



**НОВ БЪЛГАРСКИ УНИВЕРСИТЕТ**

София 1618, ул. Монтевидео 21

Департамент „Когнитивна наука и психология”

Любомир Георгиев Джалев

**СРАВНИТЕЛЕН АНАЛИЗ НА ПРИЛОЖИМОСТТА  
НА ДВЕ ПСИХОМЕТРИЧНИ ТЕОРИИ НА ТЕСТОВЕТЕ  
(ВЪРХУ ДАННИ ОТ ТЕСТА ЗА  
ОБЩООБРАЗОВАТЕЛНА ПОДГОТОВКА)**

**АВТОРЕФЕРАТ НА ДИСЕРТАЦИЯ**  
за присъждане на  
образователната и научна степен „доктор”

София, май 2013

Дисертационният труд съдържа общо 358 страници, разпределени в следните части: въведение, теоретична рамка на изследването, обща постановка, резултати, обща дискусия, цитирана литература и приложения. От общия обем на разработката цитираната литература е разположена в 12 страници, а приложенията – в 45 страници. В разработката са включени 34 фигури, от които 32 – в основния текст и 2 – в приложение, както и 73 таблици, от които 53 – в основния текст и 20 – в приложение. Цитираната литература обхваща 269 заглавия, от които 11 – на кирилица и 258 – на латиница.

Докторската теза е обсъдена и насочена за защита на заседание на департамент „Когнитивна наука и психология” в Нов български университет. Дисертантът е зачислен като докторант на самостоятелна подготовка към департамент „Когнитивна наука и психология” в Нов български университет по научна специалност 05.06.01. Обща психология за придобиване на образователна и научна степен „доктор” със заповед на Ректора на Нов български университет от 01.04.2011 г.

Защитата на докторската теза ще се състои на ..... 2013 г. от ..... ч. в заседателната зала на библиотеката на Нов български университет на открито заседание на .....

Материалите по защитата са на разположение на интересующите се в кабинет 401, корпус 1 в Нов български университет.



**НОВ БЪЛГАРСКИ УНИВЕРСИТЕТ**

София 1618, ул. Монтевидео 21

Департамент „Когнитивна наука и психология”

Любомир Георгиев Джалев

**СРАВНИТЕЛЕН АНАЛИЗ НА ПРИЛОЖИМОСТТА  
НА ДВЕ ПСИХОМЕТРИЧНИ ТЕОРИИ НА ТЕСТОВЕТЕ  
(ВЪРХУ ДАННИ ОТ ТЕСТА ЗА  
ОБЩООБРАЗОВАТЕЛНА ПОДГОТОВКА)**

**АВТОРЕФЕРАТ НА ДИСЕРТАЦИЯ**

за присъждане на  
образователната и научна степен „доктор”

Рецензенти:

проф. д-р Енчо Герганов

проф. д-р Кирил Банков

София, май 2013



## **1. Структура на дисертационния труд**

Дисертационният труд съдържа общо 358 страници, разпределени в следните части: въведение, теоретична рамка на изследването, обща постановка, резултати, обща дискусия, цитирана литература и приложения. Цитирани са общо 269 заглавия, от които 11 – на кирилица и 258 – на латиница. От общия обем на разработката цитираната литература е разположена в 12 страници, а приложенията – в 45 страници. В разработката са включени 34 фигури, от които 32 – в основния текст и 2 – в приложение, както и 73 таблици, от които 53 – в основния текст и 20 – в приложение.

## **2. Съдържание на дисертационния труд**

### **2.1. Въведение**

Образованието е един от ключовите сектори в публичната сфера на модерните общества. Като специфичен вид обществена услуга то изпълнява важни социални функции, като осъществява пренос на знания, умения, навици, нагласи и ценности от поколение към поколение. Фундаменталното значение на образованието е отразено във Всеобщата декларация за правата на човека в която то, наред с правото на живот, свобода и сигурност на личността; на свобода на мисълта, съвестта и религията, е провъзгласено за основно човешко право.

Нашето съвремие поставя пред образователната система все по настоятелно въпроса за *качеството* на образователния продукт. Образователната система, а и семейство-

то често са упрекувани, че не успяват да се справят добре с мисията си на социализиращи институции. Независимо от това образователната система е разработила редица механизми за своевременно, текущ вътрешен контрол на процесите, протичащи в нея, най ефективният сред които е *оценяването* на знания и умения. Оценкаването следва да се разглежда не просто като описващо постиженията, а като мощна движеща сила за промяна в образователната система, водеща до подобряване на качеството и до по-високи стандарти на обучение (Wolf, Bixby, Glenn & Gardner, 1991).

Освен прагматични, поставеният проблем има и специфични изследователски измерения. Натрупаният практически опит намира отражение в изграждането на психометрични теории и модели с различна степен на обобщеност. Голяма част от тях, по-конкретно ориентираните към обективното оценяване тестови теории и съставлящите ги модели, могат да бъдат класифицирани като „модели на данни”. Тяхна отличителна характеристика е, че съдържат определен набор от основни *допускания*, които могат да се възприемат като общо описание на данните, за които е предназначен съответния модел. От друга страна, при прилагането на даден модел неговите допускания следва да се разглеждат като *изисквания*, като необходими предпоставки за неговото използване. Поради това един от фундаменталните въпроси при моделите на данни е за връзката между съответния модел и емпиричните данни, който може да бъде формулиран и като въпрос за адекватността, за приложимостта на модела. Тази взаимна обусловеност поставя акцента в изследователската работа върху разработването на модели, подходящи за съответния тип данни. Това е акцентът и в емпиричната част на представената разработка, която е фокусирана върху анализ на приложимостта на две основни психомет-

рични теории – Класическата тестова теория (СТТ) и Теорията за отговор на тестов въпрос (IRT), върху резултатите от Теста по общообразователна подготовка (ТОП), който стои в основата на приемните процедури в Нов български университет.

## **2.2. Теоретична рамка на изследването**

Тази глава от дисертационния труд съдържа две части: психометрични тестови теории и преглед на изследванията по проблема за приложимостта на техните модели.

Конструирането на теста е комплексен и в някои свои фази – итеративен процес с изследователски и приложен характер, който се разгръща в няколко последователни етапа. Дейностите във всички етапи на конструиране и използване на теста се извършват в рамките на определена психометрична (тестова) теория. Напредъкът в развитието на психологическите измервания се изразява не само в интензивното развитие на теоретичните им основи, но и в екстензивното усъвършенстване на съществуващите и в разработване на нови психометрични теории и модели, чрез които измерването в поведенческите и социалните науки все повече се доближава до строгите изисквания на измерването в природните науки. Две от най-добре разработените и най-често прилагани тестови теории, които са обект на съпоставителния анализ, са Класическата тестова теория и Теорията за отговор на тестов въпрос.

СТТ е еманация на усилията на няколко поколения изследователи в областта на поведенческите и социалните науки за квантифициране на различни аспекти на индивидуалните различия. Нейните основи са поставени преди повече от 100 години от Чарлз Спийърмън в неговото знаменито изследване на интелигентността, представено в статията

„General intelligence”, objectively determined and measured” (Spearman, 1904).

Класическата теория борави с три основни конструкта: (1) наблюдаван тестов бал, който представлява „видимия” резултат от измерването, който изследователят регистрира; (2) действителен, „истински” тестов бал който се разглежда като относително устойчива величина, която не се променя при многократно измерване с един и същи тест или с различни (еквивалентни, паралелни) форми на теста и която стои във фокуса на тази теория; (3) грешка на измерването. В основното уравнение на СТТ наблюдаваният тестов бал се разглежда като сума от действителния тестов бал и грешката на измерването.

СТТ се гради на поредица от допускания („аксиоми”) относно компонентите в дясната част на основното уравнение. Голяма част от тези допускания следват от основното уравнение и поради това някои изследователи говорят не за допускания или аксиоми, а за вътрешноприсъщи особености на теорията. Това са три групи допускания относно действителния бал, грешката на измерване и връзката между тях. Формата на разпределението тестовия бал и размерността на латентните структури обикновено не се разглеждат като част от „канона” на Класическата теория. Съществуват обаче редица аргументи в полза на разглеждането на такива допускания.

В сърцевината на Класическата теория стои проблемът за надеждността, която се разглежда като степента, в която наблюдаваният резултат съдържа в себе си действителната стойност на измервания признак. Във връзка с тази основна характеристика на наблюдавания бал са разработени редица методи за нейната оценка. Въпреки че теорията е фокусирана върху оценяване на способностите на равнище тест,



в нея, освен тестови статистики, са включени и статистики, предназначени за оценка на характеристиките на отделните въпроси – трудност, дискриминативна сила и налучкване на коректния отговор.

Като най-съществено преимущество на Класическата теория може да се отбележи това, че данните, които се получават в психологическите или образователните измервания, обикновено удовлетворяват основните допускания в нейните модели. Поради това те обикновено са обозначавани като „меки“ модели, които могат да бъдат приложени в различни изследователски и приложни контексти. Като съществен недостатък на СТТ може да се посочи зависимостта на тестовите статистики, както и на индексите на въпросите, от извадката от и. л., чрез която са получени. Теорията е безразлична и към проблема за моделиране на връзката между (латентния) действителен бал и отговорите на и. л. на отделните въпроси. Нереалистично е и допускането за еднаква надеждност на измерването за всички и. л. без оглед на техния действителен бал. Като недостатък се разглежда и фокусирането към стандартната грешка на измерване, която също е константна за целия тест, като се игнорират другите възможни фактори, влияещи върху тестовите резултати.

Появата и развитието на Теорията за отговор на тестов въпрос се разглежда от мнозина изследователи като „тиха революция“ в оценяването (Embretson & Reise, 2000). В нейната основа стои концепцията за латентните черти, въведена от П. Лазарсфелд. Измерването им е индиректно, чрез наблюдение на поведението на индивиди, изпълняващи релевантни задачи (манифестирани променливи). При оценяване на когнитивното поведение латентните черти най-често се обозначават като „способности“ (abilities) (Анастаси и Урбина, 2001). IRT фокусира вниманието върху моделирането

на функционалната връзка между равнището на способностите и вероятността от коректен отговор на ниво тестов въпрос. Поради това концепцията за характеристичната крива на въпроса е основата, върху която се надграждат всички останали теоретични конструкции в нейните модели. Всеки въпрос в теста се описва чрез уникална характеристична крива, чиито свойства се определят като нейни параметри, които се атрибутират на съответния въпрос. Към трите статистики, които характеризират въпросите в СТТ, в Теорията за отговор на тестов въпрос са добавени още две - невнимателност (*carelessness*) и асиметрия на характеристичната крива. В рамките на тази теория са разработени множество теоретични модели, предназначени за моделиране на различни типове данни. Моделите се различават съобразно набора от допускания, които включват, най-съществени сред които са допусканията за размерността на пространството на латентните черти (едномерност), за вида на тяхното разпределение (нормалност), за вида на характеристичната крива (логистична), за броя на параметрите (едно-, дву- и трипараметрични) и др.

Всички изследователи, работещи в областта на психологическите измервания, са единодушни относно безспорните теоретични предимства на IRT. Новата теория предлага по-задълбочено разбиране на каузалната връзката между индивидите и айтемите, между латентните и манифестираните променливи. Тя се базира на вероятностен подход при моделиране на връзката между латентните черти и посочването на коректния отговор, както и при оценката на параметрите на модела, включително и на личностовия параметър. Едно от най-стойностните предимства на новата теория е това, че тя преодолява важни зависимости: първата от тях е зависимостта на статистиките на въпросите от групата от и.

л., въз основа на която са получени, а втората - зависимостта на тестовия бал от конкретния набор от въпроси (тест). Параметрите на въпросите и личностовия параметър в рамките на IRT са независими от средата и в този смисъл са инвариантни.

Въпреки фундаменталната значимост на въпроса за приложимостта на тестовите модели и призивите на мнозина автори като Ф. Лорд, Л. Крокер и Дж. Алгина, Р. Хамбълтън и др. за необходимостта от такива изследвания, те са твърде оскъдни и, доколкото ни е известно, няма нито едно, в което тази проблематика да е обхвана в нейната пълнота и дълбочина. По-голяма част от резултатите, които се отнасят предимно до IRT, са страничен продукт в значително по-обширната специализирана литература, посветена на параметричните статистики и на тяхната устойчивост. Поради това в тази сфера цари „емпирично безмълвие“, чувства се осезателно „липсата на емпирично знание“ (Fan, 1998) по-специално „потресаваща липса на такива данни от тестове за постижения и психометрични измервания“ (Miccieri, 1989).

Така например в свое обширно изследване С. Бреклер прави преглед на 72 статии в областта на личностовата и социалната психология, публикувани в 4 авторитетни американски психологически списания през 10-годишен (1977–1987) период. Анализирайки нарушенията на допускането за многомерна нормалност на разпределенията, авторът установява, че едва в 19% от статиите това изискване е посочено, и само в 10% от тях е обсъдено дали то е удовлетворено. В останалите над 70% от публикациите авторите избягват този въпрос, най-често приемайки *a priori* изследваните променливи като нормално разпределени (Breckler, 1990). Друг литературен преглед на статиите в 17 списания, направен от Х. Кеселман и сътрудници, показва, че авторите рядко вери-

фицират статистическите допускания и използват модели, които не са устойчиви срещу нарушенията на тези допускания (Keselman et al., 1998). Авторите обръщат внимание не само на нищожния обем на изследванията, посветени на проверката на едно или друго допускане, но и на обстоятелството, че по-голяма част от проведените изследвания са върху симулирани данни. Малко са тези, които „се осмеляват“ да работят с данни от конкретни емпирични изследвания (Micceri, 1989).

В дисертационния труд са представени две групи изследвания върху приложимостта: (1) върху данни от тестове за постижения и (2) върху данни от други източници.

В част от публикациите от първата група авторите правят по-общ сравнителен анализ на особеностите, преимуществата и недостатъците на двете психометрични тестови теории. Р. Хамбълтън и Р. Джоунс, подчертавайки предимствата на „новата“ теория, не пропускат да обърнат внимание на това, че СТТ също се развива и прилага успешно в множество тестови програми (Hambleton & Jones, 1993). Авторите отбелязват, че бидейки по-абстрактна и съдържаща понятия от по-високо ниво, една тестова теория не следва да се оценява от гледна точка на нейната полезност. И обратно, като конкретизация на дадена тестова теория, приложимостта на един теоретичен модел следва да се проверява върху конкретна съвкупност от тестови данни, с добре обмислена система от емпирични методи. Съпоставяйки двете теории и техните модели, авторите намират редица сходства, но и съществени различия между тях. Основните принципни различия, които дават огромно теоретично предимство на IRT, са тези, че статистиките на въпросите и на теста като цяло са независими от извадките от и. л. а оценките на способностите – независими от теста, чрез които са определени. Като

недостатъци авторите отбелязват сложността на моделите на IRT и проблемите, свързани с оценката на различните параметри, които рефлектират върху приложимостта на нейните модели. Според Р. Амарнани основната разлика в двете теории е в информацията, която се използва при формиране на оценка на способността – по-груба и неточна при СТТ и по-прецизна при IRT (Amarnani, 2009).

В друга част от публикациите от първата група авторите представят конкретни резултати от съпоставителни изследвания на двете теории. Най-често те са фокусирани върху статистиките на айтемите и личностовите статистики, по-точно върху взаимовръзките между сходните индекси/ параметри и доколко те са инварианти по отношение на различни извадки. К. Фан съпоставя личностовия параметър и някои от статистиките на въпросите върху емпирични данни от критерийно-ориентирана тестова батерия (Fan, 1998). Съгласно получените резултати, всички наблюдавани статистики корелират високо, с тенденция за по-слаба съпоставимост на дискриминативните индекси/ параметри, а за инвариантност на статистиките може да се говори не само при IRT, но и при СТТ. К. Фан резюмира резултатите от направените от него изследвания с мнение, че те не водят до дискредитиране на СТТ от гледна точка на приписваните ѝ слабости и не подкрепят IRT в нейното мнимо превъзходство.

Сравнително емпирично изследване на двойка съответни индекси/ параметри на въпросите (дискриминативна сила и трудност) в СТТ и 1-, 2- и 3-параметричния модел на IRT върху данни от тест за оценка на склонността към обучение, представя Р. Хернандез (Hernandez, 2009). Авторът намира висока, положителна връзка между двойките статистики, силата на която все пак варира в зависимост от съдържанието на съответния субтест и от модела на IRT.

Друго интересно изследване върху реални данни от GMAT представят група изследователи от ETS (Kingston, Leary & Wightman, 1985). Неговата цел е да се провери приложимостта на едномерния 3-параметричния логистичен модел на IRT и последствията от нарушенията на неговите допускания. Авторите подлагат на проверка допусканията за едномерност на субтестовете и за формата характеристичната крива на въпросите. Резултатите от проверката на първото допускане говорят за многомерност на латентните структури, а приложените техники за оценка на адекватността на 3-параметричната логистична функция на въпросите са противоречиви. Въпреки това авторите обобщават, че изследването е дало доказателства за приложимостта на IRT върху GMAT.

В своя публикация Р. Нандакумар представя изследване за оценка на годността на психометричния софтуер DIMTEST да разкрива едномерни латентни структури (Nandakumar, 1993). „Страничен” резултат от това изследване, проведено върху реални данни от различни тестове и тестови батерии е, че нито един от (суб)тестовете не се характеризира със стриктна едномерност.

Особено интересно е обширното изследване на М. Виберг за възможностите на СТТ и IRT при анализа на въпросите от теоретичната част на теста за придобиване на свидетелство за управление на МПС в Швеция. Авторката подлага на проверка множество допускания на 1-, 2- и 3-параметричните модели на IRT. Сред по-важните резултати са наличието на едномерна латентна структура, на локална независимост, липса на инвариантност на оценките на личностовия параметър, както и на параметрите на въпросите, наличие на висока съгласуваност между статистиките на въпросите, определени в рамките на двете теории. В заклю-

чение М. Виберг стига до извода, че нито един от анализираниите модели на IRT не съответства напълно на тестовите данни. Авторката се колебае в крайното си решение, но все пак предпочитанията ѝ са към 2- и 3-параметричния модел, Що се отнася до алтернативата IRT или СТТ, авторката смята, че двете теории са полезни в еднаква степен, тъй като носят ценна информация както за теста като инструмент за измерване, така и за изпитаните лица.

Сред изследванията, базирани върху данни от източници, различни от тестовите за постижения, особено ценно е това на Т. Мичери, който анализира характеристиките на разпределенията 440 емпирични извадки, получени в различни области на социалните и поведенческите науки (Micceri, 1989). Прилагайки разнообразни техники за изследване на формата на техните разпределения, авторът не открива никакъв устойчив модел на разпределение на данните и заключава, че твърде малка част от тях са „дори сравнително близка апроксимация на Гаусовото”.

## **2.3. Обща постановка на изследването**

### **2.3.1. Обосновка на необходимостта от изследване на приложимостта на тестовите теории (модели)**

Представяното изследване е проведено в контекста на засиления интерес към теоретичните и приложните аспекти на психологическите измервания в поведенческите и социалните, а през последните години – и в медицинските науки. То е фокусирано върху психометрични теории, представени по-горе. Те притежават две важни характеристики, от които произтичат основните направления в настоящото изследване: (а) моделите, разработени в техните рамки, са модели на данни и (б) те са функционално еквивалентни, но независи-

ми, алтернативни една на друга.

Психометричните тестови модели могат да бъдат класифицирани като модели на данни (Suppes, 1962; Frigg & Hartmann, 2006). В тях суровите, реални данни се представят в един непълен, но добре подреден, организиран и в известен смисъл идеализиран вид. Поради тези особености един от фундаменталните въпроси при прилагането на такива модели е за съответствието между даден модел, по-точно между неговите допускания, и емпиричните данни, който може да бъде формулиран като въпрос за адекватността на модела. От друга страна, двете теории могат да бъдат разглеждани като функционално еквивалентни системи за измерване. Съответствието между тях се дължи на това, че имат една и съща област на приложимост, ориентирани са към един и същи сегмент от действителността, стъпват на една и съща емпирична база от данни и са предназначени за постигане на една и съща цел - скалиране и оценка на латентни черти.

### **2.3.2. Основна цел и изследователски въпроси**

Основната изследователска цел в представената разработка е да се направи сравнително изследване на приложимостта на два модела в рамките на СТТ и IRT, върху реални данни от ТОП. Концепцията за приложимостта е третирана в два нейни аспекта:

(а) степента на съответствие между допусканията на съответния теоретичен модел и характеристиките на емпиричните тестови данни, т. е. валидността на основните допускания на теоретичния модел спрямо данните;

(б) степента, в която очакваните свойства на теоретичния модел се проявяват в емпиричните тестови данни, най-вече по отношение на очакваното „поведение” на статисти-



ките на тестовите въпроси.

В светлината на разгледаните предимства и недостатъци на двете тестови теории, в разработката са потърсени отговорите на следните два основни изследователски въпроса:

(1) Съответства ли Класическата тестова теория, в рамките на която функционира ТОП, в достатъчно висока степен на тестовите данни?

(2) Би ли довела замяната на стария психометричен модел с новата Теория за отговор на тестов въпрос, поради нейните безспорни теоретични предимства, до подобряване на измерителните качества на теста?

Тук следва да се подчертае, че акцентът в изследването е върху търсенето на различни по характер свидетелства „за“ и „против“ приложението на новата психометрична теория при разработването и анализа на резултатите от ТОП.

Изследването е фокусирано върху проверка на адекватността на следните два теоретични модела, избрани като „базови“ модели:

В теоретичната рамка на СТТ - едномерен, с нормално разпределение на действителния бал,  $\tau$ -конгенеричен модел, в теоретичната рамка на IRT - основан на дихотомични отговори, едномерен, с нормално разпределение на латентната способност  $\Theta$ , параметричен, логистичен модел.

В светлината на тези предпоставки в изследването са поставени и следните две групи от изследователски въпроси, които конкретизират формулираните по-горе общи проблеми.

Първата група произтича от разбирането на концепцията за приложимостта като съответствие между допусканията на теоретичния модел и характеристиките на тестови данни:

(1) Каква е формата на разпределенията на латентните спо-

способности? Доколко обосновано е допускането, че латентните способности следват нормалното Гаусово разпределение? (2) Каква е размерността на пространството на латентните способности, които обуславят отговорите на и. л. на въпросите от ТОП? Има ли емпирични свидетелства, които да подкрепят допускането за тяхната едномерност?

Втората група е свързана с разбирането на приложимостта като проява на очакваните свойства на теоретичния модел в емпиричните тестови данни: (3) При кой от двата теоретични модела, съответно на СТТ и IRT, статистиките на тестовите въпроси са инвариантни в различни условия, т. е. при различни извадки от индивиди? (4) Доколко статистиките на тестовите въпроси, определени в рамките на един и същи модел, в едно и също условие, т. е. при една и съща извадка от индивиди, функционират независимо една от друга? (5) Наблюдава ли се съгласуваност между индексите, определени в рамките на СТТ, и съответните им параметри, определена в рамките на IRT, в едно и също условие, т. е. при една и съща извадка от индивиди?

### **2.3.3. Методология**

В основата на емпиричната част на разработката са заложили три изследователски подхода, които спояват отделните изследвания в нейните рамки. Първият от тях е познат като Изследователски анализ на данни, предложен и обоснован от Дж. Тюки (Tukey, 1977). За разлика „класическия“ потвърдителен подход, основан на проверка на предварително формулирани хипотези (или модели), ИАД предполага последователното и многостранно изследване на събраните данни чрез разнообразни статистически техники, което да доведе до „извличане“ на модела от тях, вместо да бъде предпоставен. Вторият подход е свързан с осигуряването

на вътрешната валидност на резултатите от изследванията. В дисертационния труд е възприето разбирането за този тип валидност като статистическа валидност, която се отнася до адекватния подбор на статистическите методи и до коректността на последващата интерпретация на получените чрез тях резултати (Cohen & Swerdlik, 2005). Третият подход е свързан с осигуряване на външната валидност на резултатите от изследванията. Като мярка е приложен подходът за репликиране на ситуациите, т. е. за паралелно изследване на множество от еднотипни съвкупности от данни чрез извършване на съответните анализи върху множество от тестови варианти, използвани в различни точки от време. Свързан с него е и проблемът за екологичната валидност, обезпечена чрез използването на реални данни, получени при множество изпитни ситуации.

Източник на данни в емпиричното изследване е ТОП, който стои в основата на приемните процедури в Нов български университет. За осъществяване на целите на изследването са подбрани 15 варианта на ТОП от различни кандидатстудентски кампании през 6-годишен период, от различни етапи на тези кампании, като броят на и. л., явили се на един тестов вариант, варира от 454 до 1019.

За постигане на основната цел на изследването, както и на произтичащите от нея изследователски въпроси, е проведена серия от 5 относително самостоятелни емпирични изследвания. Изследване 1. е предназначено за анализ на формата на разпределенията на латентните променливи, а в изследване 2. се разглежда въпросът за размерността на латентните пространства с оглед на верифициране на допускането за тяхната едномерност. В изследване 3. се анализира на инвариантността на статистиките на тестовите въпроси, изследване 4. е посветено на взаимовръзките между раз-

ноименните индекси/ параметри в рамките на един и същи теоретичен модел, а изследване 5. – на съгласуваността между статистиките на въпросите, определени в рамките на СТТ, и кореспондиращите им статистики в рамките на IRT.

## 2.4. Резултати

В тази глава, разделена на две основни части, са представени резултатите от съпоставителните изследвания, направени в рамките на двете групи от изследователски въпроси. В първата част са представени две изследвания, посветени на съответствието между допусканията на теоретичните модели и характеристиките на тестовите данни.

Целта на първото изследване е да се направи анализ на формата на разпределенията на тестовите балове, главно от гледна точка на допускането за съответствие с нормалната крива. Приложени са четири групи методи - числени, основани на модел (два теста за нормалност -  $W$  на Shapiro-Wilk и  $K-S$  на Колмогоров-Смирнов, модификация на Lilliefors), числени дескриптивни методи (асиметрия и ексцес), графични методи, основани на модел (нормални вероятностни графики) и графични дескриптивни методи (хистограми на разпределенията, диаграма на Тюки тип “чувал”). Изследването е направено паралелно върху 132 съвкупности от данни, получени от 12 тестови варианта на ТОП, анализирани на равнище тест и субтест.

Резултатите от приложените методи са във висока степен консистентни и в противоречие с наложилото се мнение за нормалното разпределение като „златен стандарт”. Графичните методи, както и числените дескриптивни методи, разкриват голямото разнообразие от форми на разпределенията, с различна степен на отдалеченост от нормалното.

Свидетелствата за нормалност са сравнително малко на брой, като се очертава ясна тенденция за доминиране на разпределенията с положителна асиметрия и отрицателен ексцес. От друга страна, по-строгите тестове на съгласието водят почти еднозначно до отхвърляне на съответните нулеви хипотези за нормалност на разпределенията. Може да се направи изводът, че в общия случай разпределенията на тестовите балове в ТОП се отклоняват от нормалното Гаусово разпределение. Оттук следва, че между теоретичното допускане за нормалност на разпределението, и съответната характеристика на наблюдаваните данни, по правило липсва съответствие. Въпреки това, при цялото разнообразие от форми на наблюдаваните емпирични тестови разпределения, Гаусовата крива е тяхната най-подходяща теоретична апроксимация.

Целта на второто изследване е да се направи анализ на латентните структури на ТОП, преди всичко от гледна точка на допускането за тяхната едномерност. Приложена е комбинация от методи, която включва изследователски факторен анализ за изграждане на първоначалните модели, потвърдителен факторен анализ за проверка на хипотезите за едномерност, както и йерархичен факторен анализ за изграждане на финалните модели. Факторната структура на ТОП е изследвана върху данни от 12 варианта на теста, на две равнища – субтестово (при което компоненти на анализа са въпросите във всеки субтест) и тестово (компоненти на анализа са въпросите в цялостния тест).

На изследователския етап като метод на факторизиране е приложен анализ на главни фактори (анализ на главни оси), като методи за определяне на броя на факторите – паралелен анализ на Дж. Хорн, графичен тест на Кетел и критериите на Ф. Лорд, а като входни данни за използвани мат-

рици на тетрагоричните корелации между отговорите на и. л. на отделните въпроси.

На субтестово равнище част от латентните структури могат да се определят като еднофакторни, с един доминиращ фактор и до шест второстепенни фактора. Това е често срещаният модел, наблюдаван предимно при субтестовете по български език, математика, разсъждения и семантика. При друга част от субтестовете съответните конфигурации са с „нулева“ факторна структура, т. е. с незначителен първи фактор. Тези конфигурации са около 1/3 от всички и се наблюдават най-често при субтестовете по физика, химия и биология. Като цяло факторните решения на субтестово равнище се характеризират с чувствителност към промени в състава (въпросите) в субтеста, с ниска смислова „плътност“, т. е. с ниски факторни тегла на въпросите и като цяло – със структурната слабост, неяснота и нестабилност на латентни структури, която се изразява в различна степен на изразеност на първия фактор и различен брой на второстепенните фактори.

На равнище тест бяха установени 8 – 10 факторни решения, които не съвпадат със субтестовите скали в ТОП, тъй като въпросите от отделните субтестове се разпределят между 2 и 7 от димензиите. Въз основа на тези резултати може да се направи заключението, че латентните структури на ТОП не са организирани на тематичен (субтестов) принцип. Поради това, че ортогоналната ротация на факторите не води до постигане на прости факторни структури, беше приложен йерархичен факторен анализ. Бяха извлечени 5 фактора от втори ред, 3 от които се поддават на интерпретация. Тези фактори могат да бъдат определени като „способност за използване на правила“, „способност за обобщаване“ и „възпроизвеждане на знания“ (паметови способности).

Във втората част на тази глава са представени три изследвания, посветени на проявите на очакваните свойства на теоретичните модели в тестовите данни. По-конкретно, разгледани са въпросите, свързани със съвместното вариране на едноименните и на разноименните статистики на въпросите, поставени в едно и също или в различни условия. Статистическите анализи са предимно корелационни, с използване на параметрични и непараметрични методи в зависимост от типа на скалата на съответната статистика. Използвани са данни от три двойки варианти на ТОП, при които има почти пълно съответствие между въпросите в отделните двойки. Вариантите са от различни календарни години, от различни изпитни сесии, с различен обем на извадките.

Целта на третото изследване е да се направи анализ на инвариантността на статистиките на тестовите въпроси, разбрана като тяхна стабилност, устойчивост в различни условия, т.е. при различни извадки от и. л. Допусканията, които са подложени на проверка, са за вариативност на индексите в рамките на СТТ и за инвариантност на съответните параметри в рамките на IRT. За оценка на стабилността на изследваните статистики е приложена комбинирана критерийна мярка, включваща подходящи за съответната статистика корелационен коефициент и тест за оценка на равенството на централните тенденции на съответните разпределения. По отношение на стабилността, оценена чрез корелационните мерки, статистиките в рамките на двете теории демонстрират съпоставимо поведение. Наблюдават се изключително високи коефициенти на корелация, които при трудността на въпросите ( $p$  и  $b$ ) достигат стойности съответно 0.99 и 0.97. При останалите статистики стойностите са малко по-ниски. При въвеждане на втората мярка от критерия обаче параметрите на дискриминативна сила и налучкване на

правилния отговор в рамките на IRT се провалят, като проявяват статистически значими разлики между централните тенденции на съответните двойки разпределения. Следователно резултатите от направените изследвания сочат, че нито едно от двете допускания не е валидно напълно и в еднаква степен за всички анализирани статистики по отношение на тестовите данни. Може да се обобщи, че като цяло индексите на тестовите въпроси в рамките на СТТ са по-надеждни от параметрите в рамките на IRT.

В следващото изследване е направен анализ на взаимовръзките между (а) индексите на въпросите, определени в рамките на СТТ, и (б) параметрите, определена в рамките на IRT, в едно и също условие, т.е. при една и съща извадка от и. л. Допусканията са за наличие на взаимовръзки между индексите и за отсъствие на такива между параметрите на въпросите. Анализите са изцяло корелационни, с прилагане на подходящ за типа на съответната скала коефициент на корелация.

Резултатите от изследванията в рамките на СТТ говорят в полза на формулираното първо допускане. Между отделните двойки индекси се наблюдават умерено високи и високи, статистически значими корелации. Особено ясно изразена е корелацията между индексите на трудност и дискриминативна сила, нелинейна по своя характер (със стойност корелационното отношение  $\eta$ , равна на 0.686), апроксимирана по метода на нелинейната регресия чрез полином от втора степен. Противно на очакванията, между параметрите на въпросите в рамките на IRT също се наблюдават статистически значими, умерено до високи взаимовръзки, макар и по-слаби от тези в СТТ. Така например корелацията между параметрите на дискриминативна сила (*a*) и налучкване на коректния отговор (*c*), линейна по характер, достига



стойности от  $-0.82$ . Линейна е корелацията и между параметрите на трудност ( $b$ ) и налучкване на коректния отговор ( $c$ ) с максимална стойност от  $-0.36$ . По-сложна е взаимовръзката между параметрите на дискриминативна сила ( $a$ ) и трудност ( $b$ ), която се апроксимира с полиномна функция от 4-та степен. Особен интерес предизвиква силната, негативна взаимовръзка между параметъра за налучкване на коректния отговор и останалите два параметъра. За обяснение на този феномен е направено допускане за съществуването не на симетрична (корелационна), а на едностранна (каузална) връзка между тях. Параметърът  $c$  е разгледан не само като характеризиращ въпросите, а наред със способностите  $\Theta$ , като втори параметър в IRT, който съдържа личностов компонент и който отразява определени психични процеси.

Последното, пето изследване е посветено на съгласуваността между статистиките на въпросите, определени в рамките на СТТ, и кореспондиращите им статистики в рамките на IRT, определени в едно и също условие, т.е. при една и съща извадка от и. л. Допускането, което е подложено на проверка, е за липса на съгласуваност между разноименните статистики. Направените анализи са корелационни, с прилагане на подходящ за типа на съответната скала коефициент на корелация.

Резултатите от направените анализи са недвусмислено свидетелство за невалидността на направеното допускане по отношение на данните от ТОП. При всички двойки функционално сходни статистики се наблюдават ясно изразени взаимовръзки, които са от линеен или нелинеен тип. Особено характерна е линейната корелация между индекса на трудност ( $p$ ) и съответния параметър ( $b$ ), която в различните тестови варианти е на равнища над  $-0.90$ , достигайки до  $-0.98$ , като всички стойности са статистически значими. Коре-

лация от нелинеен тип се наблюдава между индекса на дискриминативна сила в двете му форми ( $D$  и  $r_{bis}$ ) и съответния параметър, със стойности на корелационното отношение  $\eta$  на равнища от 0.65. Тези резултати дават основание да се заключи, че двата теоретични модела могат да се разглеждат като взаимнозаменяеми, особено по отношение на трудността на въпросите.

## 2.5. Обща дискусия

Доколкото ТОП се разработва в рамките на Класическата теория, която от теоретична гледна точка страда от редица недостатъци, основният изследователски въпрос в разработката бе доколко тази теория (чрез определения базов модел) е приложима към данните от ТОП. Вторият основен изследователски въпрос бе дали преминаването от класическия модел към Теорията за отговор на тестов въпрос би довело, поради нейните безспорни теоретични предимства, до подобряване на измерителните качества на теста.

Съпоставянето на резултатите от направените анализи дава основание да се заключи, че между „поведението” на двата теоретични модела, приложени върху широка съвкупност от тестови данни, има редица очевидни сходства.

Съгласно резултатите от верификацията на основните допускания на моделите, разпределенията на тестовите баловете се характеризират с различна степен на отдалеченост от нормалната Гаусова крива. Въпреки че тази крива е тяхната най-добра апроксимация, данните не могат да се разглеждат *a priori* като нормално разпределени. Прилагането на тестовите модели би довело до определени неточности в оценките на статистиките, базирани на това разпределение. И тъй като изискването за нормалност е императивно за IRT,

при която всички статистики са основани на нормалното разпределение, нейното приложение би било по-малко обосновано от това на СТТ.

Тестовите данни се характеризират със слабо изразена, многофакторна латентна структура от общи способности. При част от анализиранияте субтестове тя може да се приеме за основно едномерна и към такива данни могат да се приложат и двата модела. На тестово равнище, въпреки наличието на силен първи фактор, латентната структура може да се разглежда като тримерна. Това поставя под въпрос използването на който и да е от двата модела.

Съгласно резултатите, представени в разработката, стабилността на индексите на въпросите, определени в рамките на СТТ, е не само напълно съпоставима с тази на параметрите по IRT, но и в много отношения я превъзхожда. Това е може би най-малко очакваната разлика в поведението на двете тестови теории. Опровергано беше и още едно допускание – между параметрите на въпросите, определени в рамките на IRT, съществуват силни взаимовръзки така, както съществуват и между индексите по СТТ. Двете теоретични рамки са сходни и поради това, че между едноименните статистиките на въпросите има висока степен на съгласуваност, особено по отношение на трудността.

Тъй като двата тестови модела бяха последователно приложени към една и съща съвкупност от тестови данни, сравнително еднаквото им представяне следва да се разглежда като свидетелство, че приеманото за даденост теоретично превъзходство на IRT не се е осъществило, поне спрямо данните от ТОП. Те са среда, към която „меката” и по-„непретенциозна” Класическата теория се оказва по-добре приспособена, отколкото новата психометрична теория. Като основна причина можем да посочим дисхармония-

та между характеристиките на реалните данни, които се отличават с определена степен на неподреденост и неорганизираност, на аморфност, и строгите изисквания на далеч по-сложната в концептуално, структурно и математическо отношение IRT. Следователно преминаването от Класическия модел към новата психометрична теория като рамка за разработването на ТОП в сегашния му вид, не би довело да съществено подобряване на неговите измерителни качества.

Като теоретична концепция IRT съществува под формата на множество модели, съдържащи различен набор от допускания, предназначени за удовлетворяване на различни видове данни. Въз основа на резултатите от изследванията бихме могли да очертаем профила на модела, който би бил по-подходящ за приложение върху ТОП. Този модел следва да бъде освободен от допускане за нормалност на разпределенията на променливите, каквито са моделите на Рамзи, непараметричните модели, както и тези, основани на асиметрични разпределения. За анализ на данните на тестово равнище по-подходящ би бил многомерен, в частност – тримерен, трипараметричен модел. За съжаление, тези модели са по-сложни и не така добре разработени както „стандартния” модел, послужил като основа на направените изследвания.

Трябва да се отбележи, че в представената разработка са засегнати далеч не всички аспекти на приложимостта на тестовите модели. Необходимо е да се извърши немалка по обем изследователска работа за изясняване на проблемите, свързани с локалната независимост на отговорите; с диференциалното функциониране на отделните айтеми и на теста; с адекватността на логистичния модел на характеристичната крива на въпросите; със съгласуваността между наблюдавания бал по СТТ и оценките на способностите по IRT;

с надеждността и валидността на резултатите на субтестово и тестово равнище, включително конструктната и предиктивната валидност; с изясняване на латентната структура на субтестово и тестово равнище и по-прецизното характеризиране на латентните способности. Търсенето на адекватен тестов модел не трябва да се разглежда като самоцел. За да изпълнява добре своите функции, ТОП следва да отговаря на четири основни характеристики:

- валидност на резултатите – особено конструктната и предиктивната валидност;
- надеждност на резултатите;
- ефективност – изпитните процедури не следва да отнемат повече ресурси (материални, човешки, финансови), отколкото е необходимо;
- приемливост – всички заинтересовани страни да се доверяват на резултатите и да бъдат удовлетворени от тяхното качество.

Изпълнението на тази нелека задача вероятно би могло да бъде постигнато чрез търсенето на приемлив компромис между подходящ, работещ теоретичен модел и промяна в дизайна на теста, който да съответства на получените емпирични резултати.

### **3. Справка за приносите в докторската теза**

#### **3.1. Приноси на изследването в теоретичен аспект**

3.1.1. В представената разработка е направено първото по рода си комплексно сравнително изследване на приложимостта на Класическата тестова теория и Теорията за отговор на тестов въпрос чрез анализиране на широк спектър от техните най-съществени особености.

3.1.2. Представените изследвания на приложимостта на двете психометрични тестови теории нарушават „емпиричното безмълвие” в тази насока. Те внасят своя дял в преодоляването на недостига от емпирично знание за приложимостта чрез получените резултати за два конкретни модела в конкретен емпиричен контекст – обстоятелство, подчертано от мнозина изследователи.

3.1.3. Беше обоснована необходимостта от изследване на формата на разпределенията и размерността на латентните структури като предпоставка за приложението на СТТ.

3.1.4. Бяха установени системни, макар и в различна степен, отклонения на разпределенията на тестовите балове от нормалната крива на равнище субтест и цялостен тест. Тези наблюдения кореспондират с някои от резултатите, публикувани в литературните източници, и като цяло се противопоставят на концепцията на Р. Фишер за „универсалната нормалност”, въз основа на която променливите в поведенческите науки най-често се приемат *a priori* като нормално разпределени. Наблюдаваните несъответствия поставят под въпрос приложимостта на двете теории към данните, полу-

чени чрез ТОП. Въпреки това, при цялото разнообразие от форми на наблюдаваните емпирични разпределения, Гаусовата крива остава тяхната най-подходяща теоретична апроксимация.

3.1.5. Изследванията на латентните структури на ТОП на субтестово и тестово равнище разкриха тяхното несъответствие с неговата планирана, „видима“ скалова структура.

3.1.6. Беше установена емпирично вариативността, динамичността на субтестовите структури по отношение на броя на латентните дименсии, включително наличието на аморфни конфигурации без доминиращ фактор (с „нулева“ факторна структура). Поради сравнително малкия брой на въпросите в субтестовете, техните латентни структури като брой фактори и обяснителна сила са чувствителни към промени във вътрешния им строеж.

3.1.7. Беше показано, че на субтестово равнище доминантността на първия фактор (едномерността на съответната латентна структура), се дължи преди всичко на големината на собствената му стойност, а не на останалите.

3.1.8. Беше установено типологично и функционално сходство между дискриминативната сила (по-конкретно - бисериалния индекс) като оценка на взаимовръзката между въпросите и суровия тестов бал (манифестирана променлива), и факторните им тегла като оценки на взаимовръзката между въпросите и съответния фактор (латентна променлива). Това води към ново разбиране на психометричното значение на дискриминативната сила като характеристика на айтемите. Може да се приеме, че тя отразява степента на принадлежност на въпроса към дадена скала, т. е. има отношение към конструктната валидност на скалата.

3.1.9. На равнище тест беше установено наличието на латентна структура, съдържаща три общи фактора от втори

ред, които могат да бъдат интерпретирани като „способност за използване на правила”, „способност за обобщаване” и „възпроизвеждане на знания” (паметови способности).

3.1.10. Беше установено, че на двете равнища на анализ съотношението между обяснителната сила на общия/общите фактори, които оказват влияние върху отговорите на всички тестови въпроси, и уникалните фактори, всеки от които влияе върху конкретен тестов въпрос, е в полза на уникалните фактори.

3.1.11. Във връзка с установената трифакторна структура от втори ред, която се проявява отчетливо на тестово и по-непоследователно – на субтестово равнище, беше поставена под въпрос конструктната валидност на ТОП и неговите субтестови компоненти.

3.1.12. Беше показано, че като цяло статистиките на тестовите въпроси, определени в рамките на СТТ, демонстрират по-висока степен на стабилност, отколкото тези, определени в рамките на новата теория. В този смисъл допусканията за зависимост на индексите от извадките от и.л., както и за независимост на параметрите, не намират достатъчно опора в емпиричните данни.

3.1.13. Беше потвърдена очакваната взаимовръзка между индексите на въпросите, определени в рамките на Класическата теория. По-конкретно, беше наблюдавана нелинейна взаимовръзка между индекса на трудност и двата индекса на дискриминативна сила (класически и бисериален коефициент), както и линейна взаимовръзка между двата дискриминативни индекса. Направена е количествена оценка на силата на нелинейната корелация между трудността и класическия индекс на дискриминативна сила, изграден е модел на полиномна функция от втора степен, описваща тази взаимовръзка.



3.1.14. Беше опровергано очакването за липса на взаимовръзки между параметрите на тестовите въпроси, определени в рамките на IRT. Между дискриминативната сила на въпросите и тяхната трудност беше наблюдавана ясно изразена нелинейна връзка. Направена е количествена оценка на силата на корелацията между тези два параметъра, изграден е модел на полиномна функция от четвърта степен, описваща тази взаимовръзка. От съществено значение са наблюдаваните негативни зависимости между параметъра на налучкване на коректния отговор и параметъра дискриминативна сила на въпросите, както и с тяхната трудност.

3.1.15. Анализът на теоретичните дефиниции на параметрите, както и на особеностите на наблюдаваните взаимовръзки между тях, дават основание да се предположи, че параметърът на налучкване на коректния отговор играе роля на предиктор по отношение на други два, и да се разглежда като статистика, която включва личностов компонент, повлияна по своеобразен начин от мотивацията на и. л. за високи постижения.

3.1.16. Беше установено емпирично, че в противоречие със схващането за ранговия характер на скалата, образувана от индексите на трудност, определени в рамките на СТТ, тази скала може да се третира като интервална.

3.1.17. В по-общ план бе показано, че някои скали, разглеждани като рангови, не са такива в строгия Стивънсов смисъл на понятието. Те са разположени в „сивата” зона между ранговия и интервалния тип скали, която може да бъде осветлена и да се направи конкретна оценка към кой скалов тип се приближава дадена скала. Тя може да съдържа както метричен (линеен), така и неметричен (рангов) компонент, като двата компонента могат да бъдат разграничени в отделни сегменти по протежението на самата скала. По-

обосновано е теоретичните конструкции от Стивънсовата класификационна система да се разглеждат като размито множество, съгласно концепцията на Л. Заде (Zadeh, 1965), според която членството на даден обект (скала) към дадено множество (скалов тип) е континуално, а не дихотомично.

## **3.2. Приноси от изследователско-методологична гледна точка**

3.2.1. Представените изследвания са осъществени чрез комплексното прилагане на три методологически подхода:

- Изследователски анализ на данни, обоснован от Дж. Тюки, в сърцевината на който стои отказа от „класическия“ потвърдителен подход, основан на предварително формулиране на хипотези и последващата им проверка със статистически тестове. ИАД предполага многостранно изучаване на събраните данни чрез системен анализ на взаимовръзките между променливите и прилагане на разнообразни методи и средства, чрез натрупване на различни свидетелства за съответствията между теоретичните модели и емпиричните данни.

- Осигуряване на вътрешната валидност на резултатите от изследванията и по-конкретно на неговата статистическа валидност, която се изразява в адекватния подбор на статистическите методи и в коректността на интерпретацията на получените чрез тях резултати.

- Осигуряване на външната валидност на резултатите, на възможността за генерализация върху други съвкупности от тестови данни. Като мярка за осигуряване на този вид валидност е приложен подходът за репликиране на ситуацияте, т. е. за анализиране на различни извадки от индивиди и от

тестови въпроси чрез паралелно изследване на множество тестови варианти на ТОП.

- Тясно свързан с външната валидност е и проблемът за осигуряване на екологичната валидност на резултатите. В представеното изследване този вид валидност е обезпечена чрез използването на реални данни, получени от 15 варианта на ТОП, използвани в кандидат-студентските кампании на НБУ през периода 2003 – 2008 година. Изследването е първото по рода си по отношение на обема на анализирания данни, получени от прилагането на тестове за постижения, и първото по рода си по отношение на ТОП.

3.2.2. В разработката е възприето по-широко третиране на проблема за приложимостта психометричните модели, който включва два аспекта: (1) степента на съответствие между допусканията на теоретичния модел и характеристиките на емпиричните тестови данни и (2) степента, в която очакваните свойства на теоретичния модел се проявяват в тези данни, по-конкретно по отношение на очакваното „поведение” на статистиките на тестовите въпроси.

3.2.3. При изследването на латентните структури на ТОП беше комбинирано приложението на изследователски факторен анализ за изграждането на подходящи модели и потвърдителен факторен анализ за проверка на тяхната адекватност.

3.2.4. При изследователския факторен анализ бе приложена комбинация от следните подходи:

- Като метод на факторизиране бе предпочетен анализ на главни фактори вместо анализ главни компоненти, който се използва обичайно в по-голяма част от изследванията в областта на поведенческите науки.

- Като основен метод за определяне на броя на факторите във факторните модели беше използван паралелният

анализ на Дж. Хорн вместо широко използваните методи като теста на Кайзер-Гутман и графичния тест на Кетел.

- Като входни данни при изследователския факторен анализ бяха използвани матрици на тетрагоричните корелации, които са подходящи за дихотомизирани бинарни данни, вместо коефициента на Пиърсън или коефициента  $\phi$  ( $\phi$ ).

- За разкриване на латентната структура на равнище тест бе приложен йерархичен факторен анализ на неортогонални фактори.

3.2.5. При определяне на инвариантността/ стабилността на статистиките на въпросите бе приложен „конюнктивен” критерий, който включва две условия, които следва да бъдат удовлетворени едновременно – надпрагова стойност на корелационния коефициент и равенство на оценките на централните тенденции на съответните разпределения чрез дисперсионен анализ с повторни измервания или знаково-ранговия тест на Уилкоксън.

3.2.6. В случаите, в които проверката на нулевите хипотези при прилагането на дисперсионен анализ с повторни измервания води до тяхното отхвърляне, беше направено системно наблюдение на размера на ефекта чрез използване на коефициента на частна корелация ета на квадрат.

### **3.3. Практически приложения на изследването и резултатите от него**

3.3.1. Беше емпирично установена по-високата степен на пригодност на Класическата тестова теория за конструиране и анализ на данните от ТОП от гледна точка на настоящия му дизайн.

3.3.2. Наблюдаваните ниски равнища на съгласуваност между въпросите както на субтестово, така и на тестово рав-

нище дават основания за очакване на ниски равнища на надеждност на тестовите резултати, особено на субтестово равнище. Това предполага засилване на вниманието към въпросите, свързани с осигуряване на надеждността на тестовите резултати.

3.3.3. В не по-малка степен следва да се обърне внимание на проблемите, свързани с валидността на ТОП на двете му равнища. Установеното несъответствие между планираната, „видима“ скалова структура и действителната му факторна структура сочат необходимостта от преосмисляне на неговия дизайн и съгласуването му с получените в настоящото изследване резултати.

3.3.4. Незадоволителното съответствие между допусканията на IRT и реалните данни, както и разминаването между очакваното и действителното „поведение“ на параметрите на тестовите въпроси не следва да се разглежда като основание за отхвърляне на възможността за нейното приложение. Тази теория има съществени предимства и огромен потенциал за решаване на различни изследователски и практически задачи. Нейните предимства обаче могат да бъдат постигнати чрез постигане на приемливо ниво на съответствие между подходящ модел и тестовите данни.

3.3.5. Въз основа на цялостния анализ на резултатите бе предложен (непълнен) модел на IRT, който в най-висока степен съответства на характеристиките на емпиричните данни. Този модел, по всяка вероятност, се нуждае от по-нататъшен детайлен анализ и прецизиране.

## Публикации по дисертационната тема

1. Джалев, Л., Богданов, Ст. (2012). *Две гледни точки към педагогическата ползваемост на електронни дидактически материали*. Raleigh, N. C.: Lulu Press Inc., 77 стр.

<http://www.lulu.com/shop/stan-bogdanov-and-ljubomir-djalev/two-viewpoints-on-pedagogical-usability-in-elearning-for-languages/paperback/product-21009445.html>

Научен електронен архив на НБУ,  
<http://eprints.nbu.bg/1384/>, 82 стр.

2. Grinberg, M. & Hristova, E., Eds. (2011). *Efficiency and Usability of E-Learning Systems: Project-Oriented Methodology Guide*. Научен електронен архив на НБУ,  
<http://eprints.nbu.bg/635/>, (143 стр., под печат)

3. Бижков, Г., Ванева, В., Танева, К., Захариева, Р., Джалев, Л., Трайкова, В. (2011). *Тест за диагностика на готовността на децата за училище. Книга за учителя*. С.: Университетско издателство „Св. Климент Охридски“ (103 стр.)

4. Бижков, Г., Ванева, В., Танева, К., Джалев, Л., Трайкова, В. (2011). *Тест за диагностика на готовността на децата за училище*. С.: Университетско издателство „Св. Климент Охридски“ (34 стр.)

5. Джалев, Л., Павлова, А. (2009). Експериментална програма “Типови игрови технологии по правата на детето”. *Дом, дете, детска градина*, бр. 5, сс. 14-35

6. Джалев, Л., Павлова, А. (2009). Изследване на ефективността на експерименталната програма “Типови игрови технологии по правата на детето”. *Дом, дете, детска градина*, бр. 6, сс. 19-41