



МЕТОДОЛОГИЯ НА ПЕДАГОГИКАТА

METHODOLOGY OF PEDAGOGY

ОБЩИ ХАРАКТЕРИСТИКИ НА ДИДАКТИЧЕСКИЯ ЕКСПЕРИМЕНТ
ЗА РАЗВИТИЕ НА ОПРЕДЕЛЕНА ОБРАЗОВАТЕЛНА
КОМПЕТЕНТНОСТ – СТАТИСТИЧЕСКИ ПОДХОД

Венка Кутева-Цветкова¹, Николай Колев²

GENERAL CHARACTERISTICS OF THE DIDACTIC EXPERIMENT
FOR DEVELOPING CERTAIN
EDUCATIONAL COMPETENCE – STATISTICAL APPROACH

Venka Kuteva-Tsvetkova, Nikolay Kolev

Abstract: *This article discusses consistent systemic aspects of the conduct of research aimed at a given teaching methodology through a didactic experiment, dedicated to certain educational competence development. Key details on the use of traditional didactic tests as a measurement tool are considered. The content is tied to a statistical approach.*

Keywords: *educational experiment; educational competence; educational measurement; reliability and validity; Item Response Theory.*

ВЪВЕДЕНИЕ

В България има добри традиции в областта на методологията и методите на педагогическите изследвания. Могат да се посочат редица автори със значим принос (без претенции за изчерпателност) като Бижков (1983, 1992, 1999), Бижков и Краевски (2007), Иванов (1990, 1993, 2006), Кацаров (1947а, 1947б) (основател на експерименталната педагогика в България), Пиръов и Цанев (1973), Радев и Александрова (2003), Стоименова (2000), Стойчев (2003), Стоянова (1996), Чавдарова-Костова, Делибалтова и Господинов (2012), Господинов (2016).

¹ **Венка Кутева-Цветкова** – професор, доктор на науките, катедра „Педагогика“, Педагогически факултет, ВТУ „Св. св. Кирил и Методий“, България, e-mail: v.kuteva@ts.uni-vt.bg

² **Николай Колев** – главен асистент, доктор, катедра „Математически анализ и приложения“, факултет „Математика и информатика“, ВТУ „Св. св. Кирил и Методий“, България, e-mail: n.g.kolev@ts.uni-vt.bg

Всички ползвани в текста базови определения имат дълга история, в която е трудно да се посочи еднозначно първоизточник. За някои неща сме се придържали като ориентир към Бижков (2006), Иванов (2006), Господинов (2016) и Цветков (2015). Целта е да се фиксира понятиен апарат в съвременна терминология.

Дидактическите експерименти имат за обичайна цел методическо усъвършенстване на учебния процес, отнесено към дадени когнитивни компетентности на определена *целева група*. Целевата група статистически се обозначава като *генерална съвкупност*. В англоезичната литература се е наложил терминът *популация*. Целевите групи предполагаме нормални статистически популации, което дава право да се ползват добре разработени статистически техники за сравнения и корелации.

Статистическите понятия имат условно две страни: теоретична – свързана с математическа структура, и практическа – свързана с физически представи за целевата група като реален сбор от *статистически единици*. Физическата популация при дидактически експеримент обикновено представлява достатъчно *еднородна* група ученици, съставена според специфична съвкупност от изисквания. Еднородността се оказва практически налице в резултат на системен административен подбор, основан на постиженията и личностния профил на ученика. Тази важна особеност представлява своеобразно предимство на дидактическия експеримент.

МЕТОДОЛОГИЯ

Научният подход за валидизиране на ефективността от даден дидактически експеримент изисква практическа проверка на подходяща хипотеза от предметната област. Тази хипотеза обикновено се формулира в следния вид: „Приложеният дидактически експеримент подобрява съществено съответните компетентности“. За нея ще се търси естествено емпирично потвърждение от статистически характер. Вътрешната логика на статистиката изисква да се формулира *нулева хипотеза* H_0 (хипотеза за нулев резултат): „Наблюдаваният емпиричен резултат има случаен характер“ – продукт на онази неизбежна случайна изменчивост, която съпътства всяко статистическо измерване. Тази хипотеза може да се *отхвърли* или *приеме* въз основа на статистически аргументи. Отхвърлянето на H_0 представлява акт на приемане на хипотезата от предметната област и съответните доказателства за отхвърляне на H_0 представляват доказателства в полза на ефективността на дидактическия експеримент. Приемането на нулевата хипотеза като правило е решение с по-ниска достоверност, затова правилният заключителен изказ тук е: „Данните от опита не дават основание за отхвърляне на нулевата хипотеза“. Отхвърлянето на H_0 се препоръчва да бъде при обща алтернатива.

Теорията предлага технически лесен критерий за отхвърляне на H_0 . Всяка нулева хипотеза притежава *проверяваща статистика*, въз основа на която се пресмята *достигнатото ниво на значимост* p -value. Стойността на p -value е между 0 и 1 и представлява величина, която със своето намаляване увеличава нашата увереност за отхвърляне на H_0 . Праговата стойност за социалните и поведенческите науки е приета да бъде 0.05 (5%). По този начин критерият за отхвърляне на H_0 приема вида: p -value < 0.05. Основната интерпретация на p -value е като оценена от опита *вероятност за грешка от първи род*. Такава грешка би настъпила при отхвърляне на фактически валидна H_0 .

Когато p -value е достатъчно малко, се казва, че наблюдаваният резултат е статистически значим (статистически достоверен). Има основания при p -value < 0.001 да се говори за *силна значимост*, при p -value между 0.001 и 0.01 да се говори само за значимост, а при p -value между 0.01 и 0.05 да се говори за *зона на неопределеност*. Стойността на p -value обаче се явява *първично понятие* и затова е добре да бъде оставено само да говори за себе си. В текста обикновено се дава стойността на проверяващата статистика заедно със стойността на p -value. Пресмятането на p -value при известна стойност (и известно разпределение) на проверяващата статистика се извършва бързо и точно посредством компютър.

Измерване. Всяка дидактическа компетентност допуска проекция в определена съвкупност от *критерии* и *показатели*, проявяващи се чрез подлежащи на измерване *индикатори*. Отделното измерване асоциира числова или знакова характеристика към дадена статистическа единица,

оформено в статистическа *променлива*. Данните представляват колекция от измервания, които се подреждат във вид на таблица. В първия ред са имената на променливите, а на всеки следващ ред стои резултат от наблюдение над някоя единица. В първия стълб има идентификатори (служебни имена) на единиците от *извадката*, а във всеки следващ стълб стоят резултатите от някоя променлива. Извадката представлява съвкупността от наблюдаваните статистически единици.

Дидактическият експеримент се нуждае от инструмент за измерване на постиженията. Такъв инструмент обикновено се явява класическата скала (тест) от предварително фиксиран брой тестови единици, всяка от които произвежда градиран резултат. Тук за измерващо количество може да се вземе суровият бал – сборът от всички балове на отделните тестови единици. При определени условия суровият бал наистина представлява добър инструмент за измерване с един (преодолим) структурен недостатък, че се явява по природа дискретна променлива, което поражда неизбежна дискусия доколко може да бъде обработван чрез методи за нормални променливи. Отговорът трябва да се търси не в предписанията, а във формата на емпирично разпределение. Ако формата има ясно изразена централна зона и симетрични опашки, то ползването на статистически техники за нормални разпределения не буди сериозни възражения.

Теория на тестовете. Статистическите методи в педагогическите измервания също имат добри традиции в България. Нещата са поставени върху вярна основа още с работата на Димитър Кацаров (1947а, 1947b) и Генчо Пиръов (1985). Отбелязваме последователните усилия на Георги Бижков (1992, 1996, 1999) относно прилагането на класическата СТТ (Classical Test Theory) и съвременната теория на тестовете IRT (Item Response Theory). Тук особено място заема „Теория и методика на дидактическите тестове“ (1996), където има раздели от теорията на бинарните IRT модели, обвързани с изследванията на Стоянова (1996, 2007). Класическата теория на тестовете СТТ намира във висока степен завършен вид в монографията на Lord и Novick (1968), съдържаща глава от Birnbaum с подробно описание на някои двоични модели от съвременната теория. В определен смисъл СТТ може да се разглежда като структурирана около концепциите за надеждност и валидност на тестовете (e.g., Golafshani 2003; Livingston 2018). Добър поглед върху еволюцията на IRT дават още монографиите на Baker и Kim (2004), Fox (2010), Hambleton и Swaminathan (1985), Linden и Hambleton (1996), Lord (1980), Rasch (1992).

Използването на дадена класическа скала предявява редица изисквания към нея, като централното място заемат *надеждността* и *валидността*. Валидността представлява атрибут, който показва доколко скалата е обвързана с онова, което е предназначена да измерва. Учебната среда е нормативно стриктно подредена, което улеснява много съществено постигането на валидност. Практически е достатъчно добрите намерения на създателя на скалата да бъдат съчетани с експертен преглед на тестовите единици.

Надеждността представлява атрибут, който показва до каква степен ансамбълът от тестови единици е групиран по същество към някакъв едномерен конструкт (в случая дидактическа компетентност). Отсъствието на надеждност представлява фатален недостатък на скалата. Тогава оценките от скалата имат фактически случаен характер. При отсъствие на надеждност, скалата не измерва нищо.

Надеждността на дадена скала може да се оценява статистически посредством пресмятане на различни коефициенти (виж например Бижков 1996), най-известният от които се явява коефициентът алфа на Кронбах. Коефициентът алфа се мени на практика между нула и едно, като благоприятни за надеждността са стойности, близки до единица. Много висока надеждност в комбинация с малко на брой тестови единици може да доведе до понижаване на валидността, когато това е свързано с висока средна корелация между тестовите единици.

Анализ на тестовите единици. Всяка тестова единица има два основни атрибута – *трудност* и *разделителна сила*. Много трудните или много лесните единици е добре да се избягват, защото носят малко информация. Трудността се отчита според успеваемостта. Разделителната сила показва доколко тестовата единица различава силните и слабите резултати и се измерва чрез корелацията между общия бал на скалата и бала на тестовата единица. Тестовите единици трябва бъдат с приблизително еднаква положителна разделителна сила.

IRT може да се разглежда в определен смисъл като конкретизация на СТТ. Използването на IRT оценка за способностите елиминира проблема за естеството на скалата от суровия бал. В българската литература по IRT, (e.g., Горчев 2021) впечатляват системните усилия на Цветков (2015a), Христов (2011) и Ангелова-Славова (2014, 2020), насочени към задачата за параметричните IRT модели. В тази област са работите на Hristov и Tsvetkov (2003, 2004, 2005) и Hristov (2006, 2007), придружени с описания на съответните алгоритми. Различни IRT схеми са описани и в Цветков и Ангелова-Славова (2010, 2015), Цветков, Христов и Ангелова-Славова (2010, 2011, 2012, 2015), Tsvetkov, Hristov и Angelova-Slavova (2016, 2017) и Ангелова-Славова, Христов и Цветков (2014). Особено място заема монографията на Стоименова „Измерителни качества на тестовите“ (Стойменова 2000), където проблематиката е осветена детайлно в много аспекти.

Експериментална и контролна група, входно и изходно измерване. Даден дидактически експеримент се валидизира като правило върху някаква *експериментална група* обучаеми, които се подлагат на обучение по дидактическата методика на експеримента. За да има база за сравнение, се формира и *контролна група*, при която обучението се провежда по приетия нормативен начин. Тези две групи трябва да имат достатъчен обем и да бъдат съществено еднородни в началото при *входното измерване*. Измерването се провежда посредством изпълнение на дидактически тест, който обхваща достатъчно пълно и хармонично целевата компетентност. Тестът трябва да покрива *критериите, показателите и индикаторите* на компетентността, което на практика го валидизира като допустим инструмент. След завършване на дидактическия експеримент се провежда *изходно измерване*, от което се очаква да регистрира статистически значима разлика в полза на експерименталната група. При изходното измерване е необходимо да се ползва *паралелна форма* (структурно и смислово еднаква) на входния вариант на теста.

При входното измерване се прави проверка за надеждност, която се повтаря накрая върху всичките резултати.

В типичния случай се наблюдава и статистически достоверен прогрес при експерименталната група, но в много случаи между входното и изходното измерване се наблюдава прогрес и при двете групи, като при експерименталната този прогрес е по-силно изразен. В качеството на убедителен статистически метод за аргументиране на последното може да се ползва *значимостта на ефекта на взаимодействие* между двата фактора на дисперсионен анализ за повтарящи се измервания. Единият фактор е принадлежността към някоя от групите (контролна или експериментална), а другият фактор е поредност на измерването (входно или изходно).

Статистическият подход изисква минимум две измервания, но по избор на експериментатора и по възможност могат да бъдат проведени и междинни или отложени контролни измервания.

Обем на извадките. Педагогическите експерименти не изискват големи обеми на извадките поради факта, че хомогенността на целевата популация се намира под перманентен административен контрол. Например контролната група може да включва изцяло класове „А“ и „Б“ от дадено представително училище, а експерименталната група може да включва класове „Г“ и „Д“. Обеми между 30 и 50 души сумарно за всяка от групите (експериментална и контролна) се приемат за достатъчни. Обеми между 20 и 30 също са приемливи. На практика голяма част от дисертационните изследвания в професионално направление 1.3. „Педагогика на обучението по...“ се придържат към посочените числа.

Статистически инструментариум. Тук например може да се използва като достатъчно изчерпателен справочник монографията на Клаус и Ебнер (1971; превод от немски), както и монографиите на Глас и Стенли (1978), Калинов (2013), Имамов (2001), Манов (1995), Хог и Крейг (1982; превод от английски) и учебника на Гнеденко и Гешев (1984). Сравнителният анализ между групите при входното и изходното измерване може да бъде проведен чрез теста на Стюдънт за две независими популации и потвърден от неговия непараметричен аналог – тест на Ман-Уитни. Сравнителният анализ между поредността на измерване може да бъде проведен чрез теста на Стюдънт за две зависими популации и потвърден от неговия непараметричен аналог – тест на Уилкоксон. Обобщаващият подход стъпва върху ефекта на взаимодействие от дисперсионен ана-

лиз за повтарящи се измервания. При необходимост се ползват и обобщения на изброени по-горе методи.

РЕЗУЛТАТИ

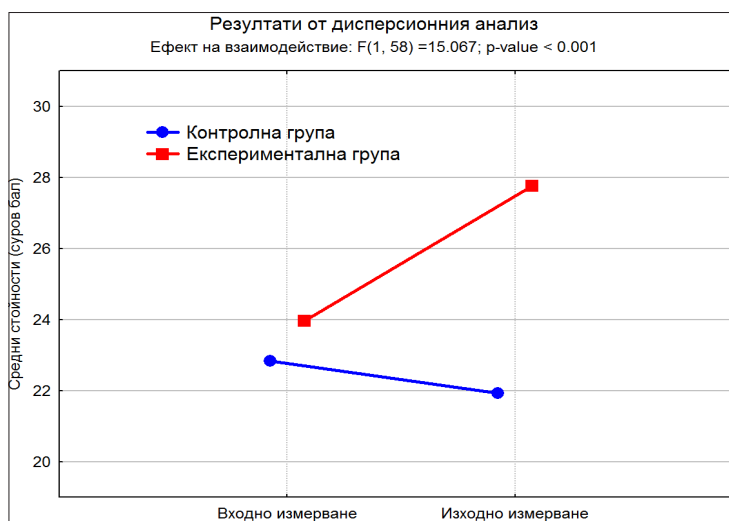
В качеството на илюстративни резултати от изложената по-горе методология предлагаме два примера на сравнително цялостно изграден статистически анализ от дидактически експеримент. Изложението е стилизирано кратко, а данните са мотивирани от реални експерименти.

Пример 1. Контролната и експерименталната групи съдържат по 30 души. Главната статистическа променлива е обвързана със специфична компетентност, измервана от скала с 16 тестови единици. По всяка тестова единица резултатите са отчитани в три подредени числови маркери, като по-високият маркер сочи към по-високо постижение.

Надеждност. Тестовите единици се съгласуват вътрешно добре в един латентен конструктор. IRT моделът Generalized Partial Credit Model (e.g., Linden и Hambleton 1997; Цветков 2015a; Hristov & Tsvetkov 2004) дава реалистични стойности за параметрите, което е белег за добра надеждност. Факторният анализ обособява водещ фактор за 43% от общата вариация, а коефициентът алфа на Кронбах има стойност 0.909. Посочените дотук индикатори гарантират в своята съвкупност много добра надеждност на скалата.

Сравнения. При сравнение между контролната и експерименталната група на входното измерване (тест на Стюдънт) не се наблюдават значими различия нито в средните ($t(58) = 0.681$; $p\text{-value} = 0.498$) нито в дисперсиите ($F(29,29) = 1.216$; $p\text{-value} = 0.602$). Резултатът се потвърждава и от теста на Ман-Уитни ($Z = 0.761$; $p\text{-value} = 0.446$). При сравнение между контролната и експерименталната група на изходното измерване (тест на Стюдънт) се наблюдава силно значимо различие ($t(58) = 4.915$; $p\text{-value} < 0.001$), без различие в дисперсиите ($F(29,29) = 1.880$; $p\text{-value} = 0.094$). Резултатът се потвърждава и от теста на Ман-Уитни ($Z = 4.103$; $p\text{-value} < 0.001$). При експерименталната група между входното и изходното измерване се наблюдава силно значим ръст на постижения ($t(29) = 4.367$; $p\text{-value} < 0.001$), докато при контролната промяната не е статистически значима ($t(29) = 1.069$; $p\text{-value} = 0.294$). Резултатът се потвърждава и от теста на Уилкоксон – съответно ($Z = 3.817$; $p\text{-value} < 0.001$) и ($Z = 0.943$; $p\text{-value} = 0.346$).

Дисперсионен анализ. Ще приложим дисперсионен анализ с повтарящи се измервания, който съдържа два фактора. Първият ще бъде принадлежност към контролна или експериментална група, а вторият ще бъде поредност на измерването – входно или изходно.



Фигура 1. Дисперсионен анализ – диаграма на средните

Проверяват се нулеви хипотези за главните ефекти на всеки фактор и трета, отнесена към нулево взаимодействие на тези два фактора. Именно отхвърлянето на тази трета хипотеза може

да се разглежда като непосредствено доказателство за наличие на статистически достоверен резултат от експерименталното дидактическо въздействие.

В резултат на анализа се наблюдава значим ефект на фактора група ($F(1,58) = 7.050$; $p\text{-value} = 0.010$), значим ефект на фактора измерване ($F(1,58) = 5.736$; $p\text{-value} = 0.020$) и силно значим ефект на взаимодействие ($F(1,58) = 15.067$; $p\text{-value} < 0.001$). Резултатът е илюстриран на фигура 1. От решаващо значение е характерното разположение на средните стойности. Фигура 1. показва, че значим прогрес между входа и изхода се наблюдава само при експерименталната група. Следващата таблица, поясняваща подробно ситуацията, съдържа $p\text{-values}$ от Post-Hoc анализа.

Таблица 1. Резултати от Post-Hoc анализа ($p\text{-values}$)

			{1}	{2}	{3}	{4}
{1}	контролна	вход		0.298	0.435	0.001
{2}	контролна	изход	0.298		0.163	0.000
{3}	експериментална	вход	0.435	0.163		0.000
{4}	експериментална	изход	0.001	0.000	0.000	

При експерименталната група се наблюдава силно значимо различие между входа и изхода ($p\text{-value} < 0.001$). При контролната група не се наблюдава значимо различие ($p\text{-value} = 0.163$). Друго силно значимо различие е между двете групи на изходното измерване ($p\text{-value} < 0.001$).

Извод. Изброените фрагменти позволяват да направим заключителен извод, че при входното измерване двете групи са съществено еднородни, предложената работна скала притежава много добра статистическа надеждност, а използваната експериментална педагогическа методика дава значим резултат, който се изразява в статистически достоверното повишаване на компетентността на експерименталната група в сравнение с контролната.

Направеният извод обединява по естествен начин няколко статистически аргумента за проверка на основната теза.

Пример 2. Контролната и експерименталната група са съставени от по 40 души. Главната статистическа променлива отново е обвързана със специфична компетентност, измервана от скала с 10 двоични тестови единици.

Надеждност. Тестовите единици се съгласуват вътрешно добре в един латентен конструкт. Моделът с два параметъра (2-PL) дава реалистични стойности за тях. Като мярка за постижение по скалата се използват MAP резултатите от IRT модела. Факторният анализ показва водещ фактор за 32% от общата вариация, а коефициентът алфа на Кронбах има стойност 0.754. Посочените индикатори осигуряват много добра надеждност на скалата.

Сравнения. При сравнение между контролната и експерименталната група на входното измерване (тест на Стюдънт) не се наблюдават значими различия нито в средните ($t(78) = 0.037$; $p\text{-value} = 0.971$) нито в дисперсиите ($F(39,39) = 1.077$; $p\text{-value} = 0.817$). Резултатът се потвърждава и от теста на Ман-Уитни ($Z = 0.101$; $p\text{-value} = 0.920$). При сравнение между контролната и експерименталната групи на изходното измерване (тест на Стюдънт) се наблюдава значимо различие ($t(78) = 2.692$; $p\text{-value} = 0.009$), без различие в дисперсиите ($F(39,39) = 1.508$; $p\text{-value} = 0.204$). Резултатът се потвърждава и от теста на Ман-Уитни ($Z = 2.656$; $p\text{-value} = 0.008$). При експерименталната група между входното и изходното измерване се наблюдава силно значим ръст на постижения ($t(39) = 5.689$; $p\text{-value} < 0.001$). Промяната за контролната група също е статистически значима ($t(39) = 3.359$; $p\text{-value} = 0.002$). Резултатът се потвърждава и от теста на Уилкоксон – съответно ($Z = 4.256$; $p\text{-value} < 0.001$) и ($Z = 3.053$; $p\text{-value} = 0.002$).

Дисперсионен анализ. Отново ще приложим дисперсионен анализ с повтарящи се измервания при два фактора – принадлежност към контролна или експериментална група и поредност на измерването.



Фигура 2. Дисперсионен анализ – диаграма на средните

Отново вниманието ще бъде насочено преди всичко към хипотезата за нулево взаимодействие. В резултат на анализа не се наблюдава значим ефект на фактора група ($F(1,78) = 2.577$; $p\text{-value} = 0.112$) (видна е само очаквана тенденция). Наблюдава се много силно значим ефект на фактора измерване ($F(1,78) = 42.766$; $p\text{-value} < 0.001$) и значим ефект на взаимодействие ($F(1,78) = 5.454$; $p\text{-value} = 0.022$). Резултатът е илюстриран на фигура 2, където виждаме значим прогрес между входа и за двете групи, но за експерименталната група ефектът е по-силно изразен. Следва поясняващата таблица на $p\text{-values}$ от Post-Hoc анализа.

Таблица 2. Резултати от Post-Hoc анализа ($p\text{-values}$)

			{1}	{2}	{3}	{4}
{1}	Контролна	вход		0.004	0.970	0.000
{2}	Контролна	изход	0.004		0.023	0.009
{3}	експериментална	вход	0.970	0.023		0.000
{4}	експериментална	изход	0.000	0.009	0.000	

При експерименталната група се наблюдава силно значимо различие между входа и изхода ($p\text{-value} < 0.001$). Друго важно значимо различие е между двете групи на изходното измерване ($p\text{-value} = 0.009$). При входното измерване не се наблюдава значимо различие ($p\text{-value} = 0.970$).

Извод. И в този случай изброените фрагменти позволяват да направим заключителен извод, че при входното измерване двете групи са съществено еднородни, предложената работна скала притежава много добра статистическа надеждност, а използваната експериментална педагогическа методика дава значим резултат, който се изразява в статистически достоверно повишаване на компетентността на експерименталната група в сравнение с контролната.

Направеният извод обаче може да се превърне в предмет на дискусия, понеже някои статистически аргументи се намират в зоната на неопределеността.

ДИСКУСИЯ

В настоящата статия е дадена своеобразна минимална, но достатъчно изчерпателна схема на ключови акценти за научно изследване, насочено към ефективност от даден дидактически експеримент. Сама по себе си такава теза се явява дискуссионна, доколкото различни автори могат да предявяват изисквания за повече структурни детайли на експеримента. Схемата винаги може

да бъде обогатена с повече дейности относно формирането на експерименталната и контролната група или времевата програма за контрол на постиженията, например с експерименталния дизайн на Соломон (e.g., Господинов 2016; Петкова 2016). Заявеният тук подход за рутинно формиране на контролна и експериментална група също бил могъл да се превърне в предмет на градивна дискусия, касаеща цялостния характер на експерименталните цели. Дискусионен е и въпросът доколко статистическата достоверност сама по себе си е достатъчна, за да гарантира обществената значимост на получените резултати. Изложената по-горе позиция на авторите по формираните дискуссионни точки е принципно консервативна.

ЗАКЛЮЧЕНИЕ

Наличието на органично спрегната с дидактически експеримент административна нормативна рамка върху целевата популация, създава в нея достатъчно висока степен на хомогенност и естествена рандомизация, което представлява голямо улеснение за коректното провеждане на експеримента.

Високото ниво на развитие на теорията на тестовете представлява хубав пример за дълбоко съдържателна и същевременно математически достъпна количествена теория в областта на социалните и поведенческите науки (e.g., Христов 2011; Ангелова-Славова 2014; Цветков 2015а).

Изработването на скала с малък брой тестови единици и висок коефициент на надеждност алфа при ниска средна корелация между тестовите единици и при умерена трудност и сходна разделителна сила представлява творческа задача в предметната област на скалата (e.g., Lord 1980; Цветков 2015b), при решаването на която статистиката има подчинена роля.

Грешката от първи род подлежи на директен количествен контрол чрез стойността на p -value, докато контролът на грешката от втори род (когато приемаме фактически невалидна H_0) е трудно обозрим и затова обикновено се избягва, както е направено и в настоящата статия. Някои автори (e.g., Cohen 1988) предлагат в качеството на заместващ аргумент да се разгледа *effect size*. Според нас обаче най-голям интерес представляват статистически достоверните и същевременно умерено силни (даже слаби) ефекти, понеже те са достъпни за оползотворяване или промяна в рамките на административния хоризонт на дадена образователна инициатива.

Технически трудната част по статистическите пресмятания и оформянето на резултатите се извършва с помощта на специализирани софтуерни платформи. Този факт, от една страна, улеснява постигането на определени цели, но от друга – завишава съществено изискванията към традиционния потребител за по-добро познаване на научния апарат на статистиката.

ЛИТЕРАТУРА

Ангелова-Славова, Р. (2014). *Приложни аспекти на вероятностни модели от теорията на тестовете: дисертация за присъждане на образователна и научна степен „доктор“*, В. Търново. // **Angelova-Slavova, R. (2014). Prilozhni aspekti na veroyatnostni modeli ot teoriyata na testovete: disertatsia za prisazhdane na obrazovatelna i nauchna stepen „doktor“, V. Tarnovo.**

Ангелова-Славова, Р., Л. Христов, Д. Цветков (2014). Информационен критерий на Акаике за избор на модел. „Дни на науката 2013“. // **Angelova-Slavova, R., L. Hristov, D. Tsvetkov (2014). Informationsionen kriteriy na Akaike za izbor na model. „Dni na naukata 2013“.**

Ангелова-Славова, Р. (2020). *Сходимост и приложения на алгоритъма на Метрополис–Хестингс*. Пловдив: Астартта. // **Angelova-Slavova, R. (2020). Shodimost i prilozhenia na algoritama na Metropolis–Hestings. Plovdiv: Astarta.**

Бижков, Г. (1983). *Методология и методи на педагогическите изследвания*. София. // **Bizhkov, G. (1983). Metodologia i metodi na pedagogicheskite izsledvania. Sofia.**

Бижков, Г. (1992). *Теория и методика на дидактическите тестове*. София: Просвета. // **Bizhkov, G. (1992). Teoria i metodika na didakticheskite testove. Sofia: Prosveta.**

Бижков, Г. (1996). *Теория и методика на дидактическите тестове*. София: Просвета. // **Bizhkov, G. (1996). Teoria i metodika na didakticheskite testove. Sofia: Prosveta.**

Бижков, Г. (1999). *Педагогическа диагностика*. София: УИ „Св. Кл. Охридски“. // **Bizhkov, G. (1999).** *Pedagogicheska diagnostika*. Sofia: UI „Sv. Kl. Ohridski“.

Бижков, Г., В. Краевски (2007). *Методология и методи на педагогическите изследвания*. София: УИ „Св. Кл. Охридски“. // **Bizhkov, G., V. Kraevski (2007).** *Metodologia i metodi na pedagogicheskite izsledvania*. Sofia: UI „Sv. Kl. Ohridski“.

Гласс, Дж., Дж. Стэнли (1976). *Статистические методы в педагогике и психологии*. Москва, Прогресс. // **Glass, G., J. Stanley (1970).** *Statistical Methods in Education and Psychology*. Englewood Cliffs, N. J.: Prentice-Hall.

Гнеденко, Б., А. Гешев (1984). *Теория на вероятностите и математическа статистика*. Пловдив: ПУ „Паисий Хилендарски“. // **Gnedenko, B., A. Geshev (1984).** *Teoria na veroyatnostite i matematcheska statistika*. Plovdiv: PU „Paisiy Hilendarski“.

Горчев, Н. (2021). Съвместно използване на класическата и съвременната теории на тестовете. *Годишна университетска научна конференция на НВУ „Васил Левски“*. // **Gorchev, N. (2021).** *Savmestno izpolzване na klasicheskata i savremennata teorii na testovete. Godishna universitetska nauchna konferentsia na NVU „Vasil Levski“*.

Господинов, Б. (2016). *Научното педагогическо изследване: методологически, технологични и методически аспекти*. София: Университетско издателство „Св. Климент Охридски“. // **Gospodinov, B. (2016).** *Nauchnoto pedagogicheskoto izsledvane: metodologicheski, tehnologichni i metodicheski aspekti*. Sofia: Universitetsko izdatelstvo „Sv. Kliment Ohridski“.

Иванов, И. (1990). *Методология и методи на педагогическите изследвания*. Шумен: ВПИ „Константин Преславски“. // **Ivanov, I. (1990).** *Metodologia i metodi na pedagogicheskite izsledvania*. Shumen: VPI „Konstantin Preslavski“.

Иванов, И. (1993). *Педагогическото изследване: методология, методи, процедура*. Шумен: Издателство „Глаукс“. // **Ivanov, I. (1993).** *Pedagogicheskoto izsledvane: metodologia, metodi, protsedura*. Shumen: Izdatelstvo „Glauks“.

Иванов, И. (2006). *Педагогическа диагностика*. Шумен: УИ „Епископ Константин Преславски“. // **Ivanov, I. (2006).** *Pedagogicheska diagnostika*. Shumen: UI „Episkop Konstantin Preslavski“.

Имамов, А. (2001). *Статистика със SPSS*. София: Тракия-М. // **Imamov, A. (2001).** *Statistika sas SPSS*. Sofia: Trakia-M.

Калинов, К. (2013). *Статистически методи в поведенческите и социалните науки*. София: НБУ. // **Kalinov, K. (2013).** *Statisticheski metodi v povedencheskite i sotsialnite nauki*. Sofia: NBU.

Кацаров, Д. (1947а). *Теория на образованието: Обща Педагогия*. София: Родопа. // **Katsarov, D. (1947a).** *Teoria na obrazovaniето: Obshta Pedagogia*. Sofia: Rodopa.

Кацаров, Д. (1947б). *Опитното изследване в педагогията; Опитна педагогия*. София. // **Katsarov, D. (1947b).** *Opitnoto izsledvane v pedagogiyata; Opitna pedagogia*. Sofia.

Клаус, Г., Х. Ебнер (1971). *Основи на статистиката за психолози, педагози и социолози*. София: Наука и изкуство. // **Klaus, G., Ebner, H. (1971).** *Osnovi na statistikata za psiholozi, pedagozi i sotsiolozi*. Sofia: Nauka i izkustvo.

Манов, А. (1995). *Статистика с компютър*. София: УИ „Стопанство“. // **Manov A. (1995).** *Statistika s kompyutar*. Sofia: UI “Stoipanstvo”.

Петкова, М. (2016). *Педагогически технологии за интегриране на GeoGebra приложения в обучението по геометрия: дисертационен труд за присъждане на образователна и научна степен „доктор“*. // **Petkova, M. (2016).** *Pedagogicheski tehnologii za integrirane na GeoGebra prilozhenia v obuchenieto po geometria: disertatsionen trud za prisazhdane na obrazovatelna i nauchna stepen „doktor“*.

Пиръов, Г. (1985). *Психология и психодиагностика на интелигентността: [Изследване]*. София: Наука и изкуство. // **Piryov, G. (1985).** *Psihologia i psihodiagnostika na inteligentnostta: [Izsledvane]*. Sofia: Nauka i izkustvo.

Пиръов, Г., Ц. Цанев (1973). *Експериментална психология*. София: „Наука и изкуство“. // **Piryov, G., Tsanev, Ts. (1973).** *Ekspierimentalna psihologia*. Sofia: Nauka i izkustvo.

Радев, П., Ал. Александрова (2003). *Дидактическа тестология*. Пловдив: „Хермес“. // **Radev, P., Al. Aleksandrova (2003).** *Didakticheska testologia*. Plovdiv: „Hermes“.

Стоименова, И. (2000). *Измерителни качества на тестовете*. София: НБУ. // **Stoimenova, I. (2000).** *Izmeritelni kachestva na testovete*. Sofia: NBU.

Стойчев, Д. (2003). *Количествени методи в педагогиката, психологията и социологията.* Пловдив: Макрос. // **Stoychev, D. (2003).** *Kolichestveni metodi v pedagogikata, psihologiyata i sotsiologiyata.* Plovdiv: Makros.

Стоянова, Ф. (1996). *Тестология за учители.* София, Атика. // **Stoyanova, F. (1996).** *Testologia za uchiteli.* Sofia: Atika.

Съйкова, Ив., С. Тодорова (1994). *Статистическото изследване: Постановка, методи, оценка на резултатите.* София: Люрен. // **Saykova, Iv., S. Todorova (1994).** *Statisticheskoto izsledvane: Postanovka, metodi, otsenka na rezultatite.* Sofia: Lyuren.

Хог, Р., А. Крейг (1982). *Увод в математическата статистика.* София: Техника. // **Hogg, R. V., Craig, A. T. (1978).** *Introduction to mathematical statistics.* New York: Collier MacMillan.

Христов, Л. (2011). *Оценка на параметрите за някои вероятностни модели от теорията на тестовете: дисертация за присъждане на образователна и научна степен „доктор“.* В. Търново. // **Hristov, L. (2011).** *Otsenka na parametrite za nyakoi veroyatnostni modeli ot teoriyata na testovete: disertatsia za prisazhdane na obrazovatelna i nauchna stepen „doktor“.* V. Tarnovo.

Цветков, Д., Р. Ангелова-Славова (2015). Оценка на IRT модели с индивидуални нива на отгатване чрез платформата BUGS. „Дни на науката 2014“, Майски четения на СУБ. // **Tsvetkov D., R. Angelova-Slavova (2015).** *Otsenka na IRT modeli s individualni niva na otgatvane chrez platformata BUGS. „Dni na naukata 2014“, Mayski chetenia na SUB.*

Цветков, Д. (2015a). *Вероятностни модели в теорията на тестовете.* Велико Търново: УИ „Св. св. Кирил и Методий“. // **Tsvetkov, D. (2015a).** *Veroyatnostni modeli v teoriyata na testovete.* Veliko Tarnovo: UI „Sv. sv. Kiril i Metodiy“.

Цветков, Д., Р. Ангелова-Славова (2010). Надеждност на дидактически тестове. *Годишник на НВУ „В. Левски“ 2010 година, Издателски комплекс на НВУ „Васил Левски“.* // **Tsvetkov, D., R. Angelova-Slavova (2010).** *Nadezhdnost na didakticheski testove. Godishnik na NVU „V. Levski“ 2010 godina, Izdatelski kompleks na NVU „Vasil Levski“.*

Цветков, Д., Л. Христов, Р. Ангелова-Славова (2012). Оценка на максимално правдоподобие в някои вероятностни модели от теория на тестовете. „Дни на науката 2012“, Майски четения на СУБ. // **Tsvetkov, D., L. Hristov, R. Angelova-Slavova (2012).** *Otsenki na maksimalno pravdopodobie v nyakoi veroyatnostni modeli ot teoria na testovete. „Dni na naukata 2012“, Mayski chetenia na SUB.*

Цветков, Д., Л. Христов, Р. Ангелова-Славова (2010). Оценка на параметрите за IRT модел с персонални нива на отгатване чрез Монте Карло методи за марковски вериги. *Сборник доклади от Научна конференция с международно участие MATTEX 2010.* // **Tsvetkov, D., L. Hristov, R. Angelova-Slavova (2010).** *Otsenka na parametrite za IRT model s personalni niva na otgatvane chrez Monte Karlo metodi za markovskii verigi. Sbornik dokladi ot Nauchna konferentsia s mezhdunarodno uchastie MATTEX 2010.*

Цветков, Д., Л. Христов, Р. Ангелова-Славова (2012). Критерии за подбор на тестови единици при компютъризирано адаптивно тестване. „Дни на науката 2012“, Майски четения на СУБ. // **Tsvetkov, D., L. Hristov, R. Angelova-Slavova (2012).** *Kriterii za podbor na testovi edinitsi pri kompyutarizirano adaptivno testvane. „Dni na naukata 2012“, Mayski chetenia na SUB.*

Цветков, Д., Л. Христов, Р. Ангелова-Славова (2011). Итеративна EAP оценка на параметрите за някои бинарни модели от теория на тестовете. *Сборник доклади от годишната университетска научна конференция с международно участие 2011.* // **Tsvetkov, D., L. Hristov, R. Angelova-Slavova (2011).** *Iterativna EAP otsenka na parametrite za nyakoi binarni modeli ot teoria na testovete. Sbornik dokladi ot godishnata universitetska nauchna konferentsia s mezhdunarodno uchastie 2011.*

Цветков, Д. (2015b). *Лекции по статистически методи в психологията.* Велико Търново: ИВИС. // **Tsvetkov, D. (2015b).** *Leksii po statisticheski metodi v psihologiyata.* Veliko Tarnovo: IVIS.

Чавдарова-Костова, С., В. Делибалтова, Б. Господинов (2012). *Педагогика.* София: УИ „Св. Климент Охридски“. // **Chavdarova-Kostova, S., V. Delibaltova, B. Gospodinov (2012).** *Pedagogika.* Sofia: UI „Sv. Kliment Ohridski“.

Baker, F., S.-H. Kim (2004). *Item response theory: parameter estimation techniques.* Marcel Dekker.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Lawrence Erlbaum Associates.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, 8(4), 597–606.

Fox, J.-P. (2010). *Bayesian item response modeling: theory and applications.* NY: Springer.

Hambleton, R. K., H. Swaminathan (1985). *Item response theory: principles and applications.* NY: Springer.

Hristov, L. (2006). Two-parameter logistic IRT model with latent classes. *International Conference on Computer Systems and Technologies – CompSysTech '2006*.

Hristov, L. (2007). EM parameter estimation for the rasch model. *36-та пролетна конференция на ЦМБ*.

Hristov, L., D. Tsvetkov (2003). Computer implementation of MMLE/EM algorithm for two-parametric logistic model. *International Conference on Computer Systems and Technologies – CompSysTech '2003*, Sofia.

Hristov, L., D. Tsvetkov (2004). A computer implementation of the generalized partial credit model. *International Conference on Computer Systems and Technologies – CompSysTech '2004*, Ruse.

Hristov, L., D. Tsvetkov (2005). IRT model with personal guessing levels. *International Conference on Computer Systems and Technologies – CompSysTech '2005*, Varna.

Linden, W. J., R. K. Hambleton (1997). *Handbook of modern item response theory*. NY: Springer.

Livingston, S. (2018). Test reliability-basic concepts. *ETS Research Memorandum Series. Educational Testing Service*. Princeton, New Jersey.

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Erlbaum.

Lord, F. M., M. R. Novick (1968). *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley.

Tsvetkov, D., L. Hristov, R. Angelova-Slavova (2016). One method to check the population homogeneity of a test. *Mathematics and Education in Mathematics 2016*.

Tsvetkov, D., L. Hristov, R. Angelova-Slavova (2017). Notes on the parameter estimation of some IRT models by means of the EM-algorithm. *Mathematics and Education in Mathematics 2017*.