

На правах рукописи

Елисеев Иван Николаевич

**МАТЕМАТИЧЕСКИЕ МОДЕЛИ И КОМПЛЕКСЫ ПРОГРАММ ДЛЯ
АВТОМАТИЗИРОВАННОЙ ОЦЕНКИ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ С
ИСПОЛЬЗОВАНИЕМ ЛАТЕНТНЫХ ПЕРЕМЕННЫХ**

Специальность 05.13.18 – Математическое моделирование, численные методы и
комплексы программ

Автореферат
диссертации на соискание ученой степени
доктора технических наук

Новочеркасск - 2013

Работа выполнена на кафедрах «Прикладная математика» федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова» и «Энергетика и безопасность жизнедеятельности» Института сферы обслуживания и предпринимательства (филиала) федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Донской государственный технический университет».

Научный консультант: доктор технических наук, профессор
Ткачёв Александр Николаевич

Официальные оппоненты: **Кравец Олег Яковлевич**,
доктор технических наук, профессор
ФГБОУ ВПО «Воронежский государственный технический университет», профессор кафедры автоматизированных и вычислительных систем
Маслак Анатолий Андреевич,
доктор технических наук, профессор филиала
ФГБОУ ВПО «Кубанский государственный университет» в г. Славянске-на-Кубани, профессор кафедры «Математика, информатика и методика их преподавания»
Наводнов Владимир Григорьевич,
доктор технических наук, профессор ФГБОУ ВПО
«Поволжский государственный технологический университет», заведующий кафедрой прикладной математики и информационных технологий

Ведущая организация: федеральное государственное автономное образовательное учреждение высшего профессионального образования «**Южный федеральный университет**»

Защита состоится 14 февраля 2014 года в 10⁰⁰ часов на заседании диссертационного совета Д 212.304.02 в ФГБОУ ВПО «Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова» по адресу: 346428, г. Новочеркасск, Ростовской обл., ул. Просвещения, 132 (главный корпус, ауд. 149).

С диссертацией можно ознакомиться в научно-технической библиотеке ЮРГПУ (НПИ). Автореферат диссертации размещён на официальном сайте ВАК vak.ed.gov.ru и на сайте www.npi-tu.ru.

Автореферат разослан «__» _____ 201__ г.

Ученый секретарь диссертационного совета
кандидат технических наук, профессор

А.Н. Иванченко

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. В современных условиях быстрого обновления техники и технологий, увеличения риска возникновения техногенных катастроф возросла роль человеческого фактора во многих сферах деятельности, и как следствие, возросли требования к надёжности оценивания уровня профессиональной подготовки кадров. Особенно это касается оборонных предприятий и взрывоопасных производств, железнодорожного, водного и воздушного транспорта, учреждений и подразделений Министерства по чрезвычайным ситуациям, Министерств обороны и внутренних дел. Уровень профессиональной подготовки специалистов и персонала служб можно оценить, прежде всего, на основе результатов их обучения в системе непрерывного образования: в ссузе, вузе, по программам повышения квалификации и переподготовки, в учебных центрах различных министерств и ведомств. Неадекватная оценка профессиональной подготовки специалиста может приводить к ошибкам в управлении персоналом, при принятии управленческих решений, большому материальному убытку и даже к человеческим жертвам. В связи с этим особую актуальность приобретает решение проблем снижения погрешности автоматизированной оценки результатов обучения, которая обеспечивается автоматизированной системой независимой оценки (АСНО).

Эффективная работа АСНО невозможна без использования математического моделирования и надёжных инструментальных средств диагностики результатов обучения. Их состав и предъявляемые требования во многом определяются используемыми подходами к оценке качества подготовки обучаемых. Специалистам и персоналу различных служб часто приходится работать в условиях быстро изменяющейся обстановки, требующей умения оперативно менять способы действия в незнакомой ситуации и находить наилучший. Поэтому для обучения и переподготовки кадров преимущественно используется компетентностный подход, согласно которому о результатах обучения судят по наличию у обучаемых компетенций, соответствующих их профессиональной деятельности. В связи с этим актуальным становится решение проблемы формализации процедуры оценивания уровня сформированности компетенций слушателей и студентов, теоретического обоснования и реализации используемых для этого моделей и инструментария, разработки программно-алгоритмических средств обработки результатов диагностики обучения (входных данных (ВД)), а также оценки качества диагностических средств (ДС) и текущего уровня подготовки обучаемых. Современные подходы к оценке учебных достижений слушателей и студентов базируются на использовании классической теории тестирования и теории латентных переменных, математический аппарат и основы которых были созданы известными зарубежными и отечественными учёными. Под термином «Латентная переменная (параметр)» принято понимать теоретический конструкт, который характеризует некое скрытое свойство или качество (например, уровень подготовки обучаемого, трудность тестового задания), которые непосредственно измерены быть не могут. Достоинствами классической теории тестирования являются обеспечение получения информации о показателях качества ДС, наглядность выполняемых расчётов и про-

стоя интерпретация данных обработки. Основной недостаток – зависимость результатов оценивания параметров обучаемых от трудности заданий ДС. Применение теории латентных переменных, базирующейся на моделях Раша, обеспечивает возможность независимости оценки вычисляемых значений латентного параметра «уровень подготовки» обучаемых θ_i от значений «трудности заданий» β_j ДС. Это способствует повышению объективности получаемых оценок уровня подготовки обучаемых. По этой причине в диссертационной работе сделан акцент на применение именно теории латентных переменных.

Создание математического аппарата классической теории тестирования и теории латентных переменных обеспечило разработку программно-алгоритмических средств обработки результатов диагностики, оценки качества ДС и уровня подготовки обучаемых. Зарубежными учёными E.B. Andersen, D. Andrich, A. Birnbaum, L.L. Guttman, J. M. Linacre, F.M. Lord, G.N. Masters, B.D. Wright и учёными России (В.С. Аванесов, А.А. Маслак, В.Г. Наводнов, М.Б. Чельщикова и др.) были созданы программные средства, реализующие возможности и теории латентных переменных, и классической теории тестирования. С их помощью можно решить многие задачи, связанные с оценкой результатов обучения. Вместе с тем, практика работы с существующим программным обеспечением (ПО) показала, что ему присущи достаточно серьёзные недостатки, связанные с наличием нерешённых проблем в самой теории латентных переменных. В частности, не обоснованы условия существования и единственности получаемых оценок латентных параметров моделей Раша, их сходимость и качество, например, состоятельность. Без решения этих проблем нельзя считать оценки латентных параметров надёжными, а полученные с их помощью оценки результатов обучения объективными.

Одной из важных для отечественного образования и не решённых в рамках теории латентных переменных задач является моделирование и параметризация диагностических тестов, в частности калибровка (определение трудности) тестовых заданий по выборкам ограниченного объёма (40-50 человек). Из-за малого объёма выборки погрешности и интервалы оценивания латентного параметра «трудность задания», рассчитываемые на основе стандартных методик параметрической статистики, могут в несколько раз превышать значения самих оценок. Одним из возможных способов решения задачи является применение методов имитационного моделирования. Однако реализация данного подхода затруднена из-за отсутствия теоретико-методологических основ создания моделей нормативных дихотомических матриц ответов заданного размера, позволяющих снизить влияние случайных факторов на параметризацию диагностических средств. Второй возможный способ решения задачи заключается в использовании непараметрических методов математической статистики, но их применение требует теоретического обоснования и экспериментальной проверки.

К перечню нерешённых задач в области разработки теоретико-методологической базы вычисления оценок латентных параметров модели Раша следует отнести и проблему математического обоснования их вычислительных процедур, хотя сами итерационные выражения для расчёта оценок в литературных источниках приводятся. Алгоритмы нахождения оценок латентных парамет-

ров на основе упомянутых процедур являются, как правило, коммерческой тайной и недоступны. По этой причине разработка отечественных программных средств невозможна без построения итерационных вычислительных алгоритмов, позволяющих создать программное обеспечение, которое обеспечивает обработку новых результатов обучения и не уступает по своим функциональным возможностям лучшим зарубежным аналогам.

В соответствии с изложенным проблема исследования определяется необходимостью устранения противоречий между требованиями(ем):

1. Автоматизации, независимости и низкой погрешности оценивания уровня подготовки обучаемых по выборкам ограниченного объёма и отсутствием теоретических основ создания необходимых для этого математических моделей, программно-алгоритмических и методических средств. Оценивания уровня сформированности общекультурных и профессиональных компетенций (ОК и ПК) обучаемых и отсутствием теоретической и методологической базы для построения моделей АСНО новых результатов обучения - компетенций;

2. Наличия качественных средств диагностики уровня сформированности компетенций и отсутствием теоретической и методической базы, обеспечивающей повышение точности расчёта характеристик и показателей качества диагностических средств по выборкам ограниченного объёма.

Актуальность диссертационного исследования обусловлена сформулированными неотложными потребностями обучения и отсутствием реальных возможностей их обеспечения.

Работа выполнена в рамках гранта по аналитической ведомственной целевой программе (АВЦП) «Развитие научного потенциала высшей школы (2009-2011 г.г.)» (РН 3.4.1/3224 и 3.4.1/10601) Минобрнауки РФ; утверждённых научных направлений «Научно-методическое обеспечение инновационного развития высшей школы» ФГБОУ ВПО «ЮРГТУ (НПИ)» и «Методы, технические и программные средства измерения латентных переменных в области образования, социальных и экономических системах» ФГБОУ ВПО «ЮРГУЭС»; ЕЗН Минобрнауки РФ «Разработка методов и программных средств для расчёта латентных переменных по экспериментальным выборкам малого объёма» (ЮРГУЭС-4.08Ф, № ГР 01.200.802798).

Целью диссертационной работы является обеспечение автоматизированной независимой оценки результатов компетентностно-ориентированного обучения на основе развития теории латентных переменных, разработки и обоснования математических моделей, создания программно-инструментальных средств.

Для достижения указанной цели были поставлены следующие **задачи исследования**:

- теоретически обосновать процедуры расчёта латентных параметров обучаемых и средств диагностики, а также разработать необходимые для этого алгоритмы и комплексы программ;

- обосновать условия существования и единственности оценок латентных параметров обучаемых и средств диагностики, исследовать состоятельность оценок, рассчитанных по дихотомической матрице ответов ограниченного размера;

- разработать методы, обеспечивающие снижение погрешности расчёта оценок латентных параметров обучаемых и средств диагностики на основе компьютерного моделирования матрицы ответов ограниченного размера;
- экспериментально и теоретически исследовать применимость непараметрических методов математической статистики для повышения точности калибровки заданий диагностических средств по дихотомическим матрицам ответов ограниченного размера; разработать необходимые для этого алгоритмы и программное обеспечение;
- разработать модель дихотомической матрицы ответов, применение которой позволит снизить влияние искажающих факторов на результаты исследования латентных параметров и характеристик диагностических средств;
- разработать методологические основы оценки уровня сформированности компетенций обучаемых и модель мониторинга, позволяющую объективно отслеживать результаты обучения и показатели качества диагностических средств;
- разработать модель системы оценки уровня подготовки выпускников учреждений по обучению, повышению квалификации и переподготовке кадров, обеспечивающую объективную оценку их профессиональной подготовки и соответствие её требованиям стандарта;
- провести вычислительные эксперименты, подтверждающие эффективность новых вычислительных алгоритмов и справедливость полученных теоретических результатов.

Объектом исследования является процесс независимой оценки результатов обучения при подготовке, повышении квалификации и переподготовке кадров в учреждениях различных министерств и ведомств.

Предмет исследования: теоретические основы, математические модели, вычислительные алгоритмы, комплексы программ систем автоматизированной оценки результатов обучения.

Методы исследования. Теоретическую основу исследования составляют: теория латентных переменных, теория вероятностей и математической статистики. Для решения поставленных задач применялись методы: вычислительной математики, имитационного моделирования, математического анализа, теории множеств и матричной алгебры, теоретического и экспериментального исследования.

На защиту выносятся:

- обоснование условий существования и единственности оценок максимального правдоподобия латентных параметров основной модели Раша и разработанный на их основе алгоритм проверки дихотомической матрицы ответов на соответствие этим условиям;
- теоретическое и экспериментальное обоснование состоятельности оценок максимального правдоподобия латентных параметров основной модели Раша, рассчитанных по допустимым квазинормативным дихотомическим матрицам ответов ограниченного размера;
- метод расчета квазигенеральных оценок латентных параметров «уровень подготовки» обучаемого и «трудность задания» ДС путём моделирования допустимой нормативной дихотомической матрицы ответов с ограниченным числом

строк и столбцов нормативными матрицами с их существенно большим количеством;

- метод калибровки заданий теста путём моделирования допустимой квазинормативной дихотомической матрицы ответов ограниченного размера аналогичными матрицами с существенно большим количеством строк и с равными значениями столбцовых сумм, позволяющий получить квазигенеральные оценки трудности заданий;

- модель нормативной дихотомической матрицы ответов, методика и алгоритм её формирования;

- модель мониторинга уровня компетенций обучаемых и оценки качества средств диагностики;

- модель системы независимой оценки уровня подготовки выпускников учреждений, осуществляющих обучение, повышение квалификации и переподготовку кадров, по уровню сформированности их компетенций.

- алгоритмы и комплексы программ для расчёта параметров и характеристик обучаемых и средств диагностики.

Научная новизна исследования:

1. Впервые получены условия допустимости, которым должны удовлетворять дихотомические матрицы результатов тестирования, используемые для расчёта оценок максимального правдоподобия латентных параметров основной модели Раша, выполнение которых обеспечивает существование и единственность оценок. Предложены алгоритм проверки дихотомической матрицы ответов на соответствие её условиям допустимости и алгоритм расчёта доли таких матриц ответов заданного размера. Показано, что доля допустимых дихотомических матриц ответов зависит от соотношения числа их строк и столбцов;

2. Впервые показано, что оценки максимального правдоподобия латентных параметров основной модели Раша, рассчитанные по квазинормативным допустимым матрицам ответов ограниченного размера ($N, L \geq 50$) являются состоятельными. Установлено, что при одновременном увеличении числа и строк, и столбцов матрицы ответов обеспечивается состоятельность оценок параметров и «уровень подготовки» обучаемого, и «уровень трудности» задания. Если увеличивается только число строк матрицы, а количество столбцов фиксировано и все столбцовые суммы одинаковы, состоятельны только оценки параметра «уровень трудности» задания, а оценки параметра «уровень подготовки» обучаемого, стремятся по вероятности к первоначальным значениям. Установлены минимально достаточные размеры нормативной допустимой матрицы ответов, по которой могут быть рассчитаны генеральные оценки латентных параметров θ_i и β_j ;

3. Предложена новая модель нормативной дихотомической матрицы ответов, методика и алгоритм её формирования по заданным законам распределения латентных параметров θ_i , β_j , отличающаяся от экспериментальных матриц ответов высокой адекватностью модели Раша и высокими показателями качества соответствующего ей виртуального диагностического средства. Предложена методика экспертизы качества диагностических средств по нормативным дихотомиче-

ским матрицам ответов ограниченного размера, отличающаяся от известных меньшей погрешностью калибровки заданий этих средств;

4. Теоретически и экспериментально обоснован новый метод калибровки заданий теста путём моделирования допустимой квазинормативной дихотомической матрицы ответов ограниченного размера $N_H \times L_H$ аналогичными матрицами с существенно большим числом строк и с одинаковыми значениями столбцовых сумм, позволяющий получить квазигенеральные оценки трудности заданий;

5. Предложен новый метод расчета квазигенеральных оценок латентных параметров θ_{i0}^* и β_{j0}^* путём моделирования допустимой нормативной дихотомической матрицы ответов ограниченного размера $N_H \times L_H$ нормативными матрицами с существенно большим числом строк и столбцов;

6. На основе предложенной методологии оценки сформированности компетенции по значениям его личностного, когнитивного и интегративно-деятельностного компонентов впервые разработана модель мониторинга, позволяющая объективно отслеживать уровень сформированности компетенций обучаемых и проводить экспертизу качества диагностических средств. Разработана новая модель АСНО уровня подготовки выпускников учреждений по обучению, повышению квалификации и переподготовке кадров, обеспечивающая объективную оценку соответствия его требованиям соответствующих стандартов;

7. Предложены новые вычислительные алгоритмы для обработки результатов диагностики обучения и проверки их качества, отличающиеся от известных надёжной сходимостью итерационных оценок к оцениваемому параметру и позволяющие создать ПО для объективной формализованной оценки параметров и характеристик обучаемых и средств диагностики.

Теоретическая значимость:

- полученные в диссертационной работе научные результаты по исследованию существования и единственности оценок максимального правдоподобия латентных параметров основной модели Раша, состоятельности этих оценок, рассчитанных по допустимой квазинормативной матрице ответов ограниченного размера, вносят вклад в дальнейшее развитие теории латентных переменных;

- получили своё дальнейшее развитие и совершенствование теория построения программно-алгоритмических средств обработки результатов диагностики обучения (входных данных) и оценки качества диагностических средств за счёт математического обоснования и реализации новых алгоритмов расчёта оценок латентных параметров моделей Раша, разработки компьютерных моделей нормативных дихотомических матриц ответов, методики и алгоритма их формирования, создания новых компьютерных алгоритмов и программного обеспечения, реализованных в программных комплексах RILP-1M, RILP-2, RILP-multi;

- созданы теоретико-методологические основы мониторинга результатов компетентностно-ориентированного обучения и оценки качества диагностических средств по матрицам ответов ограниченного размера, моделирования автоматизированных систем независимой оценки уровня подготовки выпускников учреждений, осуществляющих обучение, повышение квалификации и переподготовку кадров в рамках реализации компетентностного подхода.

Практическая значимость результатов работы:

1. Разработаны программные комплексы RILP-1М и RILP-2 который обеспечивает обработку результатов диагностики обучения, представленных квазинормативной матрицей ответов, надёжную оценку качества диагностических материалов и получение независимых оценок уровня подготовки обучаемых;

2. Использование методов калибровки тестовых заданий и расчёта квазигенеральных оценок латентных параметров путём моделирования допустимой квазинормативной матрицы ответов ограниченного размера позволяет с доверительной вероятностью 0,95 существенно снизить погрешности расчёта параметров (не менее чем в 2 раза при выборочной дисперсии уровня подготовки обучаемых, не превышающей 0,58 логит²);

3. Замена реальной допустимой нормативной дихотомической матрицы ответов её моделью с высокими показателями качества при параметризации средств диагностики обучения обеспечивает снижение влияния искажающих факторов на результаты параметризации, за счёт чего достигается повышение точности оценивания результатов обучения. С помощью моделирования нормативных дихотомических матриц ответов установлено, что погрешность калибровки тестовых заданий трудностью $|\beta_j| \geq 0,18$ логит не превышает 6 % при объёме выборки $N=200$ и 3 % при $N=300$;

4. Методологические основы оценки уровня сформированности компетенций обучаемых и методика экспертизы качества диагностических средств обеспечивают надёжное количественное оценивание результатов компетентностно-ориентированного обучения и эффективное управление его процессом;

5. Модель мониторинга уровня сформированности компетенций обучаемых позволяет объективно отслеживать уровень их подготовки в процессе обучения и получать информацию о качестве диагностических средств. Использование модели обеспечит создание в короткие сроки качественных средств диагностики, пригодных для проведения итоговой государственной аттестации выпускников;

6. Модель системы оценки уровня подготовки выпускников учреждений по обучению, повышению квалификации и переподготовке кадров, базирующаяся на методологии оценки компетенций обучаемых, позволяет получить объективную информацию о владении выпускников компетенциями и о соответствии их уровня требованиям соответствующих стандартов.

Все перечисленные в пп. 1-6 результаты позволяют обеспечить высокую надёжность независимой оценки уровня подготовки обучаемых.

Достоверность и обоснованность полученных научных результатов подтверждена соответствием фундаментальным положениям математического анализа, теории вероятностей и математической статистики; корректностью применения математического аппарата; строгим доказательством сформулированных утверждений; их обоснованной математической трактовкой и непротиворечивостью известным (опубликованным) данным; соответствием (с точностью до 2 %) ожидаемых данных эмпирическим; результатами вычислительных экспериментов; устойчивой повторяемостью результатов; представительным количеством использованных источников информации; сочетанием количественного и каче-

ственного анализа; положительной оценкой на международных и всероссийских конференциях, полученной в результате критического обсуждения. Выводы, полученные с помощью разработанных моделей и методов, находятся в логическом соответствии с основными положениями теории латентных переменных.

Реализация результатов работы. Результаты диссертационной работы внедрены и используются в следующих организациях:

- Государственное бюджетное учреждение Ростовской области «Ростовский областной центр обработки информации в сфере образования» - программный продукт RILP-1M, методика обработки результатов диагностики и расчёта характеристик и латентных параметров обучаемых и средств диагностики;

- ФГБОУ ВПО «Донской государственный технический университет» - программный комплекс RILP-1M, теоретическая модель мониторинга сформированности компетенций обучаемых и оценки качества диагностических средств, метод расчета уровня сформированности компетенций обучаемого и методология их независимой оценки, методика экспертизы качества диагностических средств;

- ФГБОУ ВПО «Ростовский государственный строительный университет» - метод и методология независимой оценки уровня сформированности общекультурных компетенций студентов, модель мониторинга сформированности компетенций обучаемых и оценки качества диагностических средств, программный комплекс RILP-1M;

- ФГБОУ ВПО «Южно-Российский государственный университет экономики и сервиса» - программные комплексы RILP-1M, RILP- multi, программы для построения психограмм и моделирования дихотомических матриц ответов, методы и методики расчёта характеристик и параметров обучаемых и средств диагностики процесса обучения, методология оценки уровня сформированности компетенций и модель его мониторинга;

- ОАО «Системный оператор единой энергетической системы» - программный комплекс RILP-1M, теоретическая модель мониторинга независимой оценки уровня сформированности компетенций обучаемых, методика экспертизы качества диагностических средств по дихотомическим матрицам ответов ограниченного размера;

- ФГОУ «Академия дополнительного профессионального образования «Учебный центр подготовки руководителей»» - методология оценки уровня сформированности компетенций студентов (использована в докладе автора диссертации на семинаре проректоров по учебной работе вузов России и положительно оценена участниками семинара (отзыв от 29.10.2011)).

Внедрение результатов работы подтверждено соответствующими актами.

Результаты диссертационного исследования использованы в учебных пособиях с грифом УМО «Методологические основы разработки и оценки качества педагогических измерительных материалов (на примере дисциплины «Основы теории цепей»))» и «Теория линейных электрических цепей в тестах» для обучения студентов ВПО направления подготовки 210300 «Радиотехника» по указанной дисциплине.

Основные результаты исследования получены при выполнении НИР «Разработка оценочных и диагностических средств для независимой оценки качества

образования в многоуровневых университетских комплексах» в рамках АВЦП «Развитие научного потенциала высшей школы (2009-2011 г.г.)» (РН 3.4.1/3224 и 3.4.1/10601) Минобрнауки РФ и НИР «Разработка методов и программных средств для расчёта латентных переменных по экспериментальным выборкам малого объёма» по ЕЗН Минобрнауки РФ (ЮРГУЭС-4.08Ф, № ГР 01.200.802798). НИР выполнены в ФГБОУ ВПО «ЮРГУЭС» под научным руководством автора настоящего диссертационного исследования.

Программное обеспечение для ЭВМ, реализующее новые методы и алгоритмы расчёта характеристик ДС и параметров обучаемых, зарегистрированы в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам (РОСПАТЕНТ).

Апробация. Основные положения и научные результаты исследований докладывались и получили одобрение на: ежегодных Международных форумах и конференциях «Современное образование: содержание, технологии, качество» (г. Санкт-Петербург, СПбГЭТУ «ЛЭТИ», 2010-2013); ежегодных Международных научно-методических конференциях НИТЭ (Астрахан. гос. техн. ун-т, 1998, 2000); Международных и Всероссийских научных конференциях в г. Таганроге «Системы и модели в информационном мире», «Методы и алгоритмы принятия эффективных решений» и др. (ТТИ ЮФУ, 2008-2011); Международной научно-практической конференции «Информатизация образования-2011» в г. Ельце (гос. ун-т им. И.А. Бунина, 2011); Международном семинаре «Физико-математическое моделирование систем» (г. Воронеж, 2011); Всероссийской и Международной научно-методической конференции «Пути совершенствования подготовки специалистов для текстильной промышленности» в МГТУ им. А.Н. Косыгина (1995, 2002); ежегодных Всероссийских научно-методических конференциях «Развитие тестовых технологий в России» (г. Москва, ФЦТ, 1999-2007); Всероссийских научно-методических семинарах «Автоматизированные системы управления учебным процессом в ВУЗе: опыт, проблемы, возможности» (г. Шахты, 2003, 2008); Всероссийской научно-методической конференции «Фундаментализация высшего технического образования» (г. Новочеркасск, НГТУ, 2000); ежегодных Всероссийских научно-практических конференциях в г. Славянске-на-Кубани «Теория и практика измерения латентных переменных в образовании и других социально-экономических системах» (2007-2010); ежегодных Всероссийских научно-практических конференциях в г. Красноярске «Тестирование в сфере образования: проблемы и перспективы развития» (СГТУ, 2009, 2010, 2013); Всероссийской научно-технической конференции «Приоритетные направления развития науки и технологий» (г. Тула, 2011).

Публикации. Результаты диссертации изложены в монографии, 52 печатных работах, в том числе в 23 статьях, опубликованных в ведущих рецензируемых научных журналах, рекомендованных ВАК, в 11 свидетельствах о регистрации программ для ЭВМ, в 18 статьях сборников материалов Международных и Всероссийских конференций, симпозиумов и семинаров.

Структура работы. Диссертация состоит из введения, шести глав, заключения, списка использованных источников из 261 наименования и двенадцати

приложений. В работе содержится 54 таблицы и 64 рисунка. Общий объем работы – 371 страница.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность исследования, сформулирована цель работы. Определены научная новизна, теоретическая и практическая значимость и основные положения, выносимые на защиту, проведён краткий анализ вопросов, подлежащих исследованию. Приводятся сведения об апробации и реализации результатов работы.

В первой главе «Теоретическое обоснование выбора математических моделей для оценки результатов обучения и постановка задач исследования» на основе анализа литературных источников выбраны подходы к оценке уровня подготовки обучаемых, сформулированы критерии и показатели качества ДС, теоретически обоснована возможность использования моделей Раша для оценки результатов обучения. Показано, что для снижения погрешности расчёта оценок уровня подготовки обучаемых необходимо повышать точность калибровки заданий средств диагностики и увеличивать их количество.

Во второй главе «Разработка алгоритмов и комплексов программ для расчёта параметров, характеризующих результаты обучения и качество диагностических средств» на базе вероятностных моделей Раша получены рекуррентные соотношения и разработаны теоретические основы построения алгоритмов и комплексов программ для расчёта оценок латентных параметров обучаемых и ДС. Искользованные допущения и основания: - латентные параметры «уровень подготовки» обучаемого θ_n и «уровень трудности» задания β_j (латентные параметры основной модели Раша) являются независимыми (в соответствии с аксиомой о локальной независимости П.Ф. Лазарсфельда);- расчёт оценок параметров осуществляется по матрице ответов (результатов тестирования) X ; - в качестве метода, на котором базируется исследование, используется метод максимального правдоподобия (ММП).

Вероятность P_{nj} принятия элементом матрицы - случайной величиной x_{nj} значения 1 в дихотомической модели Раша полностью определяется значениями θ_n, β_j . Поэтому для любой матрицы ответов X события, состоящие в том, что результат выполнения j -го задания n -м студентом окажется равным x_{nj} , независимы в совокупности. Вероятность совпадения элемента x_{nj} дихотомической матрицы ответов с соответствующим элементом матрицы вероятностей, рассчитанной по дихотомической модели Раша, определится выражением

$$P_{nj}^{\circ} = p\{x_{nj} | (\theta_n, \beta_j)\} = \frac{\exp(x_{nj}(\theta_n - \beta_j))}{1 + \exp(\theta_n - \beta_j)}.$$

В силу теоремы умножения вероятностей для независимых событий функция правдоподобия L для N тестируемых и L заданий будет равна произведению вероятностей P_{nj} , а логарифмическая функция правдоподобия λ будет иметь вид:

$$\lambda \equiv \ln \Lambda = \sum_{n=1}^N X_n \theta_n - \sum_{j=1}^L Y_j \beta_j - \sum_{n=1}^N \sum_{j=1}^L \ln[1 + \exp(\theta_n - \beta_j)], \quad (1)$$

где $\sum_{j=1}^L x_{nj} = X_n$ – первичный балл n -го участника тестирования; $\sum_{n=1}^N x_{nj} = Y_j$ – первичный балл j -го задания.

Используя необходимое условие экстремума функции нескольких переменных, на основании формулы (1) получены выражения (2), позволяющие рассчитать значения θ_n и β_j , соответствующие максимальному значению вероятности логарифмической функции λ :

$$\begin{cases} \frac{\partial \lambda}{\partial \theta_n} = \varphi_{1n}(\theta_n, \beta_j) = X_n - \sum_{j=1}^L \frac{\exp(\theta_n - \beta_j)}{1 + \exp(\theta_n - \beta_j)} = X_n - \sum_{j=1}^L P_{nj} = 0, n = \overline{1, N}, \\ \frac{\partial \lambda}{\partial \beta_j} = \varphi_{2j}(\theta_n, \beta_j) = -Y_j + \sum_{n=1}^N \frac{\exp(\theta_n - \beta_j)}{1 + \exp(\theta_n - \beta_j)} = -Y_j + \sum_{n=1}^N P_{nj} = 0, j = \overline{1, L}. \end{cases} \quad (2)$$

Сумма $\sum_{j=1}^{L_i} P_{nj}$ в первом уравнении – это ожидаемый балл $X_n^\diamond$, который получит тестируемый n за успешно выполненные задания. Сумма $\sum_{j=1}^N P_{nj}$ во втором равенстве представляет собой ожидаемый балл $Y_j^\diamond$ j -го задания.

Решая полученную нелинейную систему уравнений (2) методом секущих, получим рекуррентные выражения для расчёта латентных параметров θ_n , β_j^* :

$$\theta_n^{(1)} = \theta_n^{(0)} - \frac{(\theta_n^{(0)} - b_n) \cdot (X_n - \sum_{j=1}^L \frac{\exp(\theta_n^{(0)} - \beta_j^{(0)})}{1 + \exp(\theta_n^{(0)} - \beta_j^{(0)})})}{\sum_{j=1}^L \frac{\exp(\theta_n^{(0)} - \beta_j^{(0)})}{1 + \exp(\theta_n^{(0)} - \beta_j^{(0)})} - \sum_{j=1}^L \frac{\exp(b_n - \beta_j^{(0)})}{1 + \exp(b_n - \beta_j^{(0)})}}, \quad n = \overline{1, N} \quad (3)$$

$$\beta_j^{(1)} = \beta_j^{(0)} - \frac{(\beta_j^{(0)} - c_j) \cdot (-Y_j + \sum_{n=1}^N \frac{\exp(\theta_n^{(1)} - \beta_j^{(0)})}{1 + \exp(\theta_n^{(1)} - \beta_j^{(0)})})}{\sum_{n=1}^N \frac{\exp(\theta_n^{(1)} - \beta_j^{(0)})}{1 + \exp(\theta_n^{(1)} - \beta_j^{(0)})} - \sum_{n=1}^N \frac{\exp(\theta_n^{(1)} - c_j)}{1 + \exp(\theta_n^{(1)} - c_j)}}, \quad j = \overline{1, L} \quad (4)$$

для первых итераций и

$$\begin{aligned} \theta_n^{(t+1)} &= \theta_n^{(t)} - \frac{(\theta_n^{(t)} - \theta_n^{(t-1)}) \cdot (X_n - \sum_{j=1}^L \frac{\exp(\theta_n^{(t)} - \beta_j^{(t)})}{1 + \exp(\theta_n^{(t)} - \beta_j^{(t)})})}{\sum_{j=1}^L \frac{\exp(\theta_n^{(t)} - \beta_j^{(t)})}{1 + \exp(\theta_n^{(t)} - \beta_j^{(t)})} - \sum_{j=1}^L \frac{\exp(\theta_n^{(t-1)} - \beta_j^{(t-1)})}{1 + \exp(\theta_n^{(t-1)} - \beta_j^{(t-1)})}}, \quad n = \overline{1, N}, \\ \beta_j^{(t+1)} &= \beta_j^{(t)} - \frac{(\beta_j^{(t)} - \beta_j^{(t-1)}) \cdot (-Y_j + \sum_{n=1}^N \frac{\exp(\theta_n^{(t)} - \beta_j^{(t)})}{1 + \exp(\theta_n^{(t)} - \beta_j^{(t)})})}{\sum_{n=1}^N \frac{\exp(\theta_n^{(t)} - \beta_j^{(t)})}{1 + \exp(\theta_n^{(t)} - \beta_j^{(t)})} - \sum_{n=1}^N \frac{\exp(\theta_n^{(t-1)} - \beta_j^{(t-1)})}{1 + \exp(\theta_n^{(t-1)} - \beta_j^{(t-1)})}}, \quad j = \overline{1, L} \end{aligned}$$

для последующих итераций ($t \geq 1$). В выражениях (3), (4) $\theta_n^{(0)}$, b_n и $\beta_j^{(0)}$, c_j - начальные границы интервалов, внутри которых находятся корни уравнений.

Значения $\theta_n^{(0)}$ и $\beta_j^{(0)}$ рассчитываются на основе выражений $\theta_n^{(0)} = \ln(p_n \cdot (1 - p_n)^{-1})$, $\beta_j^{(0)} = \ln((1 - p_j) \cdot p_j^{-1})$ ($p_n = X_n \cdot L^{-1}$, $p_j = Y_j \cdot N^{-1}$). В качестве вторых начальных границ интервалов принимаются значения $b_n = \theta_n^{(0)} \pm \Delta\theta_n$ и $c_j = \beta_j^{(0)} \pm \Delta\beta_j$. Величины $\Delta\theta_n$ и $\Delta\beta_j$ выбираются минимальными так, чтобы $\text{sign}(\varphi_{1n}(b_n, \beta_j^{(0)})) \neq \text{sign}(\varphi_{1n}(\theta_n^{(0)}, \beta_j^{(0)}))$ и $\text{sign}(\varphi_{2j}(\theta_n^{(0)}, c_j)) \neq \text{sign}(\varphi_{2j}(\theta_n^{(0)}, \beta_j^{(0)}))$. Если $p_n = 1$, $p_j = 1$ или $p_n = 0$, $p_j = 0$, то профиль ответов n -го участника тестирования или j -го задания из обработки исключаются. Теоретическое обоснование алгоритмов расчёта оценок латентных параметров политомических моделей Раша для рейтинговой шкалы и частичного доверия проводится аналогично. В заключении второй главы изложены результаты разработки программных комплексов RILP- 1M и RILP- 2, реализующих полученные рекуррентные соотношения и предложенные алгоритмы, оценена достоверность полученных с их помощью результатов расчёта.

В третьей главе «Теоретическое исследование свойств оценок латентных параметров дихотомической модели Раша» теоретически и с помощью вычислительного эксперимента исследуются условия существования и единственности оценок МП латентных параметров основной модели Раша, вопросы их состоятельности при расчёте по входным данным ограниченного объёма, разрабатываются необходимые для этого алгоритмы и ПО. Для установления условий существования и единственности оценок МП латентных параметров основной модели Раша было введено новое понятие «допустимая дихотомическая матрица ответов (результатов тестирования)» X . Матрица $X \in M_{NL}$ (M_{NL} – множество всех матриц размера $N \times L$, $J_L = \{1, \dots, L\}$, $L = 1, 2, \dots$) является *допустимой*, если матрица Z , полученная из X путём упорядочивания строк по невозрастанию строчных сумм, обладает следующими свойствами: а) для каждого $j \in J_L$ найдутся номера $i, l \in J_N$, для которых $z_{ij} = 0$, $z_{lj} = 1$; б) для любого $k \in J_{N-1}$ имеются номера i, j, l такие, что $i \in J_k$, $j \in J_L$, $k < l \leq N$, причем $z_{ij} = 0$, $z_{lj} = 1$. Свойство а) означает отсутствие в матрицах Z и X столбцов, состоящих из одних нулей или из одних единиц (экстремальных столбцов). Второе требование сводится к наличию случайности в расположении элементов z_{nj} матрицы Z , связанной с ошибками в профилях ответов участников тестирования.

Исследование условий существования и единственности оценок МП латентных параметров выполнено исходя из принятых ранее допущений. При использовании основной модели Раша функция правдоподобия Λ для N тестируемых и L заданий будет равна произведению вероятностей P_{ij} :

$$\Lambda = \prod_{i=1}^N \prod_{j=1}^L \frac{\exp(x_{ij}(\theta_i - \beta_j))}{1 + \exp((\theta_i - \beta_j))}. \quad (5)$$

В силу инвариантности оценок латентных параметров модели Раша достижение функцией (5) своего наибольшего на R^{N+L} значения в некоторой точке $(\theta_1^0, \dots, \theta_N^0, \beta_1^0, \dots, \beta_L^0)$ означает, что это же значение она принимает и во всякой точке вида $(\theta_1^0 + \delta, \dots, \theta_N^0 + \delta, \beta_1^0 + \delta, \dots, \beta_L^0 + \delta)$, где δ - действительное число. Поэтому с целью обеспечения единственности точки экстремума функции Λ придадим одному из её $(N+L)$ аргументов постоянное значение, например, $\beta_L = 0$.

Показано, что необходимым и достаточным условием достижения функцией $\Lambda(\bar{\eta})$ ($\bar{\eta} = (\theta_1, \dots, \theta_N, \beta_1, \dots, \beta_{N+L-1})$) своего наибольшего значения на R^{N+L-1} является допустимость матрицы X .

Для доказательства использована логарифмическая функция максимального правдоподобия $f(\bar{\eta}) = -\ln \Lambda(\eta) = \sum_{i=1}^N \sum_{j=1}^L (\ln(1 + e^{\theta_i - \beta_j}) - x_{ij}(\theta_i - \beta_j))$. Единственной точке глобального максимума $\bar{\eta}_0 = (\theta_1^0, \dots, \theta_N^0, \beta_1^0, \dots, \beta_{L-1}^0)$ на R^{N+L-1} функции $\Lambda(\bar{\eta})$ соответствует точка глобального минимума функции $f(\bar{\eta})$. В $\bar{\eta}_0$ должно быть выполнено необходимое условие экстремума функции нескольких переменных. Для функции $f(\bar{\eta})$ оно сводится к системе уравнений

$$\begin{cases} \sum_{j=1}^L \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} = X_i, & i \in J_N \\ \sum_{i=1}^N \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} = Y_j, & j \in J_{L-1} \end{cases} \quad (6)$$

Выразим $f(\bar{\eta})$ через функцию $\varphi(\tau) = \ln(1 + e^\tau)$ при всех $\tau \in R$, учитывая, что $\varphi(\tau) - \tau = \varphi(-\tau)$:

$$f(\bar{\eta}) = \sum_{i=1}^N \sum_{j=1}^L \varphi((-1)^{x_{ij}}(\theta_i - \beta_j)). \quad (7)$$

Рассуждением от противного показано, что матрица Z удовлетворяет условиям допустимости а) и б). Нарушение условия а) возможно в двух случаях:

1). Найдётся номер $i \in J_N$ такой, что строчная сумма X_i совпадёт с одним из концов отрезка $[0; L]$, или номер $j \in J_{L-1}$, для которого столбцовая сумма Y_j совпадёт с одним из концов отрезка $[0; N]$. Однако из очевидных неравенств $0 < \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} < 1$ следует, что левые части уравнений (6) находятся внутри указанных отрезков. Таким образом, по крайней мере, одно из уравнений (6) не имеет решений; система несовместна, что противоречит выбору точки $\bar{\eta}_0$;

2). При любом $i \in J_N$ $0 < X_i < L$, а при любом $j \in J_{L-1}$ $0 < Y_j < N$. Тогда с необходимостью выполняется одно из равенств $Y_L = 0, Y_L = N$. Положив $\delta = (-1)^{x_{iL}}$, $\bar{\eta}^* = (\theta_1^0 - \delta, \dots, \theta_N^0 - \delta, \beta_1^0 - \delta, \dots, \beta_{L-1}^0 - \delta)$, $S = \sum_{i=1}^N \sum_{j=1}^{L-1} \varphi((-1)^{x_{ij}}(\theta_i^0 - \beta_j^0))$, приняв во внимание выражение (7) и возрастание функции $\varphi(\tau)$ на R , получим $f(\bar{\eta}^*) = S + \sum_{i=1}^N \varphi(\delta(\theta_i^0 - \delta)) =$

$$= S + \sum_{i=1}^N \varphi(\delta\theta_i^0 - 1) < S + \sum_{i=1}^N \varphi((-1)^{x_{il}} (\theta_i^0 - \beta_L^0)) = f(\bar{\eta}_0), \text{ что противоречит определению } \bar{\eta}_0$$

как точки глобального минимума функции $f(\bar{\eta})$. При доказательстве выполнения утверждения б) используется условие допустимости дихотомической матрицы в виде, сформулированном и доказанном в теореме: матрица $X \in M_{NL}$ допустима в том и только в том случае, если не существует множеств $Q_1 \subset J_N, S_1 \subset J_L$, удовлетворяющих следующим двум условиям: $Q_r \cup S_r \neq \emptyset, r = 1, 2$; для всех $r \in J_2, i \in Q_r, j \in S_{3-r}, x_{ij} = 2 - r$, где $Q_2 = J_N \setminus Q_1, S_2 = J_L \setminus S_1$). Пусть утверждение б) неверно. Тогда для некоторого $k \in J_{N-1}$ имеет место импликация

$$k < l \leq N, j \in J_L, \quad z_{lj} = 1 \Rightarrow z_{ij} = 1 \text{ при всех } i \in J_k. \quad (8)$$

Зададим множества Q_1 и S_2 равенствами $Q_1 = \{p_i : i \in J_k\}, S_2 = \bigcup_{l=k+1}^N \{j \in J_L : z_{lj} = 1\}$ и по-

ложим $\theta_i^{(1)} = \begin{cases} \theta_i^0 + 1, & i \in Q_1 \\ \theta_i^0, & i \in Q_2 \end{cases}, \beta_j^{(1)} = \begin{cases} \beta_j^0 + 1, & j \in S_1 \\ \beta_j^0, & j \in S_2 \end{cases}$. Пусть $L \in S_2$ (в этом случае $\beta_L^{(1)} = 0$)

и $\bar{\eta}_1 = (\theta_1^{(1)}, \dots, \theta_N^{(1)}, \beta_1^{(1)}, \dots, \beta_{L-1}^{(1)})$ - точка глобального экстремума. Тогда в результате преобразований с учётом импликации (8) получается неравенство $f(\bar{\eta}_1) < \sum_{i=1}^N \sum_{j=1}^L \varphi((-1)^{x_{ij}} \times (\theta_i^0 - \beta_j^0)) = f(\bar{\eta}_0)$, которое противоречит определению точки $\bar{\eta}_0$.

Заметив, что $\beta_L^{(2)} = 0$, а $f(\bar{\eta}_2) < f(\bar{\eta}_0)$, к этому же противоречию придём и в случае $L \in S_1$ и $\theta_i^{(2)} = \theta_i^{(1)} - 1, i \in J_N; \beta_j^{(2)} = \beta_j^{(1)} - 1, j \in J_L; \bar{\eta}_2 = (\theta_1^{(2)}, \dots, \theta_N^{(2)}, \beta_1^{(2)}, \dots, \beta_{L-1}^{(2)})$.

Показано, что $\bar{\eta}_0$ - единственная точка глобального максимума функции $\Lambda(\bar{\eta})$ для допустимой матрицы X . Как и прежде, вместо $\Lambda(\bar{\eta})$ рассматривалась функция $f(\bar{\eta}) = -\ln \Lambda(\bar{\eta})$. Выражение (7) записывалось в виде $f(\bar{\eta}) = \sum_{i=1}^N f_i(\bar{\eta}_i)$, где

$\bar{\eta}_i = (\theta_i, \beta_1, \dots, \beta_{L-1}), f_i(\bar{\eta}_i) = \sum_{j=1}^L \varphi((-1)^{x_{ij}} (\theta_i - \beta_j))$, и доказывалась строгая выпуклость

функций $f_i(\bar{\eta}_i)$ на R^L . С этой целью исследовались их вторые дифференциалы $d^2 f_i(\bar{\eta}_i)$. Введём обозначение $t_{ij} = \varphi''(\theta_i - \beta_j), i \in J_N, j \in J_L$ и заметим, что ввиду чётности функции $\varphi''(\tau) = (\ln(1 + e^\tau))''$ независимо от значения x_{ij} $\varphi''((-1)^{x_{ij}} (\theta_i - \beta_j)) = t_{ij}$.

При всех $i \in J_N, j \in J_{L-1}$ имеем $\frac{\partial^2 f_i}{\partial \theta_i^2} = \sum_{j=1}^L (-1)^{2x_{ij}} \varphi''((-1)^{x_{ij}} \times (\theta_i - \beta_j)) = \sum_{j=1}^L t_{ij}; \frac{\partial^2 f_i}{\partial \beta_j \partial \beta_l} = 0$ в

случае $l \in J_{L-1}, l \neq j; \frac{\partial^2 f_i}{\partial \beta_j^2} = (-1)^{2(x_{ij}+1)} \varphi''((-1)^{x_{ij}} \cdot (\theta_i - \beta_j)) = t_{ij}; \frac{\partial^2 f_i}{\partial \theta_i \partial \beta_j} = (-1)^{2x_{ij}+1} \varphi''((-1)^{x_{ij}} \times$

$\times (\theta_i - \beta_j)) = -t_{ij}$. Используя полученные выражения для производных, находим матрицу квадратичной формы $d^2 f_i(\bar{\eta}^{(i)})$ аргументов $d\theta_i, d\beta_1, \dots, d\beta_{L-1}$ в виде:

$$\begin{pmatrix} \sum_{j=1}^L t_{ij} & -t_{i1} & -t_{i2} & \dots & -t_{i,L-1} \\ -t_{i1} & t_{i1} & 0 & \dots & 0 \\ -t_{i2} & 0 & t_{i2} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -t_{i,L-1} & 0 & 0 & \dots & t_{i,L-1} \end{pmatrix}.$$

Все главные определители Δ_{im} этой матрицы положительны. $\Delta_{i1} = \sum_{j=1}^L t_{ij} > 0$, поскольку $\varphi''(\tau) > 0$ при всех $\tau \in R$, $t_{ij} > 0, i \in J_N, j \in J_L$. Зафиксировав произвольное $2 \leq m \leq L$ и выполнив несложные преобразования, получим $\Delta_m = \left(\sum_{j=m}^L t_{ij} \right) \cdot t_{i1} \cdot t_{i2} \cdot \dots \cdot t_{i,m-1} > 0$.

Тогда в силу критерия Сильвестра в каждой точке $\bar{\eta}^{(i)}$ пространства переменных $\theta_i, \beta_1, \dots, \beta_{L-1}$ второй дифференциал $d^2 f_i(\bar{\eta}_i)$ является положительно определённой квадратичной формой аргументов $d\theta_i, d\beta_1, \dots, d\beta_{L-1}$, из чего следует строгая выпуклость функций $f_i(\bar{\eta}_i)$. Доказана положительная определённость квадратичной формы

$$d^2 f(\bar{\eta}) = \sum_{i=1}^N d^2 f_i(\bar{\eta}_i) \quad (9)$$

аргументов $d\theta_1, \dots, d\theta_N, d\beta_1, \dots, d\beta_{L-1}$ в каждой точке $\bar{\eta} \in R^{N+L-1}$. Зафиксируем произвольный ненулевой вектор $(d\theta_1, \dots, d\theta_N, d\beta_1, \dots, d\beta_{L-1}) \in R^{N+L-1}$. Пусть $d\theta_i \neq 0$ при некотором $i \in J_N$. Тогда $d^2 f_i(\bar{\eta}_i) > 0$ при том, что все другие слагаемые в сумме (9) неотрицательны. Следовательно, $d^2 f(\bar{\eta}) > 0$. К тому же выводу приходим и при $d\beta_j \neq 0$ для некоторого $j \in J_{L-1}$: в этом случае в (9) положительны все слагаемые.

Таким образом, $d^2 f(\bar{\eta})$ - положительно определённая квадратичная форма, и $f(\bar{\eta})$ - строго выпуклая на R^{N+L-1} функция. Точка глобального минимума $\bar{\eta}_0$ функции $f(\bar{\eta})$ является для неё и точкой локального минимума, поэтому в соответствии с теоремой о единственности локального минимума у строго выпуклой функции других точек локального (и подавно, глобального) минимума функция $f(\bar{\eta})$ не имеет. Следовательно, $\bar{\eta}_0$ для функции $\Lambda(\bar{\eta})$ является единственной точкой глобального максимума.

Из приведённых доказательств следует, что оценки МП латентных параметров дихотомической модели Раша существуют и являются единственными только в том случае, когда используемая для их расчёта дихотомическая матрица ответов является допустимой. Поэтому недопустимые матрицы ответов должны удаляться из обработки результатов тестирования. Ввиду важности полученного вывода для практики вычисления надёжных оценок латентных параметров, было исследовано влияние размеров матрицы ответов на её допустимость. Для расчёта доли λ'_{2L} допустимых матриц от общего числа матриц заданного размера было получено следующее выражение:

$$\lambda'_{NL} = (1 - 2^{-L})^N + (1 - 2^{-N})^L - 1 - \sum_{n=2}^{N-1} \sum_{\ell=2}^L C_N^n C_L^\ell \lambda'_{n\ell} 2^{2n\ell - N\ell - nL} - \sum_{\ell=2}^{L-1} C_L^\ell \lambda'_{n\ell} 2^{N(\ell-L)}.$$

На его базе был разработан алгоритм расчёта величины λ'_{NL} и реализующая его программа, которые позволили с помощью вычислительного эксперимента оценить значения λ'_{NL} для некоторых дихотомических матриц, имеющих практическое значение. Было установлено, что λ'_{NL} существенно снижается для матриц с малым числом столбцов L и большим числом строк N . Например, для $N=500$ и $L=10$ значение $\lambda'_{500,10}$ может снижаться до 37,6 %. Поэтому перед обработкой матриц ответов необходимо проверять выполнение для них условий допустимости. Для этого были разработаны алгоритм и программа для выявления недопустимых дихотомических матриц ответов, которая входит в состав программного комплекса RILP-1M.

Для проверки качества оценок МП β_j^* и θ_i^* латентных параметров основной модели Раша исследовалась их состоятельность. Исследования проводились для двух случаев. В первом случае при выполнении исследований использовались следующие допущения: число столбцов L матрицы ответов ограничено; число строк N может быть бесконечно большим ($N \rightarrow \infty$); трудность всех L заданий одинакова: $\beta_1 = \beta_2 = \dots = \beta_L = 0$; элементы θ_i удовлетворяют неравенству $|\theta_i| \leq M_0, i = 1, 2, \dots, N$, где M_0 некоторая положительная постоянная; матрица ответов X удовлетворяет условиям допустимости, а её элементы x_{ij} взаимно независимы; для обеспечения единственности оценок θ_i^*, β_j^* придаём одной из них конкретное числовое значение, например, $\beta_r^* = 0$. Пусть для $N, L \geq 2$ θ_i^*, β_j^* являются оценками МП параметров θ_i, β_j , рассчитанными по дихотомическим результатам тестирования N студентов с уровнем подготовки $\theta_1, \theta_2, \dots, \theta_N$ при использовании теста с L заданиями нулевой трудности. Пусть Y - матрица, полученная из $X \in M_{NL}$ вычёркиванием экстремальных строк (если они есть). Введём событие $C = \{\text{матрица } Y \text{ допустима}\}$ и рассмотрим оценки максимального правдоподобия θ_i^*, β_j^* параметров $\theta_i, \beta_j, i \in T, j \in J_L$, рассчитанные по матрице ответов Y . Покажем, что при $N \rightarrow \infty$ вероятность наступления события C стремится к 1. Обозначим функцию успеха модели Раша через $u(\theta_i - \beta_j)$. Примем обозначения: $P(x_{ij} = 1) = u(\theta_i - \beta_j) = u(\theta_i) = p_i, P(x_{ij} = 0) = 1 - p_i = q_i, j \in J_L, i = 1, 2, \dots; p_0 = u(M_0), q_0 = 1 - p_0$. Нетрудно видеть, что $u(-M_0) = \frac{e^{-M_0}}{1 + e^{-M_0}} = \frac{1}{1 + e^{M_0}} = 1 - u(M_0) \quad 0,5 < u(M_0) < 1$. Поэтому можно записать $0 < q_0 < 0,5 < p_0 < 1, q_0 \leq p_i \leq p_0, q_0 \leq q_i \leq p_0, i = 1, 2, \dots$. Вводя вспомогательные события, благоприятствующие событию C , и оценивая вероятности их наступления, получим оценку снизу вероятности наступления события C в виде $P(C) \geq \sum_{j=1}^L (1 - \sigma^N) - L + 1 = L(1 - \sigma^N) - L + 1 = 1 - L\sigma^N$, где $\sigma = 1 - p_0 q_0 \in (0; 1)$. Так как полученная снизу оценка величины $P(C)$ стремится к 1 при $N \rightarrow \infty$, то и $P(C) \xrightarrow{N \rightarrow \infty} 1$. Следовательно, при $N \rightarrow \infty$ вероятность того, что матрица Y будет

допустимой, стремится к единице, а это означает, что существуют оценки максимального правдоподобия θ_i^*, β_j^* параметров $\theta_i, \beta_j, i \in J_N, j \in J_L$, и они являются единственными.

Введём в рассмотрение событие $E = \{ \text{наступило событие } C, \text{ и } \beta_j^* < \varepsilon \text{ при всех } j \in J_L \}$. Показано, что $P(E) \xrightarrow{N \rightarrow \infty} 1$, а значит, найдётся номер $N_0 = N_0(\varepsilon)$ такой, что $P(E) \geq 1 - \varepsilon^2$ при всех $N > N_0$. Из сходимостей $P(C)$ и $P(E)$ к 1 следует сходимость по вероятности к нулю оценок максимального правдоподобия β_j^* латентных параметров $\beta_j = 0$ ($j \in J_L$). Из теоремы об умножении вероятностей событий C и E следует $P_C(E) = \frac{P(CE)}{P(C)} = \frac{P(E)}{P(C)} \xrightarrow{N \rightarrow \infty} \frac{1}{1} = 1$. Это значит, что $\lim_{N \rightarrow \infty} P(|\beta_j^*| < \varepsilon) = 1$, что является условием состоятельности оценок β_j^* при $\beta_j = 0$. Путём введения в рассмотрение ещё нескольких вспомогательных событий и оценивания вероятности их наступления, доказываем, что оценки θ_i^* стремятся по вероятности к значениям $\theta_i^0 = \ln[x_i \cdot (L - x_i)^{-1}]$.

Для экспериментальной проверки полученных результатов использовался вычислительный эксперимент. Моделировались дихотомические матрицы с $L_H = 49$ и изменяющимся от 500 до 3007 значением N . Трудность всех виртуальных заданий в логитах задавалась равной 0 в первом эксперименте и 0,491 во втором. Анализ значений β_{j0}^* показал, что их отличия от $\beta_j = 0$ составили не более 0,002 логит, а от значения 0,491 – 0,6 %. Оценки θ_i^* практически полностью совпали с θ_i^0 во всех случаях.

Во втором случае исследовалась состоятельность оценок θ_i^* и β_j^* при $N \rightarrow \infty$ и $L \rightarrow \infty$, причём трудность заданий β_j выбиралась различной. Остальные допущения, принятые ранее, не изменялись. Состоятельность оценок θ_i^*, β_j^* параметров θ_i, β_j понимается в том смысле, что для $\varepsilon > 0, N, L \geq 2, \theta_1, \theta_2, \dots, \theta_N, \beta_1, \beta_2, \dots, \beta_L \in R$ существует событие $B = \{ \text{матрица } X \text{ допустима, а компоненты } \theta_i^*, \beta_j^* \text{ каждого главного решения системы (6) удовлетворяют неравенствам } |\theta_i^* - \theta_i| < \varepsilon, |\beta_j^* - \beta_j| < \varepsilon, i \in I_N, j \in I_L \}$. Доказано, что для любых чисел $M, \varepsilon > 0$ найдутся номер N_0 и число $\sigma > 0$ такие, что при $N, L \geq N_0, \ln \max(N, L) \leq \sigma \min(N, L), |\theta_i| \leq M, |\beta_j| \leq M, (i \in J_N, j \in J_L)$ вероятность $P(B) > 1 - \varepsilon$.

Доказательство выполнения этого условия осуществляется путём введения ряда вспомогательных событий и оценивания вероятности их наступления, на основе чего устанавливаются оценки снизу вероятностей наступления событий $F = \{ \text{матрица } X \text{ допустима} \}$, независимых в совокупности событий

$$H_j = \left\{ \left| y_j - \sum_{i=1}^N p_{ij} \right| < \alpha^2 \tau^2 N \right\}, j \in I_L \text{ и } K_i = \left\{ \left| x_i - \sum_{j=1}^L p_{ij} \right| < \alpha^2 \tau^2 L, i \in I_N \right\}. \text{ В результате по-}$$

лучаем: $P(F) > 1 - 3\tau$; для всех $j \in I_L$ $P(H_j) > 1 - be^{-Na} > 0$ и для всех $i \in I_N$

$P(K_i) > 1 - be^{-La}$. Тогда $P(H) = P\left(\prod_{j=1}^L H_j\right) = \prod_{j=1}^L P(H_j) > \prod_{j=1}^L (1 - be^{-Na}) = (1 - be^{-Na})^L$ и анало-

гичным образом найдем $P(K) > (1 - be^{-La})^N$. Здесь: $\tau = \min(q_0; 0,1\varepsilon) \in (0; 0,5)$, a, b – задаваемые положительные числа; α – наименьшее значение производной функции $u(\theta_i - \beta_j)$ на определённом отрезке $[-M_2; M_1]$, причем $\alpha > 0$. Доказано, что

$$P(H) = P\left(\bigcap_{j=1}^L H_j\right) = 1 - L + \sum_{j=1}^L P(H_j).$$

Оценивая снизу вероятность события $G = FHK$ с учётом последнего равенства и найденных оценок для вероятностей, получим $P(G) > 1 - 5\tau$. Показано, что в случае выполнения неравенств $N, L \geq N_0$, $|\theta_i| \leq M, |\beta_j| \leq M, i \in J_N, j \in J_L$, событие G благоприятствует событию B . Поэтому $P(B) \geq P(G) > 1 - 5\tau > 1 - \varepsilon$, что и требовалось доказать.

Экспериментальная проверка полученного теоретического результата проводилась на основе вычислительного эксперимента. Предполагалось, что распределения оценок β_j^* и θ_i^* являются нормальными. Допустимость данного предположения подтверждена сведениями из литературных источников и исследованиями автора работы. На основе данных обработки нормативной дихотомической матрицы ответов размером 60×49 моделировались аналогичные матрицы размером 501×500 (M1), 1002×1000 (M2), 1503×1500 (M3), 2004×2006 (M4) и 2505×2505 (M5). Смоделированные матрицы обрабатывались, после чего сравнивались между собой полученные графики зависимостей долей p_i верных ответов виртуальных тестируемых от их уровня подготовки $\theta_i^{*(k)}$ ($p_i(\theta_i^{*(k)})$), а также графики зависимостей долей p_j верных ответов на виртуальные задания от их трудности $\beta_j^{*(m)}$ ($p_j(\beta_j^{*(m)})$). Из рис. 1 видно, что наибольшие различия значений $\theta_i^{*(k)}$ ($\approx 3\%$) наблюдаются для графиков, полученных по матрицам M2 и M3 (кривые 2 и 3). Значения оценок $\theta_i^{*(k)}$ практически не изменяются, начиная с матрицы M3 размером 1500×1503 (кривые 3-5 почти совпадают).

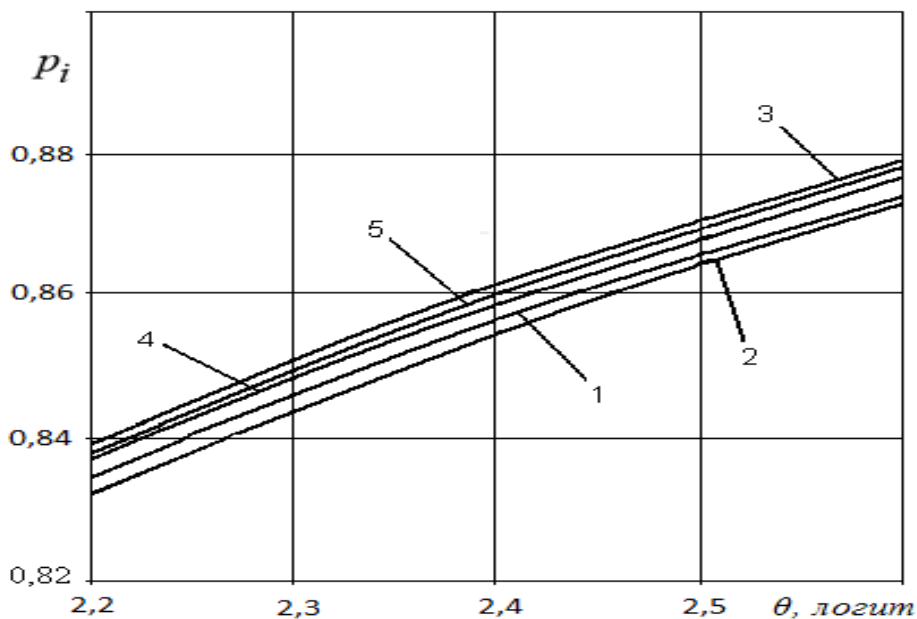


Рис. 1. Зависимости $p_i(\theta_i^{*(k)})$ для моделей матриц разного размера: 1–M1; 2–M2; 3–M3; 4–M4; 5–M5

Таким образом, оценки латентных параметров θ_i^* , β_j^* , рассчитанные по дихотомической матрице ответов ограниченного размера ($N, L \geq 50$), являются состоятельными. Квазигенеральные значения оценок могут быть вычислены по смоделированным матрицам, для которых $N, L \geq 1500$.

В четвёртой главе «Расчёт оценок латентных параметров модели Раша на основе моделирования дихотомических матриц ответов» исследовалась возможность повышения точности калибровки заданий теста по матрицам ответов ограниченного размера с помощью модифицированных бутстреп-метода и метода складного ножа, и решались задачи разработки моделей дихотомической матрицы ответов для параметризации ДС, необходимых для этого алгоритмов и ПО.

Результатами теоретического анализа и вычислительного эксперимента было показано, что полученные с помощью упомянутых методов предельные оценки трудности заданий $\bar{\beta}_{j0}^{*(CH)}$ и $\bar{\beta}_{j0}^{*(B)}$ стремятся по вероятности к оценкам МП β_j^* , рассчитанным по исходной матрице. Вследствие этого не имеет смысла использовать для повышения точности калибровки заданий бутстреп-метод и метод складного ножа, которые нередко используются за рубежом для обработки статистических данных. В процессе исследований было также выявлено, что при объёме выборки $N < 46$ и $\xi = N \cdot L^{-1} < 1$ изменение состава выборки тестируемых оказывает достаточно сильное влияние на значения оценок трудности заданий β_j^* . Поэтому для калибровки заданий диагностического теста рекомендуется использовать допустимые нормативные дихотомические матрицы результатов тестирования с числом строк не менее 46 и $\xi \geq 1$. Для проведения исследования поведения оценок $\bar{\beta}_{j0}^{*(CH)}$ и $\bar{\beta}_{j0}^{*(B)}$ были разработаны алгоритмы и программные комплексы RILP-Multi и RILP-Multi2, на которые получены свидетельства об официальной регистрации программ для ЭВМ.

Проблема повышения точности расчёта оценок латентных параметров основной модели Раша по выборкам ограниченного объёма была решена на основе компьютерного моделирования генеральной дихотомической матрицы ответов и нормативных дихотомических матриц ответов заданного размера. Предложенный алгоритм формирования таких матриц изображён на рис. 2. Значения указанных на схеме величин рассчитываются в соответствии с разработанной автором методикой. На основе моделирования дихотомических матриц ответов разных размеров показано, что относительная погрешность γ калибровки заданий трудностью $|\beta_j| \geq 0,18$ логит составляет 6 % при $N=200$ и снижается до 3 % при $N=300$.

В пятой главе «Разработка инструментария и моделей мониторинга результатов обучения и оценки качества диагностических средств» разработаны методологические основы оценки уровня сформированности компетенций обучаемых, методы расчёта квазигенеральных оценок латентных параметров основной модели Раша, модели мониторинга результатов обучения и оценки качества ДС. Теоретической основой предложенного метода расчёта квазигенеральных оценок латентных параметров по дихотомическим матрицам ответов ограниченного размера являются результаты, полученные в третьей главе.

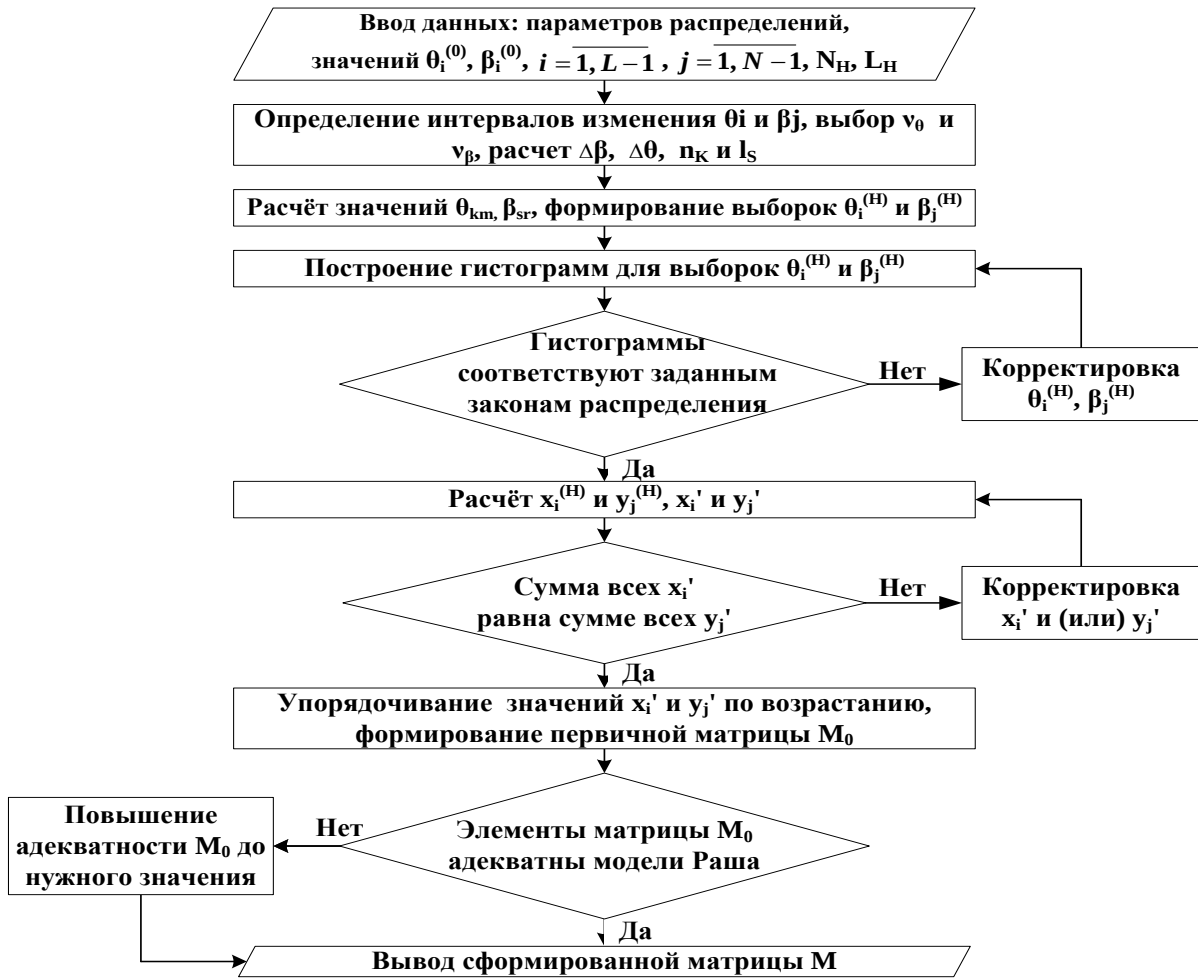


Рис. 2. Блок-схема алгоритма формирования дихотомической матрицы ответов

Основные этапы метода:

- обработка дихотомической матрицы ответов ограниченного размера $N_H \times L_H$, выявление типа и параметров распределений оценок θ_i^* и β_j^* ;
- формирование по найденным параметрам распределений оценок θ_i^* и β_j^* ($i = \overline{1, N_H}, j = \overline{1, L_H}$) модели матрицы ответов размером 1500×1500 , расчёт на её основе оценок θ_l^* и β_s^* ($l = \overline{1, 1500}, s = \overline{1, 1500}$);
- выбор из массива значений θ_l^* и β_s^* оценок $\theta_{i_0}^*$ и $\beta_{j_0}^*$, соответствующих экспериментальным значениям величин p_i и p_j , которые в свою очередь соответствуют исходным оценкам θ_i^* и β_j^* экспериментальной матрицы ответов.

Этапы метода калибровки заданий теста путём моделирования дихотомической матрицы ответов ограниченного размера $N_H \times L_H$ матрицей $N_0 \times L_H$ ($N_0 \geq 500$), в основном, сходны с рассмотренными для предыдущего метода. Отличие заключается в том, что моделирование матрицы выполняется для каждого значения β_j^* и все её столбцовые суммы одинаковы. Экспериментальная проверка методов в рамках вычислительного эксперимента, проведённого с использованием реальных матриц ответов с $N, L \geq 50$, показала их высокую эффективность: с 95 %-ой доверительной вероятностью погрешность расчёта оценок латентных параметров

снижается не менее чем в 2 раза при выборочной дисперсии уровня подготовки студентов $(\sigma_{\theta}^*)^2 \leq (0,76)^2 \approx 0,58$ логит².

На базе описанных методов и метода оценки уровня сформированности компетенций обучаемого, а также разработанной методики экспертизы качества ДС были созданы методологические основы независимой оценки качества подготовки обучаемых по значению уровня сформированности их компетенций. Базовыми положениями методологии являются: 1) уровень овладения компетенцией оценивается на основе показателей личностно - профессионального развития обучаемого по направлениям, определяемым личностным (ЛКК), когнитивным (ККК) и интегративно-деятельностным (ИДКК) компонентами компетенции, которыми представлена её структура; 2) количественная оценка уровня сформированности каждого из трёх компонентов компетенции носит вероятностный характер и оценивается своим банком качественных ДС; 3) ВД получены путём тестирования (анкетирования) обучаемых с применением методов социологических и психологических исследований; 4) обработка ВД, анализ и интерпретация её результатов проводятся с помощью ПО, работа которого базируется на использовании теории латентных переменных и классической теории тестирования. В качестве моделей измерения используются модели Раша; 5) уровень подготовки обучаемых оценивается соответствием уровня сформированности их компетенций требованиям образовательных стандартов.

На основе полученных в предыдущих главах результатов и сформулированных положений методологии, создана модель АСНО уровня подготовки выпускников учреждений, осуществляющих подготовку, повышение квалификации и переподготовку кадров. Структурная схема модели рассмотрена на примере подготовки бакалавра в соответствии с ФГОС ВПО (рис. 3).

Использование модели позволяет количественно оценить уровень подготовки выпускника и его соответствие требованиям ФГОС. Предложена модель мониторинга уровня сформированности компетенций студентов в процессе обучения, которая отличается от модели рис. 3 наличием дополнительного блока по проверке качества средств диагностики и их доработке. Это позволяет создать банк качественных ДС для проверки владения компетенциями к моменту итоговой государственной аттестации выпускников.

В шестой главе «Применение результатов диссертационного исследования для независимой оценки результатов обучения» рассматриваются примеры применения разработанного инструментария к оценке уровня сформированности компетенций студентов, обучающихся по программам высшего и среднего профессионального образования. В таблицах 1 и 2 представлены фрагменты результатов расчёта оценок уровня знаний (когнитивного компонента компетенции) тестируемых и трудности заданий педагогического теста, который использовался для получения входных данных. Расчёт исходных оценок θ_i^* , β_j^* производился по экспериментальной матрице ответов М1 размером 51х50. Оценки θ_{i1}^* , β_{j1}^* рассчитаны по матрице М4 размером 50х45, полученной путём удаления из М1 неадекватных элементов.

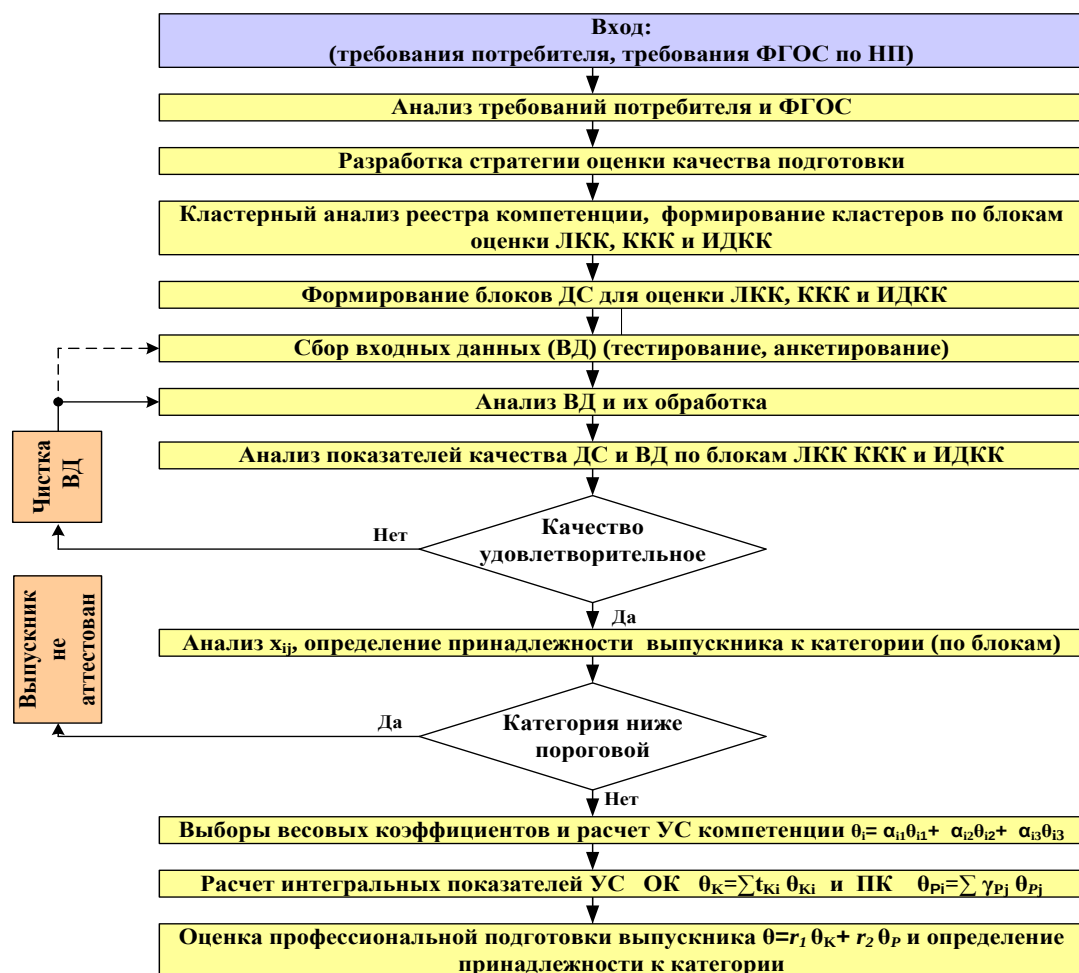


Рис. 3. Модель АСНО качества подготовки выпускников

Таблица 1 – Исходные θ_{i1}^* и квазигенеральные θ_{i0}^* оценки параметров θ_i уровня подготовки студентов в логитах ($\sigma_{\beta_{\max}}^* = 1,014$ логит, $\bar{\beta}_{pr} = 0,2952$ логит)

i	θ_{i1}^*	$\Delta_{\theta_{i1}}$	θ_{i0}^*	$\theta_{ipr}^{(2)}$	$\Delta_{2\theta_{pr}}$	n_{pr}
16	-1,278	0,7396	-1,2823	-1,0584	0,2239	3
18	-0,786	0,6812	-0,7921	-0,5425	0,2496	3
17	-0,457	0,657	-0,468	-0,196	0,272	2
8	-0,145	0,6449	-0,1612	0,133	0,2942	2
4	0,161	0,6449	0,1413	0,4574	0,3161	2
7	0,367	0,649	0,3455	0,6755	0,33	2
10	0,797	0,6752	0,772	1,1325	0,3605	2
19	1,152	0,7155	1,1249	1,5111	0,3862	2
50	1,563	0,784	1,5273	1,9493	0,422	2
33	2,565	1,0782	2,5222	3,008	0,4858	2

Оценки β_{j0}^* и β_{j0}^* , θ_{i0}^* найдены методами калибровки тестовых заданий и расчёта квазигенеральных оценок латентных параметров путём моделирования матрицы М4. В таблицах 1 и 2: Δ_{θ} , Δ_{β} и $\Delta_{2\theta_{pr}}$, $\Delta_{2\beta_{pr}}$ - полуширина 95 %-х доверительных интервалов оценивания значений θ_i^* , β_j^* и β_{j0}^* , θ_{i0}^* соответственно; n_{pr} - кратность отношения $\Delta_{\theta_{i1}}$ к $\Delta_{2\theta_{pr}}$ и $\Delta_{\beta_{j1}}$ к $\Delta_{2\beta_{pr}}$; γ_{β_j} - относительные различия оценок β_j^* и β_{j0}^* .

Из таблиц видно, что применение описанных в пятой главе методов расчёта оценок латентных параметров по матрицам ответов ограниченного размера позволяет с доверительной вероятностью 0,95 снизить погрешности их расчёта не менее чем в 2 раза.

Таблица 2 - Исходные β_j^* , β_{j1}^* и квазигенеральные β_{j0}^* , β_{j0y}^* оценки латентных параметров β_j в логитах ($\sigma_{\theta_{\max}}^*=1,06$ логит, $\bar{\theta}_{pr}=0,6884$ логит)

j	β_j^*	$\Delta_{\beta_{j1}}$	β_{j1}^*	β_{j0}^*	$\Delta_{2\beta_{pr}}$	n_{pr}	$ \gamma_{\beta j} \%$	β_{j0y}^*
05	-1,611	0,846	-1,652	-1,6606	0,1708	5	3,0	-1,6566
25	-1,023	0,7194	-1,057	-1,0566	0,1989	4	3,2	-1,0324
40	-0,674	0,6692	-0,701	-0,6968	0,2178	3	3,3	-0,678
22	-0,172	0,625	-0,189	-0,1796	0,247	2	4,2	-0,1716
44	0,014	0,6149	0,001	0,0112	0,2582	2	25,0	0,0117
26	0,195	0,6109	0,187	0,1978	0,2691	2	1,4	0,209
01	0,554	0,6129	0,557	0,5662	0,2918	2	2,2	0,574
13	1,017	0,637	1,037	1,0423	0,3195	2	2,4	1,0528
27	1,651	0,7094	1,699	1,6954	0,3957	2	2,6	1,7142

На примере оценивания личностных качеств обучаемых с помощью опросников Т. Лири для исследования межличностных отношений в коллективе и опросника Л.П. Калининского «Профиль личностных свойств» показано, что созданные во 2-5 главах диссертации алгоритмы, программное обеспечение и инструментарий для обработки результатов обучения позволяют количественно оценить степень сформированности ОК обучаемых, выявить аномалии в их личностных качествах, мешающие формированию компетенций.

Оценка интегративно-деятельностного компонента компетенции производилась на основе измерения качества выполнения и защиты дипломной работы. В таблице 3 представлены оценки θ_i^* параметра θ_i «качество дипломной работы» и погрешности их расчёта ε_{θ_i} , полученные по политомической матрице ответов, сформированной на основе оценок эксперта. Для расчёта значений θ_i^* использовался алгоритм, базирующийся на политомической модели частичного доверия. В качестве диагностического теста применялся разработанный с участием автора набор индикаторов, который приведён в диссертации. Категория G качества дипломной работы определялась по значению индивидуального балла X_i ($X_{i\max}=48$) с помощью представленной в диссертационной работе таблицы. Качество оценивается положительно, если $G \geq 2$. Представленная в таблице 3 информация свидетельствует о хорошем качестве дипломных работ ($G \geq 2$, $\theta_i^* > 0$) у большинства выпускников, что является признаком достаточно высокого уровня сформированности интегративно-деятельностного компонента компетенции.

Таким образом, результаты шестой главы позволяют сделать вывод о том, что созданные модели АСНО результатов обучения, методология и инструментальные средства обеспечивают независимую количественную оценку новых результатов образования в виде компетенций обучаемых и позволяют судить о соответствии их уровня сформированности требованиям стандарта.

Таблица 3 – Оценки θ_i^* параметра «качество дипломной работы» (логит)

i	θ_i^*	ε_{θ_i}	X_i	G	i	θ_i^*	ε_{θ_i}	X_i	G
1	3,556	0,43	41	4	17	1,699	0,40	28	3
2	3,961	0,49	43	4	18	2,003	0,38	30	3
3	2,812	0,37	36	3	19	-1,779	0,58	13	2
4	4,522	0,59	45	4	20	-1,158	0,54	15	2
5	5,496	0,87	47	4	21	3,556	0,43	41	4
6	3,746	0,45	42	4	22	-0,876	0,53	16	2
7	3,556	0,43	41	4	23	2,284	0,37	32	3
8	4,920	0,69	46	4	24	1,855	0,39	29	3
9	4,214	0,53	44	4	25	-0,349	0,50	18	2
10	4,522	0,59	45	4	26	2,946	0,38	37	4
11	4,522	0,59	45	4	27	4,214	0,53	44	4
12	4,214	0,53	44	4	28	0,996	0,44	24	3
13	3,085	0,39	38	4	29	0,137	0,48	20	2
14	4,522	0,59	45	4	30	-1,158	0,54	15	2
15	3,746	0,45	42	4	31	1,537	0,41	27	3
16	1,537	0,41	27	3	32	4,522	0,59	45	4

В заключении изложены итоги выполненного исследования.

В приложениях приведены руководства пользователя для работы с программными комплексами RILP-1M, RILP-multi, RILP-2, акты внедрения результатов диссертационного исследования, иллюстративные и другие материалы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИОННОЙ РАБОТЫ

1. Математически обоснована допустимость использования моделей Раша для оценки результатов обучения. На базе модели Раша для дихотомической переменной и метода максимального правдоподобия получены рекуррентные соотношения, обеспечивающие надёжную сходимость итерационного процесса расчёта оценок латентных параметров – «трудности заданий (индикаторов)» β_j диагностического средства и «уровня подготовки» обучаемых θ_i по результатам их диагностики (тестирования, анкетирования). Предложен алгоритм расчёта оценок максимального правдоподобия латентных параметров, обеспечивающих независимую оценку результатов обучения и проверку качества диагностических средств.

2. Получены рекуррентные выражения для расчёта значений уровня подготовки студентов θ_n , уровня трудности тестовых заданий β_i и уровня трудности каждого из шагов задания τ_{ij} по результатам диагностики, базирующиеся на использовании метода максимального правдоподобия и политомических моделей Раша для рейтинговой шкалы и частичного доверия. Разработаны алгоритмы итерационных процедур для расчёта оценок указанных латентных параметров, которые позволяют количественно оценить качество диагностических материалов и личностные параметры обучаемого, определяющие уровень владения компетенциями.

3. Впервые на основе результатов теоретических исследований и вычислительного эксперимента показано, что для существования и единственности оце-

нок максимального правдоподобия латентных параметров основной модели Раша дихотомическая матрица ответов, по которой рассчитываются оценки, должна удовлетворять установленным условиям допустимости, которые теоретически обоснованы и сформулированы. Предложен алгоритм проверки дихотомической матрицы ответов на соответствие её условиям допустимости, позволяющий выявлять и удалять из обработки недопустимые матрицы, что обеспечит повышение объективности оценки результатов обучения. Теоретически и результатами вычислительного эксперимента доказано, что доля недопустимых дихотомических матриц ответов в общем объёме матриц результатов тестирования может достигать больших значений ($\approx 62,4\%$), когда число строк в матрице велико ($N \approx 500$), а число столбцов мало ($L \approx 10$). При увеличении числа столбцов L до 50 доля недопустимых матриц снижается до 3 %. Предложен алгоритм расчёта доли допустимых матриц результатов тестирования заданного размера.

4. Впервые показано, что оценки максимального правдоподобия латентных параметров основной модели Раша, рассчитанные по квазинормативным допустимым матрицам ответов ограниченного размера ($N, L \geq 50$) являются состоятельными. Установлено, что при одновременном увеличении числа и строк, и столбцов матрицы ответов обеспечивается состоятельность оценок параметров и «уровень подготовки» обучаемого, и «уровень трудности» задания. Если увеличивается только число строк матрицы, а количество столбцов фиксировано и все столбцовые суммы одинаковы, состоятельны только оценки параметра «уровень трудности» задания, а оценки параметра «уровень подготовки» обучаемого стремятся по вероятности к первоначальным значениям $\theta_i^0 = \ln(p_i q_i^{-1})$. Установлены минимально достаточные размеры нормативной допустимой матрицы ответов, по которой могут быть рассчитаны генеральные оценки латентных параметров θ_i и β_j .

5. Предложен новый метод калибровки заданий теста путём моделирования допустимой квазинормативной дихотомической матрицы ответов ограниченного размера $N_H \times L_H$ аналогичными матрицами с существенно большим числом строк и с одинаковыми значениями столбцовых сумм, позволяющий получить квазигенеральные оценки трудности заданий. Предложен новый метод расчета квазигенеральных оценок максимального правдоподобия латентных параметров θ_i и β_j основной модели Раша путём моделирования матрицы ответов $N_H \times L_H$ нормативными матрицами $N_1 \times L_1$ с существенно большим числом строк и столбцов ($N_1, L_1 \geq 1500$). С доверительной вероятностью 0,95 методы обеспечивает снижение погрешностей расчёта оценок не менее чем в 2 раза, если выборочная дисперсия уровня подготовки обучаемых $(\sigma_o^*)^2$ не превышает 0,58 логит².

6. Впервые на основе результатов вычислительного эксперимента и теоретических исследований показано, что использование модифицированных бутстреп-метода и метода складного ножа не обеспечивает достижения выигрыша в точности расчёта заданий диагностического средства по единственной квазинормативной дихотомической матрице ответов ограниченного размера.

Предложена модель генеральной дихотомической матрицы ответов, формирование которой осуществляется на базе основной модели Раша по заданным за-

конам распределения латентных параметров θ_i , β_j . Оценки латентных параметров θ_{i0}^* , β_{j0}^* , рассчитанные по этой модели матрицы, могут использоваться в качестве генеральных при решении задач моделирования и параметризации диагностических тестов. Предложены методика и алгоритм формирования модели нормативной дихотомической матрицы ответов, позволяющие осуществлять моделирование качественных матриц заданного размера, использование которых обеспечивает снижение влияния искажающих факторов на результаты параметризации и моделирования диагностических тестов. С использованием моделей допустимых дихотомических матриц ответов показано, что погрешность калибровки тестовых заданий трудностью $|\beta_j| > 0,18$ логит не превышает 6 % при $N=200$ и 3 % при $N=300$.

7. Разработаны методологические основы независимой оценки уровня сформированности компетенций обучаемых, базирующиеся на принятых подходах к диагностике компетенций, использовании вероятностных моделей Раша, методов математической статистики и теории латентных переменных, методов социологических и психодиагностических исследований. Предложена методика экспертизы качества диагностических средств по матрицам ответов ограниченного размера на основе дихотомической модели Раша и теории латентных переменных в сочетании с классической теорией тестов. Предложен метод расчёта уровня сформированности компетенции обучаемого по значениям оценок её личностного, когнитивного и интегративно-деятельностного компонентов. Применение перечисленного инструментария обеспечивает надёжную диагностику результатов компетентностно-ориентированного обучения.

8. Предложена модель мониторинга уровня сформированности компетенций обучаемых, позволяющая объективно отслеживать уровень их подготовки получать информацию о качестве диагностических средств. Использование модели обеспечит создание в процессе обучения слушателей и студентов средства диагностики с высокими показателями качества, необходимые для проведения итоговой государственной аттестации выпускников. Предложена модель системы оценки уровня подготовки выпускников учреждений по обучению, повышению квалификации и переподготовке кадров, позволяющая объективно оценить профессиональную подготовку выпускников и соответствие её требованиям соответствующих стандартов.

9. Разработаны пакеты и комплексы программ (RILP-1, RILP-2, RILP-Multi и их модификации), реализующие предложенные модели и алгоритмы расчёта оценок латентных параметров обучаемых и средств диагностики, которые отличаются от известных надёжной сходимостью итерационных оценок к оцениваемому параметру. Применение созданного ПО обеспечивает автоматизацию и повышение эффективности работы АСНО результатов обучения. С использованием применяемых в образовании и социологических исследованиях диагностических тестов и опросников показано, что созданные в процессе диссертационного исследования модели и инструментарий АСНО обеспечивают независимую оценку новых результатов обучения - компетенций обучаемых.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Монография

1. **Елисеев, И.Н.** Методы, алгоритмы и программные комплексы для расчёта характеристик диагностических средств независимой оценки качества образования: монография / И.Н. Елисеев. – 2 изд., перераб. и доп. - Новочеркасск: Лик, 2013. - 285 с.

Публикации в ведущих научных журналах из списка ВАК

2. **Елисеев, И.Н.** Диагностика индивидуальных свойств личности в студенческой среде на основе модели Раша / И.Н. Елисеев // Стандарты и мониторинг в образовании. - 2009. - № 3. - С. 35-38.

3. **Елисеев, И.Н.** Алгоритмы итерационных процедур вычисления оценок уровня подготовки студентов / И.Н. Елисеев // Известия вузов. Электромеханика. - 2013. № 3. С. 83-90.

4. **Елисеев, И.Н.** Исследование существования и единственности оценок максимального правдоподобия параметров латентных переменных однопараметрической дихотомической модели Раша / И.Н. Елисеев, И.С. Шрайфель // Информатизация образования и науки. - 2011. - № 3 (11). - С. 117-129.

5. **Елисеев, И.Н.** Исследование погрешности расчёта трудности заданий теста на основе моделирования дихотомической матрицы ответов / И.Н. Елисеев // Педагогическая информатика. - 2011. - № 4. - С. 92-101.

6. **Елисеев, И.Н.** Калибровка заданий теста с использованием бутстреп-метода / И.Н. Елисеев // Программные продукты и системы. 2010. № 2. С 96-99.

7. **Елисеев, И.Н.** Методология оценки уровня компетенций студента / И.Н. Елисеев // Информатика и образование. - 2012. - №4. - С.80-85.

8. **Елисеев, И.Н.** Модель мониторинга уровня компетенций студентов и оценки качества диагностических средств / И.Н. Елисеев // Известия вузов. Северо-Кавказский регион. Технические науки. - 2012. - №3. - С. 126-132.

9. **Елисеев, И.Н.** Модель системы оценки качества подготовки выпускников бакалавриата и магистратуры / И.Н. Елисеев // Информатизация образования и науки. - 2012. - №3(15). - С.106-115.

10. **Елисеев, И.Н.** Модель дихотомической матрицы результатов тестирования / И.Н. Елисеев // Программные продукты и системы. - 2011. - № 3.- С. 80-86.

11. **Елисеев, И.Н.** Модель оценивания латентных параметров дихотомической модели Раша / И.Н. Елисеев, И.С. Шрайфель // Известия вузов. Северо-Кавказский регион. Технические науки. - 2011.-№ 6.-С. 37-46.

12. **Елисеев, И.Н.** Оценка уровня подготовленности выпускников колледжа на основе измерения качества квалификационных работ / И.Н. Елисеев // Стандарты и мониторинг в образовании. - 2010. - № 1. - С. 47-49.

13. **Елисеев, И.Н.** О состоятельности дихотомической модели Раша / И.Н. Елисеев, И.С. Шрайфель // Известия вузов. Северо-Кавказский регион. Технические науки. - 2012. - №5. - С. 127-136.

14. Ларина, Т.Н. Оценка уровня компетенций студента на основе модели Раша [Электронный ресурс] / Т.Н. Ларина, **И.Н. Елисеев** // Инженерный вестник

Дона: электронный научный журнал - 2012. - № 4. Ч.1. Режим доступа: <http://www.ivdon.ru/magazine/latest/> (дата обращения 26.10.2012)

15. **Елисеев, И.Н.** Программный комплекс RILP-1 / И.Н. Елисеев, И.И. Елисеев, А.В. Фисунов // Программные продукты и системы. - 2009. - № 2. С. 178-181.

16. **Елисеев, И.Н.** Расчёт параметров тестовых заданий при представлении результатов тестирования в виде полиномической переменной / И.Н. Елисеев, И.И. Елисеев // Известия вузов. Северо-Кавказский регион. Технические науки. - 2002. - №3. - С. 123–125.

17. **Елисеев, И.Н.** RILP-Multi для расчёта предельных оценок параметров индикаторов бутстреп-методом / И.Н. Елисеев, И.И. Елисеев, А.В. Фисунов // Программные продукты и системы. - 2011. - №2. - С. 114-117.

18. **Елисеев, И.Н.** Теоретические основы алгоритма расчёта латентных переменных программным комплексом RILP-1М / И.Н. Елисеев // Программные продукты и системы. - 2011. - №2. - С. 67-72.

19. **Елисеев, И.Н.** Теоретические основы алгоритма расчёта латентных переменных программным комплексом RILP-2 / И.Н. Елисеев // Известия вузов. Северо-Кавказский регион. Технические науки. - 2011. - №3. - С. 3-8.

20. **Елисеев, И.Н.** Теоретическое обоснование алгоритма расчёта латентных переменных программным комплексом RILP-2 на основе модели рейтинговой шкалы / И.Н. Елисеев, И.И. Елисеев // Известия вузов. Северо-Кавказский регион. Технические науки. - 2011. - №4. - С. 14-19.

21. **Елисеев, И.Н.** Теоретические основы расчёта доли допустимых дихотомических матриц ответов / И.Н. Елисеев, И.С. Шрайфель // Информатизация образования и науки. – 2011, - № 4. - С. 60-74.

22. **Елисеев, И.Н.** Экспериментальные исследования состоятельности оценок латентных параметров модели Раша / И.Н. Елисеев // Программные продукты и системы. - 2012. - №3. - С. 166-171.

23. **Елисеев, И.Н.** Экспериментальное подтверждение состоятельности оценок трудности заданий теста / И.Н. Елисеев // Программные продукты и системы. - 2012. - №2. - С. 153-156.

24. **Елисеев, И.Н.** Экспертиза качества тестов по электротехническим дисциплинам / И.Н. Елисеев // Известия вузов. Электромеханика. - 2011. - №2. - С. 75-80.

Свидетельства о регистрации программ для ПЭВМ

25. Свидетельство об официальной регистрации программы для ЭВМ, №2009610643. Измеритель латентных переменных RILP-1 / **И.Н. Елисеев**, А.В. Фисунов, И.И. Елисеев. - Заявл. 02.12.2008, г. Москва. - Оpubл. 28.01.2009.

26. Свидетельство об официальной регистрации программы для ЭВМ, № 2010611109. RILP-2 / **И. Н. Елисеев**, А.И. Шерстобитов, А.В. Фисунов, И.И. Елисеев. - Заявл. 08.12.2009. Москва. - Оpubл. 05.02.2010.

27. Свидетельство об официальной регистрации программы для ЭВМ, №2011615213. RILP-1М / **И. Н. Елисеев**, А.В. Фисунов, И.И. Елисеев. - Заявл. 12.05.20011, г. Москва. - Оpubл. 01.07.2011.

28. Свидетельство об официальной регистрации программы для ЭВМ,

№ 2010611111. RILP-Multi / **И. Н. Елисеев**, А.В.Фисунов, И.И. Елисеев. – Заявл. 08.12.2009, г. Москва. – Оpubл. 05.02.2010.

29. Свидетельство об официальной регистрации программы для ЭВМ, № 2011615220. RILP-Multi2 / **И.Н. Елисеев**, А.В. Фисунов, И.И. Елисеев. - Заявл. 12.05.2011, г. Москва. - Оpubл. 01.07.2011.

30. Свидетельство об официальной регистрации программы для ЭВМ, №2011615214. Программа для повышения адекватности элементов дихотомической матрицы принятой модели измерения / **И.Н. Елисеев**, А.В. Фисунов, А.С. Папян. - Заявл. 12.05.2011, г. Москва. - Оpubл. 01.07.2011.

31. Свидетельство об официальной регистрации программы для ЭВМ, №2011615215. Программа формирования генеральных выборок латентных переменных модели Раша / **И.Н. Елисеев**, А.В. Фисунов, А.С. Папян. - Заявл. 12.05.2011, г. Москва. - Оpubл. 01.07.2011.

32. Свидетельство об официальной регистрации программы для ЭВМ, №2011615218. Программа моделирования дихотомической матрицы результатов тестирования / **И.Н. Елисеев**, А.В. Фисунов, А.С. Папян. - Заявл. 12.05.2011, г. Москва. - Оpubл. 01.07.2011.

33. Свидетельство об официальной регистрации программы для ЭВМ, №2009610893. Программа для построения психогрaмм / **И.Н. Елисеев**, А.В. Фисунов, И.И. Елисеев. - Заявл. 16.12.2008, г. Москва. - Оpubл. 09.02.2008.

34. Свидетельство об официальной регистрации программы для ЭВМ, №2011615217. Программа для выявления недопустимых дихотомических матриц результатов тестирования / **И.Н. Елисеев**, И.С. Шрайфель, А.В. Фисунов. - Заявл. 12.05.2011, г. Москва. - Оpubл. 01.07.2011.

35. Свидетельство об официальной регистрации программы для ЭВМ, №2011615216. Программа расчета доли недопустимых дихотомических матриц результатов тестирования / **И.Н. Елисеев**, И.С. Шрайфель, А.В. Фисунов. - Заявл. 12.05.2011, г. Москва. - Оpubл. 01.07.2011.

Публикации в сборниках научных статей, трудов и материалов конференций

36. **Елисеев, И.Н.** Алгоритм проверки дихотомической матрицы результатов тестирования на соответствие условиям допустимости / И.Н. Елисеев, А.В. Фисунов // Перспективы развития гуманитарных и технических систем (Перспективы_ГТС_11): материалы всерос. науч. конф. - Таганрог, 2011. - С. 22-25.

37. **Елисеев, И.Н.** Алгоритм расчёта уровня подготовки студентов с использованием латентных переменных / И.Н. Елисеев // Современный учебно-воспитательный процесс: теория и практика: материалы V всероссийской (с международным участием) заочной научно-практической конференции. - Красноярск, СГТУ, 2013. Ч. 2. - С. 212-219.

38. **Елисеев, И.Н.** Диагностика коммуникативных компетенций студентов на основе политомической модели Раша / И.Н. Елисеев, Т.Н. Ларина, Л.Ю. Литвиненко // Современное образование: содержание, технологии, качество: материалы XVII международной научно-методической конференции. - Санкт-Петербург, СПбГЭТУ «ЛЭТИ», 2011. Т. 2. - С. 7-9.

39. **Елисеев, И.Н.** Диагностика формально-динамических свойств личности на основе политомической модели Раша / И.Н. Елисеев, Л.Ю. Литвиненко // Информационное общество: идеи, технологии, системы: материалы международной научной конференции. Таганрог, ТТИ ЮФУ, 2010. Ч. 2. - С. 30-38.

40. **Елисеев, И.Н.** Измерение качества квалификационных работ выпускников колледжа на основе модели Раша / И.Н. Елисеев, Л.Н. Скореева, Ю.С. Пономарева // Системы и модели в информационном мире (СМИ-2009) (Таганрог, март 2009г.): сб. статей международной научной конференции. - Таганрог, ТТИ ЮФУ, 2009. Ч. 1. - С. 33-39.

41. **Елисеев, И.Н.** Исследование погрешности калибровки заданий теста на модели дихотомической матрицы ответов / И.Н. Елисеев, А.В. Фисунов, А.С. Папаян // Актуальные вопросы исследования общественных и технических систем (АВИОТС-2011): материалы Всероссийской научной конференции. - Таганрог, ТТИ ЮФУ, 2011. Ч. 1. - С. 36-42.

42. **Елисеев, И.Н.** Исследование темперамента личности на основе методов теории латентных переменных / И.Н. Елисеев, Л.Ю. Литвиненко // Системы и модели в информационном мире (СМИ-2009) (Таганрог, март 2009 г.): сб. статей между. научной конференции. - Таганрог, ТТИ ЮФУ, 2009. Ч. 1. - С. 27-33

43. **Елисеев, И.Н.** Использование модели Раша для диагностики социально-личностных компетенций / И.Н. Елисеев // Современное образование: содержание, технологии, качество: материалы международного форума. - Санкт-Петербург, СПбГЭТУ «ЛЭТИ», 2010. - Т.2. - С. 239-241.

44. **Елисеев, И.Н.** Метод оценки уровня сформированности компетенций и качества подготовки выпускника / И.Н. Елисеев // Современное образование: содержание, технологии, качество: материалы XIX международной научно-методической конференции. - Санкт-Петербург, СПбГЭТУ «ЛЭТИ», 2013. - Т. 1. - С. 83-84.

45. **Елисеев, И. Н.** Метод чистки теста путём последовательного удаления неадекватных респондентов и заданий / И.Н. Елисеев // Тестирование в сфере образования: проблемы и перспективы развития: материалы III всероссийской (с международным участием) научно-практической конференции. - Красноярск, СГТУ, 2010. - С. 27-31.

46. **Елисеев, И.Н.** Методологические основы оценки уровня сформированности компетенций выпускника бакалавриата (магистратуры) / И.Н. Елисеев // Современное образование: содержание, технологии, качество: материалы XVIII международной научно-методической конференции. - Санкт-Петербург, СПбГЭТУ «ЛЭТИ», 2012. - Т. 2. - С. 139-141.

47. **Елисеев, И.Н.** Расчёт латентных параметров диагностических материалов на основе дихотомической модели Раша / И.Н. Елисеев, И.С. Шрайфель // Физико-математическое моделирование систем: материалы VII междунар. семинара. - Воронеж: ГОУ ВПО «Воронежский гос. техн. ун-т», 2011. Ч. 3. - С. 80-85.

48. **Елисеев, И. Н.** Оценка качества опросника «Уровень сформированности гражданственности и патриотизма» для измерения латентной переменной «Гражданственность и патриотизм» / И.Н. Елисеев, Л.В. Литвиненко, А.С. Папаян // Методы и алгоритмы принятия эффективных решений» (МАПР - 09) (Таганрог, ок-

тябрь 2009 г.): материалы международной научной конференции. - Таганрог, ТТИ ЮФУ, 2009. Ч. 1. – С. 31-35.

49. **Елисеев, И.Н.** Оценка общекультурных компетенций с помощью теста-опросника Лири / И.Н. Елисеев, Т.Н. Ларина, Л.В. Носкова // Перспективы развития гуманитарных и технических систем (Перспективы_ГТС_11): материалы всероссийской научной конференции. - Таганрог, ТТИ ЮФУ, 2011. Ч.1. - С. 16-21.

50. **Елисеев, И.Н.** Программный комплекс RILP-1 для измерения латентных переменных / И.Н. Елисеев, И.И. Елисеев, А.В. Фисунов // Инновации в обществе, технике и культуре: материалы международной научной конференции. - Таганрог, ТТИ ЮФУ, 2008, Ч.2. - С. 22-28.

51. **Елисеев, И.Н.** Программный комплекс RILP-MULTI для расчета предельных оценок латентных переменных / И.Н. Елисеев, И.И. Елисеев, А.В. Фисунов // Методы и алгоритмы принятия эффективных решений (МАПР-09): материалы межд. научной конференции. - Таганрог, ТТИ ЮФУ, 2009. Ч. 2. –С. 35-38.

52. **Елисеев, И.Н.** Расчёт параметров опросника Русалова для исследования формально-динамических качеств индивидуальности / И.Н. Елисеев, Л.Ю. Литвиненко // Информатизация образования-2011: материалы международной научно-практической конференции. - Елец, Елецкий государственный университет им. И.А. Бунина, 2011. Т. 2. - С. 107-114.

53. **Елисеев, И.Н.** Чистка теста с использованием специализированных программных средств для расчёта латентных переменных / И.Н. Елисеев, И.И. Елисеев, Т.Н. Ларина // Теория и практика измерения латентных переменных в образовании и других социально-экономических системах: сб. ст. IX всероссийской (с международным участием) научно-практической конференции. - Славянск-на-Кубани, 2007. - С.66-69.

Личный вклад автора в работах, опубликованных в соавторстве

Постановка задач исследований, планирование натурного и вычислительного экспериментов, обработка их данных и анализ полученных результатов, их обобщение в виде выводов и рекомендаций, интерпретация, изложение содержания работ в виде научных публикаций, апробация являются личным вкладом автора во все работы, выполненные в соавторстве. Кроме того, личный вклад автора состоит в том, что в работах: [4; 11; 13; 21; 47] обоснованы процедуры исследования свойств оценок латентных параметров модели Раша, теоретические результаты подтверждены данными натурного и вычислительного экспериментов и интерпретированы, разработаны алгоритмы и методы расчёта квазигенеральных оценок латентных параметров; [15-17; 20; 25-33; 36; 50; 51] теоретически обоснованы и разработаны алгоритмы работы компьютерных программ, структура программных комплексов, выполнен анализ качества их работы по результатам натурного и вычислительного экспериментов; [14; 38-42; 34; 35; 48; 49; 52; 53] теоретически обоснованы и разработаны методики и алгоритмы расчёта оценок латентных параметров и оценивания качества диагностических средств.