիան օգտագործվում է չափումների ստանդարտ սխալի գնահատման համար:

Առաջադրանքի ինֆորմացիոն (Ֆիշերի) ֆունկցիան [13] արտահայտվում է հետևյալ բանաձևի տեսքով՝

$$I_i(\theta) = \frac{(P'_i(\theta))^2}{P_i(\theta) \cdot Q_i(\theta)}, \quad (16)$$

որտեղ $P_i(\theta)$ -ն i -րդ առաջադրանքին ճիշտ պատասխանելու հավանականության ֆունկցիան է (item response function), $Q_i(\theta)$ -ն սխալ պատասխանելու հավանականության ֆունկցիան է, որը որոշվում է հետևյալ կերպ.

$$Q_i(\theta) = 1 - P_i(\theta) = \frac{1}{1 + e^{D(\theta - b_i)}},$$

իսկ $P'_i(\theta) = D * P_i(\theta) * Q_i(\theta)$:

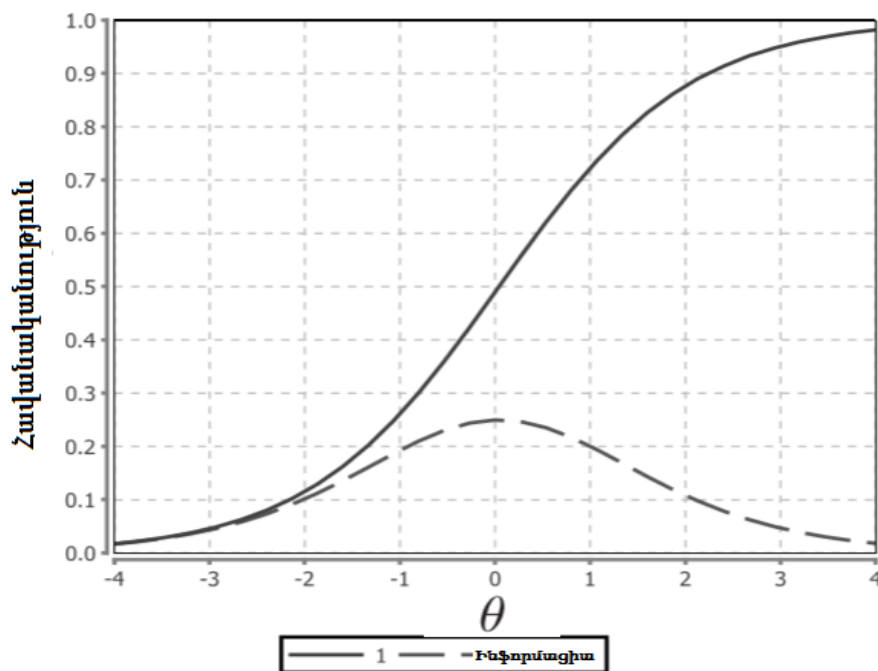
Մեկ պարամետրով մոդելի առաջադրանքի ինֆորմացիոն ֆունկցիան նկարագրվում է հետևյալ բանաձևով՝

$$I_i(\theta, b_i) = P_i(\theta, b_i) Q_i(\theta, b_i) \quad (17)$$

Շատ հեշտ է տեսնել, որ առաջադրանքի ինֆորմացիոն ֆունկցիայի մաքսիմում արժեքը 0,25 է: Դա համապատասխանում է այն կետին, որի դեպքում ճիշտ պատասխանելու և սխալ պատասխանելու հավանականությունների արժեքները միաժամանակ 0,5 են: Քանի որ առաջադրանքի բարդությունը որոշվում է ճիշտ

պատասխանելու հավանականության ֆունկցիայի 0,5 արժեքին համապատասխան, ուստի մեկ պարամետրով մոդելի ցանկացած առաջադրանք առավել ինֆորմատիվ է այն փորձարկվողների համար, որոնց պատրաստվածության մակարդակը համարժեք է առաջադրանքի բարդության մակարդակին (0,5): Եթե պատրաստվածության մակարդակը ավելի մեծ է կամ ավելի փոքր է առաջադրանքի բարդությունից, ապա առաջադրանքի ինֆորմացիան նվազում է: Այս կախվածությունը բերված է նկար 7-ում:

Նկար 7-ում պատկերված առաջադրանքի բարդությունը 0,04 է ($P_i(\theta) = 0.5$), և ինֆորմացիոն ֆունկցիան իր առավելագույն արժեքն ընդունում է, երբ փորձարկվողի պատրաստվածության մակարդակը 0,04 է: Այսպիսով՝ նկար 7-ում ներկայացված առաջադրանքը լավագույն ձևով ճիշտ է չափում այն փորձարկվողներին, որոնց պատրաստվածության մակարդակը 0,04 է: Իհարկե, առաջադրանքը տալիս է օգտակար ինֆորմացիա նաև այն փորձարկվողների մասին, որոնց պատրաստվածության մակարդակը գտնվում է -1,0-ից 1,0 միջակայքում:



Նկար 7. Առաջադրանքի բնութագրական կորը և ինֆորմացիոն ֆունկցիան:

Երկու պարամետրերով մոդելի դեպքում ինֆորմացիոն ֆունկցիան ունի հետևյալ տեսքը՝

$$I_i(\theta, b_i, a_i) = a_i^2 P_i(\theta, b_i) Q_i(\theta, b_i) \quad (18)$$

Երկու պարամետրերով մոդելի դեպքում ևս առաջադրանքի ինֆորմացիոն ֆունկցիան իր մաքսիմում արժեքն ընդունում է առաջադրանքի բարդության արժեքի դեպքում: Սակայն ֆունկցիայի տեսքը և մաքսիմումի արժեքները խիստ կախված են տարբերակիչ ունակության a պարամետրի արժեքներից: Այն դեպքերում, երբ տարբերակիչ ունակությունը բարձր է, առաջադրանքը տալիս է ավելի շատ ինֆորմացիա փորձարկվողի պատրաստվածության վերաբերյալ և ինֆորմացիան կենտրոնանում է առաջադրանքի բարդության արժեքի շուրջ: Ցածր տարբերակիչ ունակությամբ առաջադրանքները ավելի քիչ ինֆորմատիվ են և ինֆորմացիան ցրվում է պատրաստվածության մակարդակի միջակայքի գրեթե ամբողջ երկայնքով:

Երեք պարամետրերով մոդելի դեպքում ինֆորմացիոն ֆունկցիան արտահայտվում է հետևյալ բանաձևով [33]՝

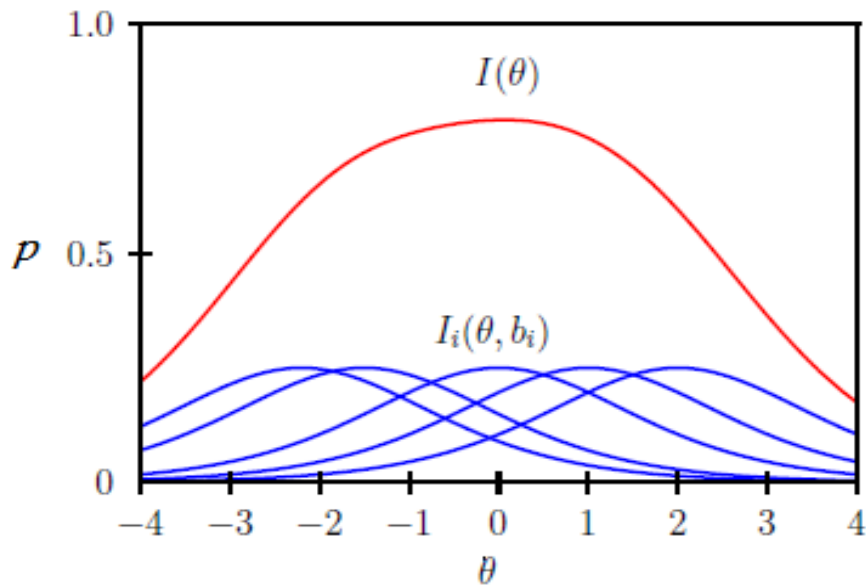
$$I_i(\theta; a_i; b_i; c_i) = a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[\frac{P_i(\theta) - c_i}{1 - c_i} \right]^2 : \quad (19)$$

Այս դեպքում առաջադրանքն ավելի շատ ինֆորմացիա է տալիս փորձարկվողի պատրաստվածության վերաբերյալ c_i պարամետրի փոքր արժեքների դեպքում:

Ամբողջ թեստի ինֆորմացիոն ֆունկցիան (test information function-TIF) որոշվում է տվյալ թեստի առաջադրանքների ինֆորմացիոն ֆունկցիաների միավորման միջոցով: Թեստի ինֆորմացիոն ֆունկցիան հավասար է առաջադրանքների ինֆորմացիոն ֆունկցիաների գումարին՝

$$I(\theta) = \sum_{i=1}^J I_i(\theta):$$

Նկար 8-ում ներկայացված է մեկ պարամետրով հինգ առաջադրանքների ինֆորմացիոն ֆունկցիաները և դրանց հիման վրա կառուցված թեստի ինֆորմացիոն ֆունկցիան:



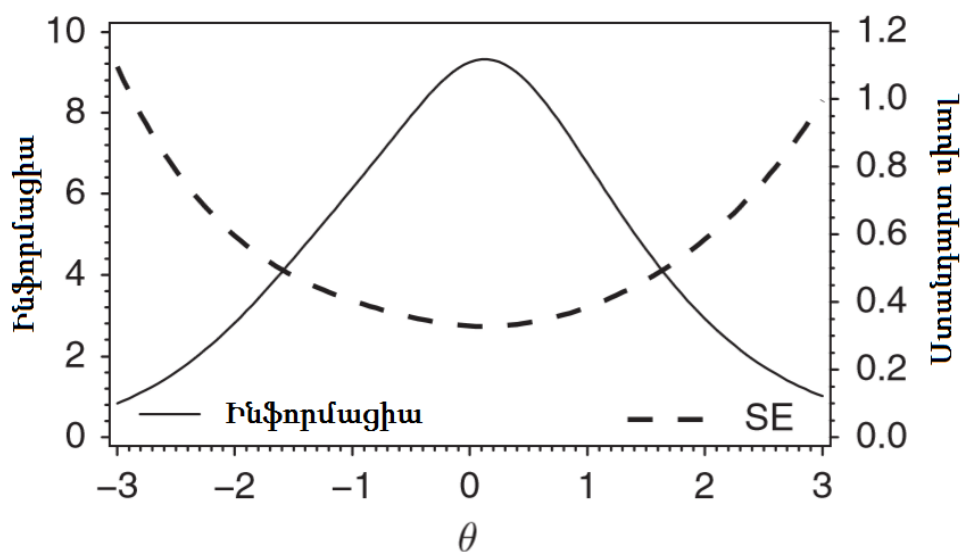
Նկար 8. Առաջադրանքների և թեստի ինֆորմացիոն ֆունկցիան:

Փորձարկվողի պատրաստվածության մակարդակի θ արժեքների ցրվածքը ($Var(\theta)$) արտահայտվում է ինֆորմացիոն ֆունկցիայի միջոցով՝

$$Var(\theta) = \frac{1}{I(\theta)},$$

իսկ θ արժեքների չափման ստանդարտ սխալը ($SE(\theta)$) արտահայտվում է հետևյալ բանաձևով՝

$$SE(\theta) = \sqrt{Var(\theta)} = \frac{1}{\sqrt{I(\theta)}} :$$



Նկար 9. Ինֆորմացիոն ֆունկցիայի և չափման ստանդարտ սխալի ֆունկցիայի միջև կապը

Ինֆորմացիայի քանակի փոքր արժեքներին համապատասխանում են չափման ստանդարտ սխալի մեծ արժեքներ և հակառակը: Ինչպես երևում է նկար 9-ում ներկայացված թեստի ինֆորմացիոն ֆունկցիայի և չափման ստանդարտ սխալի ֆունկցիայի գրաֆիկներից, այն կետերում, որտեղ ինֆորմացիան մեծ է, չափման ստանդարտ սխալը փոքր է:

1.4. Եզրակացություններ

Գոյություն ունի կախվածություն թեստերի դասական և մաթեմատիկական տեսությամբ որակական պարամետրերի միջև [68][69]: Այսպես՝ կախվածություն կա CTT կետային բիսերիալ գործակցի և IRT տարբերակիչ ունակության a պարամետրի միջև, ինչպես նաև CTT բարդության ինդեքսի և IRT բարդության b պարամետրի միջև:

Թեստերի դասական և մաթեմատիկական տեսություններից յուրաքանչյուրն ունի իր առավելություններն ու թերությունները [53][74]:

Թեստերի դասական տեսության կարևոր առավելություններից են.

- վերլուծության համար համեմատաբար քիչ քանակությամբ թեստավորման արդյունքներ են հարկավոր,
- առավել պարզ մաթեմատիկական վերլուծություն է կատարվում համեմատած մաթեմատիկական տեսության հետ,
- բնութագրիչ պարամետրերի գնահատման մոդելը շատ պարզ է,
- վերլուծությունների համար անհրաժեշտություն չկա խորը ուսումնասիրություններ կատարել մոդելի ճիշտ ընտրության հարցում:

Թեստերի մաթեմատիկական տեսության կարևոր առավելություններից են.

- թեստային առաջադրանքի բարդության գնահատականը կախված չէ փորձարկվողների խմբի ընտրությունից,
- փորձարկվողների պատրաստվածության մակարդակի գնահատականը կախված չէ կիրառվող թեստային առաջադրանքների բանկից,
- IRT մոդելները թույլ են տալիս նկարագրել թեստային առաջադրանքի բարդության և փորձարկվողի պատրաստվածության մակարդակի միջև կապը,

- առաջադրանքների և փորձարկվողների պարամետրերը չափվում են նույն սանդղակով:

Ռաշի մոդելի առավելություններն ամբողջությամբ ներկայացված են Մ. Վիլսոնի աշխատանքում [88]:

Չնայած մի շարք առավելություններին և արդյունքների մշակման պարզությանը, թեստերի դասական տեսությունն ունի որոշակի թերություններ.

- թեստային առաջադրանքի բարդությունը և տարբերակիչ ունակությունը, որոնց արժեքները ստացվում են վիճակագրական տվյալների վերլուծության հիման վրա, կախված են փորձարկվողների խմբի ընտրությունից,
- փորձարկվողների պատրաստվածության մակարդակի գնահատականը կախված է թեստում թեստային առաջադրանքների բարդությունից:

Թեստերի մաթեմատիկական տեսության թերություններից կարելի է նշել.

- բարդ մաթեմատիկական ապարատը,
- մոդելների բազմազանությունը, ինչն էլ պահանջում է խորը գիտելիքներ՝ կախված թեստի տեսակից համապատասխան ճիշտ մոդելի ընտրությունը վավեր վերլուծության համար:

ԱՌԿԱ ՀԱՄԱԿԱՐԳԵՐԻ ՀԵՏԱԶՈՏՈՒԹՅՈՒՆ

**2.1. Թեստավորման արդյունքների վերլուծության ժամանակակից առավել
լայն կիրառվող ծրագրերի հետազոտություն**

Թեստերի վիճակագրական փորձաքննության անցկացման համար անհրաժեշտություն է դառնում կիրառել համակարգեր, ծրագրային փաթեթներ [32], որոնք թեստերի տեսություններից որևէ մեկի կամ երկուսի հիման վրա կիրականացնեն թեստավորման արդյունքների վերլուծություն և որակական բնութագրերի գնահատում: Կատարվել է նմանատիպ առավել լայն կիրառվող համակարգերի հետազոտություն, առանձնահատկությունների, առավելությունների և թերությունների բացահայտում:

IRT տեսությամբ վերլուծություն իրականացնող համակարգչային ծրագրերը (IRT ծրագրեր) անվանվում են նաև Rasch Measurement Analyse Software [87]: IRT ծրագրերը մշակվել են դեռևս 1970-ականներին և աշխատել են DOS միջավայրում: Այդպիսի ծրագրերից են Bigsteps-ը [36], Facets-ը [50], GENIRV-ը [34], RESCEN-ը [73] և այլն: Ներկայումս DOS միջավայրում աշխատող IRT ծրագրերը կապված մի շարք սահմանափակումների հետ շատ նեղ կիրառություն ունեն ` ցածր արագագործություն (16 բիթ), համակարգի միջոցների սահմանափակ օգտագործում, անհամատեղելիություն 32 և 64 բիթանոց Windows օպերացիոն համակարգերի հետ, դեպի ոչ օգտվող-կողմնորոշված ինտերֆեյս՝ հրամանային տողի ինտերֆեյս (Command Line Interface):

Ներկայումս լայն կիրառություն են գտել Windows-ի հենքով, օգտվող կողմնորոշված ինտերֆեյսով IRT ծրագրերը:

CITAS-ը [40] (Classical Item and Test Analysis Spreadsheet) Excel ձևաչափի հասարակ ֆայլ է, որի միջոցով անցկացվում է թեստավորման արդյունքների վերլուծություն հիմնված թեստերի դասական տեսության վրա: Ստացվում են այնպիսի ցուցանիշներ, ինչպիսիք են թեստի միավորների միջին արժեքը և միջին շեղումը,

հուսալիությունը, չափումների սխալների ստանդարտ արժեքը (SEM), թեստային առաջադրանքների կետային-բիսերիալ գործակիցները, ինչպես նաև ստացվում է առաջադրանքներին տրված տարբերակների վերլուծությունը: CITAS-ով կարելի է վերլուծել ոչ ավել, քան 50 փորձարկվողի և 50 միաճյուղ թեստային առաջադրանքից կազմված թեստի արդյունքներ: Այն հասանելի է նաև OpenOffice.org ձևաչափով և մանկավարժների համար շատ հարմար, պարզ, շատ ժամանակ և այլ միջոցներ չպահանջող գործիք է՝ թեստավորման արդյունքների վերլուծության կազմակերպման նպատակով: CITAS-ը անվճար է և հասանելի է www.assess.com կայքում:

Iteman 4-ը [58] Windows հավելված է, որը հնարավորություն է տալիս ստանալ թեստերի և թեստային առաջադրանքների մանրամասն վերլուծության հաշվետվություն՝ հիմնված թեստերի դասական տեսության վրա: Հաշվետվության նպատակն է օգտագործել ստացված ցուցանիշները թեստերի որակական բնութագրերը գնահատելու համար:

Թեստավորման արդյունքները և թեստային առաջադրանքների տվյալները վերբեռնվում են .txt կամ .csv ձևաչափի ֆայլերի տեսքով, իսկ վերլուծության արդյունքը ստացվում է RTF (Word) ձևաչափով առանձին ֆայլի տեսքով, որն ունի խմբագրման հնարավորություն: Հիմնական առանձնահատկություններն են.

- արդյունքների վերլուծությունը տրվում է Word ձևաչափով,
- տրվում է գրաֆիկների տեսքով վերլուծություն,
- ապահովվում է թեստի, թեստային առաջադրանքների, առաջադրանքների տարբերակների մակարդակով վերլուծություն,
- իրականացվում է հուսալիության բազմազան գործակիցների հաշվարկ:

Թեստավորում անցկացնող մի շարք կազմակերպություններ օգտագործում են այս ծրագիրը իրենց հաճախորդներին թեստավորման արդյունքների վերլուծության հաշվետվություններ տրամադրելու համար: Ծրագիրը վճարովի է, տրամադրվում է մի քանի լիցենզիաներով, նվազագույն գինը կազմում է \$495 և հասանելի է www.assess.com կայքում: Վերլուծվող արդյունքներում փորձարկվողների թվի սահմանափակում չկա, իսկ առաջադրանքների առավելագույն քանակը կազմում է 10.000: Հասանելի է նաև անվճար ցուցադրական տարբերակը, որի միջոցով կարելի է

իրականացնել ոչ ավել, քան 50 փորձարկվողի և 50 թեստային առաջադրանքից կազմված թեստերի վերլուծություն:

Xcalibre 4 - ը [92] IRT տեսությամբ թեստային առաջադրանքների վերլուծության իրականացման Windows հավելված է: Xcalibre 4 -ով թեստի վերլուծության դեպքում կիրառվում են 4 միաճյուղ և 5 բազմաճյուղ IRT մոդելներ, տրվում է մանրամասն և ամփոփ վերլուծություն, որն իր մեջ ներառում է աղյուսակներ և գրաֆիկներ: Այս ծրագիրը շատ հեշտ է կիրառման համար, ունի օգտվող-կողմնորոշված ինտերֆեյս, չկա ծրագրավորման կոդերով աշխատելու անհրաժեշտություն: Թեստավորման արդյունքները և թեստային առաջադրանքների տվյալները վերբեռնվում են .txt կամ .csv ձևաչափի ֆայլերի տեսքով: Վերլուծության հաշվետվությունը ստացվում է rich text file (RTF) ձևաչափով, այսինքն՝ վերլուծության արդյունքների հիման վրա հաշվետվություն մշակելու անհրաժեշտություն չկա: Գրաֆիկները ներառում են առաջադրանքի ինֆորմացիոն ֆունկցիան (IIF), թեստի ինֆորմացիոն ֆունկցիան (TIF), չափման սխալի պայմանական ստանդարտ արժեքը (CSEM), բազմաթիվ հաճախությունների բաշխման ֆունկցիաներ և այլն: Սատարվող IRT մոդելներն են.

- 1 պարամետրով միաճյուղ մոդել (1PLM),
- 2 պարամետրով միաճյուղ մոդել (2PLM),
- 3 պարամետրով միաճյուղ մոդել (3PLM),
- Ռաշի միաճյուղ մոդել, որը սանդղակավորում է թեստային առաջադրանքի և ոչ թե փորձարկվողի բնութագրերը,
- Ռաշի միավորների սանդղակավորման մոդել (RRSM կամ RSM),
- Ռաշի մասնակի միավորներով մոդել (RPCM կամ PCM),
- Ընդհանրացված միավորների սանդղակավորման մոդել (GRSM),
- Ընդհանրացված մասնակի միավորներով մոդել (GPCM),
- Samejima-ի մոդել (SGRM կամ GRM):

Առանձնահատկություններից է նաև այն, որ վերլուծության որոշ թվային տվյալներ ստացվում են նաև CSV ձևաչափով, որը շատ հեշտությամբ կարելի է կիրառել էլեկտրոնային աղյուսակների հետ աշխատող ծրագրերում: Թույլ է տալիս

վերլուծել 1500 թեստային առաջադրանք: Xcalibre 4 ծրագիրը մշակված է և պարբերաբար թարմացվում է «[Assessment Systems Corporation](http://www.assessment-systems.com)» [31] կազմակերպության կողմից, վճարովի է և կան մի շարք լիցենզիոն տարբերակներ, նվազագույն գինը կազմում է \$495: www.assessment.com կայքում հասանելի է ծրագրի ցուցադրական տարբերակը, որը կարելի է կիրառել ոչ ավել, քան 50 փորձարկվողի և 50 թեստային առաջադրանքից կազմված թեստերի վերլուծության դեպքերում:

Winsteps-ը [64] Windows-ի հենքով ծրագիր է, որի միջոցով IRT տեսության հիման վրա անցկացվում է Ռաշի մոդելով վերլուծություն (Rasch Analysis) [80]: Մշակվել է Benjamin Wright-ի և John Michael Linacre-ի [66] կողմից 1980-ականներին Չիկագոյի համալսարանում: Կիրառվում է ուսումնական թեստերի, հանրային հարցումների, ռեյտինգավորման սանդղակավորման վերլուծությունների կազմակերպման համար: Առաջադրանքների վերլուծությունն իր մեջ ներառում է միաճյուղ Ռաշի մոդել, IRT մեկ պարամետրով բազմակի ընտրությամբ (MCQ), միավորների սանդղակավորման (RSM), մասնակի միավորներով (PCM) մոդելները, յուրաքանչյուրն իր մեջ մինչ 255 կատեգորիայով: Ներկայացվում է վերլուծությունը ըստ տարբեր կատեգորիաների ինչպես առաջադրանքների, այնպես էլ փորձարկվողների մակարդակով: Վերլուծությունն իր մեջ ներառում է և թվային տվյալներ, աղյուսակներ և բազմաթիվ գրաֆիկներ:

Ծրագրի վերաբերյալ առկա են բազմաբովանդակ և ծավալուն մի շարք ուղեցույցներ՝ ինչպես առցանց էլեկտրոնային ձեռնարկի տեսքով, այնպես էլ առանձին ներբեռնվող ձևաչափով: Ծրագիրը փոխգործակցում է Excel, R, SAS, SPSS, STATA, Txt ֆայլերի հետ, որը հնարավորություն է տալիս թեստավորման արդյունքները վերբեռնել վերոնշյալ ֆայլերի տեսքով: Ծրագիրն ապահովվում է շուրջ 40 տիպի աղյուսակների, ֆայլերի, գրաֆիկների ստացում: Առանձնահատուկ ֆունկցիոնալություն ունի գրաֆիկների ներկայացման բաժինը, որտեղ առկա են գրաֆիկների հետ աշխատելու մի շարք ղեկավարման էլեմենտներ: Գրաֆիկների միջոցով ստացվում են թեստերի, թեստային առաջադրանքների, փորձարկվողների մի շարք բնութագրեր: Winsteps-ի մշակողների խմբից այն մշտապես կատարելագործվում է, վերջին տարբերակը Winsteps 3.81.0-ն է, որը թողարկվել է 2014թ. փետրվարին: Ծրագրի միջոցով կարելի է

իրականացնել 1.000.000 փորձարկվողների և 30.000 առաջադրանքի վերլուծություն: Ծրագիրը վճարովի է, սակայն կա նաև ցուցադրական տարբերակը՝ MiniStep [72], որով կարելի է իրականացնել ոչ ավել, քան 75 փորձարկվողի և 25 առաջադրանքի վերլուծություն: Ծրագիրը ներկայացված է <http://www.winsteps.com> պաշտոնական կայքում, որտեղ առկա են մի շարք նկարագրություններ, օգտվողի ուղեցույցներ [64], Rasch մոդելի չափումների վերաբերյալ հրատարակություններ:

Ինչպես և Winsteps-ը, **Facets-ը** [65] ևս Windows-ի հենքով ծրագիր է, որը նույնպես մշակվել է John Michael Linacre -ի կողմից: Նախատեսված է Rasch չափումներով ավելի բարդ կիրառությունների համար: Հզոր և ճկուն ծրագիր է, իր մեջ ներառում է բոլոր այն մոդելները և հնարավորությունները, որոնք առկա են Winsteps-ում և ապահովում է Winsteps-ում բացակայող many-facet Rasch մոդելը: Many-facet Rasch մոդելը կիրառվում է, երբ անհրաժեշտ է լինում կոմբինացված վերլուծել այնպիսի փորձարկումների արդյունքներ, որտեղ կիրառվում են հետերոգեն, տարբեր թեստեր և առաջադրանքներ: Facets ծրագիրը հնարավորություն է տալիս շատ ճկուն ձևով մեկ վերլուծության մեջ ներկայացնել տարբեր ձևաչափի առաջադրանքների արդյունքները:

Facets-ի միջոցով վերլուծության արդյունքում կարելի է ստանալ համապատասխան սանդղակով արդյունքների ֆայլերը, որոնք որպես մուտքային ֆայլեր շատ հեշտությամբ մուտքագրվում են Winsteps-ում ավելի բարդ Rasch մոդելներով վերլուծություն կազմակերպելու նպատակով: Facets-ի կիրառելը Winstep-ի համեմատությամբ ավելի բարդ է, հեղինակի կողմից խորհուրդ է տրվում Facets-ը կիրառել հատկապես Many-facet Rasch մոդելով վերլուծություն կատարելու համար [89]: Facets-ը ևս վճարովի է, արժեքը կազմում է 149\$, հնարավորություն է տալիս վերլուծել շուրջ 1.000.000 փորձարկվողի տվյալ: Անվճար ցուցադրական տարբերակը կոչվում է Minifac [71], որով կարելի է վերլուծել մինչև 2000 փորձարկվող/առաջադրանք պատասխաններ: Facets-ում համեմատաբար բարդ մոդելների հաշվարկների ժամանակ պահանջվում է նշանակալից ժամանակահատված, խորհուրդ է տրվում կիրառել նվազագույնը 1GB օպերատիվ հիշողություն: Օգտվողի ուղեցույցները և Minifac ծրագիրը հասանելի են www.winsteps.com կայքում:

jMetrik-ն անվճար ծրագիր է՝ նախատեսված psychometric արդյունքների վերլուծություններ կատարելու համար [62]: JMetrik-ը Java հավելված է և գործարկվում է Windows, Mac OSX և Linux օպերացիոն համակարգերի միջավայրերում: Ունի օգտվողին կողմնորոշված ինտերֆեյս, ինտեգրված տվյալների բազա և ներառում է բազմազան վիճակագրական մեթոդներ և գրաֆիկներ: Թեստավորման արդյունքները մուտքագրվում են առանձին .txt, .csv ձևաչափի ֆայլերի տեսքով, որոնք վիզուալ միջավայրում աղյուսակի տեսքով դառնում են հասանելի նաև խմբագրման համար: Մուտքագրված տվյալների հիման վրա էլ ստեղծվում է տվյալների բազա: Թեստավորման փորձարկվողների և թեստային առաջադրանքների քանակի նկատմամբ չկա սահմանափակում, դա կախված է մեքենայի հիշողությունից: Թեստը վերլուծվում է և CTT և IRT տեսությունների հիման վրա: Ապահովվում է մի շարք մոդելներ՝ Ռա2, 2PL, 3PL, PCM, RSM, GRM, GPCM, ինչպես նաև վերջիններիս կոմբինացիաները: Հաշվարկվում են թեստի հուսալիության, կոռելյացիայի, չափումների ստանդարտ շեղման մի շարք գործակիցներ:

Ծրագիրն ապահովում է թեստերի, թեստային առաջադրանքների, փորձարկվողների վերաբերյալ գրաֆիկների ստացումը, ինչպես նաև խմբագրումը:

Վերլուծության արդյունքները ստացվում են ծրագրի միջավայրում և կարելի է բեռնել .txt ձևաչափով, իսկ գրաֆիկները՝ JPG, PNG ձևաչափով: Ծրագիրը, թեստավորման արդյունքների նմուշային ֆայլերը, ուղեցույցը հասանելի են <http://www.jmetrik.com/> կայքում: Վճարովի տարբերակով հասանելի է նաև «Applied Measurement with jMetrik» [70] գիրքը:

RUMM 2030 ծրագիրը Windows հավելված է, որը թույլ է տալիս իրականացնել վերլուծություն Ռա2ի մոդելի հիման վրա [26]: Այն մշակվել է RASCH-Andrich սանդղակավորման մոդելի հեղինակ Դ. Էնդրիչի և Ավստրալիայի Perth լաբորատորիայի աշխատակիցների կողմից [25]: Կիրառվում է ինչպես միաճյուղ, այնպես էլ բազմաճյուղ սանդղակներով թեստերի դեպքում, ապահովում է ըստ Ռա2ի մոդելի անհրաժեշտ տվյալների և գրաֆիկների ստացում: Այս ծրագիրը շատ կիրառելի է հատկապես ուսումնական գործընթացում՝ Ռա2ի մոդելով չափումների տեսությունը ուսումնասիրելիս: Հնարավորություն է տրվում տարբեր հայտնի էլեկտրոնային

աղյուսակների ձևաչափերից (Excel կամ SPSS) հեշտությամբ պատճենել և տեղադրել տվյալներ: Թերություն է համարվում լիցենզիայի գինը և ժամկետայնությունը, մեկ ակադեմիական տարվա համար այն կազմում է 700\$, ուսուցման համար անհրաժեշտ է մասնակցել դասընթացների:

ACER ConQuest[23] ծրագիրն իր մեջ համակցում է item response և latent regression մոդելները: Մշակվել է «Australian Council for Educational Research (ACER)»-ում Լ. Մարագարետի, Զ. Ադամսի, Մ. Վիլսոնի կողմից [91]: Ծրագիրն ապահովում է վերլուծություն և գնահատում հետևյալ մոդելներով.

- Ռաշ մոդելներ,
- Միավորների սանդղակավորման մոդել,
- Մասնակի միավորներով մոդել,
- Կարգավորված միջնորմային մոդել (Ordered Partition Model),
- Գծային տրամաբանական թեստային մոդել,
- Բազմակողմանի մոդելներ (Multifaceted),
- Ընդհանրացված միաչափ մոդելներ,
- Բազմաչափ IRT մոդելներ,
- Լատենտային նվազումով մոդելներ (Latent Regression Models):

Կարևոր առանձնահատկություններից մեկն այն է, որ այն կարող է գնահատել ոչ միայն միաչափ, այլ նաև բազմաչափ IRT մոդելով [39]: Մուտքային տվյալները կարելի է վերբեռնել անմիջապես SPSS ֆայլերի տեսքով: Արդյունքները կարելի է ստանալ Excel կամ SPSS ձևաչափերով: ConQuest-ն ունի և գրաֆիկական ինտերֆեյս և Console հրամանային տողի ինտերֆեյս: Ապահովում է բազմազան գունավոր, ինֆորմացիոն գրաֆիկների և քարտեզների ստացում: Սակայն կապված այն հանգամանքի հետ, որ այստեղ կիրառվում են չափման բազմազան մոդելներ և սանդղակներ, չկա տվյալների քանակի սահմանափակություն, ունի մեծ գնային արժեք: ConQuest -ը գործարկվում է Windows-ի միջավայրում և գինը կազմում է \$699 : Australian Council for Educational Research-ի պաշտոնական <http://www.acer.edu.au/> կայքում առկա է ծրագրի կարճ նկարագրությունը, ինչպես նաև ներբեռնելու համար հասանելի է ցուցադրական տարբերակը:

IRTPRO-ն [57] Windows հավելված է, որը մշակվել է SSI (Scientific Software International) –ի կողմից: Նախատեսված է IRT վերլուծություն իրականացնելու համար: Ապահովում է հետևյալ մոդելներով վերլուծություն.

- 1 պարամետրով միաճյուղ մոդել (1PLM),
- 2 պարամետրով միաճյուղ մոդել (2PLM),
- 3 պարամետրով միաճյուղ մոդել (3PLM),
- ընդհանրացված մասնակի միավորներով մոդել (GPCM),
- գնահատումով մոդել,
- նոմինալ մոդել:

IRTPRO-ում տվյալները կարելի է մուտքագրել .csv, .fixed, .txt, .xls ձևաչափերով: Մուտքագրված տվյալները պահվում են IRTPRO տվյալների ֆայլի տեսքով (.ssig), որը ծրագրի վիզուալ միջավայրում ցուցադրվում է էլեկտրոնային աղյուսակի տեսքով: Ծրագրի միջավայրում կարելի է ստանալ բազմաթիվ գունավոր գրաֆիկներ: Ունի օգտվողի ծավալուն ուղեցույց, որտեղ նկարագրվում է նաև IRT տեսությունը: Ծրագիրը վճարովի է , տրվում է լիցենզիայի մի քանի տարբերակ, նվազագույն վճարը կազմում է 495\$: Կարելի է պատվիրել նաև ցուցադրական 15 օրվա ժամկետով տարբերակը, որտեղ կան հետևյալ սահմանափակումները.

- թեստային առաջադրանքների թիվը 25,
- փորձարկվողների թիվը՝ 1000:

ConstructMap [43] ծրագիրը մշակվել է The Berkeley Evaluation and Assessment Research (BEAR) կենտրոնում (ԱՄՆ) [35]: Առկա է ծրագրի երկու տարբերակ՝ նախատեսված Windows և MacOS օպերացիոն համակարգերի համար : Ծրագրի միջոցով կարելի է իրականացնել վերլուծություն ինչպես միաճյուղ գնահատումով Ռաշի կիրառությամբ, այնպես էլ բազմաճյուղ գնահատումով մասնակի միավորներով կամ ռեյտինգային սանդղակներով մոդելների կիրառությամբ վերլուծություն:

Թեստավորման արդյունքները կարելի է մուտքագրել txt կամ Excel ֆայլերի տեսքով: Ծրագրի փաթեթում առկա են մի շարք պատրաստի օրինակներ: Հնարավորություն է տրվում ստանալ վերլուծության բազմաթիվ արդյունքներ, որոնցում առկա են ինչպես IRT, այնպես էլ CTT մի շարք ցուցանիշներ: Կարելի է ստանալ

թեստի, թեստային առաջադրանքների, փորձարկվողների վերաբերյալ բնութագծերը՝ (TIF, ICC և այլն) գրաֆիկների տեսքով: Վերլուծության արդյունքները կարելի է պահպանել .txt ֆայլի տեսքով, իսկ գրաֆիկները՝ PNG կամ JPG ձևաչափերով: Ծրագիրը, նկարագրությունը, օգտվողի ուղեցույցները հասանելի են պաշտոնական կայքում: ConstructMap-ը անվճար է, տեղադրման համար պահանջվում է Java միջավայր:

IATA ծրագիրը [78] մշակվել է Կանադայում գտնվող «Polymetrika International» ծրագրավորման կազմակերպության կողմից: Այս ծրագիրը desktop հավելված է, աշխատում է Windows միջավայրում և տեղադրման համար պահանջվում է Net Framework միջավայր: Ծրագիրը հնարավորություն է տալիս վերլուծել հոգեբանական և ուսումնական թեստերի տվյալներ՝ ապահովել IRT, DIF վերլուծություն: IATA-ում տվյալները կարելի է մուտքագրել .xls, .xlsx, .csv, .xml, .sav, .txt ձևաչափերով: Գրաֆիկները, հաշվարկվող ցուցանիշների աղյուսակները ստացվում են ինտերակտիվ ռեժիմում՝ ծրագրի պատուհանների սահմաններում: Ծրագիրն անվճար է, այն և օգտվողի ուղեցույցը հասանելի են <http://polymetrika.com/home/IATA> կայքում: ՀՀ կառավարության աշխատակազմի «Գնահատման և թեստավորման կենտրոն» ՊՈԱԿ-ում [2] կիրառվում է այս ծրագրի հայերեն տարբերակը:

2.2. Եզրակացություններ

Այսպիսով, կատարված հետազոտությունից երևում է, որ թեստերի որակական փորձաքննության անցկացման ծրագրերի հիմնական տարբերությունը սատարվող IRT մոդելների ցանկն է, վերլուծության ենթարկվող փորձարկվողների և առաջադրանքների թիվը, ստացվող տվյալների աղյուսակների, գրաֆիկների քանակը, արդյունքների հաշվետվությունների ստացման ձևաչափերը և այլն:

Winsteps և Facets ծրագրերն առանձնանում են վերլուծության արդյունքում ստացվող աղյուսակների բազմազանությամբ և քանակով, ստացվող գրաֆիկների հետ աշխատելու վիզուալ միջավայրի հնարավորությամբ, ինչպես նաև բազմաբովանդակ ուղեցույցների առկայությամբ:

Կարևոր է նաև արդյունքների ստացման ձևաչափերը: Այս տեսանկյունից առանձնացվում են հատկապես IteMan 4, Xcalibre 4 ծրագրերը, որոնց միջոցով վերլուծության արդյունքները ներկայացվում են պատրաստի ամբողջական հաշվետվությունների տեսքով՝ rtf ձևաչափով: Հաշվետվությունն իր մեջ ներառում է ինչպես թվային տվյալներով աղյուսակներ, այնպես էլ գրաֆիկներ, նաև բնութագրերի վերաբերյալ տեղեկատվություն: Նմանատիպ հաշվետվությունները շատ ավելի մատչելի են մանկավարժների համար: Իսկ մնացած ծրագրերում թվային տվյալները հասանելի են txt ձևաչափով, գրաֆիկները՝ png, jpg ձևաչափերով:

Ուսումնասիրված ծրագրերից մեծամասնությունում սահմանափակում է դրվում վերլուծվող թեստի փորձարկվողների, առաջադրանքների քանակի վրա: Այս տեսանկյունից առանձնանում է jMetrik ծրագիրը, որտեղ չկա նմանատիպ սահմանափակում և որում ներմուծվող տվյալների հիման վրա ստեղծվում է վիզուալ միջավայրում ցուցադրվող աղյուսակ, որն էլ պահպանվում է տվյալների բազայում:

Կարևոր առանձնահատկություն է նաև մուտքագրվող և ստացվող ֆայլերի ձևաչափերի սատարման բազմազանությունը: Հիմնականում թեստավորման արդյունքները և թեստային առաջադրանքների պատասխանների ֆայլերը մուտքագրվում են txt, երբեմն նաև Excel կամ CSV ձևաչափով: Մի շարք ծրագրերում (Winsteps, Facets, RUMM 2030) հնարավորություն է տրվում փոխգործակցել նաև R, SAS, SPSS, STATA ձևաչափի ֆայլերի հետ:

IRT վերլուծություն իրականացնելու ժամանակ կարևոր է տվյալ ծրագրի կողմից սատարվող մոդելների բազմազանությունը: Գրեթե բոլոր IRT ծրագրերում գոնե կիրառվում են 1PL, 2PL, 3PL մոդելները: Acer ConQuest ծրագիրն առանձնանում է նրանով, որ այստեղ հնարավորություն է ստեղծված վերլուծել նաև բազմաչափ IRT մոդելով: Սատարվող մոդելների բազմազանությամբ աչքի են ընկնում հատկապես Xcalibre 4, WinSteps, ACER ConQuest ծրագրերը:

Ուսումնասիրված ծրագրերը հիմնականում Windows-ի հենքով աշխատող ծրագրեր են, շատ հազվադեպ են այն ծրագրերը, որոնք աշխատում են նաև Linux, MacOS օպերացիոն համակարգերի միջավայրերում:

Ծրագրերից շատերը կիրառվում են ոչ միայն թեստավորման արդյունքների վերլուծության նպատակով, ինչն էլ ավելի է բարդացնում վերջիններիս ֆունկցիոնալությունը, ինչպես նաև բազմազան են դառնում չափվող սանդղակները:

Ուսումնասիրված ծրագրերը հիմնականում անգլերեն և վճարովի են, ունեն առաջարկվող տարբեր լիցենզիոն փաթեթներ, հասանելի են նաև այդ ծրագրերի ցուցադրական փորձնական տարբերակները, որոնցից շատերում առկա են ֆունկցիոնալ հնարավորությունների տեսանկյունից սահմանափակումներ:

Ժամանակակից առավել լայն կիրառվող ծրագրերի թերություններից կարելի է նշել մի քանիսը:

- Նման համակարգերը հիմնականում չունեն օգտվողին կողմնորոշված պարզ և մատչելի ինտերֆեյս, երբեմն պահանջում են հրամանային տողով աշխատելու հմտություններ, որոշ դեպքերում նաև տեխնիկական խորը գիտելիքներ:
- Հիմնականում վճարովի են և կապված բազմագործառության հետ ունեն բարձր արժեքներ:
- IRT –ով վերլուծություն անցկացնող ծրագրերը բազմագործառային են և կիրառվում են տարբեր չափումներ գնահատելու համար, իսկ միայն թեստավորման վերաբերյալ վերլուծություն անցկացնելու համար անհրաժեշտ է ծրագրից գտնել, առանձնացնել և խմբավորել միայն թեստավորման մոդուլները, որը այդքան էլ դյուրին գործ չէ:
- Հիմնականում անգլերեն լեզվով են, շատ հազվադեպ նաև ռուսերեն լեզվով:
- Ունեն բարդ մաթեմատիկական ապարատ, ինչը կիրառվում է ոչ միայն թեստավորման վերլուծության համար: Մանկավարժների կողմից բարդ է նման ապարատի, բազմազան բնութագրերի ընկալումը:
- Թեստավորման արդյունքների վերլուծությունը հիմնականում ստացվում է բազմաթիվ աղյուսակների տեսքով, որոնք պահպանվում են txt ձևաչափով: Գրաֆիկներն էլ իրենց հերթին ստացվում են առանձին ֆայլերի տեսքով՝ jpg կամ png ձևաչափով: Այսինքն՝ մեկ միասնական ֆայլի տեսքով հաշվետվություն

ստանալու համար անհրաժեշտ է լինում տարբեր ֆայլերից կատարել խմբագրումներ և ստանալ նոր հաշվետվություն, որը շատ ավելի հարմար և կիրառելի է մանկավարժի համար:

- Արդյունքների վերլուծությունից բացի, չի տրամադրվում գնահատվող որակական բնութագրերի մանրամասն նկարագրությունը և մեթոդական ցուցումներ՝ որակական բնութագրերի արժեքների փոփոխման վերաբերյալ: Այսպիսով՝ խնդիր է առաջադրվում ունենալ հայկական շուկայի համար նմանատիպ մի համակարգ, որը
- կիրականացնի թեստավորման արդյունքների վերլուծություն և որակի գնահատում CTT-ի և IRT-ի հիման վրա,
- կունենա այն առանձնահատկությունները, որոնք բնորոշ են նմանատիպ համակարգերին,
- չի ունենա վերը նկարագրված թերությունները,
- կլինի հասանելի հայերեն լեզվով, կունենա շատ պարզ և հասկանալի ինտերֆեյս մանկավարժների կողմից օգտագործելու տեսանկյունից,
- վերլուծության արդյունքները կներկայացվեն խմբագրվող rtf ձևաչափի հաշվետվության տեսքով և կտրվի գնահատված որակական բնութագրերի մանրամասն նկարագրություն, ինչպես նաև մանկավարժների համար կտրամադրի մեթոդական ցուցումներ այս կամ այն բնութագրի արժեքը փոփոխելու համար:

ԳԼՈՒԽ III.

ԹԵՍՏԵՐԻ ՈՐԱԿԻ ԳՆԱՀԱՏՄԱՆ ՆՈՐ ՄՈԴԵԼԻ ՄՇԱԿՈՒՄ

Մենք առաջարկում ենք թեստերի որակի գնահատման նոր մոդել [55]՝ հիմնված ինֆորմացիայի չափի մեծությունների վրա, մասնավորապես, Շենոնի էնտրոպիայի, պայմանական էնտրոպիայի և միջին փոխադարձ ինֆորմացիայի վրա [44]:

3.1. Շենոնի ինֆորմացիայի չափի մեծությունները և դրանց հատկությունները

Ենթադրենք թեստը կազմված է N թեստային առաջադրանքներից և յուրաքանչյուր առաջադրանք դիտարկվում է որպես պատահական մեծություն: Այս թեստային առաջադրանքները (փոփոխականները) կնշանակենք $X_1, X_2, \dots, X_N$: Յուրաքանչյուր փոփոխական մեծություն կարող է ընդունել 2 արժեք՝ 1 ճիշտ պատասխանելու դեպքում և 0 սխալ պատասխանելու դեպքում՝ համապատասխանաբար p և $1-p$ հաճախականություններով, որտեղ p -ն թեստին տրված ճիշտ պատասխանների քանակի հարաբերությունն է փորձարկողների ընդհանուր քանակին՝

$$X_i = \begin{cases} 1 & p_i \text{ հաճախականությամբ,} \\ 0 & 1 - p_i \text{ հաճախականությամբ,} \end{cases} \quad i = \overline{1, N}$$

X պատահական մեծության էնտրոպիան տրվում է հետևյալ բանաձևով՝

$$H(X) = - \sum_{i=1}^N p(x_i) \log p(x_i) \quad (20)$$

Այստեղ մենք դիտարկում ենք լոգարիթմի հիմքը 2, և էնտրոպիան արտահայտվում է բիթերով: Ընդունված է, որ $0 \log 0 = 0$, ինչը հեշտությամբ հիմնավորվում է անընդհատությամբ, քանի որ $x \log x \rightarrow 0$, երբ $x \rightarrow 0$: Այսպիսով, զրոյական հավանականությամբ արժեքները ավելացնելուց էնտրոպիան չի փոխվում:

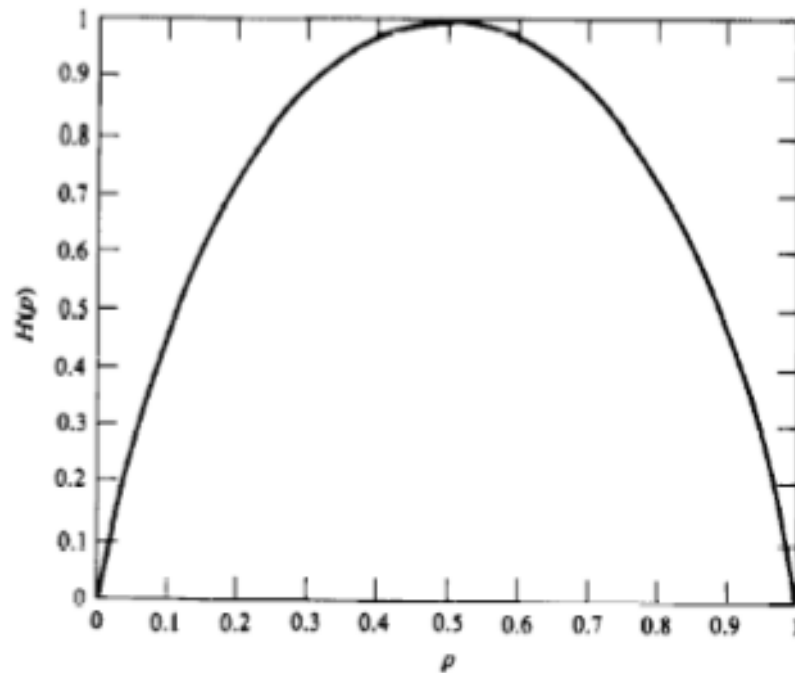
Ելնելով սեփական ինֆորմացիայի մեկնաբանությունից՝ էնտրոպիան կարելի է մեկնաբանել մի կողմից՝ որպես X պատահական մեծության անորոշության չափանիշ, մյուս կողմից՝ որպես մեկ այլ պատահական մեծության մասին միջինում տրվող ինֆորմացիայի ամենամեծ քանակ:

Քանի որ մեր դեպքում X պատահական մեծությունը ընդունում է երկու արժեք՝ p և $1-p$ հավանականություններով:

Էնտրոպիան կնշանակենք $H(p_i)$, որը հավասար է հետևյալ արտահայտությանը

$$H(X) = -p_i \log p_i - (1 - p_i) \log(1 - p_i) = H(p_i) \quad (21)$$

$H(p)$ ֆունկցիայի գրաֆիկը ներկայացված է նկար 10-ում:



Նկար 10. $H(p)$ ֆունկցիայի գրաֆիկը:

Մասնավոր դեպքում, երբ $p = 1/2$, էնտրոպիան հավասար է 1-ի: Գրաֆիկի վրա կարելի է տեսնել էնտրոպիայի հիմնական հատկությունները. էնտրոպիան p -ի ուռուցիկ ֆունկցիա է և հավասարվում է 0-ի, երբ $p = 0$ կամ $p = 1$: Սա իմաստ ունի, քանի որ $p = 0$ կամ $p = 1$ դեպքում X -ը պատահական չէ և, հետևաբար, անորոշությունը 0 է: Մյուս կողմից՝ անորոշությունը մեծագույնն է, եթե $p = 1/2$:

Նաև դիտարկվել է X_i մեծության պայմանական էնտրոպիան X_j ընդհատ մեծության պայմանում՝

$$H(X_i|X_j) = \sum_{x_i, x_j} p(x_i, x_j) \log \frac{1}{p(x_i|x_j)}, \quad (22)$$

որտեղ $p(x_i, x_j)$ -ն երկու առաջադրանքների համատեղ հավանականությունն է, իսկ $p(x_i|x_j) = p(x_i, x_j)/p(x_j)$ -ը պայմանական հավանականությունն է:

Երկու առաջադրանքների միջին փոխադարձ ինֆորմացիան հաշվարկվում է հետևյալ բանաձևով՝

$$I(X_i \wedge X_j) = \sum_{x_i, x_j} p(x_i, x_j) \log \frac{p(x_i, x_j)}{p(x_i) * p(x_j)} = \quad (23)$$

$$H(X_i) - H(X_i | X_j) = H(X_j) - H(X_j | X_i):$$

Նոր մոդելի ձևակերպման համար օգտագործվել են էնտրոպիայի և միջին փոխադարձ ինֆորմացիայի հետևյալ հատկությունները՝

Հատկություն 1. Երկուական տվյալների դեպքում (0 կամ 1) տեղի ունեն հետևյալ անհավասարությունները՝

$$0 \leq H(X) \leq 1,$$

ընդ որում՝ հավասարությունը ձախից տեղի ունի այն և միայն այն դեպքում, երբ $p_i = 0$ կամ $p_i = 1$, իսկ աջից հավասարությունը տեղի ունի այն և միայն այն դեպքում, երբ $p_i = 1/2$:

Հատկություն 2. Տեղի ունեն հետևյալ անհավասարությունները՝

$$0 \leq I(X_i \wedge X_j) \leq \min[H(X_i), H(X_j)], \quad i, j = \overline{1, N}$$

ընդ որում՝ հավասարությունը ձախից տեղի ունի այն և միայն այն դեպքում, երբ X_i և X_j մեծությունները միմյանցից անկախ են:

Հատկություն 3. Պայմանական էնտրոպիան չի գերազանցում ոչ պայմանականին՝

$$H(X_i|X_j) \leq H(X_i), \quad i, j = \overline{1, N}$$

ընդ որում՝ հավասարությունը տեղի ունի այն և միայն այն դեպքում, երբ X_i և X_j մեծությունները միմյանցից անկախ են:

3.2. Թեստերի որակի գնահատման նոր մոդելի նկարագրությունը

Մեր կողմից առաջարկված մոդելը հիմնված է նախորդ ենթազրկում նկարագրված Շենոնի ինֆորմացիայի չափի մեծություններ հատկությունների վրա: Ձևակերպվել են հետևյալ մեթոդները [41]:

Մեթոդ 1. Եթե $H(X_i)$ մեծության արժեքը մոտ է 0-ին, ապա թեստային առաջադրանքը վատն է, այսինքն՝ կամ շատ բարդ է կամ շատ հեշտ: Եթե $H(X_i)$ մեծության արժեքը մոտ է 1-ին, ապա թեստային առաջադրանքը լավն է, այսինքն՝ ունի միջին բարդություն:

Մեթոդ 2. Եթե $I(X_i \wedge X_j)$ մեծության արժեքը մոտ է 0-ին, ապա X_i և X_j թեստային առաջադրանքները միմյանցից անկախ են: Իսկ եթե արժեքը մոտ է $\min[H(X_i), H(X_j)]$ -ին, ապա X_i և X_j թեստային առաջադրանքները կրկնում են մեկը մյուսին և փոքր էնտրոպիա ունեցողը պետք է հեռացվի թեստից:

Մեթոդ 3. $H(X_i|X_j)$ պայմանական էնտրոպիայի արժեքը ցանկալի է մոտ լինի $H(X_i)$ -ին, այսինքն X_j առաջադրանքը չնչին ազդեցություն ունենա X_i -ի անորոշության վրա:

Ըստ նկարագրված մոդելի ներկայացնենք թեստի որակական պարամետրերի հաշվարկի օրինակ: Հաշվարկի համար ընտրված են 10 թեստային առաջադրանքից բաղկացած 1000 փորձարկվողի կողմից կատարված թեստավորման արդյունքներ:

Թեստավորման արդյունքների տվյալները և հաշվարկված առաջադրանքների $H(P)$, CTT բարդության և IRT և պարամետրի արժեքները ներկայացված են աղյուսակ 6-ում:

Ըստ CTT տեսության առաջադրանքները լավն են, եթե բարդության արժեքը գտնվում է 0,3-ից 0,74 միջակայքում: Ըստ ներկայացված տվյալների X_4 և X_8 առաջադրանքները հեշտ են, իսկ մնացածների բարդությունը միջին է: Նույն արդյունքին կարող ենք հանգել, եթե վերլուծենք $H(X)$ -ի արժեքները: Ընդ որում՝ ինչպես CTT տեսության, այնպես էլ մեր առաջարկված մոդելի հիման վրա կարող ենք ասել, որ X_4 առաջադրանքն ավելի հեշտ է, քան X_8 -ը: $H(X)$ -ի արժեքները նաև համադրելի են IRT և բարդության պարամետրերի արժեքների հետ: Ըստ IRT

տեսության X_4 և X_8 առաջադրանքները համեմատաբար ավելի հեշտ առաջադրանքներ են:

Աղյուսակ 6.

Առաջադրանքների հաշվարկված $H(X)$, CTT բարդության և IRT b պարամետրի արժեքները

Առաջադրանք	p_i ճիշտ պատասխանների համար	$1 - p_i$ սխալ պատասխանների համար	$H(X)$ էնտրոպիա	CTT բարդություն	IRT b պարամետր
X_1	0.558	0.442	0.99	0.558	0.36
X_2	0.408	0.592	0.98	0.408	1.52
X_3	0.633	0.367	0.95	0.633	-0.19
X_4	0.850	0.150	0.61	0.85	-1.03
X_5	0.597	0.403	0.97	0.597	0.07
X_6	0.446	0.554	0.99	0.446	1.61
X_7	0.486	0.514	1.00	0.486	0.42
X_8	0.754	0.264	0.80	0.754	-0.75
X_9	0.516	0.484	1.00	0.516	0.46
X_{10}	0.532	0.468	1.00	0.532	0.74

CTT տեսության շրջանակներում հաշվարկվել են առաջադրանքների միջև կոռելյացիայի գործակցի արժեքները: Ինչպես հայտնի է, գործակիցը արժեքներ է ընդունում -1-ից +1 միջակայքից, և ցանկալի է թեստում կիրառել 0,3-ից փոքր կամ հավասար կոռելյացիայի գործակցի արժեք ունեցող առաջադրանքներ: Եթե կոռելյացիայի գործակիցը ընդունում է +1-ին մոտ արժեքներ, դա նշանակում է, որ այդ թեստային առաջադրանքները անկախ չեն և կրկնում են մեկը մյուսին: Ըստ աղյուսակ 7-ում ներկայացված տվյալների X_6 և X_{10} առաջադրանքներն ունեն բացասական կոռելյացիա, այսինքն՝ այստեղ նկատվում է առաջադրանքների մեծ անկախություն:

CTT կոռելյացիայի ($R(X_i, X_j)$) արժեքները:

$R(X_i, X_j)$	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.0000	0.0915	0.1244	0.0716	0.0898	0.0816	0.1926	0.1462	0.1212	0.0530
X2	0.0915	1.0000	0.0411	0.0923	0.0059	0.0370	0.1495	0.0868	0.0630	0.0528
X3	0.1244	0.0411	1.0000	0.0927	0.1104	0.0404	0.1509	0.2106	0.1261	0.1216
X4	0.0716	0.0923	0.0927	1.0000	0.1173	0.0727	0.0835	0.2087	0.0863	0.1111
X5	0.0898	0.0059	0.1104	0.1173	1.0000	0.0810	0.1381	0.1366	0.1099	0.0752
X6	0.0816	0.0370	0.0404	0.0727	0.0810	1.0000	0.1258	0.0781	0.0196	-0.025
X7	0.1926	0.1495	0.1509	0.0835	0.1381	0.1258	1.0000	0.1745	0.1290	0.0740
X8	0.1462	0.0868	0.2106	0.2087	0.1366	0.0781	0.1745	1.0000	0.1763	0.0925
X9	0.1212	0.0630	0.1261	0.0863	0.1099	0.0196	0.1290	0.1763	1.0000	0.0701
X10	0.0530	0.0528	0.1216	0.1111	0.0752	-0.025	0.0740	0.0925	0.0701	1.0000

Նույն եզրահանգումը կարող ենք կատարել, եթե վերլուծենք աղյուսակ 8-ում ներկայացված միջին փոխադարձ ինֆորմացիայի ($I(X_i \wedge X_j)$) արժեքները:

Շատ հեշտ է տեսնել, որ X_6 և X_{10} առաջադրանքների միջին փոխադարձ ինֆորմացիան շատ մոտ է 0-ին, ինչն էլ նշանակում է, որ վերջիններս համարյա անկախ են (մեթոդ 2):

Աղյուսակ 9-ում ներկայացված են առաջադրանքների պայմանական էնտրոպիայի ($H(X_i \wedge X_j)$) արժեքները:

Ըստ առաջարկված մեթոդ 3-ի X_6 և X_{10} առաջադրանքների միջև ևս նկատվում է անկախություն, քանի որ ըստ աղյուսակ 9-ի տվյալների X_6 և X_{10} առաջադրանքների պայմանական էնտրոպիայի արժեքը մոտ է X_{10} առաջադրանքի էնտրոպիայի արժեքին:

Աղյուսակ 8. Առաջադրանքների միջին փոխադարձ ինֆորմացիայի $(I(X_i \wedge X_j))$ արժեքները:

$(I(X_i \wedge X_j))$	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	-	0.006068	0.011149	0.003675	0.00581	0.00481	0.026966	0.015342	0.010612	0.002029
X2	0.006068	-	0.001223	0.00634	0.000025	0.000985	0.016165	0.005522	0.002866	0.002013
X3	0.011149	0.001223	-	0.006039	0.008746	0.00118	0.016544	0.031166	0.011492	0.01067
X4	0.003675	0.00634	0.006039	-	0.009746	0.003868	0.005073	0.028188	0.005386	0.008915
X5	0.00581	0.000025	0.008746	0.009746	-	0.004745	0.013825	0.013286	0.00873	0.004075
X6	0.00481	0.000985	0.00118	0.003868	0.004745	-	0.011435	0.004439	0.000277	0.000461
X7	0.026966	0.016165	0.016544	0.005073	0.013825	0.011435	-	0.022326	0.012042	0.003952
X8	0.015342	0.005522	0.031166	0.028188	0.013286	0.004439	0.022326	-	0.022594	0.006161
X9	0.010612	0.002866	0.011492	0.005386	0.00873	0.000277	0.012042	0.022594	-	0.00355
X10	0.002029	0.002013	0.01067	0.008915	0.004075	0.000461	0.003952	0.006161	0.00355	-

Աղյուսակ 9. Առաջադրանքների պայմանական էնտրոպիայի $(H(X_i \wedge X_j))$ արժեքները:

$(H(X_i \wedge X_j))$	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	-	0.96937 1	0.93719 2	0.606165	0.966869	0.98676	0.97246 8	0.78953 4	0.98865	0.995014
X2	0.984204	-	0.94711 8	0.6035	0.972653	0.99058 4	0.98326 9	0.79935 5	0.99639 5	0.995031
X3	0.979123	0.97421 6	-	0.603801	0.963932	0.99038 9	0.98289	0.77371 1	0.98776 9	0.986374
X4	0.986597	0.96909 8	0.94230 2	-	0.962933	0.98770 2	0.99436 2	0.77668 8	0.99387 6	0.988129
X5	0.984462	0.97541 3	0.93959 4	0.600094	-	0.98682 5	0.98561	0.79159	0.99053 1	0.992969
X6	0.985462	0.97445 3	0.94716	0.605973	0.967934	-	0.988	0.80043 7	0.99898 5	0.996582
X7	0.963305	0.95927 3	0.93179 6	0.604768	0.958854	0.98013 5	-	0.78255 1	0.98721 9	0.993092
X8	0.974929	0.96991 7	0.91717 5	0.581652	0.959392	0.98713 1	0.97710 9	-	0.97666 8	0.990883
X9	0.97966	0.97257 2	0.93684 9	0.604455	0.963948	0.99129 3	0.98739 2	0.78228 3	-	0.993493
X10	0.988242	0.97342 6	0.93767 1	0.600926	0.968604	0.99110 9	0.99548 3	0.79871 6	0.99571 1	-

3.3. Եզրակացություններ

Մշակվել է թեստի որակի գնահատման նոր մոդել, որը հիմնված է Շենոնի էնտրոպիայի, պայմանական էնտրոպիայի և միջին փոխադարձ ինֆորմացիայի և այս մեծությունների հատկությունների վրա:

Նշված մեթոդն ունի IRT-ի համեմատությամբ ավելի պարզ մաթեմատիկական ապարատ և հաշվարկները կատարվում են ավելի արագ:

Թեստային առաջադրանքի էնտրոպիայի $H(X_i)$ -ի արժեքները համադրելի են CTT բարդության ինդեքսի U_m և IRT-ի պարամետրի արժեքների հետ:

Միջին և բավարար բարդության առաջադրանքների դեպքում $H(X_i)$ էնտրոպիան ընդունում է 1-ին մոտ արժեքներ, իսկ շատ հեշտ կամ բարդ առաջադրանքների դեպքում 0-ին մոտ արժեքներ:

Առաջադրանքների միջին փոխադարձ ինֆորմացիան $(I(X_i \wedge X_j))$ բնութագրում է առաջադրանքների փոխկապակցվածությունը, որն ունի իրեն համարժեք մեծությունը CTT-ում՝ $R(X_i, X_j)$, իսկ IRT-ում բացակայում է:

Առաջադրանքների միջին փոխադարձ ինֆորմացիան ընդունում է 0-ին մոտ արժեքներ, երբ առաջադրանքները միմյանցից համարյա անկախ են, հակառակ դեպքում, երբ այս արժեքը մոտ է $\min[H(X_i), H(X_j)]$ -ին, ապա X_i և X_j թեստային առաջադրանքները կրկնում են մեկը մյուսին և փոքր էնտրոպիա ունեցողը պետք է հեռացվի թեստից:

ԳԼՈՒԽ IV.

ԹԵՍՏԵՐԻ ՈՐԱԿԻ ԳՆԱՀԱՏՄԱՆ ԳՈՐԾԻՔԱՅԻՆ ԾՐԱԳՐԱՅԻՆ ՀԱՄԱԿԱՐԳԻ ՄՇԱԿՈՒՄ

Ատենախոսության շրջանակներում թեստերի որակի ստուգման, գնահատման գործընթացի կազմակերպման նպատակով մշակվել է «Թեստերի որակի վերլուծություն» ծրագրային համակարգը [54]: Համակարգն ունի հետևյալ առանձնահատկությունները.

- իրականացնում է թեստերի որակի բազմակողմանի վերլուծություն՝ հիմնված CTT, IRT և մեր կողմից մշակված մեթոդի վրա,
- ունի օգտվողին կողմնորոշված պարզ web ինտերֆեյս,
- թեստավորման արդյունքները մուտքագրվում են օգտվողի համար շատ մատչելի CSV ձևաչափի ֆայլերի տեսքով,
- համակարգը գեներացնում է թեստի որակի վերլուծության ամբողջական rtf ձևաչափի հաշվետվության ֆայլ,
- ստացվող հաշվետվության ֆայլն իր մեջ բովանդակում է ինչպես հաշվարկվող ցուցանիշների աղյուսակները, գրաֆիկները, այնպես էլ նկարագրություններ դրանց վերաբերյալ,
- հաշվետվության ֆայլի մեջ գունային ձևով տարբերակվում են անբավարար ցուցանիշներով թեստային առաջադրանքները,
- օգտվողի ինտերֆեյսում հասանելի են թեստերի որակական բնութագրերի վերաբերյալ նկարագրություններ, թեստավորման արդյունքների ցուցադրական օրինակներ,
- համակարգը հասանելի է հայերեն և անգլերեն լեզուներով,
- գործարկվում է սերվերի վրա, ինչն էլ հնարավորություն է տալիս լրկալ ցանցի կամ համացանցի միջոցով դիմել համակարգին web զննարկիչի միջավայրում առանց լրացուցիչ ծրագրային միջոցների կիրառման,
- գործարկվում է տարբեր օպերացիոն համակարգերի միջավայրերում՝ Windows, MacOS, Linux:

4.1. «Թեստերի որակի վերլուծություն» համակարգի հնարավորությունները

«Թեստերի որակի վերլուծություն» համակարգը հնարավորություն է տալիս իրականացնել թեստերի որակի վերլուծություն և գնահատում թեստերի երկու՝ CTT և IRT տեսությունների և Շենոնի ինֆորմացիայի տեսության հիման վրա: Համակարգը սատարում է հետևյալ IRT մոդելները.

- 1 պարամետրով միաճյուղ մոդել (Ռաշի մոդել - 1PLM),
- 2 պարամետրով միաճյուղ մոդել (2PLM),
- 3 պարամետրով միաճյուղ մոդել (3PLM),
- մասնակի միավորներով բազմաճյուղ մոդել (PCM),
- ընդհանրացված մասնակի միավորներով բազմաճյուղ մոդել (GPCM):

CTT տեսության շրջանակներում ապահովվում է հետևյալ մեծությունների հաշվարկը.

- թեստի հուսալիության մի շարք գործակիցներ,
- թեստի միավորների ստանդարտ շեղումը,
- թեստային առաջադրանքների բարդությունը,
- թեստային առաջադրանքների տարբերակիչ ունակությունը,
- թեստային առաջադրանքների միավորների ստանդարտ շեղումը,
- թեստային առաջադրանքների միջև կոռելյացիան,
- թեստային առաջադրանքի և անհատական միավորի միջև կոռելյացիան:

IRT տեսության շրջանակներում հաշվարկվում են հետևյալ բնութագրերը և ֆունկցիաները՝

- թեստային առաջադրանքների IRT «a պարամետրը»,
- թեստային առաջադրանքների IRT «b պարամետրը»,
- թեստային առաջադրանքների IRT «c պարամետրը»,
- թեստային առաջադրանքների պատասխանների ֆունկցիան,
- թեստային առաջադրանքների ինֆորմացիոն ֆունկցիան,
- թեստի պատասխանների ֆունկցիան,
- թեստի ինֆորմացիոն ֆունկցիան,
- թեստի չափումների ստանդարտ սխալի ֆունկցիան:

Մշակված նոր մոտեցման հիման վրա ստացվում են հետևյալ ցուցանիշների արժեքները՝

- թեստային առաջադրանքների էնտրոպիան,
- թեստային առաջադրանքների միջև պայմանական էնտրոպիան,
- թեստային առաջադրանքների միջև միջին փոխադարձ ինֆորմացիան:

Բացի վերոնշյալ բնութագրերի հաշվարկից համակարգի միջոցով իրականացվում է մի շարք վիճակագրական տվյալների ստացում՝

- թեստավորման արդյունքների առավելագույն, միջին և նվազագույն միավորները,
- թեստավորման արդյունքների միավորների հաճախությունները,
- թեստավորման արդյունքների տարբերակների միավորների հաճախությունները:

Գեներացվում են հետևյալ գրաֆիկները՝

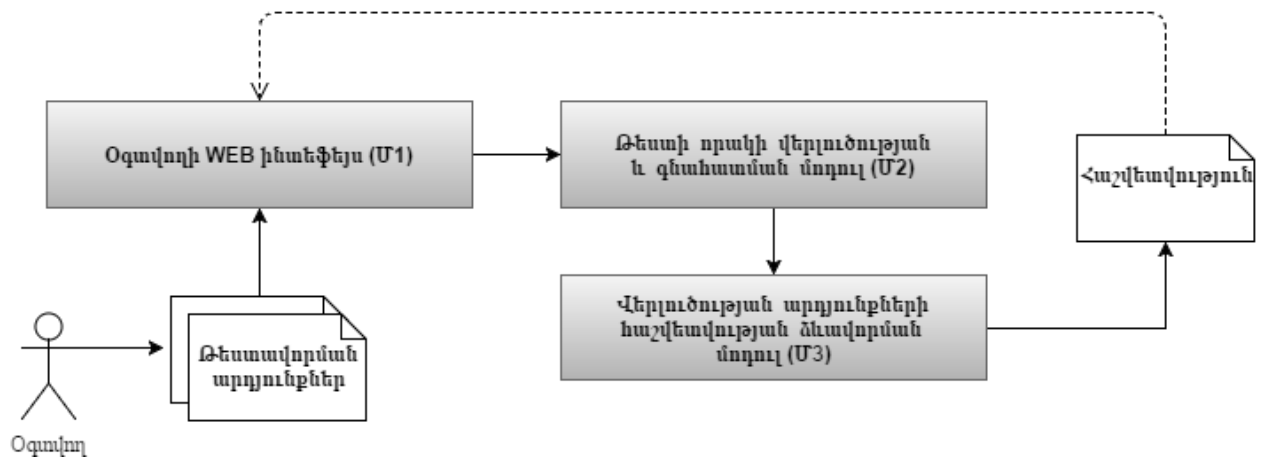
- թեստային առաջադրանքի բնութագրական կորը,
- թեստային առաջադրանքի ինֆորմացիոն ֆունկցիայի գրաֆիկը,
- թեստի բնութագրական կորը,
- թեստի ինֆորմացիոն ֆունկցիայի գրաֆիկը,
- թեստի չափումների ստանդարտ սխալի ֆունկցիայի գրաֆիկը:

4.2. Համակարգի կառուցվածքը

Համակարգի կառուցվածքի և գործառնության ընդհանրացված սխեման բերված է նկար 11-ում:

Համակարգը բաղկացած է երեք հիմնական մոդուլից.

- օգտվողի ինտերֆեյս,
- թեստի որակի վերլուծության և գնահատման մոդուլ,
- արդյունքների մշակման և հաշվետվության ձևավորման մոդուլ:



Նկար 11. Համակարգի կառուցվածքի և գործառնության ընդհանրացված սխեման:

Օգտվողի web ինտերֆեյսը նախատեսված է օգտվողի հետ փոխգործակցության ապահովման համար, մասնավորապես այս մոդուլի միջոցով մուտքագրվում են վերլուծության ենթարկվող թեստավորման արդյունքների ֆայլերը:

Թեստի որակի վերլուծության և գնահատման մոդուլի միջոցով օգտվողի web ինտերֆեյսից ստացված տվյալների հիման վրա իրականացվում է թեստերի որակական ցուցանիշների հաշվարկ և գնահատում, համապատասխան աղյուսակների ստեղծում և գրանցում տվյալների բազայում:

Ստեղծված տվյալների բազայի հիման վրա **վերլուծության արդյունքների հաշվետվության ձևավորման մոդուլի** միջոցով գեներացվում են հաշվետվության ֆայլերը, որոնք ներբեռնելու տարբերակով հասանելի են դառնում օգտվողի ինտերֆեյսում:

Վերլուծության համար անհրաժեշտ թեստավորման արդյունքները, որոնք հանդիսանում են մուտքային ինֆորմացիա, մուտքագրվում են txt կամ csv ձևաչափի ֆայլերի տեսքով: Թեստի որակական վերլուծության և գնահատման արդյունքները՝ որպես ելքային ինֆորմացիա, ստացվում են մեկ ամբողջական հաշվետվության՝ rtf ձևաչափի խմբագրվող ֆայլի տեսքով: Բացի այդ առանձին csv ձևաչափով ֆայլի տեսքով ստացվում են հաշվարկման արդյունքների աղյուսակները և գրաֆիկների նկարները:

4.3. Օգտվողի ինտերֆեյսը

Համակարգի օգտվողի ինտերֆեյսը web ինտերֆեյս է, որը գործարկվում է web զննարկիչի միջավայրում: Օգտվողի ինտերֆեյսն ապահովում է հետևյալ հիմնական հնարավորությունները.

- օգտվողը կարող է վերբեռնել (մուտքագրել) վերլուծվող թեստի փորձարկման արդյունքները՝ txt կամ csv ձևաչափի ֆայլերի տեսքով,
- վերբեռնել թեստի պատասխանների բանալիների ֆայլը՝ txt կամ csv ձևաչափով,
- ընտրել վերլուծության համար անհրաժեշտ համապատասխան մոդելները,
- վերլուծությունից հետո ներբեռնել թեստի որակի վերլուծության և գնահատման արդյունքները՝ մեկ արխիվային ֆայլի տեսքով,
- ծանոթանալ թեստերի որակական բնութագրերի հակիրճ նկարագրությանը,
- ներբեռնել փորձարկման համար անհրաժեշտ թեստերի արդյունքների և պատասխանների ցուցադրական օրինակներ,
- ծանոթանալ համակարգի մասին օգտվողի ուղեցույցին:

Նշված հնարավորությունների ապահովման համար օգտվողի ինտերֆեյսում մշակվել են հետևյալ բաժինները.

- թեստի վերլուծություն,
- որակական բնութագրեր,
- օրինակներ,
- օգնություն,
- լեզվի ընտրություն:

Համակարգի օգտվողի ինտերֆեյսի գլխավոր էջի տեսքը ներկայացված է նկար 12-ում:

Համակարգը գործարկելիս **«Թեստի վերլուծություն»** բաժինը արտապատկերվում է գլխավոր էջում: Այս բաժնում ներկայացված են թեստի վերլուծության համար նախատեսված ֆայլերի մուտքագրման համար անհրաժեշտ օգտվողի ինտերֆեյսի ղեկավարման տարրերը և դաշտերը:

ԹԵՄՏԻ ՎԵՐԼՈՒԾՈՒԹՅՈՒՆ

ՈՐԱՐԱՆԱԿԱՆ ԲՆՈՒԹԱԳՐԵՐ

ՕՐԻՆԱԿՆԵՐ

ՕԳՆՈՒԹՅՈՒՆ

ՀԱՅ

ENG

Թեստի վերլուծություն

Թեստի արդյունքների ֆայլ

ՎԵՐԲԵՈՒՆԵԼ

Թեստի պատասխանների ֆայլ

ՎԵՐԲԵՈՒՆԵԼ

Թեստի վերնագիր

Թեստի վերնագիր

Թեստի գնահատման տեսակը

Միանյութ (Բինար կամ տարբերակի ընտրություն)

Բազմանյութ (Polytomous)

☒ Խառը

IRT մոդելներ

1 պարամետրով միանյութ մոդել (Ռաշի մոդել - 1PLM),

2 պարամետրով միանյութ մոդել (2PLM),

☒ 3 պարամետրով միանյութ մոդել (3PLM),

Մասնակի միավորներով բազմանյութ մոդել (PCM),

☒ Ընդհանրացված մասնակի միավորներով բազմանյութ մոդել (GPCM):

ՎԵՐԼՈՒԾՆԵԼ

Նկար 12. Համակարգի օգտվողի ինտերֆեյսի գլխավոր էջի տեսքը:

Վերլուծության կազմակերպման համար անհրաժեշտ է մուտքագրել երկու ֆայլեր.

- **թեստի փորձարկումների արդյունքների ֆայլ**, որտեղ աղյուսակի տեսքով ներկայացված է փորձնական ճանապարհով ստացված թեստավորման արդյունքների մատրիցան, այսինքն առաջադրանքներից յուրաքանչյուրին տրված փորձարկվողների պատասխանների աղյուսակը:
- **թեստի առաջադրանքների բանալիների ֆայլ**, որտեղ աղյուսակի տեսքով ներկայացվում են առաջադրանքները, դրանց տեսակը, գնահատման ձևը, հնարավոր տարբերակների քանակը և ճիշտ պատասխանը:

Այս երկու ֆայլերը մուտքագրվում են txt կամ CSV ձևաչափով: CSV (Comma Separated Values)-ն տվյալների պահպանման ֆայլի ձևաչափ է, որում տվյալները պահպանվում են աղյուսակների նման և յուրաքանչյուր տող համապատասխանում է մեկ գրանցման, իսկ դաշտերն էլ միմյանցից առանձնացվում են ստորակետով (comma): Ֆայլի txt ձևաչափը տվյալներ պահելու հիմնական միջոց է, որտեղ ազատ ոճով կարելի է պահպանել տվյալները, սահմանել տվյալների դաշտերի առանձնացման

71

նշիչները և այլն: Թերություն է այն, որ օգտվողի համար դժվար է ազատ ոճի կառուցվածքի հետ աշխատելը, շատ դեպքերում լրացուցիչ բացատրների կամ նոր տողերի ավելացումը բերում է տվյալ ֆայլից տվյալների ընթերցման սխալի:

CSV ձևաչափով ֆայլերի ներկայացումը ապահովում է մի շարք առավելություններ.

- օգտվողը, որը մեր խնդրի դեպքում առավել հաճախ մանկավարժն է, կարող է հեշտությամբ հասկանալ ֆայլում ներկայացվող տվյալների կառուցվածքը և բովանդակությունը,
- օգտվողը հեշտությամբ կարող է խմբագրել ֆայլը, քանի որ խմբագրման համար կարելի է կիրառել մի շարք պարզագույն միջոցներ՝ տեքստային խմբագրիչ (Notepad), Excel աղյուսակային խմբագրիչ, Google Docs ամպային տեխնոլոգիա, առցանց CSV խմբագրիչներ:

Թեստավորման արդյունքների մատրիցան $N+1$ տողից և $M+1$ սյունից բաղկացած աղյուսակ է, որտեղ N -ը փորձարկվողների քանակն է, իսկ M -ը առաջադրանքների քանակը: Աղյուսակի առաջին տողում գրանցվում է առաջադրանքների անվանումները, իսկ առաջին սյան մեջ փորձարկվողների անունները: Առաջադրանքին և փորձարկվողին համապատասխանող դաշտում գրանցվում է տվյալ առաջադրանքից տրված փորձարկվողի ստացած միավորը:

Աղյուսակ 10-ում բերված է թեստավորման արդյունքների մատրիցայի օրինակ:

Առաջադրանքների բանալիների ֆայլում M առաջադրանքի դեպքում ստեղծվում է $M+1$ տող և 4 սյուն ունեցող աղյուսակ, որի առաջին տողը ներկայացնում է աղյուսակի վերնագիրը՝ ID, AnswerKey, AlternativeCount, Type:

Աղյուսակ 10.

Միաճյուղ բինար գնահատումով թեստավորման արդյունքների մատրիցայի օրինակ

ID	Առաջադրանք1	Առաջադրանք2	Առաջադրանք3	ԱռաջադրանքN
Փորձարկվող1	1	1	1	0
Փորձարկվող2	1	0	1	1
Փորձարկվող3	1	0	1	1
ՓորձարկվողN	1	1	0	0

**Թեստի միաճյուղ բինար գնահատումով առաջադրանքների բանալիների
ֆայլի բովանդակության օրինակ**

ID	AnswerKey	AlternativeCount	Type
Առաջադրանք1	0	2	B
Առաջադրանք2	0	2	B
Առաջադրանք3	0	2	B
ԱռաջադրանքN	0	2	B

Առաջին սյան մեջ նշվում են առաջադրանքների համարները:

Առանձնացվում է երեք տիպի գնահատումով առաջադրանքներ.

- **միաճյուղ բինար գնահատումով առաջադրանքներ,**
- **միաճյուղ տարբերակի ընտրությամբ գնահատումով առաջադրանքներ,**
- **բազմաճյուղ ազատ միավորների գնահատումով առաջադրանքներ:**

Վերոնշյալ տիպերից կախված տարբեր է թեստավորման արդյունքների մատրիցայի և առաջադրանքների բանալիների ֆայլի բովանդակությունը:

Միաճյուղ բինար գնահատումով առաջադրանքների դեպքում թեստավորման արդյունքները ներկայացվում են բինար տեսքով՝ 1-տվյալ առաջադրանքին փորձարկվողի տված ճիշտ պատասխանի դեպքում և 0-համապատասխանաբար սխալ պատասխանի դեպքում:

Առաջադրանքների բանալիների ֆայլի AnswerKey դաշտում միաճյուղ բինար գնահատման դեպքում գրվում է 0, AlternativeCount դաշտում նշվում է առաջադրանքի հնարավոր տարբերակների քանակը՝ 2, Type դաշտում գրվում է առաջադրանքի տեսակը՝ «B» (Binary): Նկար 13-ում ներկայացված է միաճյուղ բինար գնահատումով 10 առաջադրանքից (A1, A2, A3, ..., A_n) և 10 փորձարկվողի (Ex1, Ex2, ..., Ex_n) արդյունքներով թեստի փորձարկումների արդյունքների ֆայլի օրինակ:

1	ID, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10
2	Ex1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1
3	Ex2, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1
4	Ex3, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0
5	Ex4, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0
6	Ex5, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0
7	Ex6, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1
8	Ex7, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1
9	Ex8, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1
10	Ex9, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0

Նկար 13. Միաճյուղ բինար գնահատումով 10 առաջադրանքից և 10 փորձարկվողի արդյունքներով թեստի փորձարկումների արդյունքների ֆայլի օրինակ:

1	ID, AnswerKey, AlternativeCount, Type
2	A1, 0, 2, B
3	A2, 0, 2, B
4	A3, 0, 2, B
5	A4, 0, 2, B
6	A5, 0, 2, B
7	A6, 0, 2, B
8	A7, 0, 2, B
9	A8, 0, 2, B
10	A9, 0, 2, B
11	A10, 0, 2, B

Նկար 14. Միաճյուղ բինար գնահատումով 10 առաջադրանքով թեստի առաջադրանքների բանալիների ֆայլի օրինակ:

Նկար 14-ում ներկայացված է նույն թեստի առաջադրանքների բանալիների ֆայլը:

Միաճյուղ տարբերակի ընտրությամբ գնահատումով առաջադրանքների դեպքում թեստավորման արդյունքների մատրիցայի դաշտում նշվում է փորձարկվողի կողմից տվյալ առաջադրանքին տրված տարբերակը (A,B,C,D կամ 1,2,3,4 և այլն): Նկար 15-ում ներկայացված է միաճյուղ բինար գնահատումով 10 առաջադրանքից (A1, A2, A3, ..., A_n) և 10 փորձարկվողի (Ex1, Ex2, ..., Ex_n) արդյունքներով թեստի

փորձարկումների արդյունքների ֆայլի օրինակ: Յուրաքանչյուր առաջադրանք ունի հնարավոր 4 տարբերակ (A,B,C,D) և մեկ հնարավոր ճիշտ պատասխան:

1	ID, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10
2	Ex1, A, A, D, D, C, B, C, D, D, D
3	Ex2, A, B, C, C, C, B, A, D, A, D
4	Ex3, A, C, D, B, D, D, D, D, A, D
5	Ex4, A, B, D, B, D, D, C, D, A, C
6	Ex5, A, D, C, B, C, B, C, D, B, D
7	Ex6, D, C, B, C, A, B, C, D, A, D
8	Ex7, A, A, C, B, C, B, D, D, D, D
9	Ex8, A, B, C, B, C, B, C, D, D, D
10	Ex9, A, D, C, D, D, B, D, D, A, D
11	Ex10, B, A, C, B, C, B, B, D, D, D

Նկար 15. Միաճյուղ տարբերակի ընտրությամբ գնահատումով 10 առաջադրանքից և 10 փորձարկվողի արդյունքներով թեստի փորձարկումների արդյունքների ֆայլի օրինակ:

Առաջադրանքների բանալիների ֆայլի AnswerKey դաշտում գրանցվում է տվյալ առաջադրանքի ճիշտ տարբերակի անվանումը (օրինակ A, B կամ 1,2), AlternativeCount դաշտում նշվում է տվյալ առաջադրանքի հնարավոր տարբերակների քանակը (օրինակ 4), Type դաշտում գրվում է առաջադրանքի տեսակը՝ «M» (Multiple Choice): Նկար 16-ում ներկայացված է նույն թեստի առաջադրանքների բանալիների ֆայլի բովանդակությունը:

Բազմաճյուղ ազատ միավորների գնահատումով առաջադրանքների դեպքում գնահատականը ներկայացվում է որևէ թվային միջակայքով: Այս տիպի գնահատումը կարող է կիրառվել մի քանի հնարավոր ճիշտ պատասխաններով առաջադրանքների դեպքում, երբ կախված տարբերակների քանակից տրվում են տարբեր միավորներ: Կիրառելի են նաև ազատ պատասխանների հնարավորությամբ առաջադրանքները, օրինակ շարադրություն կամ կարճ տեքստային պատասխան: Թեստավորման արդյունքների մատրիցայի միավորների դաշտում նշվում է առաջադրանքից տրված փորձարկվողի ստացած միավորը:

1	ID, AnswerKey, AlternativeCount, Type
2	A1, A, 4, M
3	A2, D, 4, M
4	A3, C, 4, M
5	A4, B, 4, M
6	A5, C, 4, M
7	A6, B, 4, M
8	A7, C, 4, M
9	A8, D, 4, M
10	A9, A, 4, M
11	A10, D, 4, M

Նկար 16. Միաճյուղ տարբերակի ընտրությամբ գնահատումով 10 առաջադրանքով թեստի առաջադրանքների բանալիների ֆայլի օրինակ:

Թեստի առաջադրանքների բանալիների ֆայլի AnswerKey դաշտում գրվում է 0, AlternativeCount դաշտում գրվում է տվյալ առաջադրանքին տրվող միավորի կարգերի քանակը՝ ընդունենք 4, որին կհամապատասխանեն 0,1,2,3 միավորները, իսկ Type դաշտում գրվում է «P» (Polytomous):

Նկար 17-ում ներկայացված է բազմաճյուղ ազատ միավորների գնահատումով 10 առաջադրանքից ($A_1, A_2, A_3, \dots, A_n$) և 10 փորձարկվողի ($Ex_1, Ex_2, \dots, Ex_n$) արդյունքներով թեստի փորձարկումների արդյունքների ֆայլի օրինակ: Յուրաքանչյուր առաջադրանք ունի գնահատման 4 հնարավոր արժեք՝ 0,1,2,3:

Նկար 18-ում բերված է նույն թեստի առաջադրանքների բանալիների ֆայլի բովանդակությունը:

1	ID, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10
2	Ex1, 0, 1, 1, 1, 1, 3, 2, 0, 1, 1
3	Ex2, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1
4	Ex3, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0
5	Ex4, 3, 1, 2, 3, 2, 2, 3, 3, 3, 1
6	Ex5, 2, 2, 2, 1, 2, 1, 2, 2, 2, 1
7	Ex6, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1
8	Ex7, 3, 1, 3, 2, 2, 3, 1, 3, 2, 3
9	Ex8, 1, 1, 2, 2, 1, 0, 1, 1, 1, 2
10	Ex9, 0, 2, 2, 1, 1, 0, 2, 2, 2, 0
11	Ex10, 0, 0, 1, 1, 1, 1, 1, 1, 0, 2

Նկար 17. Բազմաճյուղ ազատ միավորների գնահատումով (0,1,2,3) 10 առաջադրանքից և 10 փորձարկվողի արդյունքներով թեստի փորձարկումների արդյունքների ֆայլի օրինակ:

1	ID, AnswerKey, AlternativeCount, Type
2	A1, 0, 4, P
3	A2, 0, 4, P
4	A3, 0, 4, P
5	A4, 0, 4, P
6	A5, 0, 4, P
7	A6, 0, 4, P
8	A7, 0, 4, P
9	A8, 0, 4, P
10	A9, 0, 4, P
11	A10, 0, 4, P

Նկար 18. Բազմաճյուղ ազատ միավորների գնահատումով (0,1,2,3) 10 առաջադրանքից և 10 փորձարկվողի արդյունքներով թեստի պատասխանների բանալիների ֆայլի օրինակ:

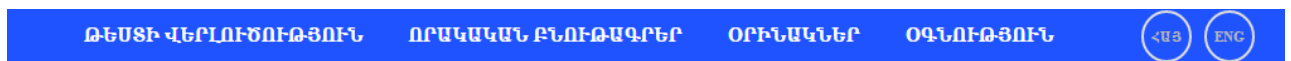
Թեստի ֆայլերը մուտքագրելուց հետո օգտվողի ինտերֆեյսի «Թեստի վերնագիր» դաշտում լրացվում է թեստի անվանումը, որը երևալու է համակարգի միջոցով գեներացվող հաշվետվության ֆայլի մեջ: «Թեստի գնահատման տեսակը» դաշտից օգտվողը ընտրում է հնարավոր գնահատման տեսակը, իսկ «IRT մոդելներ»

դաշտից՝ IRT վերլուծության համար նախատեսված համապատասխան մոդելը կամ մոդելները:

«Որակական բնութագրեր» բաժնում բերված են թեստերի CTT, IRT տեսությունների շրջանակներում գնահատվող որակական բնութագրերի կարճ նկարագրությունները: Մշակված նկարագրությունը հասանելի է նաև pdf ձևաչափով՝ նախատեսված ներբեռնելու համար: Բերված նկարագրությունները կարող են ծառայել այն օգտվողներին, որոնք ցանկանում են տեղեկատվություն ստանալ թեստերի որակի բնութագրերի, դրանց նշանակության մասին:

«Օրինակներ» բաժնից կարելի է ներբեռնել տաբեր տեսակի գնահատմամբ թեստերի փորձարկումների և պատասխանների բանալիների ֆայլերի օրինակներ: Այս օրինակները հնարավորություն են տալիս օգտվողին ծանոթանալ մուտքագրման ֆայլերի կառուցվածքին և մուտքագրելով այդ օրինակներից ստանալ տվյալ թեստի որակական վերլուծության և գնահատման հաշվետվությունը: Նկար 19-ում ներկայացված է «Օրինակներ» բաժնի տեսքը:

«Օգնություն» բաժինը նախատեսված է համակարգի օգտագործման նպատակով օգտվողին տեղեկատվություն տրամադրելու համար: Այստեղ բերված է թեստերի փորձարկումների արդյունքների և պատասխանների բանալիների ֆայլերի կառուցվածքի վերաբերյալ մանրամասն նկարագրություն:



Օրինակներ

-  [Exam1-1500 փորձարկվողի և 20 միաճյուղ բինար առաջադրանքներից բաղկացած թեստավորման արդյունքներ:](#)
-  [Exam2-6000 փորձարկվողի և 20 միաճյուղ՝ պատասխաններից մեկի բնութագրամբ առաջադրանքներից բաղկացած թեստավորման արդյունքներ:](#)
-  [Exam3-1000 փորձարկվողի և 20 բազմաճյուղ՝ 4 հնարավոր միավորներով \(0.1.2.3\) առաջադրանքներից բաղկացած թեստավորման արդյունքներ:](#)
-  [Exam4 \(խառը\)- 1500 փորձարկվող և 35 տարբեր առաջադրանքներ: 1- 30 առաջադրանքներ բինար առաջադրանքներ են իսկ 31-35 բազմաճյուղ՝ երեք հնարավոր միավորներով \(0.1,2\):](#)

Նկար 19. Համակարգի օգտվողի ինտերֆեյսի «Օրինակներ» բաժնի տեսքը:

Համակարգի օգտվողի ինտերֆեյսի լեզուների պատկերանկարների միջոցով օգտվողը կարող է փոխել համակարգի լեզուն: Ներկայումս համակարգը հասանելի է հայերեն և անգլերեն լեզուներով:

4.4. Որակի վերլուծության և գնահատման մոդուլը

Թեստի որակի վերլուծության և գնահատման մոդուլը առանձնացված է օգտվողի ինտերֆեյսից և գործարկվում է «Վերլուծել» կոճակի սեղմումով: Այս մոդուլը որպես մուտքային ինֆորմացիա ընդունում է օգտվողի ինտերֆեյսից օգտվողի կողմից վերբեռնված թեստի համապատասխան ֆայլերը, ինչպես նաև ընտրված որոշակի պարամետրերի արժեքները: Որպես ելքային արդյունք, այստեղից կատարված հաշվարկների, վերլուծությունների արդյունքները տրամադրվում է արդյունքների մշակման և հաշվետվության ձևավորման մոդուլին: Թեստի որակի վերլուծության և գնահատման մոդուլը իրենից ներկայացնում է Java հավելված, որի հիմնական մասը կազմում է բաց ծրագրային կոդով Java psychometrics API [59] գրադարանը: Ստորև ներկայացված է այս մոդուլի կողմից կատարվող գործողությունների ամբողջական նկարագրությունը.

1. կատարվում է մուտքագրված թեստավորման արդյունքների և թեստի պատասխանների ֆայլերի ընթերցում,
2. ըստ ֆայլերի կառուցվածքի տվյալների բազայում ստեղծվում են համապատասխան աղյուսակները, մասնավորապես՝ թեստավորման արդյունքների մատրիցան և պատասխանների աղյուսակը, որոնցում գրանցվում է ֆայլերի պարունակությունը,
3. հաշվարկվում են մի շարք վիճակագրական ցուցանիշներ, Java psychometrics API գրադարանի մեթոդների միջոցով հաշվարկվում են CTT և IRT ցուցանիշները, ինչպես նաև հաշվարկվում են Շենոնի ցուցանիշները,
4. կատարվում է սահմանային ընդունելի արժեքներ ունեցող ցուցանիշների գնահատում և առանձնացում,

5. հաշվարկված ցուցանիշների հիման վրա տվյալների բազայում ստեղծվում են աղյուսակներ և գրանցվում հաշվարկված ցուցանիշների արժեքները,

6. գործարկվում է հաշվետվության ձևավորման մոդուլը, որին որպես մուտքային պարամետր ուղարկվում է տվյալների բազայում գրանցված աղյուսակների անվանումները:

Կարևոր է նշել, որ նկարագրված մոդուլը կարող է գործարկվել առանձնացված համակարգչի վրա որպես desktop հավելված: Այս տարբերակում օգտվողն անմիջապես առնչվում է մոդուլի GUI ինտերֆեյսի հետ, որն ապահովում է հետևյալ հնարավորությունները.

- օգտվողը ծրագրի պատուհանում կարողանում է վիզուալ տեսնել թեստավորման արդյունքների մատրիցան, աշխատել վերջինիս հետ,
- ընտրել վերլուծության տարբեր պարամետրեր, վիզուալ միջավայրում կատարել վերլուծություն,
- առանձնացնել առաջադրանքները և դիտարկել ամբողջական վերլուծության արդյունքներն այդ առաջադրանքների հեռացումից հետո,
- գեներացնել մի շարք գրաֆիկներ,
- վիզուալ միջավայրում աշխատել տվյալների բազայի աղյուսակների հետ,
- առանձին ֆայլերի տեսքով պահել վերլուծության տարբեր չափանիշներով ստացված արդյունքները, ինչպես նաև գրաֆիկները:

Նշված տարբերակով կիրառումը հարմար է նեղ տիրույթի մասնագետների համար, ովքեր ունեն տեխնիկական գիտելիքներ և տիրապետում են թեստերի վերլուծության տեսություններին: Առանձնացված գործարկելու համար չի պահանջվում լրացուցիչ ծրագրային միջոցների բարդ կարգաբերումներ, ընդամենը գործարկվող օպերացիոն համակարգի միջավայրում պետք է լինի JRE սատարում, ինչն էլ թույլ է տալիս հավելվածը կիրառել մի շարք օպերացիոն համակարգերի միջավայրերում՝ Windows, Linux, MacOS:

4.5. Արդյունքների մշակման և հաշվետվության ձևավորման մոդուլը

Արդյունքների մշակման և հաշվետվության ձևավորման մոդուլի միջոցով ձևավորվում են վերջնական հաշվետվության ֆայլերը, որոնք օգտվողը ներբեռնում է օգտվողի ինտերֆեյսից տրամադրվող հղման միջոցով: Մոդուլի միջոցով հաշվետվության ձևավորումը կատարվում է հետևյալ փուլերով.

- ստացված մուտքային պարամետրերի հիման վրա համապատասխան աղյուսակներից կատարվում է տվյալների ընթերցում,
- ընթերցված տվյալների հիման վրա CSV ձևաչափով գեներացվում են անհրաժեշտ աղյուսակների ֆայլերը և պահպանվում հաշվետվության առանձին ֆայլային տիրույթում (դիրեկտորիա),
- որպես առանձին ֆայլեր գեներացվում են համապատասխան գրաֆիկները և պահպանվում նույն հաշվետվության ֆայլային տիրույթում,
- ըստ աղյուսակների կառուցվածքի ստեղծվում է rtf ձևաչափի հաշվետվության ֆայլը,
- մշակվում է ֆայլի բովանդակությունը, որտեղ ներառվում են տարբեր աղյուսակներ, գեներացված գրաֆիկներ, նկարագրություններ, ընդ որում այս փուլում ըստ ընթերցված տվյալների գունային տարբերակով առանձնացվում են անբավարար ցուցանիշները:
- ձևավորվում է հաշվետվությունների արխիվային ֆայլ-փաթեթը, որն իր մեջ ներառում է հաշվետվության rtf ֆայլը, գեներացված գրաֆիկները, աղյուսակների CSV ձևաչափով ֆայլերը,
- որպես ելքային արդյունք այս ամբողջ փաթեթը ուղարկվում է օգտվողի web ինտերֆեյս՝ ներբեռնման հղման հնարավորությամբ:

Հաշվետվության ֆայլի rtf ձևաչափն ունի մի շարք առավելություններ, որոնցից կարելի է նշել մի քանիսը.

- հեշտ կիրառելի է, կարելի այն բացել տաբեր տեքստային խմբագրիչներով,
- առանց որևէ լրացուցիչ փոփոխությունների կարելի է հեշտությամբ տպել,

- հեշտությամբ կարելի է խմբագրել ֆայլի պարունակությունը:
Հաշվետվությունների rtf ֆայլը բաղկացած է հետևյալ բաժիններից.
- ներածություն,
- ընդհանրացված վիճակագրություն,
- CTT ցուցանիշներ,
- IRT ցուցանիշներ,
- ինֆորմացիայի տեսության ցուցանիշներ,
- գրաֆիկներ:

Ներածություն բաժնում բերված է ընդհանուր տեղեկություններ հաշվետվության բովանդակության, ինչպես նաև թեստերի որակական ցուցանիշների վերաբերյալ: Յուրաքանչյուր բաժնի սկզբում բերված է այդ բաժնում ներկայացվող աղյուսակների և տվյալների վերաբերյալ ընդհանուր տեղեկատվություն: Այն ցուցանիշները, որոնք ունեն սահմանային արժեքներ, դրանց համապատասխան տրվում է գունավորում (կարմիր գույն), ինչը թույլ է տալիս օգտվողին միանգամից տարբերակել խնդրահարույց առաջադրանքները կամ թեստի այլ ցուցանիշներ:

Ընդհանրացված վիճակագրություն բաժնում ներկայացված են մի քանի աղյուսակներ: Առաջին աղյուսակը պարունակում է ընդհանրական տվյալներ ամբողջ թեստի վերաբերյալ: Այստեղ ներկայացված են հետևյալ վիճակագրական ցուցանիշները՝

- թեստի անվանումը,
- թեստի փորձարկվողների և առաջադրանքների քանակները,
- թեստավորման նվազագույն, միջին և առավելագույն միավորները,

Այս բաժնում ներկայացված է նաև թեստի հուսալիության 5 գործակիցների արժեքները.

- Գութմանի L2 (Guttman's L2)
- Ալֆա (Alpha),
- Ֆելդ-Գիլմեր (Feldt-Gilmer),
- Ֆելդ-Բրեննան (Feldt-Brennan),
- Ռաջուի բետա (Raju's beta):

Այս ցուցանիշների նկարագրությունը ներկայացված է այս աշխատանքի առաջին գլխի առաջին բաժնում:

Առանձին աղյուսակի տեսքով ներկայացված են հուսալիության արժեքները յուրաքանչյուր առաջադրանքի հեռացումից հետո:

Փորձարկվողների անհատական միավորների բաշխումը, թեստավորումից ստացված միավորների բաշխումը ևս ներկայացված է այս բաժնում: Յուրաքանչյուր առաջադրանքի մակարդակով ներկայացված է նաև տվյալ առաջադրանքի կատարման (պատասխանելու) ցուցանիշները, ընդ որում մի քանի տարբերակներով առաջադրանքների դեպքում ներկայացվում է նաև այդ տարբերակի ընտրության ցուցանիշները:

CTT ցուցանիշներ բաժնում բերված աղյուսակներից առաջինը պարունակում է հետևյալ տվյալները՝

- թեստային առաջադրանքների բարդությունը,
- թեստային առաջադրանքների տարբերակիչ ունակությունը,
- թեստային առաջադրանքների միավորների ստանդարտ շեղումը:

Առանձին աղյուսակի տեսքով ներկայացված է նաև թեստային առաջադրանքների միջև կոռելյացիան, թեստային առաջադրանքի և անհատական միավորի միջև կոռելյացիան:

IRT ցուցանիշներ բաժնում ներկայացված է IRT տեսության շատ հակիրճ նկարագրությունը և հաշվարկվող ցուցանիշների աղյուսակները: Նկարագրության մեջ ներկայացված է այս տեսության շրջանակներում գնահատվող առաջադրանքի բարդության «b», տարբերակիչ ունակության «a» և պատահականությամբ գուշակելու «c» պարամետրերի նկարագրությունները:

Առաջին աղյուսակում ներկայացված են առաջադրանքները և նրանց IRT «a», «b», «c» պարամետրերի արժեքները ու համապատասխան չափման ստանդարտ սխալների արժեքները:

Երկրորդ աղյուսակում ներկայացված է փորձարկվողների պատրաստվածության մակարդակը բնութագրող պարամետրի բաշխումը:

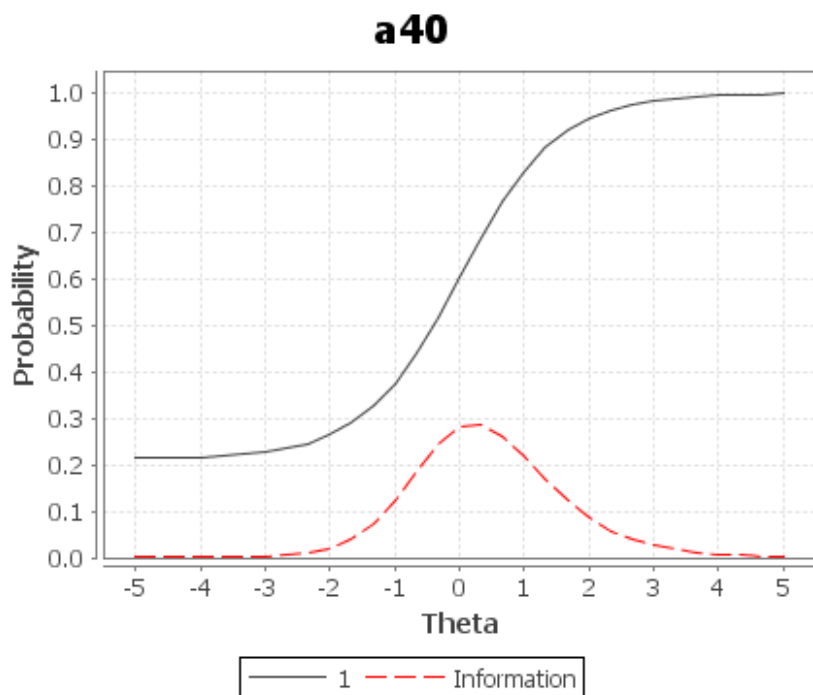
Ինֆորմացիայի տեսության ցուցանիշներ բաժնում նկարագրված են առաջադրանքների էնտրոպիայի, միջին փոխադարձ ինֆորմացիայի, պայմանական էնտրոպիայի մեծությունները: Ներկայացված են համապատասխան հաշվարկված ցուցանիշների աղյուսակները:

Գրաֆիկներ բաժնում ներկայացված են թեստի և թեստային առաջադրանքների IRT գրաֆիկները: Հաշվետվության մեջ ներառված են հետևյալ IRT գրաֆիկները՝

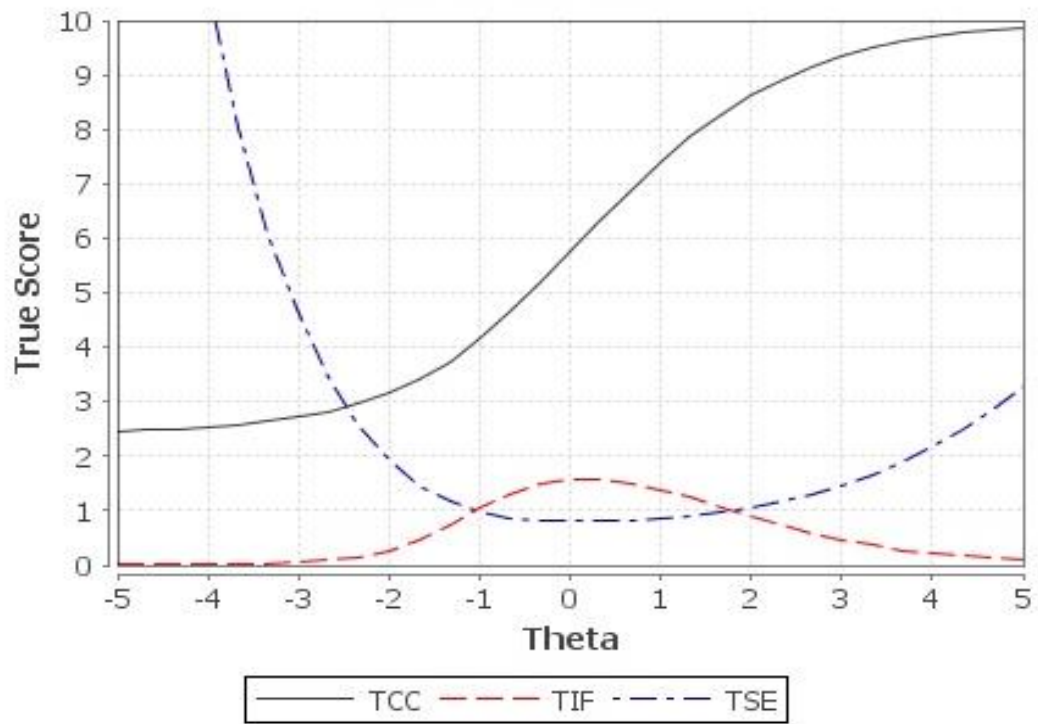
- թեստային առաջադրանքի բնութագրական կորը,
- թեստային առաջադրանքի ինֆորմացիոն ֆունկցիայի գրաֆիկը,
- թեստի բնութագրական կորը,
- թեստի ինֆորմացիոն ֆունկցիայի գրաֆիկը,
- թեստի չափումների ստանդարտ սխալի ֆունկցիայի գրաֆիկը:

Թեստի առաջադրանքների բարդության բաշխումը ըստ փորձարկվողների պատրաստվածության մակարդակի ներկայացվում է առաջադրանքների քարտեզի (item map) միջոցով, որը նույնպես ներառված է այս բաժնում:

Նկար 20 և 21-ում բերված են մի քանի գրաֆիկների օրինակներ:



Նկար 20. Միաճյուղ գնահատումով առաջադրանքի բնութագրական կորը և ինֆորմացիոն ֆունկցիայի գրաֆիկը:



Նկար 21. Թեստի պատասխանների ֆունկցիայի (TCC), ինֆորմացիոն ֆունկցիայի(TIF) և չափումների ստանդարտ սխալի ֆունկցիայի (TSE) գրաֆիկների պատկերը:

4.6. Նոր տեխնոլոգիական լուծումները

Խնդիր էր դրվել համակարգի մշակման փուլում ծրագրավորման լեզուների և տեխնոլոգիաների ընտրությունը կատարել հաշվի առնելով հետևյալ առանձնահատկությունները:

- Ցանկալի էր, որ համակարգն ունենա օգտվողին կողմնորոշված պարզ ինտերֆեյս:
- Ցանկալի էր, որ օգտվողը չառնչվեր համակարգի անբաժանելի մաս կազմող մաթեմատիկական և վիճակագրական բարդ ապարատին:
- Համակարգը պետք է լիներ պլատֆորմից անկախ, որը թույլ կտա այն գործարկել ցանկացած օպերացիոն համակարգի միջավայրում՝ Windows, Linux կամ MacOS:
- Ցանկալի էր, որ համակարգն ունենա web հենքով օգտվողի ինտերֆեյս և գործարկվի սերվերի վրա, ինչը թույլ կտար օգտվողին դիմել համակարգին լոկալ ցանցի միջոցով կամ համացանցի միջոցով՝ գործարկելով այն web զննարկիչի միջավայրում:

Համակարգի ծրագրային ճարտարապետությունը

Ծրագրային ապահովումը կլիենտ-սերվերային ճարտարապետությամբ համակարգ է [48], ինչը թույլ է տալիս համակարգը բաժանել առանձին մոդուլների: Նման ճարտարապետության դեպքում համակարգի բաղադրիչները ներկայացվում են տարբեր պլատֆորմներով, և տարբեր բաղադրիչներ նույն միասնական խնդրի լուծման նպատակով կարող են փոխգործակցել մեկը մյուսի հետ ցանցի միջոցով:

Կլիենտ-սերվերային ճարտարապետության առավելություններից կարելի է նշել:

- Նույն համակագում գործարկվում են տարբեր տեխնոլոգիաներով մշակված և տարբեր միջավայրերում աշխատող բաղադրիչներ:
- Օգտվողից քողարկվում են համակարգում օգտագործված տարբեր տեխնոլոգիաները՝ պլատֆորմներ, ծրագրավորման լեզուներ, օպերացիոն համակարգեր և այլն:

- Նմանատիպ համակարգերում շատ ճկուն են նոր միջոցների, բաղադրիչների ավելացման հնարավորությունները:

Մեր խնդրի շրջանակներում կլիենտ-սերվերային ճարտարապետությամբ մշակելու մոտեցումը ապահովել է հետևյալ առավելությունները:

- Օգտվողն առնչվում է միայն համակարգի բաղադրիչ հանդիսացող օգտվողի web ինտերֆեյսի հետ (client), որը մշակված է HTML, CSS, JavaScript տեխնոլոգիաներով:
- Համակարգի սերվերային մասն առանձնացված է կլիենտային մասից և գրված է PHP սերվերային ծրագրավորման լեզվով [77]:
- Սերվերի վրա առանձնացված գործարկվում է թեստի որակի վերլուծության և գնահատման մոդուլը, որը մշակվել է այլ տեխնոլոգիայով՝ Java լեզվով և կարող է կիրառվել որպես առանձին հավելված նեղ տիրույթի մասնագետների կողմից ինտերակտիվ ռեժիմում բազմակողմանի որակական վերլուծություն կատարելու համար:
- Օգտվողից քողարկված է որակի վերլուծության և գնահատման բարդ մոդուլը, որի աշխատանքի արդյունքը օգտվողը տեսնում է միայն web ինտերֆեյսում:

Համակարգի մոդուլների միջև փոխգործակցությունը կատարվում է հետևյալ սխեմայով:

- Համակարգի մուտքային ինֆորմացիան՝ վերլուծության ենթարկվող թեստավորման արդյունքների ֆայլերը, ինչպես նաև վերլուծության հետ կապված տարբեր պարամետրերի արժեքները, մուտքագրվում են օգտվողի web ինտերֆեյսից:
- Սերվերի վրա իրականացվում է մուտքագրված ֆայլերի պատճենում, վավերականացում, սխալների դեպքում հաղորդագրությունների արտապատկերում:
- Սերվերային սկրիպտի միջոցով գործարկվում է թեստի որակի վերլուծության և գնահատման մոդուլը, որպես արտաքին ծրագիր: այս մոդուլին որպես մուտքային ինֆորմացիա ուղարկվում է մուտքագրված ֆայլերի

հասցեները, իսկ պարամետրերի արժեքներն ուղարկվում են JSON ձևաչափով [60]՝ բանալի-արժեք սխեմայով:

- Գործարկված վերլուծության մոդելը իրականացնում է ֆայլերի ընթերցում, տվյալների բազայում տվյալների գրանցում, թեստի որակի բոլոր ցուցանիշների հաշվարկ, աղյուսակների ստեղծում:
- Ավարտից հետո ծրագրի կատարումը փոխանցվում է հաշվետվության ձևավորման մոդուլին:
- Հաշվետվության ձևավորման մոդուլի միջոցով ստեղծվում է հաշվետվությունների արխիվային ֆայլը, որն իր մեջ պարունակում է ցուցանիշների աղյուսակների ֆայլերը, գրաֆիկները և խմբագրվող տեքստային վերջնական հաշվետվության ֆայլը (rtf) :
- Հաշվետվության ձևավորումից հետո ծրագրի կատարումը փոխանցվում է քիբ սերվերային սկրիպտին:
- PHP սերվերային սկրիպտը հաշվետվության ֆայլի հղումը հասանելի է դարձնում օգտվողին web ինտերֆեյսի միջոցով:

Համակարգի բաղադրիչների մշակման մոտեցումները

Նկարագրությունից պարզ է դառնում, որ web սերվերը, որի վրա պետք է տեղադրվի համակարգը, պետք է ունենա PHP ինտերպրետատոր և Java ծրագրերի գործարկման համար անհրաժեշտ միջավայր՝ Java Runtime Environment (JRE) : Գոյություն ունեն մի շարք բաց ծրագրային կոդով և անվճար լիզենցիայով տարածվող վեբ սերվերներ [42], որոնց միջավայրում ինտեգրվում է JRE միջավայրը:

Web սերվերի վրա տեղադրելու նման մոտեցումը բավարարում է համակարգի պլատֆորմից անկախ լինելու և լոկալ ցանցի կամ համացանցի միջոցով հասանելի լինելու պահանջը: Web օգտվողի ինտերֆեյսի մշակման համար օգտագործվել են հետևյալ տեխնոլոգիաները՝ HTML, CSS, JavaScript:

Համակարգի հաջորդ բաղադրիչը PHP սերվերային սկրիպտն է, որը հանդիսանում է կապող կամուրջ օգտվողի web ինտերֆեյսի և մաթեմատիկական և վիճակագրական հաշվարկներն իրականացնող Java հավելվածի միջև: PHP սերվերային սկրիպտի միջոցով իրականացվում է մուտքագրված ֆայլերի վավերացում,

ընթերցում, անհրաժեշտության դեպքում ձևափոխությունների իրականացում: Մուտքագրված թեստավորման ֆայլերի համար ընդունելի է CSV և TXT ձևաչափերը, որոնց կազմման սկզբունքները, ինչպես նաև օրինակները հասանելի են օգտվողի web ինտերֆեյսում: Գոյություն ունեն մի շարք տվյալների պահպանման ձևաչափեր՝ CSV, XML, TXT: CSV (Comma Separated Values)-ն [41] տվյալների պահպանման ֆայլի ձևաչափ է, որում տվյալները պահպանվում են աղյուսակների նման և յուրաքանչյուր տող համապատասխանում է մեկ գրանցման, իսկ դաշտերն էլ միմյանցից առանձնացվում են ստորակետով (comma): XML ձևաչափում տվյալները ներկայացվում են տեգերի, դրանց բնութագրիչների և արժեքների սխեմայի տեսքով, որը մշակվում է XML գծանշման լեզվով (eXtensible Markup Language) [49]: Ծրագրավորման տեխնոլոգիաների կիրառման տեսանկյունից այս ձևով տվյալների պահպանումը ունի շատ մեծ առավելություններ: Սակայն լայն բնագավառի օգտվողների կողմից ավելի հեշտ է CSV ձևաչափի կիրառությունը: Ֆայլի TXT ձևաչափը [84] տվյալներ պահելու հիմնական միջոց է, որտեղ ազատ ոճով կարելի է պահպանել տվյալները, սահմանել տվյալների դաշտերի առանձնացման նշիչները և այլն: TXT ձևաչափի թերություններից կարելի է նշել այն, որ օգտվողի համար դժվար է ազատ ոճի կառուցվածքի հետ աշխատելը, շատ դեպքերում լրացուցիչ բացատրների կամ նոր տողերի ավելացումը բերում է տվյալ ֆայլից տվյալների ընթերցման սխալի:

CSV ձևաչափի առավելություններից կարելի է նշել.

- մարդու կողմից հեշտությամբ է ընկալվում տվյալների կառուցվածքը և բովանդակությունը,
- հեշտ է տվյալների խմբագրելը, խմբագրման համար կարելի է կիրառել մի շարք պարզագույն միջոցներ՝ տեքստային խմբագրիչ (Notepad), Excel աղյուսակային խմբագրիչ, Google Docs ամպային տեխնոլոգիա, առցանց CSV խմբագրիչներ,
- CSV բովանդակությունն ավելի սեղմ է և պարզ, XML-ի դեպքում յուրաքանչյուր տողի յուրաքանչյուր սյան համար պետք է գրվի տեգ, իսկ այս դեպքում տվյալների դաշտերը գրանցվում են միայն աղյուսակի առաջին տողում,

- տարբեր տեխնոլոգիաներ ունեն պարզ և բազմազան միջոցներ այս տիպի ֆայլերի հետ աշխատելու համար:

Մեր խնդրի շրջանակներում CSV ֆայլի ընտրությունը պայմանավորված էր հետևյալ առանձնահատկություններով.

- օգտվողը, որը հիմնականում մանկավարժն է, կկարողանա հետությամբ ընկալի ֆայլում տվյալների կառուցվածքը և բովանդակությունը,

- օգտվողը կկարողանա հեշտությամբ խմբագրել կամ ստեղծել թեստավորման արդյունքների ֆայլ Excel կամ որևէ այլ տեքստային խմբագրիչով,

- շատ հեշտ կլինի տարբեր տեխնոլոգիաներով մշակված թեստավորման համակարգերից թեստավորման արդյունքների ֆայլերի գեներացումը CSV ձևաչափով:

CSV ֆայլերի ընթերցումը, ինչպես նաև գեներացումը կատարվում է և PHP և Java ծրագրային կոդերում: PHP ծրագրային կոդում CSV ֆայլերի հետ աշխատելու համար օգտագործվում է ֆայլերը բացելու `fopen()` ֆունկցիան, CSV տվյալները ընթերցելու `fgetcsv()` ֆունկցիան և ընթերցված տվյալների զանգվածի հետ աշխատելու `array` ֆունկցիաները:

Սերվերից մաթեմատիկական և վիճակագրական հաշվարկներն իրականացնող Java հավելվածի գործարկման համար PHP ծրագրային կոդում կիրառվում է `exec()` ֆունկցիան: `exec()` ֆունկցիան թույլ է տալիս սերվերի վրա գործարկել արտաքին ծրագրեր, `shell` հրամաններ՝ ուղարկելով ծրագրային հրաման, մուտքային արգումենտներ և ստանալ ելքային արժեքները: `exec()` ֆունկցիայի արգումենտ ծրագրային հրամանն ունի հետևյալ տեսքը՝ `java - jar TestAnalyzer.jar`, որտեղ `TestAnalyzer.jar`-ը մաթեմատիկական և վիճակագրական հաշվարկներն իրականացնող Java հավելվածն է: Կարևոր է նշել, որ `exec()` ֆունկցիայի ակտիվացման համար պետք է `php.ini` կոնֆիգուրացիոն ֆայլում փոխել համապատասխան նշիչի արժեքը:

Մաթեմատիկական և վիճակագրական հաշվարկներն իրականացնող Java հավելվածն առանձին բաղադրիչ է: Այս բաղադրիչը մշակված է Java լեզվով, որը տալիս է հետևյալ առավելությունները.

- մշակված մոդուլը բազմապլատֆորմային է, այսինքն կարող է գործարկվել Java Virtual Machine [86] սատարող տարբեր օպերացիոն համակարգերի վրա՝ Windows, Linux, MacOS,
- Java մոդուլը կարող է օգտագործվել առանձնացված, որպես desktop հավելված, թեստավորման որակի գնահատման ամբողջ գործընթացը ինտերակտիվ ռեժիմում կազմակերպելու նպատակով,
- IRT և CTT ցուցանիշների հաշվարկների համար կարելի է ինտեգրել java լեզվով գրված Java psychometrics API բաց ծրագրային կոդով գրադարանը,
- Java լեզուն շատ ճկուն է աշխատում մի շարք բաց ծրագրային կոդով տվյալների բազաների կառավարման համակարգերի հետ՝ օրինակ Java Apache Derby:

Այս մոդուլում IRT և CTT ցուցանիշները իրականացվում են բաց ծրագրային կոդով Java psychometrics API գրադարանի հիման վրա, որը բաղկացած է 21 package-ից, ավելի քանի 200 դասից (class): Java psychometrics API գրադարանի package-ների ամբողջական ցանկը բերված է աղյուսակ 12-ում: Java psychometrics API գրադարանում առկա են մի շարք մաթեմատիկական և վիճակագրական ֆունկցիաներ, որոնք ներառված չեն Apache Commons Math գրադարանում [85] : Այս գրադարանը պարունակում է IRT ցուցանիշների գնահատման մի շարք դասեր, մասնավորապես, սատարվում է IRT միաճյուղ գնահատման մեկ, երկու, երեք պարամետրանոց մոդելները, բազմաճյուղ գնահատման մասնակի միավորներով և ընդհանրացված մասնակի միավորներով մոդելները: CTT ցուցանիշների գնահատումը սատարող կլասների միջոցով կարելի է հաշվարկել հուսալիության մի շարք գործակիցներ, տարբեր բաշխումներ, բարդություն, տարբերակիչ ունակություն և այլն:

Այս մոդուլի միջոցով ևս կատարվում է CSV ֆայլերի ընթերցում, տվյալները գրանցվում է Java Apache Derby [27] բաց ծրագրային կոդով ռեյացիոն տվյալների բազայում: Java Apache Derby տվյալների բազայի կառավարման համակարգը տարածվում է անվճար Apache 2.0 [28] ստանդարտով: Այստեղ ստեղծվում է թեստավորման արդյունքների, առաջադրանքների բանալիների աղյուսակները, ինչպես նաև ծրագրի կատարման ընթացքում գեներացվող տարբեր ցուցանիշների

արժեքների աղյուսակները: Առաջարկված նոր մեթոդի հիման վրա էնտրոպիայի, պայմանական էնտրոպիայի և փոխադարձ միջին ինֆորմացիայի մեծությունները ևս հաշվարկվում են այս մոդուլի միջոցով: Այս ցուցանիշների համար կիրառվում են Java լեզվի Math դասի մեթոդները:

Աղյուսակ 12.

Java psychometrics API գրադարանի package-ների ամբողջական ցանկը

<code>com.itemanalysis.psychometrics.analysis</code>
<code>com.itemanalysis.psychometrics.cfa</code>
<code>com.itemanalysis.psychometrics.cmh</code>
<code>com.itemanalysis.psychometrics.data</code>
<code>com.itemanalysis.psychometrics.distribution</code>
<code>com.itemanalysis.psychometrics.factoranalysis</code>
<code>com.itemanalysis.psychometrics.histogram</code>
<code>com.itemanalysis.psychometrics.irt.equating</code>
<code>com.itemanalysis.psychometrics.irt.estimate</code>
<code>com.itemanalysis.psychometrics.irt.model</code>
<code>com.itemanalysis.psychometrics.kernel</code>
<code>com.itemanalysis.psychometrics.measurement</code>
<code>com.itemanalysis.psychometrics.mixture</code>
<code>com.itemanalysis.psychometrics.optimization</code>
<code>com.itemanalysis.psychometrics.polycor</code>
<code>com.itemanalysis.psychometrics.rasch</code>
<code>com.itemanalysis.psychometrics.reliability</code>
<code>com.itemanalysis.psychometrics.scaling</code>
<code>com.itemanalysis.psychometrics.statistics</code>
<code>com.itemanalysis.psychometrics.texttable</code>
<code>com.itemanalysis.psychometrics.tools</code>

Վերջնական հաշվետվությունների ֆայլը գեներացվում և ձևավորվում է հաշվետվությունների ձևավորման մոդուլի միջոցով, որը ևս մշակված է Java լեզվով: Հաշվետվությունների ֆայլը տրվում է rtf ձևաչափով, որի ձևավորման համար ներառվել է Apache POI API [29] գրադարանը: Apache POI API գրադարանն ունի Microsoft Office տարբեր ընդլայնումով փաստաթղթերի հետ աշխատելու միջոցներ՝ Word, Excel, Power Point, Visio, Publisher և այլն: Հաշվետվությունների ֆայլը ներկայացվում է rich text format-ով [81], որի մշակման համար կիրառվել է Apache POI API գրադարանի XWPF (XML Word Processor Format) [30] բաղադրիչը: Հաշվետվության ֆայլում տեքստային հատվածներ տեղադրելու համար օգտագործվում է XWPFDocument կլասի createParagraph() մեթոդը, իսկ աղյուսակներ ստեղծելու համար՝ createTable() մեթոդը: XWPFRun կլասի մեթոդների միջոցով սահմանվում են տեքստի տառաչափերը, գույնը, անցումները, ոճերը: Այն ցուցանիշները, որոնք հաշվարկվել են և դրանց արժեքները անբավարար են, ըստ նշիչի հաշվետվության մեջ ներկայացվում են կարմիր գունավորումով: Գրաֆիկները ևս ներդրվում են հաշվետվության ֆայլի մեջ: Ներդրման համար կիրառվում է ֆայլերի և դիրեկտորիաների հետ աշխատելու Java լեզվի միջոցները XWPFDocument կլասի addPicture() մեթոդը: Գրաֆիկների գեներացման համար ներառվել է JFreeChart անվճար գրադարանը [61], որը տարածվում է GNU Lesser General Public Licence լիզենցիայով: JFreeChart գրադարանը տարամադրում է շատ ճկուն միջոցներ գրաֆիկներ ստեղծելու համար: Այն սատարում է տարատեսակ երկչափ և եռաչափ գրաֆիկներ, գրաֆիկների ֆայլերի JPG, PNG, SVG և այլ ձևաչափեր, ունի շատ մանրամասն ու մատչելի նկարագրություններ: Այս գրադարանի միջոցով գունավոր տեսքով գեներացվում են առաջադրանքների բնութագրական կորերը և ինֆորմացիոն ֆունկցիայի գրաֆիկները, թեստի բնութագրական կորը, ինֆորմացիոն ֆունկցիայի և չափումների ստանդարտ սխալի գրաֆիկը, թեստի քարտեզը, ինչպես նաև այլ բաշխումների գրաֆիկներ:

4.7. Եզրակացություններ

Մշակվել է համակարգ, որը բացի նմանատիպ համակարգերին բնորոշ առանձնահատկություններից, ունի մի շարք առավելություններ, որոնք չկան այդ համակարգերում: Այդպիսի առավելություններն են.

- Իրականացնում է թեստերի որակի գնահատում միաժամանակ CTT-ի, IRT-ի տեսությունների և մեր կողմից մշակված նոր մոտեցման հիման վրա:
- Օգտվողի ինտերֆեյսը web հենքով է, ինչը թույլ է տալիս օգտվողին դիմել համակարգին լոկալ ցանցի միջոցով կամ համացանցի միջոցով՝ գործարկելով այն web զննարկիչի միջավայրում:
- Համակարգը պլատֆորմից անկախ է և որպես արդյունք այն կարող է գործարկվել տարբեր օպերացիոն համակարգերի միջավայրերում:
- Օգտվողի ինտերֆեյսը, ինչպես նաև այլ փաստաթղթեր հասանելի են հայերեն լեզվով:
- Գեներացվող հաշվետվության ֆայլը ձևավորված rtf փաստաթուղթ է, որն իր մեջ ներառում է ձևավորված աղյուսակներ, գրաֆիկներ:
- Առանձին csv ձևաչափի ֆայլերի տեսքով ստացվում են նաև մի շարք աղյուսակներ, ինչպես նաև JPG ձևաչափով ստացվում են բոլոր գրաֆիկները:
- Քանի որ Java հավելվածները պլատֆորմից անկախ են, համակարգի բաղադրիչ հանդիսացող վիճակագրական և մաթեմատիկական ցուցանիշները հաշվարկող մոդուլը կարող է կիրառվել առանձին, որպես desktop հավելված:
- Պարզեցված օգտվողի ինտերֆեյսը հնարավոր է դարձնում համակարգի կիրառումը օգտվողների լայն շրջանակների համար:

ԳԼՈՒԽ V.

ՆՈՐ ՄԵԹՈԴԻ ԲՆՈՒԹԱԳՐԻՉՆԵՐԻ ԸՆԴՈՒՆԵԼԻ

ԱՐԺԵՔՆԵՐԻ ՈՐՈՇՈՒՄ

Թեստերի որակի գնահատման նոր մոդելի շրջանակներում հաշվարկվող ցուցանիշների սահմանային թույլատրելի արժեքների, ինչպես նաև վավեր վերլուծության իրականացման համար փորձարկվողների բավարար քանակի որոշման նպատակով մոդելը փորձարկվել է մեծածավալ թեստավորման արդյունքների համար [56]: Մասնավորապես, փորձնական եղանակով ստացվել են հետևյալ հարցերի պատասխանները.

- որքա՞ն ճշգրտությամբ է առաջարկված նոր մոդելը գնահատում թեստերի որակը,
- որքանո՞վ է այն համադրելի CTT և IRT գնահատման մեթոդների հետ,
- որո՞նք են հաշվարկվող $H(X_i)$ և $I(X_i X_j)$ մեծությունների սահմանային թույլատրելի ճշգրիտ արժեքները,
- նվազագույնը ի՞նչ քանակությամբ փորձարկվողների խումբ է անհրաժեշտ մոդելով վավեր վերլուծություն և գնահատում իրականացնելու համար:

5.1. Թեստերի որակի գնահատման նոր մոդելի փորձարկումների արդյունքները

Փորձարկման համար ընտրվել են 2008 թվականի «Հայոց լեզու և գրականություն» առարկայի ավարտական միասնական քննությունների թեստավորման արդյունքները, որոնք ծածկագրված տեսքով տրամադրվել է ՀՀ կառավարության աշխատակազմի «Գնահատման և թեստավորման կենտրոն» ՊՈԱԿ-ի [2] կողմից: Վերլուծության համար ընտրվել է 4 թեստ, որոնցից յուրաքանչյուրը պարունակում է 80 առաջադրանք, իսկ թեստավորմանը մասնակցած փորձարկվողների ընդհանուր թիվը 2000 է: Առաջադրանքներն ունեն անվանումներ. առաջին 50 X_i առաջադրանքների անվանումներն են համապատասխանաբար $A_1, A_2, \dots, A_{50}$, իսկ վերջին 30 X_i

առաջադրանքներինն է՝ $B_1, B_2, \dots, B_{30}$: Վերլուծության համար օգտագործվել է մեր կողմից մշակված «Թեստերի որակի վերլուծություն» համակարգը [40]:

Չորս թեստերի յուրաքանչյուր առաջադրանքի համար հաշվարկվել են $H(X_i)$ էնտրոպիայի (17), U_m CTT բարդության ինդեքսի (7) և IRT b պարամետրի(9) արժեքները: Ուսումնասիրվել է նշված պարամետրերի համադրելիությունը և որոշվել է $H(X_i)$ պարամետրի թույլատրելի սահմանային արժեքները:

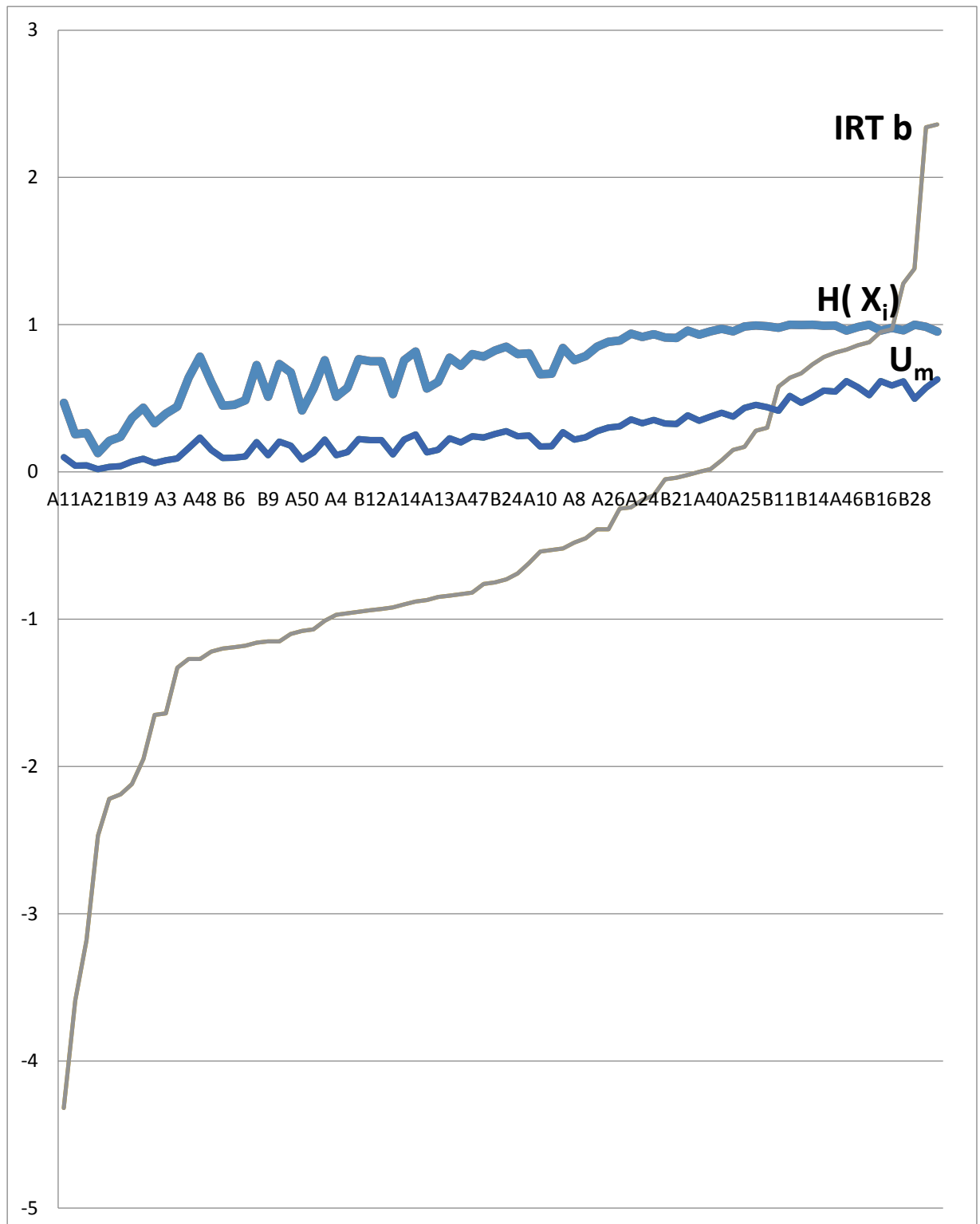
$H(X_i)$ և $I(X_i \wedge X_j)$ մեծությունների արժեքների կախվածությունը փորձարկվողների խմբի քանակից ուսումնասիրելու և համապատասխան փորձարկվողների բավարար քանակը որոշելու համար յուրաքանչյուր թեստի համար կատարվել է հինգ փորձարկում: Յուրաքանչյուր փորձարկման համար պատահական ընտրությամբ առանձնացվել է համապատասխանաբար 500, 300, 200, 100 և 50 փորձարկվողի արդյունք:

Հաշվարկվել և վերլուծվել են յուրաքանչյուր թեստի X_i և X_j առաջադրանքների միջև $I(X_i \wedge X_j)$ փոխադարձ ինֆորմացիայի և CTT $R(X_i, X_j)$ կոռելյացիոն գործակցի արժեքները: Որոշվել է $I(X_i \wedge X_j)$ փոխադարձ ինֆորմացիայի մեծության թույլատրելի սահմանային արժեքները:

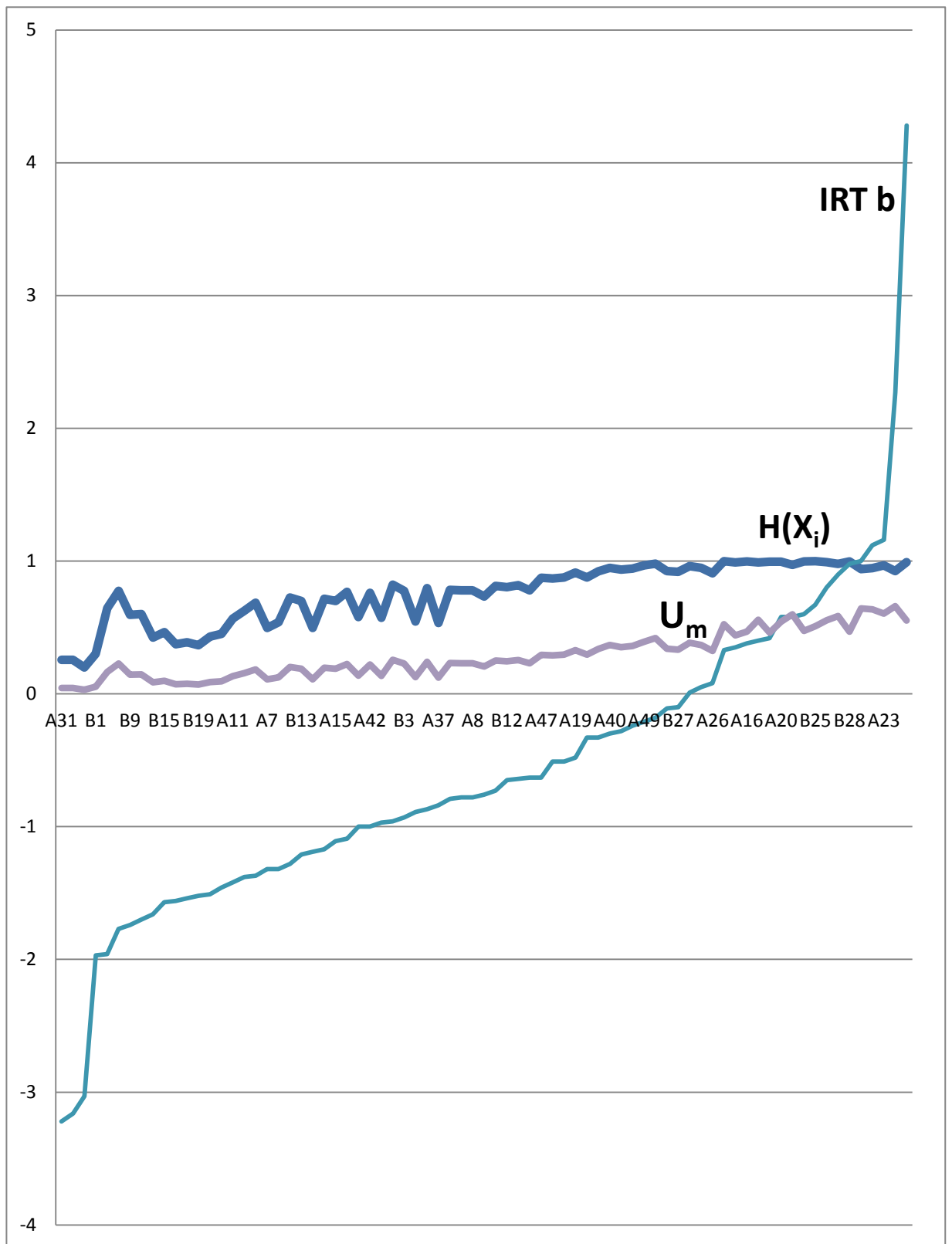
Թեստ 1-ի առաջին փորձարկման արդյունքների հիման վրա $H(X_i)$, CTT բարդության ինդեքսի և IRT b պարամետրի արժեքների համեմատությունը ներկայացված է նկար 22-ում:

Ըստ CTT տեսության թեստային առաջադրանքները, որոնց բարդությունը արժեքներ է ընդունում 0,3-0,74 միջակայքից, լավ առաջադրանքներ են, այսինքն՝ ոչ շատ բարդ և ոչ էլ շատ հեշտ: Այդպիսի առաջադրանքների քանակը թեստում 34 –ն է, որոնց համապատասխան $H(X_i)$ մեծությունը արժեքներ է ընդունում 0,82-1,0 միջակայքից: Հեշտ առաջադրանքների քանակը 29 է, որոնց բարդության արժեքը գտնվում է 0,75-0,9 միջակայքում, իսկ $H(X_i)$ մեծության արժեքը 0,48-0,81 միջակայքում: Շատ հեշտ առաջադրանքներին, որոնց բարդությունը արժեքներ է ընդունում 0,9-1,0 միջակայքից, համապատասխանում է $H(X_i)$ մեծության 0,12-0,47 միջակայքի արժեքներ: Այսպիսի առաջադրանքների թիվը 14 է: Ակնհայտ է նաև, որ

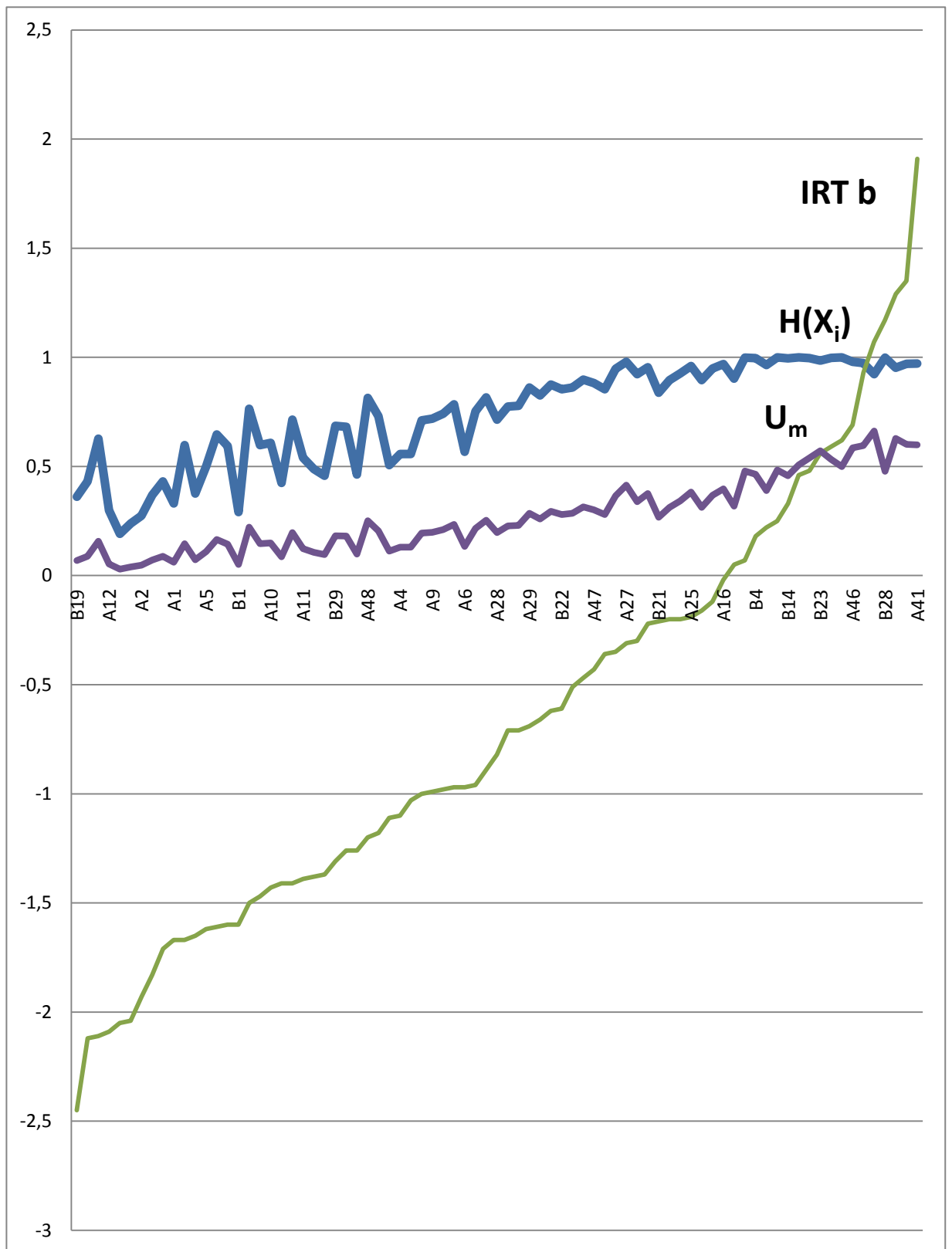
IRT b պարամետրի աճին համապատասխանում է $H(X_i)$ մեծության աճ:
Մոտավորապես նույն արդյունքներն են ստացվել նաև մյուս 3 թեստերը վերլուծելիս:



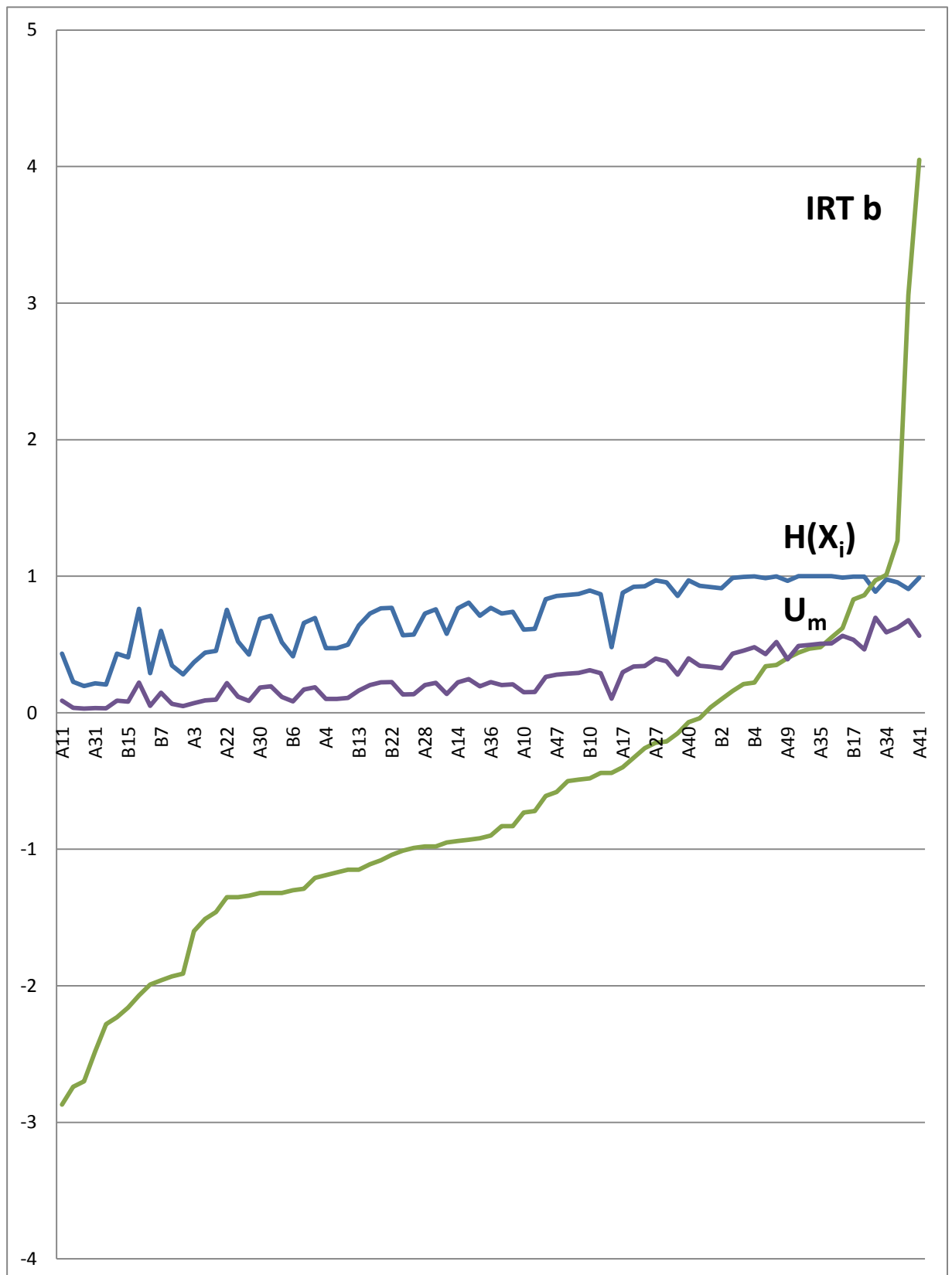
Նկար 22. Թեստ 1-ի $H(X_i)$, CTT բարդության ինդեքսի (U_m) և IRT b պարամետրի գրաֆիկները:



Նկար 23. Թեստ 2-ի $H(X_i)$, CTT բարդության ինդեքսի (U_m) և IRT b պարամետրի գրաֆիկները:

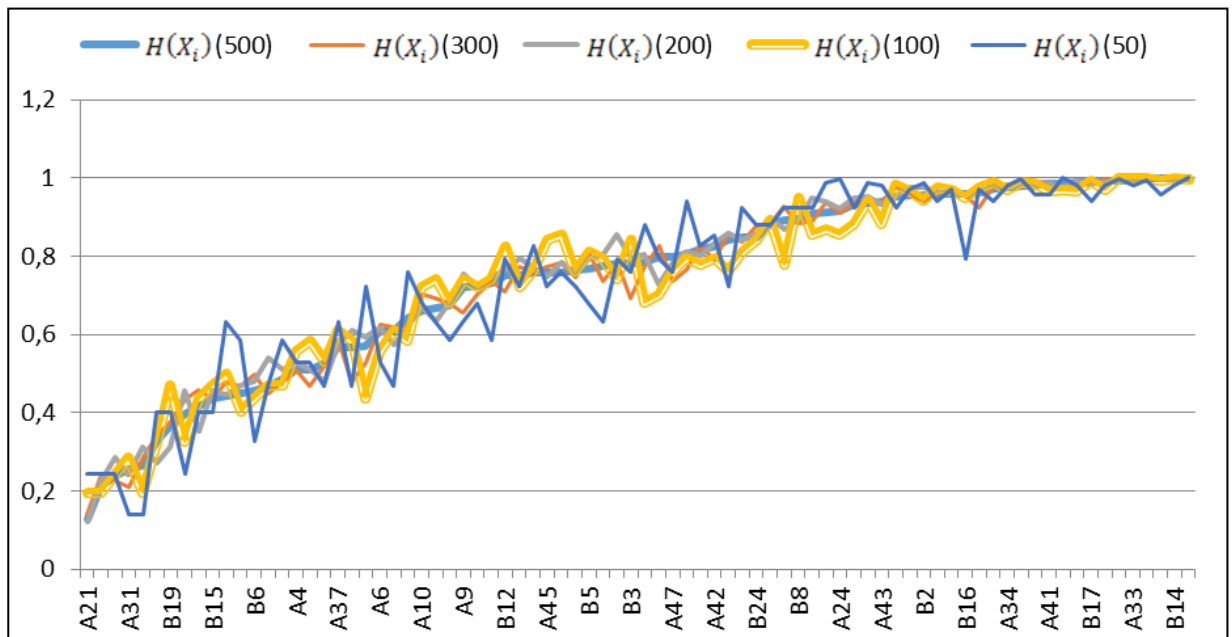


Նկար 24. Թեստ 3-ի $H(X_i)$, CTT բարդության ինդեքսի (U_m) և IRT b պարամետրի գրաֆիկները:



Նկար 25. Թեստ 4-ի $H(X_i)$, CTT բարդության ինդեքսի (U_m) և IRT b պարամետրի 5 գրաֆիկները:

$H(X_i)$ մեծության արժեքների կախվածությունը փորձարկվողների խմբի քանակից ուսումնասիրելու նպատակով յուրաքանչյուր թեստի համար հինգ փորձարկումների արդյունքների հիման վրա կառուցվել է $H(X_i)$ գրաֆիկը:



Նկար 26. Թեստ 1-ի $H(X_i)$ մեծության 5 փորձարկումների գրաֆիկը:

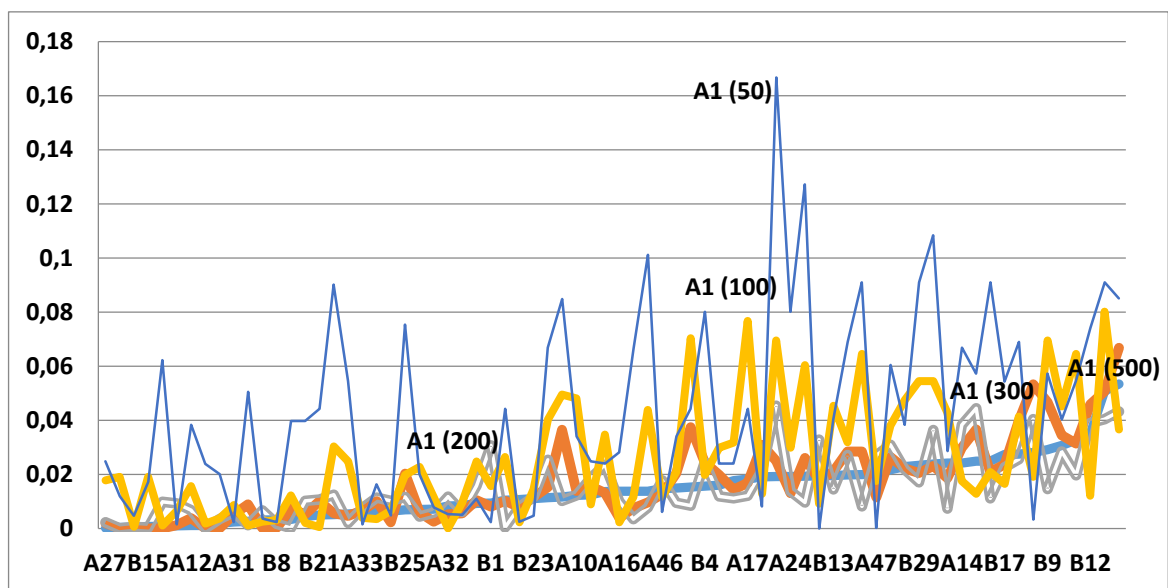
Վերլուծվել են $H(X_i)$ մեծության արժեքների շեղումները՝ ըստ յուրաքանչյուր փորձարկման արդյունքների: $H(X_i)$ մեծության արժեքների առավելագույն շեղումները ներկայացված են աղյուսակ 13-ում:

Արդյունքների վերլուծությունը ցույց է տալիս, որ փորձարկվողների քանակը մինչև 100 նվազեցնելիս $H(X_i)$ մեծության արժեքների շեղումները փոքր են և առավելագույն շեղումը 0,15 է: Սակայն 100-ից ավելի փոքր քանակի դեպքում, մասնավորապես 50-ի, արժեքների շեղումը մեծանում է և հասնում է մոտ 0,3:

Նույն սկզբունքով էլ յուրաքանչյուր թեստի համար հաշվարկվել է $I(X_i \wedge X_j)$ միջին փոխադարձ ինֆորմացիայի արժեքները և ուսումնասիրվել այդ արժեքների կախվածությունը փորձարկվողների քանակից: Նկար 27 - ում ներկայացված է թեստ1-ի A1 առաջադրանքի միջին փոխադարձ ինֆորմացիայի գրաֆիկը տաբեր փորձարկումների հիման վրա:

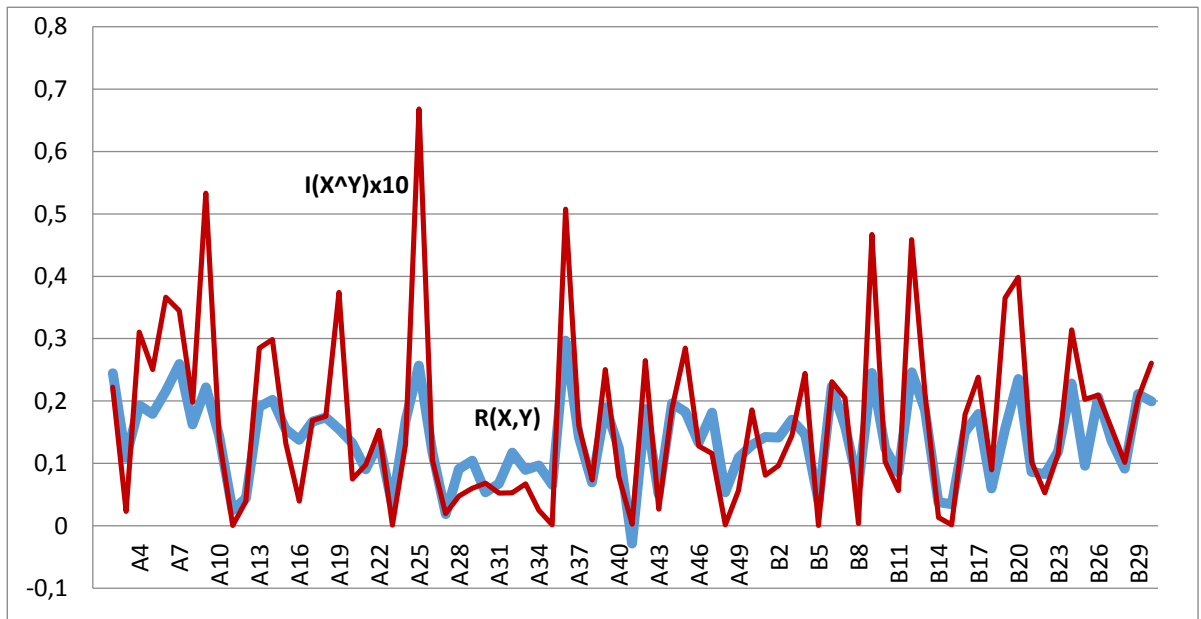
Թեստ 1-ի $H(X_i)$ մեծության արժեքների շեղումները՝ ըստ յուրաքանչյուր փորձարկման արդյունքների

$H(X_i)$ մեծության արժեքների առավելագույն շեղումները	500	300	200	100	50
500	-	0.089	0.07	0.135	0.2
300	0.089	-	0.13	0.15	0.19
200	0.07	0.13	-	0.16	-
100	0.135	0.15	0.16	-	0.3
50	0.2	0.19	0.23	0.3	-

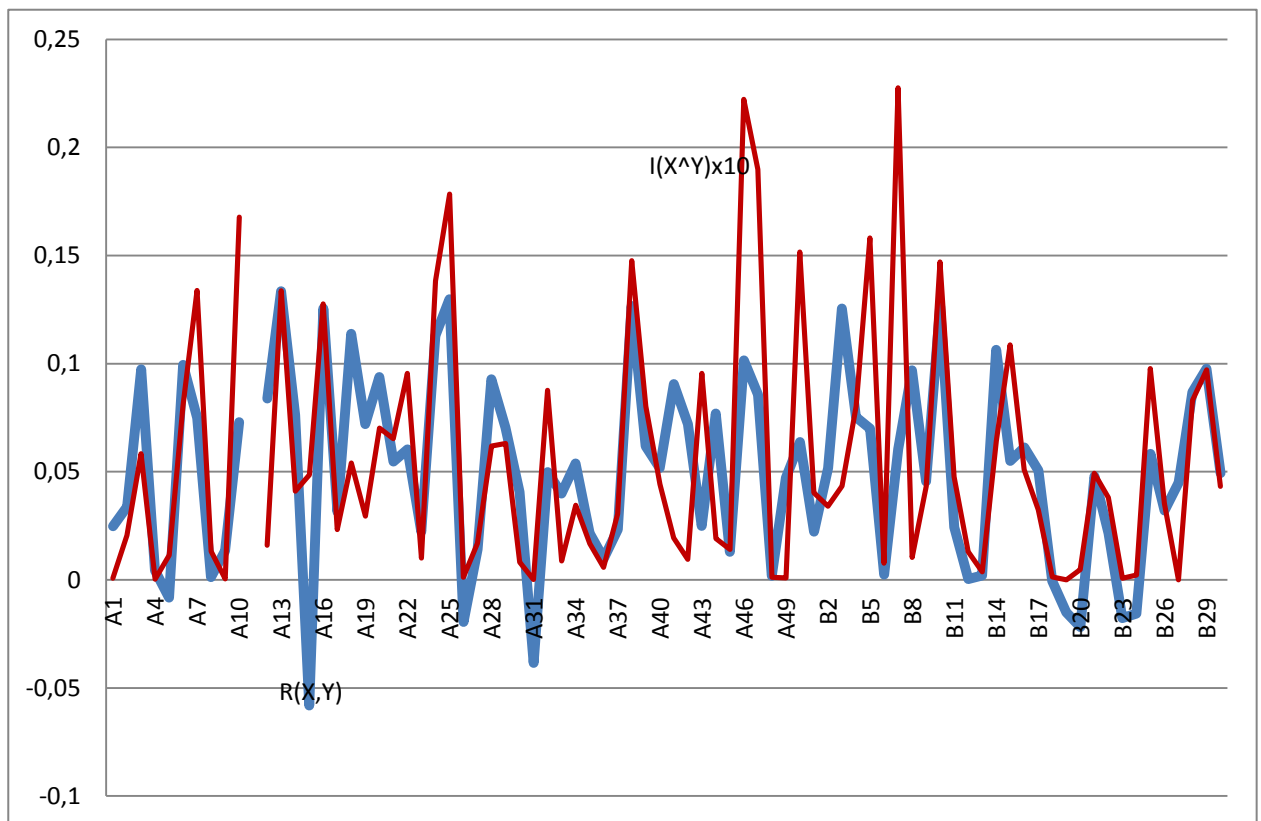


Նկար 27. Թեստ 1-ի A1 առաջադրանքի $I(X_i \wedge X_j)$ մեծության 5 փորձարկումների գրաֆիկները:

Առաջադրանքների միջին փոխադարձ ինֆորմացիայի և CTT կոռելյացիայի գործակցի միջև կապը ևս վերլուծվել է: Որոշ առաջադրանքների գրաֆիկները ներկայացված են նկար 28-ում և նկար 29-ում: Քանի որ միջին փոխադարձ ինֆորմացիայի արժեքները շուրջ 10 անգամ փոքր են CTT կոռելյացիայի գործակցի արժեքների նկատմամբ և համադրությունը ավելի ակնառու երևալու նպատակով այս երկու մեծությունները ցուցադրվում են նույն գրաֆիկի վրա, այդ իսկ պատճառով էլ գրաֆիկի վրա միջին փոխադարձ ինֆորմացիայի մեծության արժեքը 10 անգամ մեծացվել է:



Նկար 28. A1 առաջադրանքի և մնացած առաջադրանքների միջև CTT կոռելյացիայի ($R(X_i, X_j)$) և միջին փոխադարձ ինֆորմացիայի ($I(X_i \wedge X_j)$) գրաֆիկները:



Նկար 29. A11 առաջադրանքի և մնացած առաջադրանքների միջև CTT կոռելյացիայի ($R(X_i, X_j)$) և միջին փոխադարձ ինֆորմացիայի ($I(X_i \wedge X_j)$) գրաֆիկները:

Վերլուծությունից և գրաֆիկներից երևում է, որ առաջադրանքների կոռելյացիայի գործակցի և միջին փոխադարձ ինֆորմացիայի մեծությունների արժեքները համադրելի են: Այսպես, A11 առաջադրանքը ունի բացասական կոռելյացիա մի շարք առաջադրանքների հետ և միևնույն ժամանակ այդ առաջադրանքներին համապատասխանող միջին փոխադարձ ինֆորմացիան ևս ընդունում է շատ փոքր արժեքներ: Այսինքն՝ անհրաժեշտություն է առաջանում այս առաջադրանքը հեռացնել թեստերի բանկից: Վերլուծելով տվյալները նաև որոշվել է նաև առաջադրանքի միջին փոխադարձ ինֆորմացիայի մեծության նվազագույն շեմը՝ 0,005:

5.2. Եզրակացություններ

Առաջարկված թեստերի որակի գնահատման նոր մոդելի փորձարկումներից և կատարված արդյունքների վերլուծությունից հետևում է.

1. Մեր կողմից առաջարկված մեթոդները ճշգրիտ գնահատում են թեստերի որակը և համադրելի են CTT և IRT գնահատման մեթոդների հետ:
2. IRT տեսության համեմատությամբ առավել պարզ մաթեմատիկական վերլուծություն է կատարվում:
3. Առաջադրանքների փոխկապակցվածությունը բնութագրող մեծությունն ունի համարժեքը CTT-ում, իսկ IRT-ում համարժեքը բացակայում է:
4. Լավ առաջադրանքների համար $H(X_i)$ մեծությունը արժեքներ է ընդունում 0,8-1,0 միջակայքից, հեշտ (բավարար բարդությամբ) առաջադրանքների համար միջակայքը հետևյալն է՝ 0,45-0,8, իսկ վատ առաջադրանքների համար (շատ հեշտ կամ շատ բարդ)՝ 0-0,44:
5. Լավ առաջադրանքի միջին փոխադարձ ինֆորմացիայի մեծության ստորին և վերին սահմանային արժեքները որոշված են՝

$$0.005 \leq I(X_i \wedge X_j) \leq 0.05$$

6. IRT-ի և CTT-ի համեմատությամբ ավելի փոքր քանակությամբ թեստավորվողների խումբ է հարկավոր վավեր վերլուծություն և գնահատում

իրականացնելու համար: Փորձարկվողների նվազագույն քանակը 100-ն է, իսկ CTT-ում նախընտրելի է 200-500, IRT-ի որոշ մոդելների դեպքում 500 և ավել:

ԵԶՐԱԿԱՑՈՒԹՅՈՒՆՆԵՐ

Այսպիսով, աշխատանքում ստացվել են հետևյալ հիմնական արդյունքները.

- Մշակվել է թեստերի որակի գնահատման նոր մեթոդ՝ հիմնված ինֆորմացիայի տեսության տարրերի վրա [41]:
- Որոշվել են առաջարկված մեթոդի ցուցանիշների ընդունելի սահմանները և վավեր վերլուծության համար անհրաժեշտ թեստավորման արդյունքների քանակը, որն էականորեն ավելի փոքր է մյուս մեթոդների հետ համեմատած [1],[42]:
- Մշակվել է CTT, IRT տեսությունների և առաջարկված նոր մեթոդի հիման վրա թեստերի որակի վերլուծություն և գնահատում իրականացնող «Թեստերի որակի վերլուծություն» ծրագրային համակարգը, որը գործարկվում է սերվերի վրա, ունի web հենքով օգտվողի պարզ և մատչելի ինտերֆեյս, տրամադրում է rtf ձևաչափի վերլուծության հաշվետվության ֆայլ [26],[40]:
- Մշակվել է վիճակագրական և մաթեմատիկական ցուցանիշները հաշվարկող մոդուլի GUI ինտերֆեյս, ինչի շնորհիվ էլ այն կարող է կիրառվել նաև առանձին համակարգչի վրա որպես desktop ծրագրային հավելված: Մշակված մոդուլը բազմապլատֆորմ է և գործարկվում է տարբեր օպերացիոն համակարգերի միջավայրերում [40]:

ՕԳՏԱԳՈՐԾԱԾ ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. Ավետիսյան Վ.Կ., Գիտելիքների ստուգման թեստերի որակական բնութագրերի հետազոտություն, Հայաստանի Ճարտարագիտական Ակադեմիայի Լրաբեր, Հ.12, N1, Երևան, 2015, էջ 156-163
2. Գնահատման և թեստավորման կենտրոն, <http://www.atc.am/>
3. Մելիքյան Հ.Ա., Թեստային առաջադրանքների մշակման սկզբունքների և փուլերի մասին, Գավառի պետական համալսարանի գիտական հոդվածների ժողովածու, Հ.14, Երևան, 2012, էջ 107-111
4. Аванесов В.С., Истоки и основные понятия математической теории измерений (Item Response Theory), Педагогические Измерения, №3, 2007, с. 3-37.
5. Аванесов В.С., Композиция тестовых заданий, Учебная книга для преподавателей вузов, учителей школ, аспирантов и студентов педвузов, 2 изд., испр. и доп., Москва, 1998, 217 с.
6. Аванесов В.С., Методологические и теоретические основы тестового и педагогического контроля, автореф. дис. ... докт. пед. наук., Санкт-Петербург, Гос. ун-т, 1994, 32 с.
7. Аванесов В.С., Основы научной организации педагогического контроля в высшей школе, Москва, 1989, 167 с.
8. Аванесов В.С., Item Response Theory: основные понятия и положения, Педагогические измерения, № 2, Москва, 2007, с.3-28.
9. Анастаси А., Урбина С., Психологическое тестирование, Санкт-Петербург, Питер, 2006, 688 с.
10. Беспалько В.П., Основы теории педагогических систем, Воронеж, Издательство Воронежского университета, 1977, 304 с.
11. Глас Дж., Стэнли Дж., Статистические методы в педагогике и психологии. Москва, Прогресс, 1976, 496 с.
12. Карпенко А. П., Домников А. С., Белоус В. В, Тестовый метод контроля качества обучения и критерии качества образовательных тестов. Обзор, Наука и образование: научное издание МГТУ им. Н.Э. Баумана, №04, Москва, 2011, с.1.

13. Ким В.С., Тестирование учебных достижений, Монография, Уссурийск, Издательство УГПИ, 2007, 214 с.
14. Клайн П., Справочное руководство по конструированию тестов, Киев, 1994, 184 с.
15. Майоров А.Н., Теория и практика создания тестов для системы образования, Москва, «Интеллект-центр», 2001, 296 с.
16. Маслак А.А., Измерение латентных переменных в социально-экономических системах: Монография, Славянск-на-Кубани, Изд.центр СГПИ, 2006, 333 с.
17. Маслак А.А., Критерии оценки качества тестов, Труды международного симпозиума, Томск, Изд-во ТПУ, 2004, с. 58-60.
18. Нейман Ю.М., Хлебников В.А., Введение в теорию моделирования и параметризации педагогических тестов, Москва, 2000, 168 с.
19. Олейник Н.М., Тест как инструмент измерения уровня знаний и трудности заданий в современной технологии обучения, Учеб. пособие по спец. курсу, Донецк, ДонГУ, 1991, 168 с.
20. Переверзев В. Ю., Критериально-ориентированные педагогические тесты для итоговой аттестации студентов, Москва, НМЦ СПО Минобразования РФ, 1999, 152 с.
21. Чельшкова М.Б., Разработка педагогических тестов на основе современных математических моделей: Учебное пособие, Москва, 1995, 32 с.
22. Чельшкова М.Б., Теория и практика конструирования педагогических тестов: Учебное пособие, Москва, Логос, 2002, 432 с.
23. ACER ConQuest 3.0.1 computer program, <http://www.acer.edu.au/conquest>,
24. Andrich D., Rasch Models for Development, London, Sage Publications, 1988, 94 p.
25. Andrich D., Sheridan B. & Luo G., Rumm 2030: Rasch Unidimensional Measurement Models (software), Western Australia, RUMM Laboratory Perth, 2012.
26. Andrich D., Sheridan B., Lyne A. & Luo G., Welcome to RUMM. User's guide : a Windows-based item analysis program employing Rasch Unidimensional Measurement Models, Perth, Rasch Analyst, 1996, 44 p.
27. Apache Derby, <https://db.apache.org/derby/>

28. Apache License Version 2.0, <http://www.apache.org/licenses/LICENSE-2.0>
29. Apache POI - the Java API for Microsoft Documents, <https://www.poi.apache.org/>
30. Apache POI Word Tutorial, http://www.tutorialspoint.com//apache_poi_word/index.htm
31. Assessment Systems Corporation, <http://www.assess.com>
32. Avetisyan V., Survey of software for the test quality analysis, International Journal "Information Content and Processing", Volume 2, Number1, Bulgaria, 2015, pp. 82-92.
33. Ayala R.J. , The Theory and Practice of Item Response Theory, New York, The Guilford Press, 2009, p.144,
34. Baker F. B., GENIRV: A program to generate item response vectors (Unpublished manuscript), Madison, WI: University of Wisconsin, Laboratory of Experimental Design. 1989.
35. Berkeley Evaluation and Assessment Research (BEAR) Center, <http://www.bearcenter.berkeley.edu>
36. Bigsteps-DOS precursor to WINSTEPS. Final Version: 2.82, <http://www.winsteps.com/bigsteps.htm>
37. Birnbaum A., Some Latent Trait Models and Their Use in Inferring an Examinee's Ability, In F.M. Lord and M.R. Novick, Statistical Theories of Mental Test Scores, Reading (Mass.) Addison-Wesley, 1968, pp.397-479.
38. Bock R.D., Aitkin M., Marginal maximum likelihood estimation of item parameters: application of an EM algorithm, Psychometrika 46 (4), 1981, pp. 443-459.
39. Briggs D. C. & Wilson M. R., An Introduction to Multidimensional Measurement using Rasch Models, 4(1), 2003, 87-100.
40. CITAS - free item analysis with classical test theory, <http://www.assess.com/xcart/product.php?productid=407&cat=25&page=1>
41. Comma-separated values, https://en.wikipedia.org/wiki/Comma-separated_values,
42. Comparison of web server software, https://en.wikipedia.org/wiki/Comparison_of_web_server_software
43. ConstructMap software, <http://www.bearcenter.berkeley.edu/software/constructmap>

44. Cover T. M., Thomas J. A., Elements of Information Theory, 2nd Edition, New York, Wiley, 2006, 776p.
45. Crocker L., Algina J., Introduction to Classical and Modern Test Theory, New York, Wadsworth Pub Co, 2006, 527 p.
46. Cronbach L. J., Coefficient alpha and the internal structure of tests, Psychometrika, 16, 1951, pp. 297-334.
47. DeMars C., Item Response Theory, New York, Oxford University Press, 2010, 144 p.
48. Distributed Architecture,
http://www.tutorialspoint.com/software_architecture_design/distributed_architecture.htm
49. Extensible Markup Language (XML), <https://www.w3.org/XML/>
50. Facets: DOS precursor to the current Windows-based Facets. Final Version: 3.22,
<http://www.winsteps.com/facdost.htm>
51. Guttman L., A basis for analyzing test-retest reliability, Psychometrika, 10:4, 1945, pp. 255-282.
52. Hambleton R. K., Swaminathan H. & Rogers H. J., Fundamentals of item response theory, Newbury Park, CA: Sage, 1991, 184 p.
53. Hambleton R. K. and Jones R. W., Comparison of classical test theory and item response theory and their applications to test development, Educational Measurement: Issues and Practice, vol. 12, no.3, pp. 38-47, 1993.
54. Haroutunian M., Avetisyan. V., Development of the test quality evaluation system, Proceedings of the International Conference on Computer Science and Information Technologies (CSIT 2015), Yerevan, 2015, pp. 372-375.
55. Haroutunian M., Avetisyan. V., New approach for test quality evaluation based on Shannon Information measures, Mathematical Problems of Computer Science, vol. 44, 2015, pp. 7-21.
56. Haroutunian M., Avetisyan. V., Analysis of the Experiments of a New Approach for TestQuality Evaluation, Mathematical Problems of Computer Science, vol. 45, 2016, pp. 35-43.

57. IRTPRO 2.1 for Windows by Li Cai, David Thissen & Stephen du Toit, <http://www.ssicentral.com/irt>
58. Iteman 4 - Test and item analysis software with classical test theory, <http://www.assess.com/xcart/product.php?productid=417&cat=25&page=1>,
59. Java library for psychometric methods, <https://www.java.net/projects/psychometrics>,
60. JavaScript Object Notation, <http://www.json.org>
61. JFreeChart free software, <http://www.jfree.org/jfreechart>
62. JMetrik-computer program for psychometric analysis, <http://www.jmetrik.com/index.php>
63. Kuder G. F., Richardson M. W., The theory of the estimation of test reliability, *Psychometrika*, Volume 2, issue 3, 1937, pp. 151-160.
64. Linacre J. M., Winsteps® Rasch measurement computer program User's Guide, Beaverton, Oregon: Winsteps.com
65. Linacre J. M., A User's Guide to FACETS Rasch-Model Computer Programs, 2006, 172 p.
66. Linacre J. M., From Microscale to Winsteps: 20 years of Rasch Software development, *Rasch Measurement Transactions*, 17:4, 2004, 958 p.
67. Linden W. J., Hambleton R. K., Handbook of modern item response theory, New York, Springer-Verlag, 1997, 512 p.
68. Lord F.M., Applications of item response theory to practical testing problems. Mahwah, NJ: Lawrence Erlbaum Associates, 1980, 274 p.
69. Lord F.M., The relation of the reliability of multiple-choice tests to the distribution of item difficulties, *Psychometrika*, 17, 1952, pp. 181-194.
70. Meyer J. P., Applied Measurement with jMetrik, Routledge, 2014, 149 p.
71. MINIFAC- Evaluation, Student and Demonstration (Demo) Version of FACETS, <http://www.winsteps.com/minifac.htm>
72. Ministep-Evaluation, Student and Demonstration (Demo) Version of WINSTEPS, <http://www.winsteps.com/ministep.htm>

73. Muraki E., RESGEN: Item response generator. Princeton, NJ: Educational Testing Service, 1992, 60 p.
74. Novick M. R., The axioms and principal results of classical test theory, Journal of mathematical psychology, Volume 3, Issue 1, 1966, pp. 1-18.
75. Ostini R., Michael L. N., Handbook of polytomous item response theory models, Routledge, 2010, 306 p.
76. Ostini R., Michael L. N., Polytomous Item Response Theory Models. Quantitative Applications in the Social Sciences (Book 144), SAGE Publications, 2005, 120 p.
77. PHP: HyperText preprocessor, <https://secure.php.net/>
78. Polymetrika software, <http://polymetrika.com/home/IATA>
79. Rasch G., Probabilistic Models for Some Intelligence and Attainment Tests, Expanded edition, Chicago, The University of Chicago Press, 1980.
80. Rasch Measurement Analysis Software Directory, <http://www.rasch-analysis.com/index.htm>
81. Rich Text Format, https://en.wikipedia.org/wiki/Rich_Text_Format
82. Spearman C. , Correlation calculated from faulty data, British Journal of Psychology, Vol.3, N2, 1910, pp. 271-295.
83. Spearman C., The proof and measurement of association between two things, American Journal of Psychology, vol. 15, 1904, pp. 72-101.
84. Text file, https://en.wikipedia.org/wiki/Text_file
85. The Apache Commons Mathematics Library, <https://www.commons.apache.org/proper/commons-math>
86. The Java Virtual Machine Specification, <https://docs.oracle.com/javase/specs/jvms/se7/html/index.html>
87. Types of Reliability, The Research Methods Knowledge Base, <http://www.socialresearchmethods.net/kb/reotypes.php>
88. Wilson M., Constructing Measures: An Item Response Modeling Approach, Mahwah, New Jersey, Lawrence Erlbaum associates, 2005, 228 p.
89. Winsteps and Facets Comparison, <http://www.winsteps.com/winfac.htm>

90. Wright B.D. & Stone M.H., Best Test Design, Chicago, MESA PRESS, 1979, 222 p.
91. Wu L. M, Adams R. J. and Wilson M. R., ACER conquest: Generalised item response modelling software manual. Melbourne, 1998.
92. Xcalibre 4 - Software for IRT analysis and calibration,
<http://www.assess.com/xcart/product.php?productid=415&cat=22&page=1>