

DRAFT

**IRSA**

**ТЕХНИЧЕСКИЙ ОТЧЕТ:**

**ЛОКАЛИЗАЦИЯ ИНСТРУМЕНТАРИЯ (2018)**

**ОСНОВНОЕ ИССЛЕДОВАНИЕ (2019)**

**(INTERNATIONAL RESEARCH OF STUDENTS ACHIEVEMENTS)**

[ВВЕДЕНИЕ 3](#_Toc22115372)

[ГЛАВА 1. ОПИСАНИЕ ИНСТРУМЕНТА STUDENT ACHIEVEMENTS MONITORING 5](#_Toc22115373)

[1.1. Теоретические основания и конструктивные особенности SAM 5](#_Toc22115374)

[1.2. Содержание теста по математике 10](#_Toc22115375)

[ГЛАВА 2. ОПРОСНИКИ ДЛЯ СБОРА КОНТЕКСТНОЙ ИНФОРМАЦИИ 18](#_Toc22115376)

[2.1. Особенности разработки контекстных анкет 18](#_Toc22115377)

[2.2. Содержательные и структурные особенности анкеты для учеников 22](#_Toc22115378)

[ГЛАВА 3. ПРОЦЕДУРА ИССЛЕДОВАНИЯ 26](#_Toc22115379)

[3.1. Выборка исследования 26](#_Toc22115380)

[3.1.1. Выборка учеников 26](#_Toc22115381)

[3.1.2. Выборка учителей 35](#_Toc22115382)

[3.2. Процедура локализации инструмента SAM 44](#_Toc22115383)

[3.3. Этапы локализации SAM на армянском языке и результаты её выполнения 45](#_Toc22115384)

[3.4. Результаты полномасштабной апробации 51](#_Toc22115385)

[3.5. Шкалирование результатов 53](#_Toc22115386)

[3.6. Установление пороговых баллов по данным пилотного исследования 2018 54](#_Toc22115387)

[3.7. Анализ различительного функционирования заданий 55](#_Toc22115388)

[ГЛАВА 4. ПСИХОМЕТРИЧЕСКИЕ ХАРАКТЕРИСТИКИ ТЕСТОВ SAM 56](#_Toc22115389)

[ГЛАВА 5. ПСИХОМЕТРИЧЕСКИЕ ХАРАКТЕРИСТИКИ ИССЛЕДУЕМЫХ ПСИХОЛОГИЧЕСКИХ КОНСТРУКТОВ, 2019 ГОД 60](#_Toc22115390)

[5.1. Психометрические характеристики конструктов из контекстной анкеты ученик 60](#_Toc22115391)

[5.2. Психометрические характеристики конструктов из контекстной анкеты учителя 74](#_Toc22115392)

[СПИСОК ЛИТЕРАТУРЫ 82](#_Toc22115393)

[ПРИЛОЖЕНИЕ 1 83](#_Toc22115394)

[ПРИЛОЖЕНИЕ 2 86](#_Toc22115395)

[ПРИЛОЖЕНИЕ 3 88](#_Toc22115396)

[ПРИЛОЖЕНИЕ 4 94](#_Toc22115397)

[ПРИЛОЖЕНИЕ 5 98](#_Toc22115398)

[ПРИЛОЖЕНИЕ 6 104](#_Toc22115399)

[ПРИЛОЖЕНИЕ 7 112](#_Toc22115400)

[ПРИЛОЖЕНИЕ 8 118](#_Toc22115401)

[ПРИЛОЖЕНИЕ 9 124](#_Toc22115402)

[ПРИЛОЖЕНИЕ 10 129](#_Toc22115403)

[ПРИЛОЖЕНИЕ 11 130](#_Toc22115404)

[ПРИЛОЖЕНИЕ 12 132](#_Toc22115405)

[ПРИЛОЖЕНИЕ 13 136](#_Toc22115406)

[ПРИЛОЖЕНИЕ 14 139](#_Toc22115407)

[ПРИЛОЖЕНИЕ 15 140](#_Toc22115408)

[ПРИЛОЖЕНИЕ 16 141](#_Toc22115409)

[ПРИЛОЖЕНИЕ 17 142](#_Toc22115410)

[ПРИЛОЖЕНИЕ 18 143](#_Toc22115411)

ВВЕДЕНИЕ

В последние годы в образовательной среде по всему миру проводятся исследования, модернизация диагностического и педагогического инструментария в русле компетентностного подхода. Запуск данного процесса обусловлен необходимостью приблизить образовательные практики к повседневной жизни, чтобы у выпускников учебного учреждения была возможность эффективно включиться в социальные, культурные, экономические и политические подсистемы государственного и мирового масштаба в качестве полноправного субъекта. В этом смысле создание, апробация и распространение диагностических образовательных инструментов является задачей наивысшего уровня.

Если спуститься на ступень ниже и рассмотреть данную проблему как международную или межстрановую, то можно обнаружить, что ресурсы для участия в образовательных диагностических проектах мирового масштаба, таких как PISA или TIMSS присутствуют далеко не у каждого заинтересованного государства. В данном контексте особенно актуальным становится способность более развитых и обеспеченных ресурсами стран выступать в качестве образовательного донора. Это, с одной стороны, позволит ускорить развитие нуждающихся в поддержке государств в сфере образования, а с другой стороны, обеспечит государство-донор укреплением международных дружеских контактов и откроет новые возможности для кросс-культурного взаимообмена.

Переходя на локальный уровень, обратимся к наиболее заинтересованной социальной группе – учителям. Отметим, что особенно важным и полезным исследование диагностических образовательных инструментов является именно для педагогического сообщества, как государственного, так и международного масштаба. Каждый профессионал в своей области заинтересован в развитии квалификации и осуществлении своей деятельности максимально продуктивно и результативно. В этом смысле предоставление педагогу инструментария, на основании которого он сможет выстраивать образовательный процесс более эффективно, может рассматриваться как ключевая цель исследовательского сообщества в области образования по всему миру и в России, в частности.

Наконец, на уровне отдельно взятой личности ­– личности учащегося – изучение особенностей диагностический образовательных инструментов также является необходимым и актуальным, так как может рассматриваться как залог качественного социо-культурного развития человека.

Все вышеперечисленные аспекты определяют актуальность проведения международного исследования IRSA (International Research of Students Achievements). Исследование учебно-предметных компетенций учеников начальной школы проводится в Армении, Беларуси, Киргизии, России и Таджикистана. Участие в совместном международном проекте предоставляет возможности для наращивания их экспертного потенциала, формирования опытных национальных команд для проведения международных исследований качества образования и стимулирования обмена знаниями и опытом.

В соответствии с ограничениями программы READ определено: проводить исследование в школах из крупных городов. Выборка для исследования представляет города в соответствии с областями и типом школы (школа, лицей). Такой подход позволит сравнить тенденции в образовании в крупных городах, которые функционируют лучше благодаря большим финансовым ассигнованиям. Таким образом каждую страну представляет примерно 1000 учеников в возрасте от 10 до11 лет. Учитывая ограничения формирования выборки исследование носит описательно-сравнительный характер.

В центре исследования находится вопрос: какие факторы поддерживают математическую компетентность школьников, а также каковы их сходства и различия в странах, участвующих в исследовании?

Основные задачи исследования:

* локализация инструмента SAM в странах, участвующих в исследовании;
* оценить уровень развития познавательных способностей учащихся начальной школы, их математическую компетентность;
* рассмотреть образовательный контекст посредством обращения к информации о социально-экономической ситуации и культурной среде учащихся в школе;
* заполнить пробелы в данных и собрать информацию о состоянии образовательных систем в исследуемых странах для удовлетворения потребностей всех заинтересованных сторон, включая международные организации и учреждения-доноры.

Исследование проводится в рамках программы READ 2 с 2016 по 2019 год. Применение единой методологии в каждой стране позволяет сравнивать качество образования в разных странах, анализировать содержание, корректно оценивать образовательную ситуацию в начальной школе.

ГЛАВА 1. ОПИСАНИЕ ИНСТРУМЕНТА STUDENT ACHIEVEMENTS MONITORING

**1.1. Теоретические основания и конструктивные особенности SAM**

Student Achievement Monitoring (SAM) представляет собой инструмент для оценки образовательных компетенций школьника. Ключевым преимуществом SAM, выделяющим его среди других средств измерения школьных достижений учащихся, является опора на систему представлений о возможных качественных уровнях (ступенях) образовательных результатов. Основанием данной системы являются разработки модели учебного процесса, созданные силами отечественных психологов: в первую очередь, Л. С. Выготским, а также его последователями – П. Я. Гальпериным, В. В. Давыдовым, Д. Б. Элькониным. На сегодняшний день наиболее подробно и структурировано механизм работы данной модели учебного процесса изложен в книге «Диагностика учебной успешности в начальной школе» под редакцией П. Г. Нежнова (Диагностика учебной успешности в начальной школе, 2009).

Авторы SAM создавали данный диагностический инструмент, опираясь на компетентностный подход, который в последние десятилетия стал основным вектором трансформации и преобразования систем образования во всем мире и в России, в частности. Понятие «компетенция» в наиболее общем виде можно интерпретировать как целостное свойство человека, состоящее, с одной стороны, из способности осуществлять какую-либо деятельность под воздействием персональных социальных требований (внешний аспект), а с другой стороны, из умений и склонностей, имеющих личностное происхождение (внутренний аспект). Следует отметить, что существенной характеристикой компетенции является единовременная мобилизация внутренне и внешне заданных способностей, которая позволяет эффективно действовать в самых разных социально-практических контекстах. Компетенция как конструкт обозначила фундаментальную проблему современного образовательного процесса во всём мире: разрыв между образованием и социальной практикой. Это ознаменовало начало нового этапа в разработке диагностических образовательных инструментов.

В России целью создания нового образовательного диагностического средства ставилась разработка такого комплекса тестов, который, с одной стороны, позволит содержательно характеризовать особенности протекания учебного процесса на разных его этапах, а с другой стороны, станет ориентиром для выстраивания стратегии профессиональной деятельности педагогов и администрации школ. При этом компетенция стала рассматриваться как способность осмысленно и осознанно осваивать культурные образцы действия с целью дальнейшего участия в развитии общества и личностного развития. Результатом образовательного процесса, рассматриваемого в подобном русле, должно стать органичное формирование субъекта продуктивного социального действия.

Базовым для разработки нового диагностического инструмента стало положение, согласно которому, обучение считается одним из важнейших элементов целостного психического развития ребенка, с одной стороны, а также видится процессом освоения развивающейся личностью ключевых знаковых структур, в которых в обобщенном виде заключены способы жизнедеятельности, характерные для данной культуры. Логическим следствием из данного положения является восприятие образовательного процесса как направляемого саморазвития ребенка, облеченного в форму образовательных циклов, на каждом из которых ребенок овладевает тем или иным «слоем» культурных орудий или способов жизнедеятельности. Каждый из «слоев» содержит в себе одновременно три вида ориентиров, которые осваиваются ребенком поступательно от наиболее простого и распространенного к наиболее сложному и доступному ограниченному кругу одаренных представителей данного общества.

Первый вид ориентиров – формальный – заключает в себе внешние характеристики определенной совокупности предметных ситуаций и соответствующих им алгоритмов действий. Мера обобщенности данных образцов жизнедеятельности минимальна и применима к ограниченному количеству однотипных, стандартизированных задач. Овладение данным видом ориентиров достигается за счет культурных образцов, представленных в форме правил и стереотипов действия.

Второй вид ориентиров – рефлексивный – подразумевает формирование у личности понятия о существенном, ключевом отношении, которое лежит в основе того или иного типа предметных ситуаций и позволяет обнаружить и обобщить основополагающий принцип действий в рамках аналогичных обстоятельств. Это становится фактором экстраполяции схемы решения задач на всю совокупность ситуаций, к которым может быть применим выявленный фундаментальный принцип действий. Присвоение ребенком данного вида ориентиров может быть осуществлено на основании культурных образцов, существующих в виде моделей, схем, образов, которые содержат в себе основание сущностных преобразований в той или иной сфере.

Третий вид ориентиров – функциональный – открывает перед личностью возможность сформировать представление о границах применения способа действий, выявить поле и специфику его потенциального осуществления. Данный вид ориентиров является наиболее трудоемким в освоении, но и наиболее широко применимым, т.к. имеет свойство подстраиваться под различные ситуативные контексты. Культурные образцы, необходимые для постижения ребенком данного вида ориентиров, включают в себя разветвленные смысловые структуры, открывающие личности возможность для решения самых разнообразных, в том числе, не относящихся напрямую к той или иной предметной сфере, задач.

Все предметные тесты, в частности, тесты по математической грамотности, включенные в пакет SAM, строятся из тематических задачных блоков, состоящих из трех иерархически выстроенных заданий. Тестовые материалы разрабатываются и анализируются на основе трех предложенных выше уровней освоения способа действия, которые разрабатывались в работах В. В. Давыдова и его последователей. Задание первого уровня является наиболее простым из трех, задание третьего уровня – наиболее сложным. В функциональном плане каждый блок является средством, которое позволяет определить уровень усвоения учеником соответствующего раздела выбранной учебной дисциплины. Уровень определяется по самой сложной задаче, решенной школьником. Систематизировав результаты по каждому тематическому блоку, исследователь получает возможность составить структурную компетентностную модель усвоения участниками исследования школьной программы и сформулировать соответствующие рекомендации по дальнейшему обучению.

Уровень 1. **Формальный**

Уровень 2. **Рефлексивный**

Уровень 3. **Функциональный**

*Рисунок 1. Уровни освоения обобщенного способа действия*

Для создания задач и оценки результатов их решения школьниками разработаны критерии, соответствующие каждому из компетентностных уровней усвоения школьниками образовательной программы по дисциплине.

**Критерии первого (формального) компетентностного уровня**

Наиболее общей характеристикой первого компетентностного уровня является способность школьника действовать по заданному образцу, выбираемому по внешним признакам задания, присутствующим в условии. Школьник, усвоивший данный компетентностный уровень, успешно применяет стандартные схемы и правила к соответствующим типам задач, а также выполняет алгоритмы, которые описаны в задаче прямым текстом.

Задачи-индикаторы:

* содержат ясно сформулированные условия, однозначно указывающие на их принадлежность к типу задач, которые имеют стандартный, универсальный способ решения;
* содержат подсказки, которые напрямую ведут учащегося к выбору верной схемы решения;
* содержат свойства и отношения, существенные для определения алгоритма решения, в очевидных внешних признаках (условиях).

**Задачная ситуация**

**Схема действия**

*Рисунок 3. Ориентировка действия на формальном уровне*

**Критерии второго (рефлексивного) компетентностного уровня**

Ключевая обобщающая характеристика второго компетентностного уровня – это способность школьника проводить содержательный анализ проблемной ситуации, представленной в задаче, а также сопоставлять освоенные схемы и алгоритмы действий с представлением о том, что является смысловым ядром задачи. Задания на данном уровне могут быть решены учеником только при условии верной интерпретации и исчерпывающего понимания представленных формулировок. Речь идет о выделении основополагающего отношения/признака/свойства из всех имеющихся условий, на основе которого школьник определяет стратегию решения задания.

Задачи-индикаторы:

* не содержат условий-подсказок, которые открывают перед учеником возможность использовать в решении стандартные схемы;
* требуют понимания их предметной сущности и специфики для верного выбора направления в поиске решения;
* предполагают наличие усложненных внешних признаков, которые затрудняют отнесение их к тому или иному типу: условия требуют преобразования, условия представлены в абстрактной форме, требующей конкретизации, условия даны “от обратного” и т. д.

**Схема действия**

**Задачная ситуация**

**Существенное**

**отношение**

*Рисунок 4. Ориентировка действия на рефлексивном уровне*

**Критерии третьего (функционального) компетентностного уровня**

Основной компетенцией данного уровня является способность школьника уверенно ориентироваться в пространстве потенциальных схем осуществления усвоенного на предыдущих уровнях способа действия, обобщать, преобразовывать и расширять его за счет ясного видения возможностей и границ его эффективного применения.

Задачи-индикаторы:

* содержат условия, предполагающие восстановление и реорганизацию сущностно значимых смыслов, а также последующую экспериментальную апробацию различных вариантов осуществления способа действия с целью выбора наиболее оптимального из них;
* предполагают подробный содержательный анализ применимости имеющихся схем/формул/алгоритмов решения на предмет их эффективности;
* требуют определения точно очерченных границ использования основного способа решения для последующего выбора дополнительных вспомогательных инструментов;
* имеют несколько содержательно-смысловых пластов, при этом центральный способ действия является инструментом решения на наиболее высоком уровне;
* включают в себя необходимость творческой адаптации способа действия к уникальным условиям, содержащимся в описании предметной ситуации.

**Схема действия**

**Существенное**

**отношение**

**Схема действия**

**Схема действия**

**Задачная ситуация**

*Рисунок 5. Ориентировка действия на функциональном уровня*

**Общие структурно-конструктивные характеристики теста SAM**

основой теста являются два подхода: нормативно-ориентированный (позволяющий сравнивать между собой результаты разных участников тестирования и результаты одного участника, полученные на разных этапах его обучения, посредством определения интегрального балла) и критериально-ориентированный (позволяющий качественно оценить компетентностный уровень усвоения предметного содержания учебной дисциплины участником посредством ступенчатой шкалы достижений, состоящей из интегральных баллов);

* все задачи теста (45) объединены в блоки (15);
* каждый блок состоит из трех задач, выстроенных в иерархической последовательности по уровню сложности;
* задачи предъявляются школьнику по блокам, при этом порядок предъявления может быть случайным;
* тесты могут быть предложены школьнику в бланковой или компьютерной форме;
* совокупность задач одного уровня из каждого блока является субтестом, в котором представлены темы из различных разделов образовательной программы;

**ТЕСТ**

**Блок 2**

**Блок 3**

**Блок N**

**Блок 1**

**Задания 1-го уровня (= субтест 1)**

**Задания 2-го уровня (= субтест 2)**

**Задания 1-го уровня (= субтест 3)**

*Рисунок 6. Структура теста*

* в тест включены задания различных форм: открытой (с кратким ответом), закрытой (с выбором одного ответа из нескольких предложенных), задания с графическим ответом (требующие построений);
* рекомендуемое время выполнения теста – 90 минут (два урока по 45 минут);
* проведение тестирования может быть осуществлено в два этапа: 8 блоков – в первый день, 8 блоков – во второй;
* примерное время выполнения заданий первого уровня – 1 минута, второго уровня – 2 минуты, третьего – 3 минуты;
* тест разработан в нескольких вариантах, в каждом варианте использованы близкие статистические характеристики и представлены не менее шести одинаковых заданий;
* при оценке решения используется дихотомия, где 1 – верный ответ, 0 – неверный ответ;
* за каждый субтест, демонстрирующий компетентностный уровень учащегося в усвоении учебной дисциплины, участник тестирования может набрать максимум 15 баллов, за тест в целом – 45 баллов.

**1.2. Содержание теста по математике**

При формулировке и отборе тестовых заданий создатели ориентировались на стандарты общего образования в выбранных для исследования странах. В целом, содержание теста по математике для начальной школы сфокусировано вокруг понятия “натуральное число” и включает в себя всевозможные вариации в рамках данной тематики (порядок цифр в числе, разряды многозначного числа, свойства чисел и действий, операции с числами и последовательность их осуществления). Несмотря на то, что тематически тест дифференцирован на пять обособленных разделов, каждый из них основывается на вычислениях, даже если речь идет о геометрическом (пространственном) материале. Выделенные области знания (таблица 1) включают в себя содержание образовательных программ по математике в начальной школе, реализуемых как в отечественных, так и в зарубежных учебных заведениях. Количественное соотношение заданий из каждого раздела представлено в таблице 2.

Таблица 1. Содержание теста по математике (SAM: Инструмент мониторинга учебных достижений школьников. Framework, 2011)

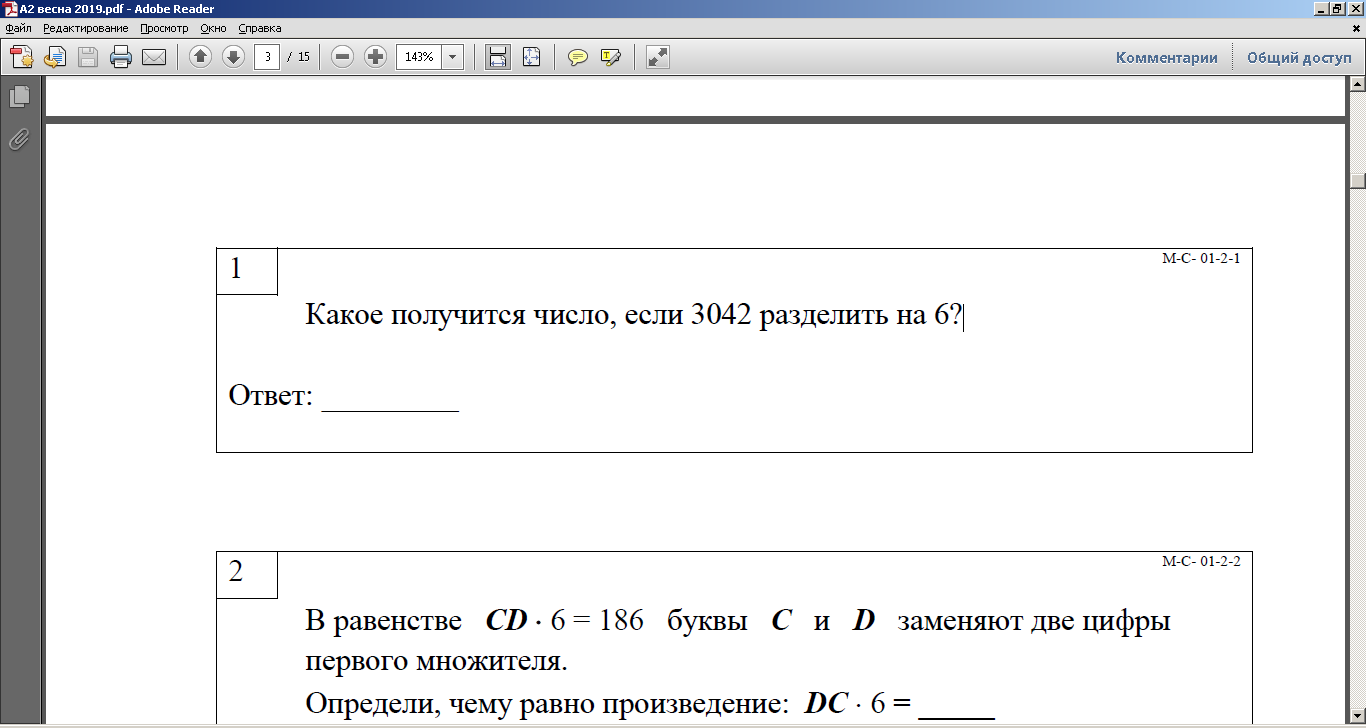
|  |  |  |
| --- | --- | --- |
| **Разделы содержания** | **Средства ориентировки математических действий** | **Математические действия** |
| Числа и вычисления | * последовательность натуральных чисел; * числовая прямая; * позиционный принцип; * свойства арифметических действий; * порядок действий. | * сравнение многозначных чисел; * выполнение алгоритмических действий с многозначными числами; * прикидка; * элементы рационального счета. |
| Измерение величин | * отношение между числом, величиной и единицей; * отношение “целого и частей”; * формула площади прямоугольника. | * прямое измерение длин линий и площадей фигур (непосредственное «укладывание» единицы, «укладывание» единицы с предварительной перегруппировкой частей объекта); * косвенное измерение (измерение с помощью приборов, вычисление по формулам). |
| Закономерности | * “индукционный шаг”; * повторяемость (периодичность). | * выявление закономерности в числовых и геометрических последовательностях и других структурированных объектах; * вычисление количества элементов в структурированном объекте. |
| Зависимости | * отношения между однородными величинами (равенство, неравенство, кратности, разностное, “целого и частей”); * прямая пропорциональная зависимость между величинами; * производные величины: скорость, производительность труда и пр.; * соотношения между единицами. | * решение текстовых задач; * описание зависимостей между величинами на различных математических языках (представление зависимостей между величинами на чертежах, схемами, формулами и пр.); * действия с именованными числами. |
| Элементы геометрии | * форма и другие свойства фигур (основные виды геометрических фигур); * пространственные отношения между фигурами; * симметрия. | * распознавание геометрических фигур; * определение взаимного расположения геометрических фигур. |

Таблица 2. Соотношение разных заданий в тестовой тетради

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Разделы** | **Количество блоков** | **Количество задач** | | | |
| **ВСЕГО** | **1-й уровень** | **2-й уровень** | **3-й уровень** |
| Числа и вычисления | 3 | 9 | 3 | 3 | 3 |
| Измерение величин | 4 | 12 | 4 | 4 | 4 |
| Закономерности | 2 | 6 | 2 | 2 | 2 |
| Зависимости | 4 | 12 | 4 | 4 | 4 |
| Элементы геометрии | 2 | 6 | 2 | 2 | 2 |
| **ВСЕГО** | **15** | **45** | **15** | **15** | **15** |

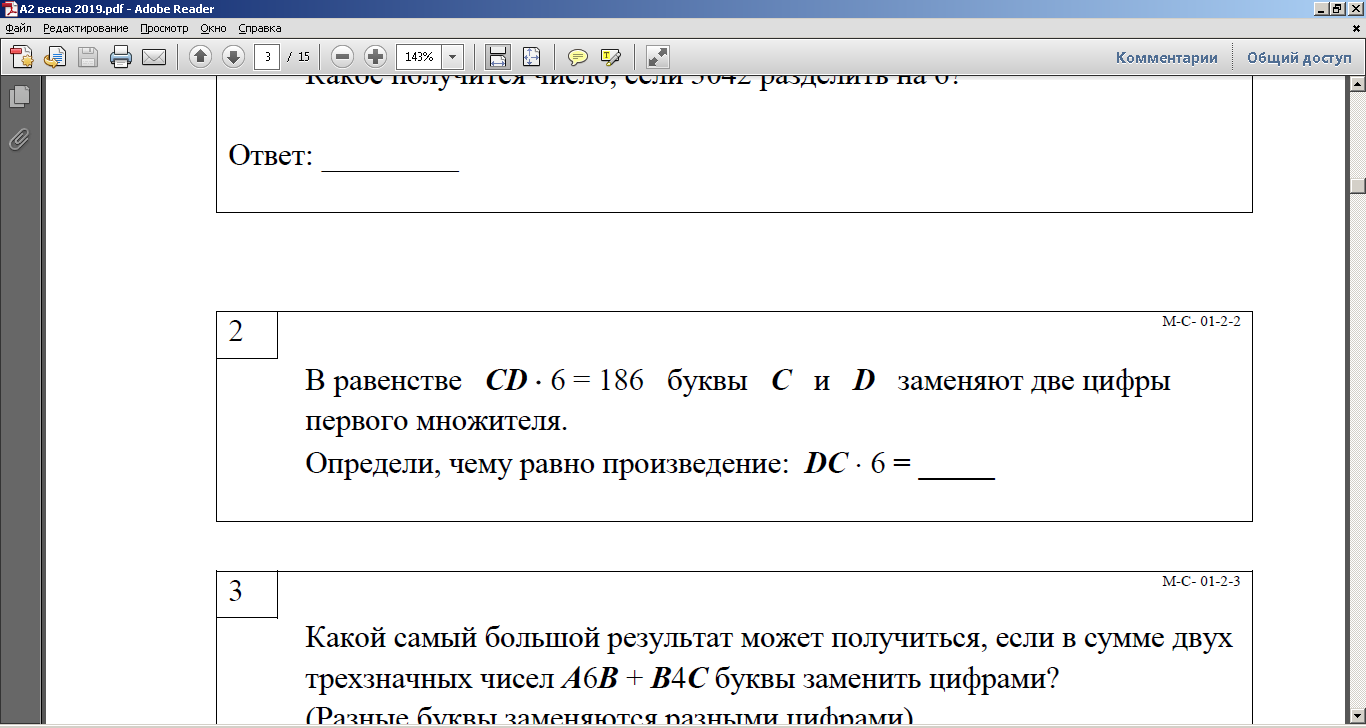
В разделе “***Числа и вычисления***” задания первого уровня подразумевают единичное выполнение простых арифметических действий (деление, умножение) и применение знаний относительно состава многозначного числа.

Пример задания:



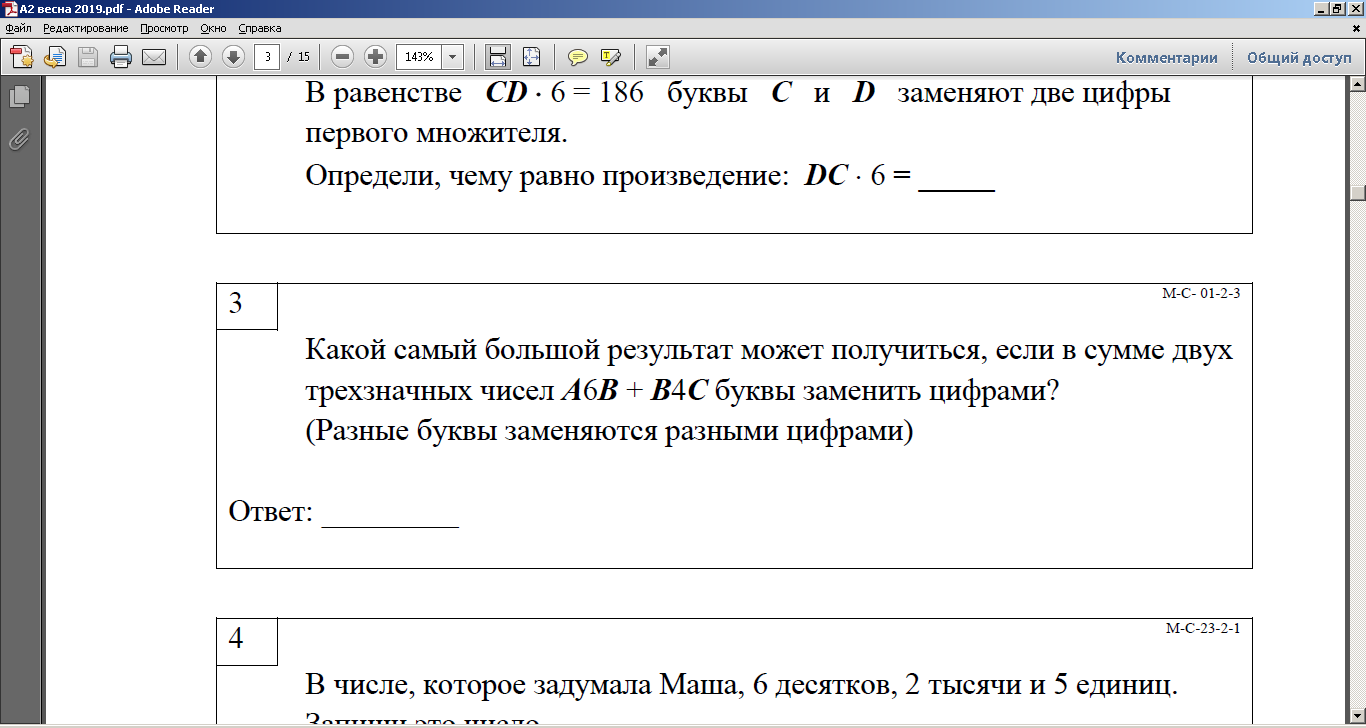
Задания второго уровня предполагают осуществление нескольких последовательных разнохарактерных арифметических действий с целью приведения числа к виду, который позволит школьнику применить известный алгоритм решения задачи и вычислить правильный ответ.

Пример задания:



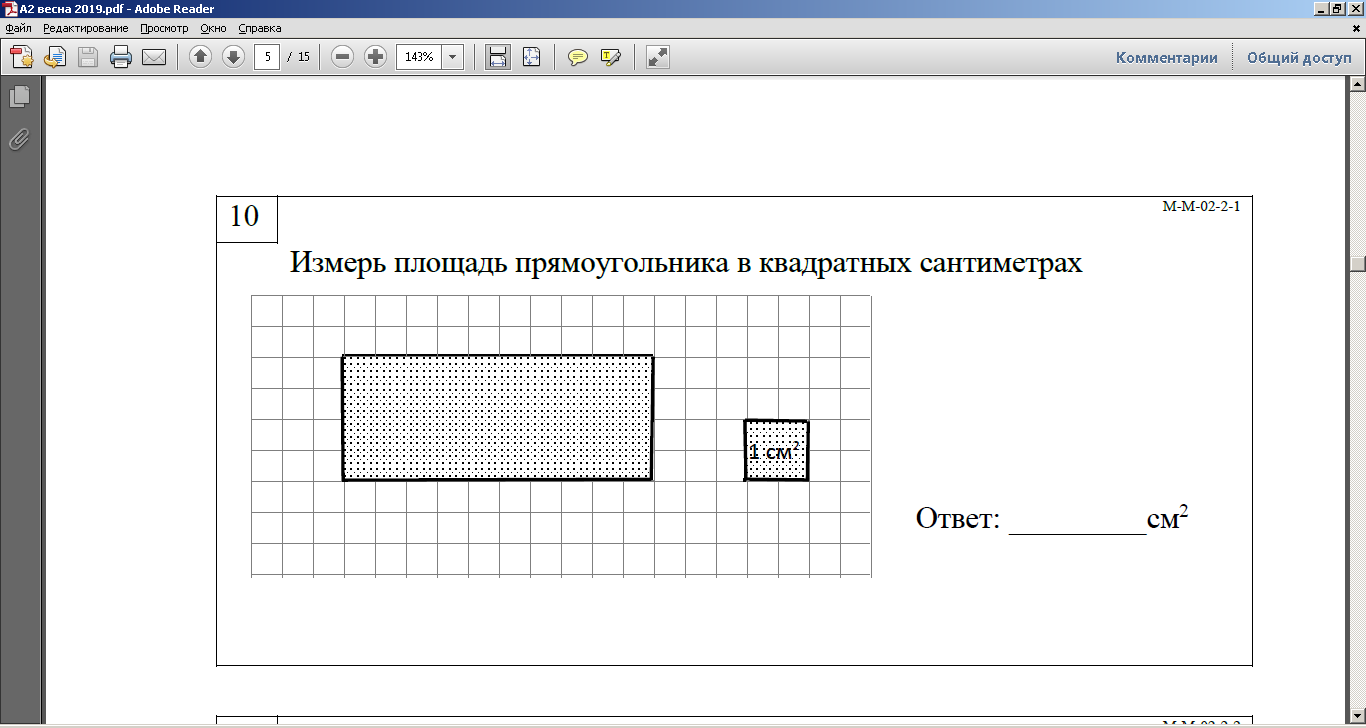
Для решения заданий третьего уровня школьник должен владеть знаниями о качественных особенностях строения многозначного числа (его свойствах и разрядах), специфике арифметических действий в контексте использования для решения задачи метода подбора ответов.

Пример задания:



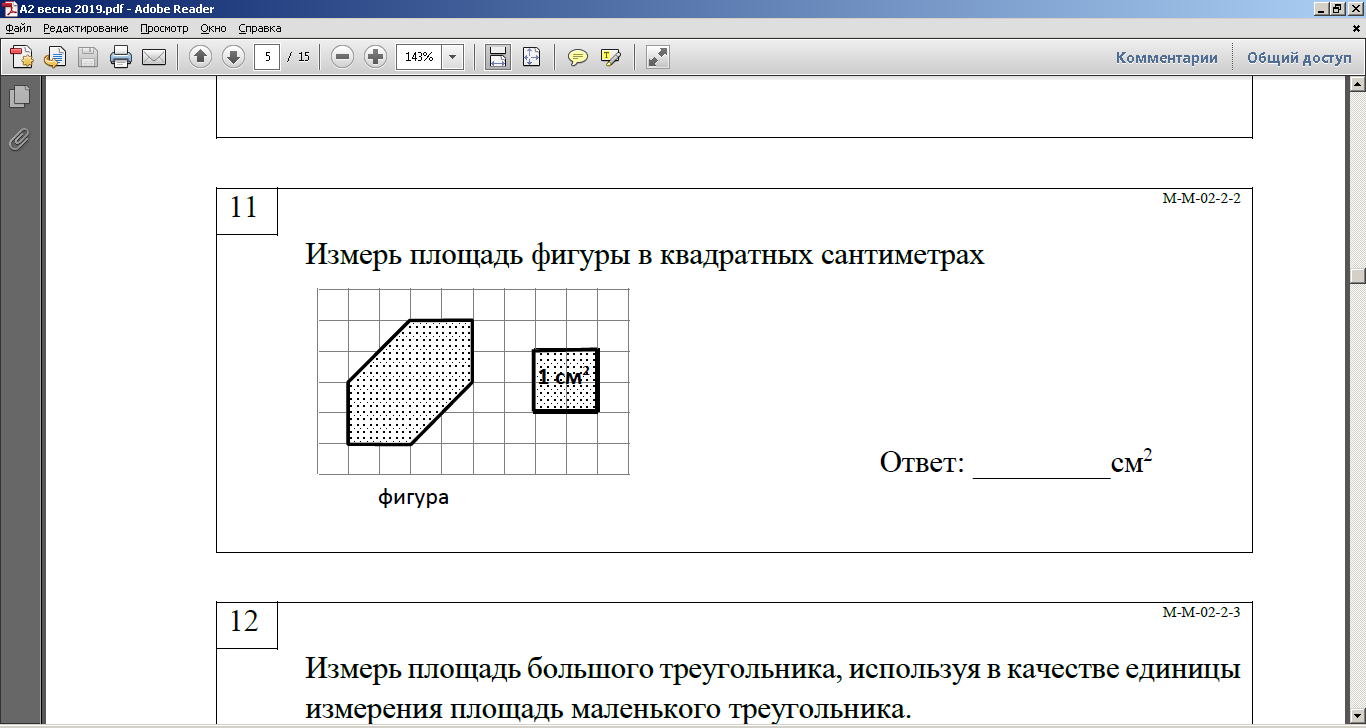
В разделе “***Измерения величин***” задания первого уровня связаны с простыми измерениями, которые могут быть осуществлены посредством а) использования известных измерительных приборов (линеек); б) применения арифметических действий к указанным/изображенным графически единицам измерения; в) использования простейших формул.

Пример задания:



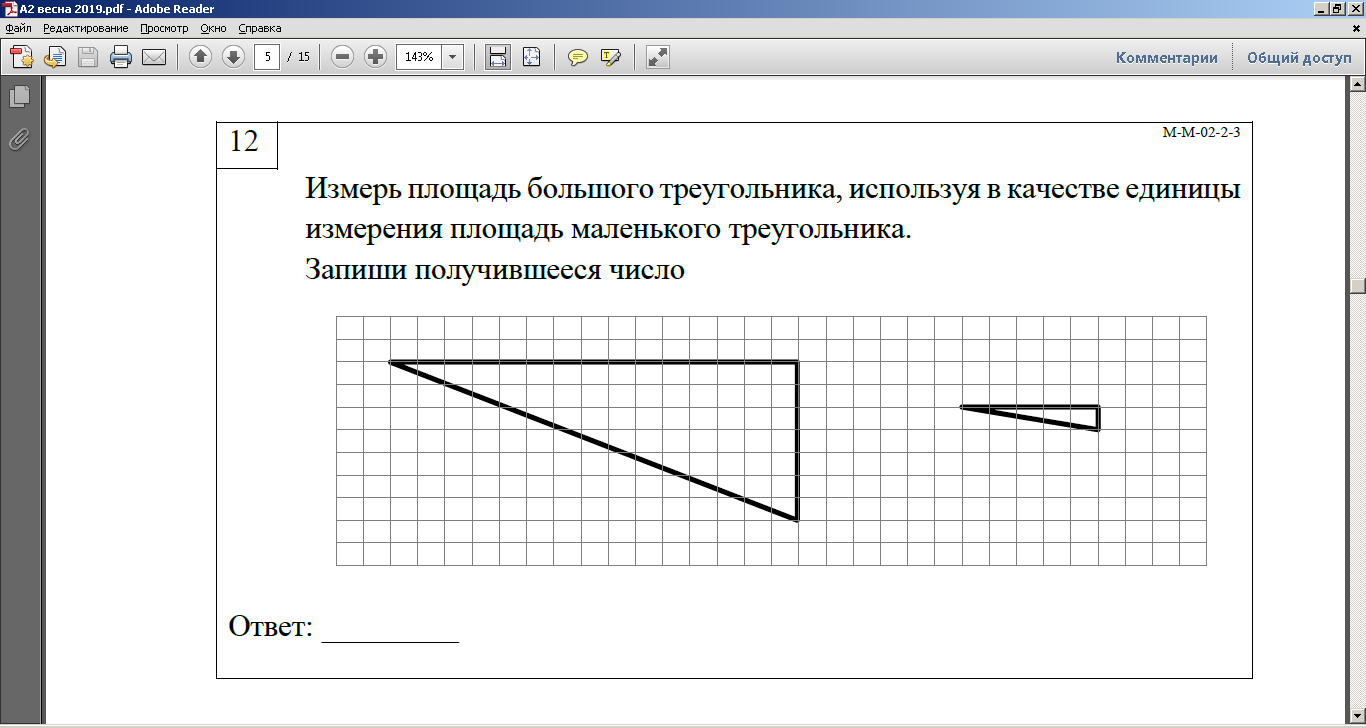
Задачи второго уровня предполагают осуществление нескольких действий, первое из которых имеет целью приведение предметной ситуации к типовой, подразумевающей применение стандартного алгоритма решения.

Пример задания:



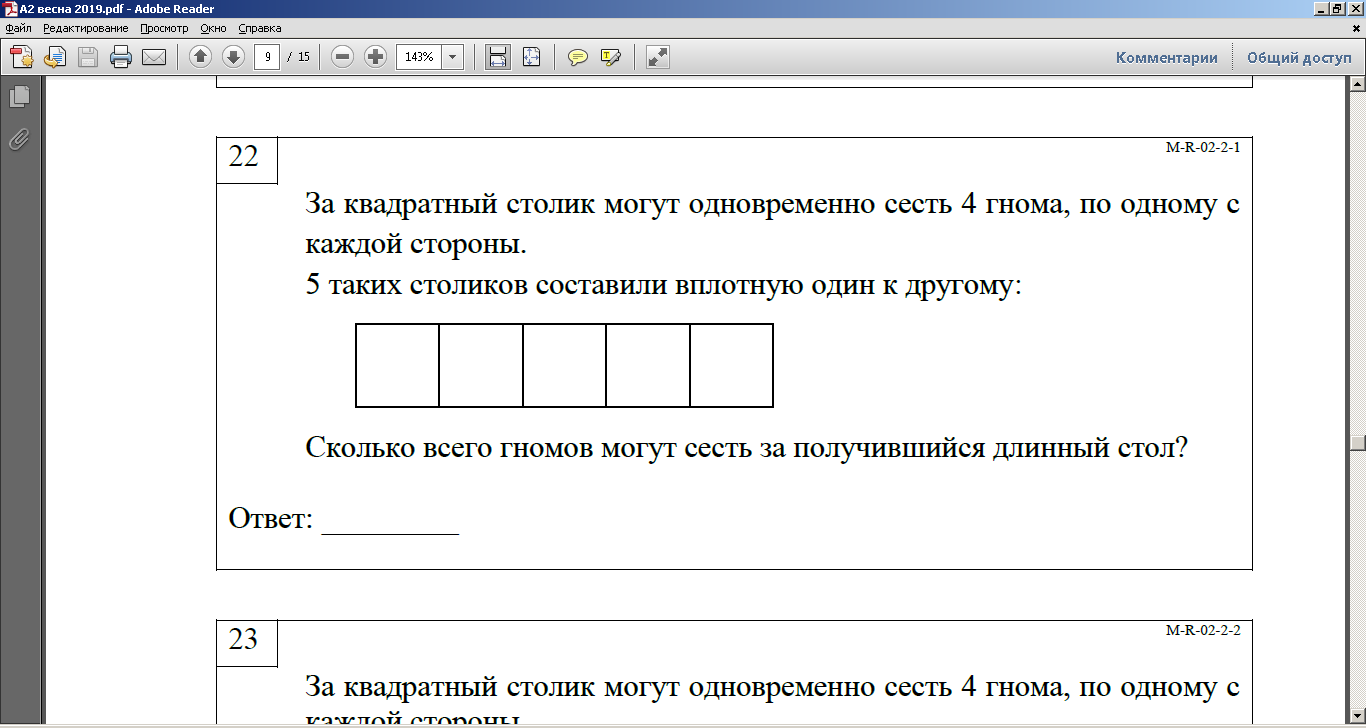
На третьем уровне в заданиях содержатся указания использовать для поиска решения нестандартные измерительные единицы, предлагается составить формулу для измерения на основе некоторых известных данных, а также проводить абстрактные измерения с помощью анализа графических изображений.

Пример задания:



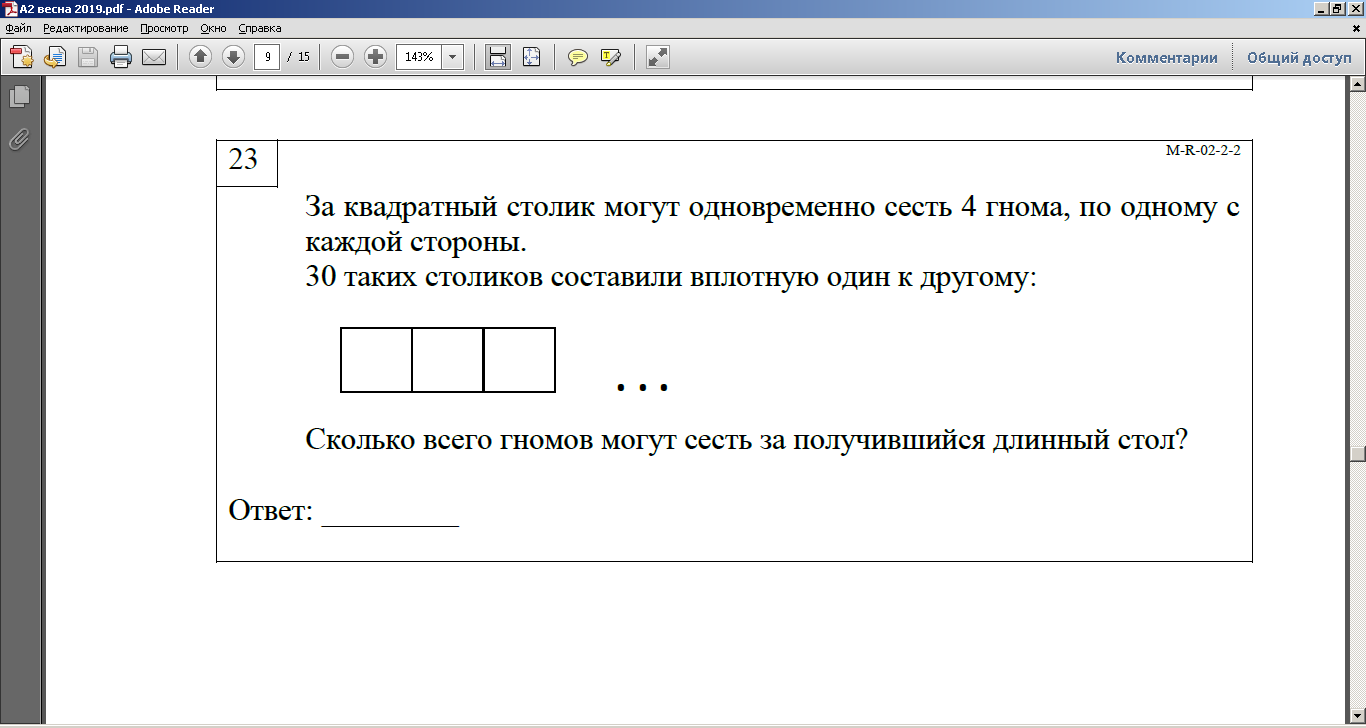
В разделе “***Закономерности***” задачи первого уровня содержат в себе простые условия, которые позволяют на основе внешних признаков (точно определенных “шага” и структуры последовательности) прийти к верному решению.

Пример задания:



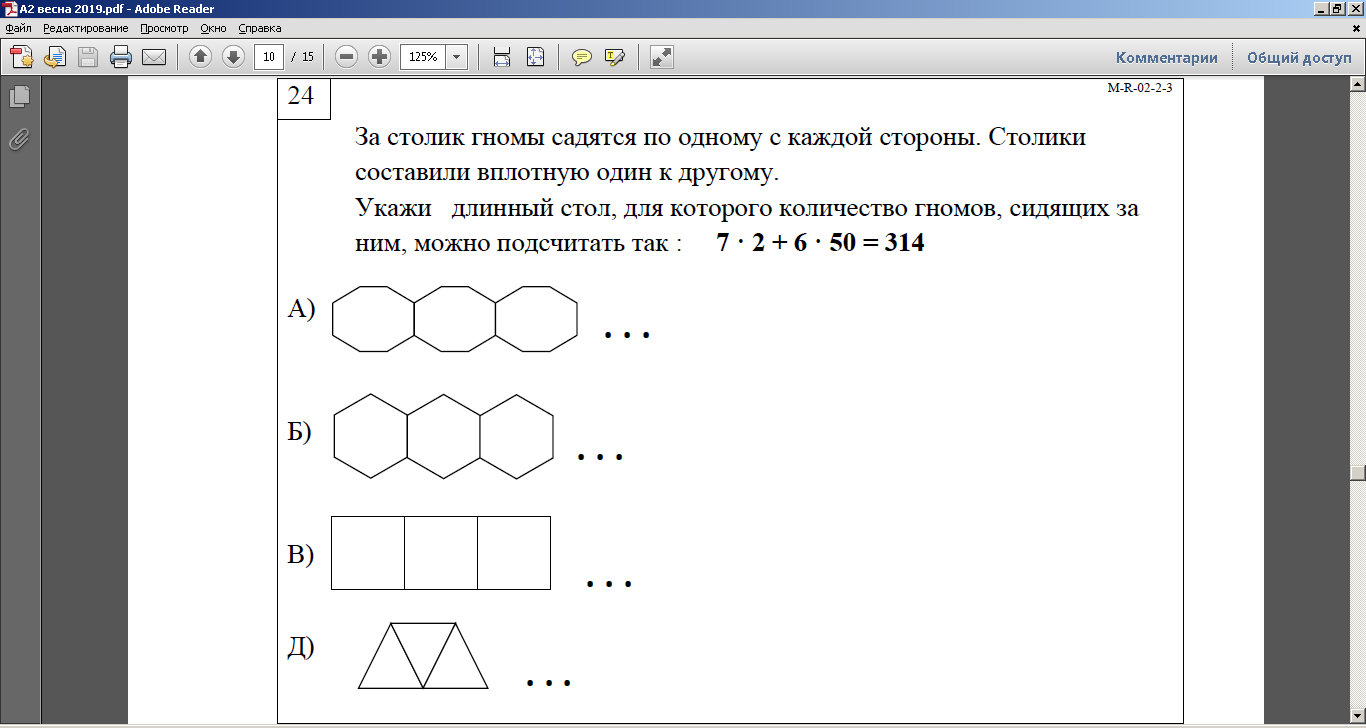
Задания второго уровня могут быть решены школьником в случае, если ему удастся выявить “скрытую” закономерность, которая является основополагающей для выбора верного направления поиска ответа, и применить её на том отрезке последовательности, который не изображен в задаче.

Пример задания:



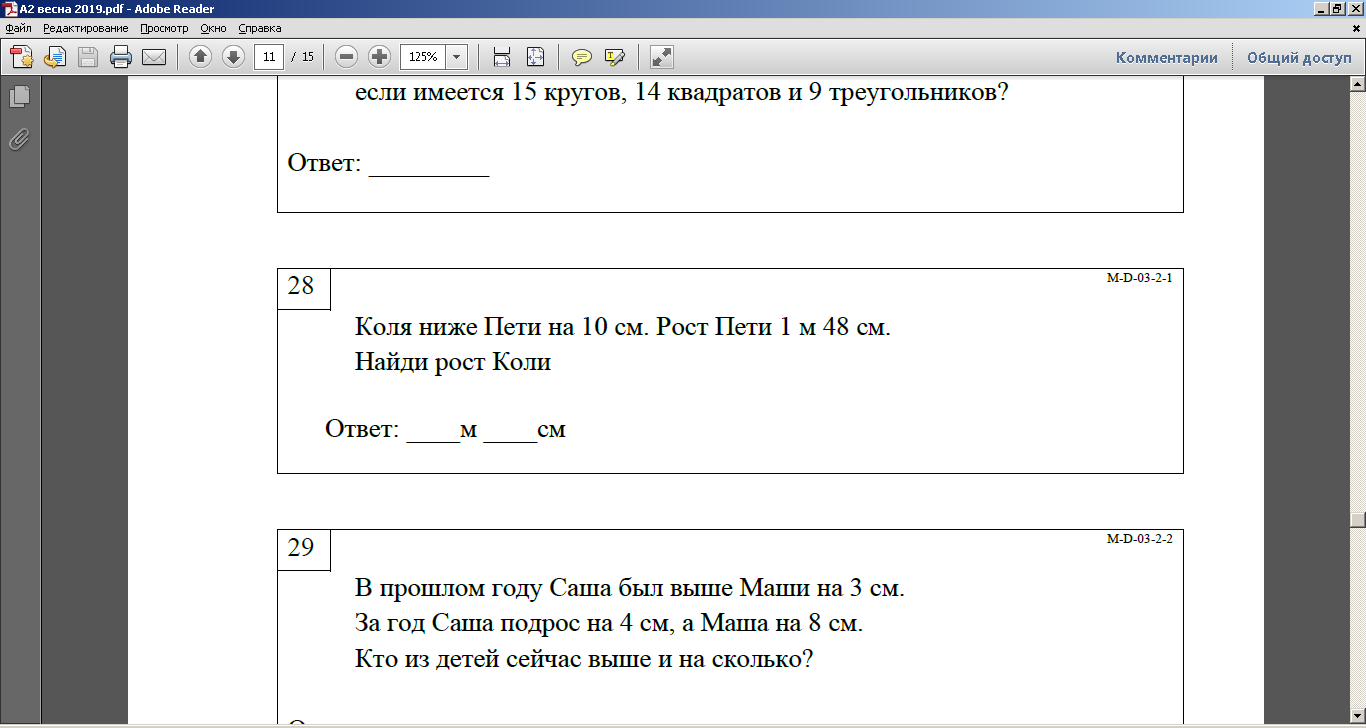
Для решения задач на третьем уровне школьнику необходимо ясно понимать специфику нестандартного “шага” закономерности и на этой основе определить формулу для вычисления правильного ответа.

Пример задания:



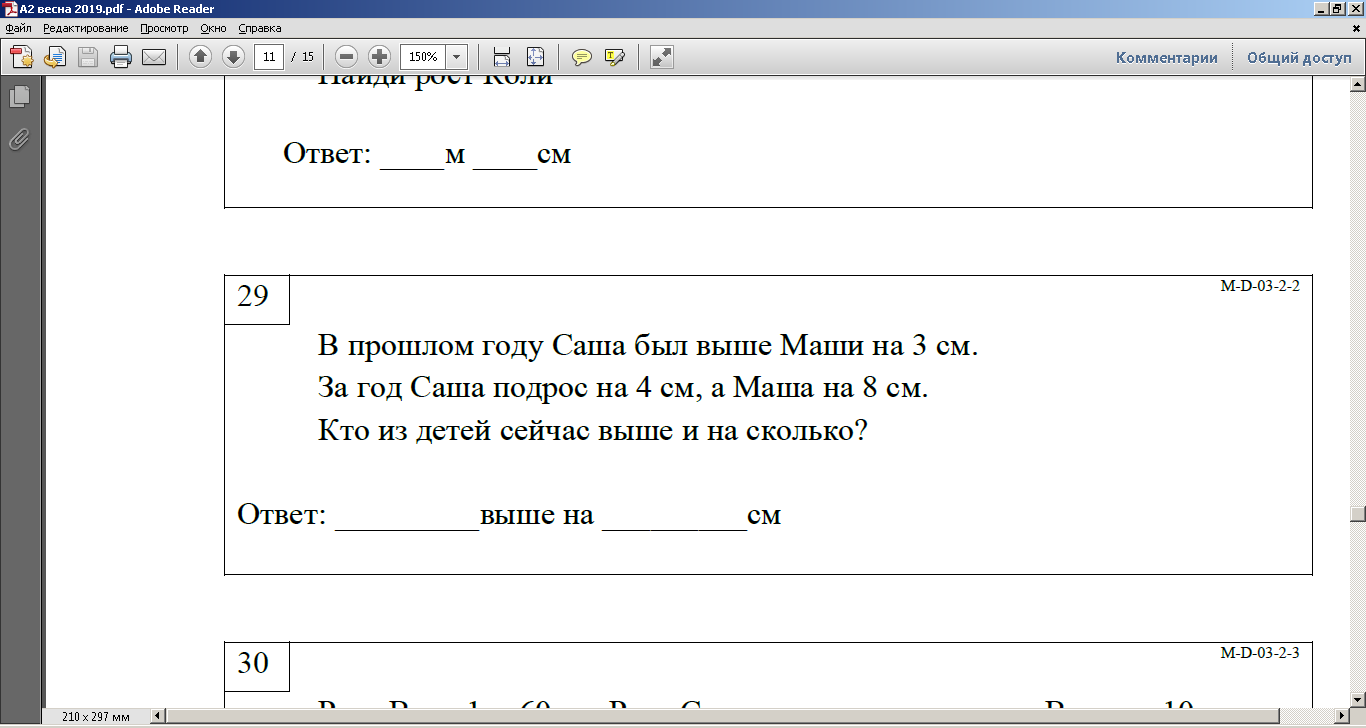
В разделе “***Зависимости***” первый компетентностный уровень представлен простыми текстовыми задачами, условия которых ясно демонстрируют ключевые отношения, необходимые для поиска верного ответа.

Пример задания:



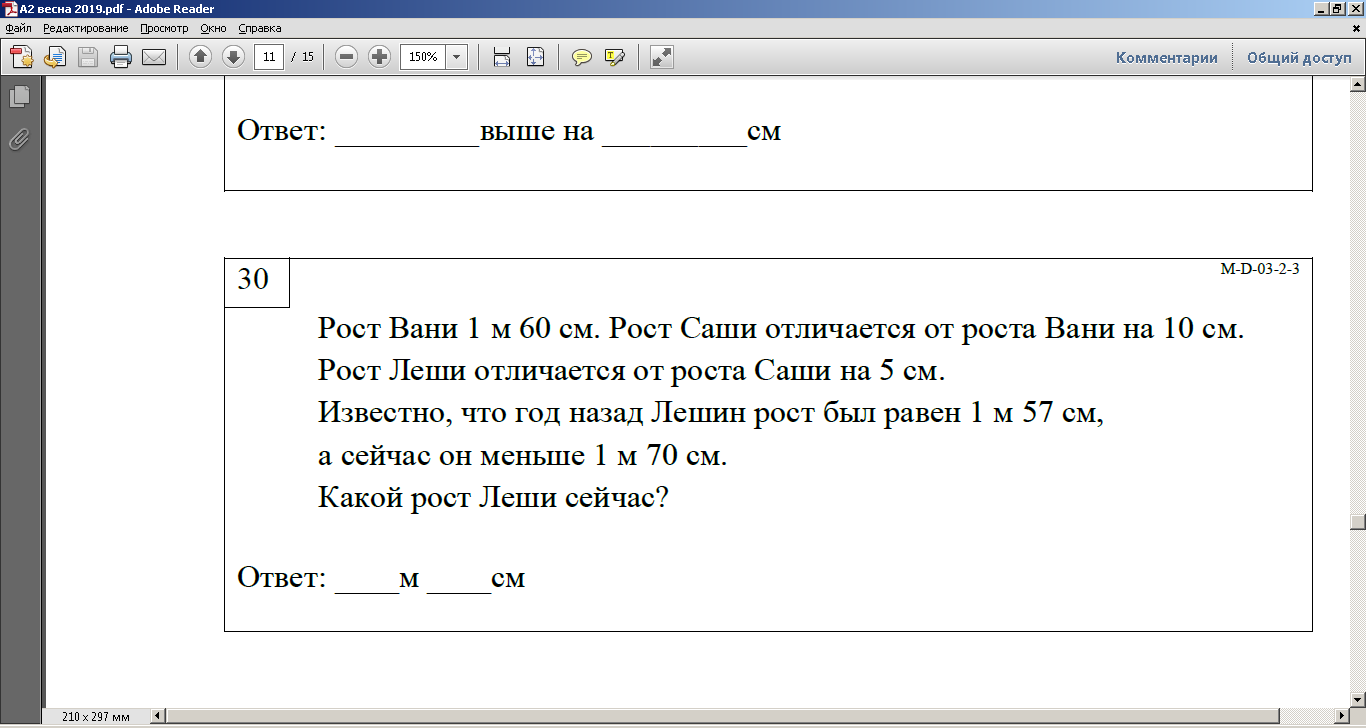
Задания второго уровня предполагают выделение сущностного отношения из нескольких представленных в описании условий и выполнение нескольких последовательных действий на основании выявленной специфики.

Пример задания:



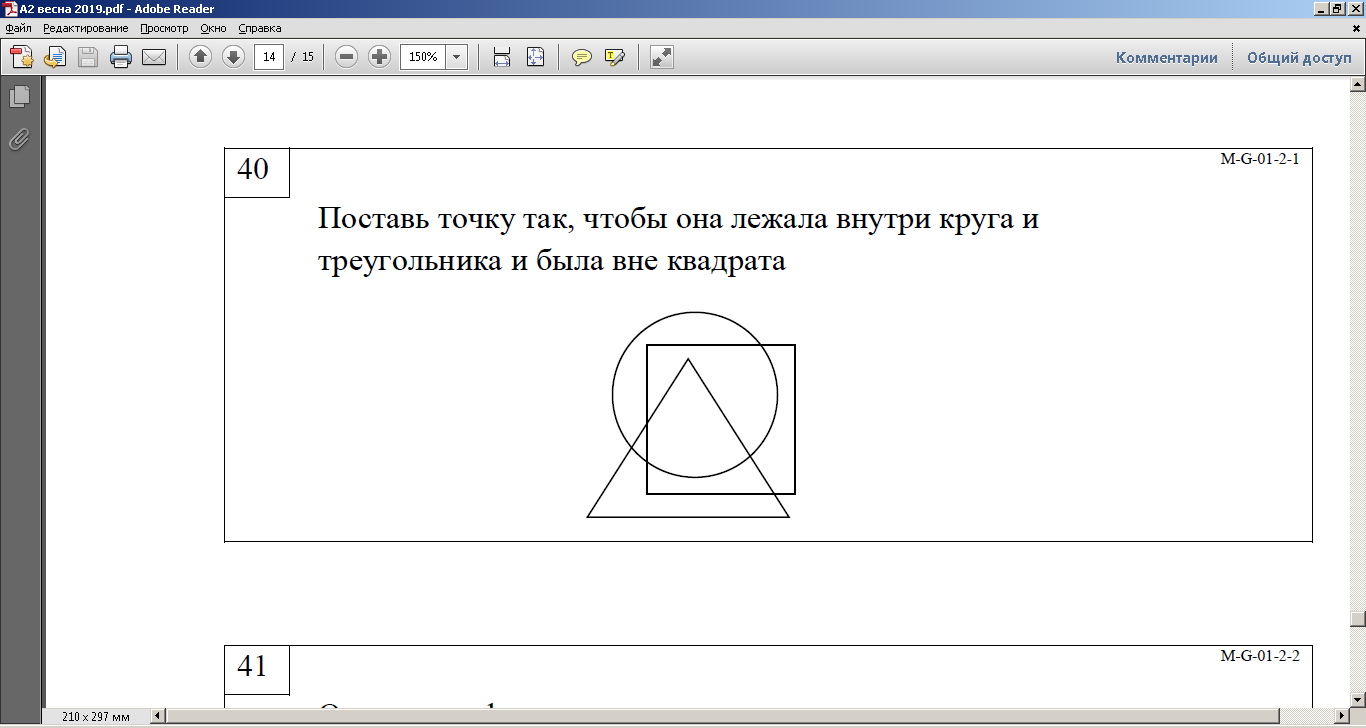
В задачах третьего уровня школьник должен выполнить арифметическое действие, учитывая одновременно несколько одинаково значимых отношений, представленных в условиях.

Пример задания:



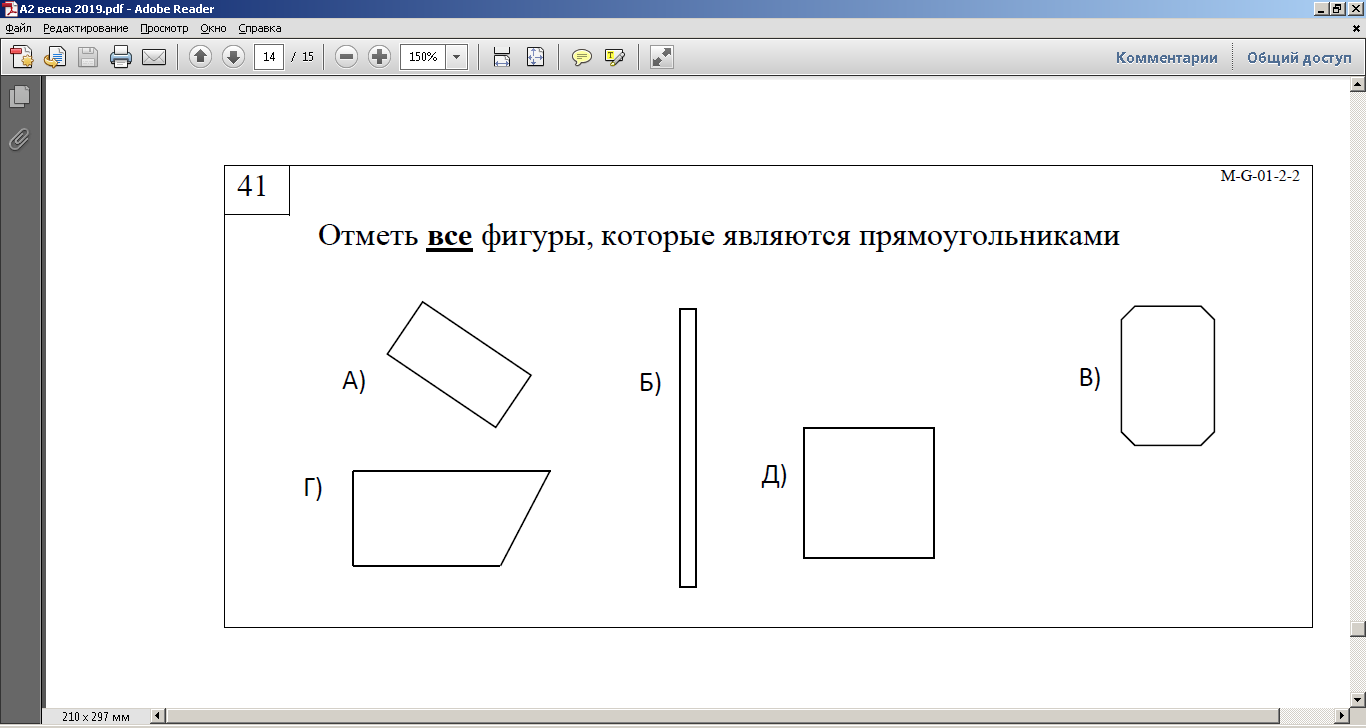
В разделе “***Элементы геометрии***” задания первого уровня направлены на применение знаний о простых геометрических фигурах.

Пример задания:



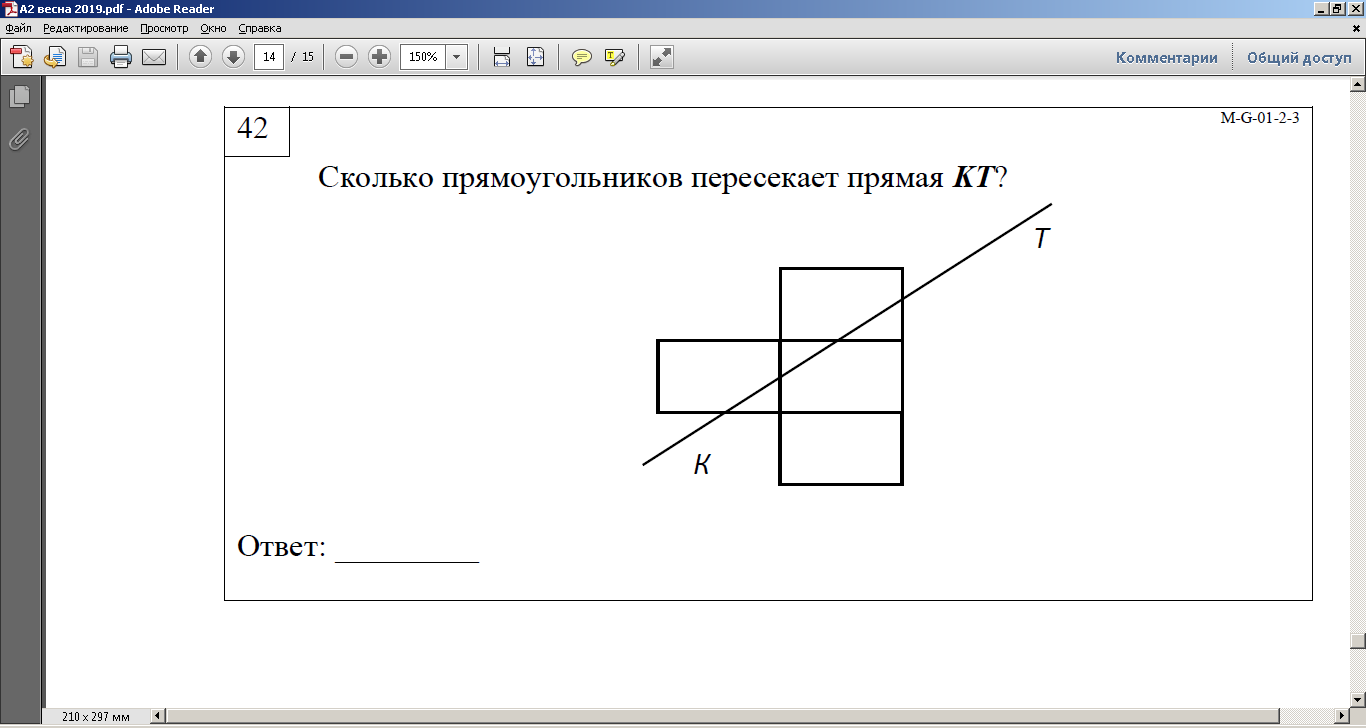
Задачи второго уровня содержат в себе изображения геометрических фигур, имеющих нестандартную форму или положение, что усложняет поиск и выбор правильного ответа учащимся.

Пример задания:



Задания третьего уровня предполагают, что школьник обладает навыками использования абстрактного и пространственного мышления.

Пример задания:



ГЛАВА 2. ОПРОСНИКИ ДЛЯ СБОРА КОНТЕКСТНОЙ ИНФОРМАЦИИ

Исследование уровня математической грамотности SAM включает в себя комплект контекстных анкет, состоящих из опросника для учителя и опросника для ученика. Контекстные анкеты разработаны специалистами Института Образования НИУ «Высшая Школа Экономики».

Разработка данного элемента диагностического образовательного инструмента SAM была осуществлена в связи с необходимостью, во-первых, получить информацию о фоновом контексте, в котором находятся испытуемые посредством сбора социологических, организационных и социо-культурных данных. Совокупность подобных материалов, обработанная посредством статистического анализа, позволяет сделать вывод о том, какие факторы являются особенно значимыми для достижения академической успешности в данной школе, в данном классе. Данный аспект является значимым преимуществом SAM, т.к. в результатах анкетирования педагог может обнаружить полезную информацию, сбор которой, как правило, не осуществляется школьными специалистами самостоятельно.

Кроме того, следует отметить, что идейное ядро инструмента SAM базируется на стремлении создать связующее звено между диагностикой и образовательной практикой, т.е. разработать такое диагностическое средство, которое станет для работников школы способом ориентации в актуальном образовательном процессе в данной школе, в данном классе. В этом смысле формирование контекстных опросников разработчиками SAM было обусловлено потребностью школьных специалистов иметь набор ясно определенных маркеров, посредством которых можно было бы направлять образовательный процесс в то или иное русло. Анализ результатов анкетирования позволяет сделать выводы о сущностных характеристиках образовательной среды, которые требуют развития (если оказывают позитивное, конструктивное влияние на ход и участников процесса обучения) или коррекции (если оказывают негативное, деструктивное влияние процесса обучения).

Каждая анкета в комплекте имеет собственные содержательные и структурные особенности, заданные для того, чтобы получить возможность объемного рассмотрения образовательного процесса (глазами ученика и глазами учителя). Подобный подход к диагностике в сфере образования видится эффективным, т.к. предоставляет шанс приобрести более достоверную информацию о ходе и специфике обучения, принимать осмысленные педагогические и административные решения, а также наметить наиболее пути совершенствования образовательного процесса.

**2.1. Особенности разработки контекстных анкет**

Создание контекстных опросников в их первоначальном формате осуществлялось с опорой на смежный опыт авторов диагностического инструмента PISA, которые фокусировали свое внимание на решении аналогичной проблемы – сборе дополнительной информации о характеристиках образовательного процесса – но на более широкой целевой аудитории (основная школа). В этом смысле значимым преимуществом SAM является ориентация на узкую группу потенциальных респондентов (начальная школа) и, как следствие, более детальная проработка элементов анкеты.

В центре сформированных анкет лежат следующие основополагающие гипотетические представления о специфике культурно-образовательной среды как ключевого фактора академической успешности учащегося начальной школы:

1. Культурно-образовательная среда – это многогранный, создаваемый взрослыми конструкт, который складывается из взаимоотношений, в которые ребенок вступает в школе (с ровесниками, старшеклассниками, учащимися младших классов, учителями, администрацией школы) и вне школы (иное социальное окружение ребенка – родственники, друзья и знакомые, единомышленники и т. д.).

2. Культурно-образовательная среда может быть рассмотрена как целостность, состоящая из двух уровней. На содержательном уровне культурно-образовательная среда включает в себя комплекс учебных дисциплин, внешкольных занятий, внеучебной игровой и творческой деятельности, информационных потоков, культурных образцов. Перечисленная совокупность факторов создают ядро системного культурного контекста, в который учащийся погружен постоянно. Иные элементы являются случайными или же заданными социально-демографическими параметрами (место проживания, состав и специфические особенности семьи и т. д.). На организационном уровне культурно-образовательная среда может быть рассмотрена как коммуникация учащегося со взрослыми (учителями, администрацией школы, родственниками) и другими школьниками разного возраста. Предполагается, что общение вне школьных стен носит более эмоциональный, интимный характер, т.к. не имеет прямой ассоциации с ведущим видом деятельности ребенка – обучением. В свою очередь, общение в школьной среде опосредовано специфическими формами и структурами, которые задаются учебной ситуацией, конструируемой работниками школы.

3. Немаловажными свойствами культурно-образовательной среды являются осознанность деятельности как школьных специалистов, так и взрослого внешкольного окружения ребенка, в процессе актуализации и использования тех или иных средств и инструментов для формирования культурно-образовательной среды, а также ценностная солидарность взрослых в этом процессе.

4. Качество и сущностные особенности культурно-образовательной среды детерминируют способность ученика начальной школы к решению тестов по математической грамотности на высоком уровне.

5. Для осуществления конструктивного, развивающего влияния культурно-образовательной среды на академическую успешность ребенка она должна создавать стимулирующие условия и позволять ребенку проявлять инициативу, совершать поисковую деятельность, взаимодействовать и вступать в диалог с другими детьми и взрослыми, получать эмоциональную и содержательную поддержку и помощь.

**Учебные дисциплины**

**Культурные образцы**

**Внешкольные занятия**

**Внеучебная деятельность**

**Информационные потоки**

**СОДЕРЖАТЕЛЬНЫЙ УРОВЕНЬ**

**УЧАЩИЙСЯ**

**Взаимоотношения**

**в**

**школе**

**Взаимоотношения**

**вне**

**школы**

**ОРГАНИЗАЦИОННЫЙ УРОВЕНЬ**

*Рисунок 7. Структура культурно-образовательной среды*

На основании перечисленных теоретических положений были также выделены центральные требования к структуре и содержанию анкет (набору респондентов, выбору тем, типам вопросов):

1. Испытуемые должны быть подобраны таким образом, чтобы позволить исследователю извлечь достаточно полную и разнохарактерную информацию о специфике культурно-образовательной среды, в которой находятся школьники-участники тестирования.

2. Содержательно анкеты должны состоять из вопросов, которые позволяют исследователю получить комплексную и многообразную информацию о разных формах активности учащегося в школе и за ее пределами. Это могут темы, связанные с кругом общения ребенка и сферами его интересов, причем, как практических, так и информационных.

3. Структурно вопросы анкеты должны базироваться на представленных выше гипотетических положениях и включать в себя термины, отражающие различные формы активности, с одной стороны, и просьбу оценить степень их значимости, привлекательности, длительности, предпочтительности, во-вторых.

В первоначальной версии анкета для учеников состояла из 22 вопросов, которые были условно разделены на три тематических блока: первый блок вопросов был посвящен учебному процессу, второй блок – внеучебной жизни в школе, третий блок – внешкольной деятельности. Информация, полученная из ответов учащихся на первый блок вопросов, составляла характеристику образовательного процесса и включала в себя данные о степени содержательной насыщенности, особенностях коммуникации в учебном коллективе, специфике эмоционального климата в классе и в школе. Ответы учащихся на второй блок вопросов расширяли представления о культурно-образовательной среде школы, сформированные на основании материалов, полученных из первого блока. Третий блок вопросов был направлен на определение степени включенности семьи в функционирование культурно-образовательной среды. В каждом из тематических блоков присутствовали вопросы фактического (например, «Как часто на уроках ваш учитель снижает отметку за исправления или за небрежную запись в тетради?») и субъективно-оценочного типа (например, «Бывает ли, что на уроках тебе скучно или не интересно?»). При этом следует отметить, что содержательно вопросы фактического характера повторялись в анкетах учеников и учителей, что позволяло получить более достоверную и объемную информацию о специфике культурно-образовательной среды, во-первых, и выявить ее противоречия, несовершенства, а значит – наметить оптимальные пути развития, во-вторых. Вопросы фактического характера предоставляли исследователю возможность на основе анализа полученных результатов анкетирования сделать вывод об особенностях образовательного процесса как воздействия, вопросы субъективно-оценочного характера – о том, как учащийся чувствует себя в связи с подобным воздействием (насколько ему эмоционально комфортно, насколько он мотивирован учиться).

Анкета для учителей состояла из 30 вопросов, и была аналогично ученической разделена на четыре группы вопросов по следующим темам: а) уроки (организация учебного процесса и общения с детьми); б) организация внеурочной жизни школы; в) профессиональные и личностные установки, а также общение с коллегами; г) общение с родителями. Анализ ответов на вопросы из каждого блока производится с опорой на ответы учеников, что позволяет выстроить расширенную и связную картину, культурно-образовательной среды. Так же, как и в опроснике для учеников, в текст анкеты для учителей были включены вопросы фактического характера (например, «Как часто на своих уроках Вы организуете работу детей в небольших группах?»), так и субъективно-оценочные (например, «Когда Вы испытываете наибольшее удовлетворение от урока?»).

Следует отметить, что несмотря на содержательное наличие в текстах опросников тематических блоков, вопросы на одну и ту же тему были разгруппированы и перемешаны в анкете. Более того, каждый показатель в рамках одного и того же диагностического тематизма оценивался посредством собственных критериев. Подобный прием был использован для повышения уровня надежности и достоверности полученных данных. Еще один прием, который был применен с целью получения высококачественных результатов, уже был обозначен выше: речь идет о пересечении тематизмов и формулировок некоторых вопросов в анкетах обеих групп испытуемых.

Несмотря на то, что первоначальный вариант анкет зарекомендовал себя как надежный и валидный, в рамках проведения данного исследования было принято решение разработать новые варианты опросников с целью более узкой исследовательской фокусировки и получения углубленных представлений о психологических характеристиках культурно-образовательной среды как факторах, которые оказывают существенное влияние на академическую успешность в области математической грамотности. Рассмотрим каждую из анкет более подробно.

**2.2. Содержательные и структурные особенности анкеты для учеников**

Структурно опросник состоит из нескольких больших блоков вопросов. Вопросы из первого блока являются традиционными и направленными на сбор социально-демографической информации об учащемся. Второй блок вопросов представлен конструктом, который позволяет оценить уровень субъективного благополучия испытуемого. Данный блок включает в себя три тематических раздела, каждый из которых рассматривается в текущем исследовании как фактор, влияющий на уровень достижений в сфере математической грамотности: а) физическое благополучие; б) удовлетворенность школой; в) удовлетворенность друзьями. Третий и четвертый вопросные блоки направлены на оценку уровня мотивации к чтению и уровня мотивации к математике соответственно и так же содержат три маркированных рубрики: а) внутренняя мотивация; б) идентифицированная регуляция; в) контролируемая регуляция.

**АНКЕТА УЧЕНИКА**

**Блок 1.**

**Социально-демографическая информация**

**Блок 2.**

**Субъективное**

**благополучие**

**Блок 3.**

**Мотивация**

**к**

**чтению**

**Блок 3.**

**Мотивация**

**к**

**математике**

**Физическое благополучие**

**Удовлетворенность**

**школой**

**Удовлетворенность**

**друзьями**

**Внутренняя мотивация**

**Идентифицированная регуляция**

**Контролируемая регуляция**

*Рисунок 8. Содержательная структура анкеты ученика*

Блок 2, направленный на оценку субъективного благополучия, состоит из 15 утверждений. Первые два вопроса представлены разделом, условно обозначенным как «физическое благополучие». Они позволяют выяснить, как школьник ощущает себя физически, как оценивает свое здоровье. Информация о субъективном физическом благополучии необходима, т.к. в силу возрастных особенностей у ребенка могут проявляться психосоматические симптомы, которые могут стать маркером неподходящей культурно-образовательной среды. Кроме того, при рассмотрении субъективного физического благополучия как фактора, влияющего на уровень достижений в области математической грамотности, можно получить информацию о том, как испытуемый чувствовал себя в период, когда проводилось тестирование. Это важно, т.к. физическое состояние оказывает непосредственное влияние на функционирование когнитивных функций психики, мобилизация которых является необходимым условием успешного прохождения теста по математике.

Примеры вопросов:

**За последнюю неделю как часто ты чувствовал себя полным сил:**

* всегда
* часто
* иногда
* редко
* никогда

**Как ты оцениваешь свое здоровье:**

* отличное
* хорошее
* нормальное
* не очень хорошее
* плохое

Второй и третий разделы направлены на определение уровня удовлетворенности учащегося школой и удовлетворенности друзьями. Следует отметить, что вопросы из данных разделов перемешаны и объединены в общий список, чтобы повысить надежность полученных ответов. При формулировке вопросов этих разделов использовалась адаптированная для России версия шкалы удовлетворенности школой и шкалы удовлетворенности друзьями Многокомпонентного опросника удовлетворенности жизнью для школьников (Мultidimensional Students’ Life Satisfaction Scale). Предполагалось, что участники исследования при ответе на вопросы данных разделов будут оценивать частотность ситуаций, предложенных в утверждениях опросника, по четырехбалльной равноплечной шкале без нейтральной ответной категории («Никогда» — «Редко» — «Часто» — «Почти всегда»). Примеры вопросов представлены в Таблице 3.

Данные конструкты были выбраны для включения в структуру анкеты, т.к. предполагается, что то, как учащийся психологически ощущает себя в школе (воспринимает ли школу позитивно или негативно, чувствует ли пользу от обучения, и т. д.), и то, каковы его отношения с ближайшим окружением (есть ли у него друзья, насколько ценными и надежными учащийся их считает и т.д.), может существенно сказаться на академической успешности испытуемого. Субъективное благополучие в социальной сфере может интерпретироваться как значимая характеристика общего эмоционального состояния, которое наравне с физическим оказывает непосредственное влияние на эффективность работы когнитивных функций респондента и может рассматриваться в качестве маркера уровня адаптации ребенка к культурно-образовательной среде.

Таблица 3. Примеры вопросов, направленных на определение уровня

субъективного благополучия респондентов (социальные элементы)

|  |  |
| --- | --- |
| Утверждение | Фактор |
| 1. Я с охотой хожу в школу. | Удовлетворенность школой |
| 2. Мои друзья хорошо ко мне относятся. | Удовлетворенность друзьями |

Третий и четвертый блоки анкеты для учеников посвящены исследованию уровня мотивации учащихся к чтению и математике соответственно. Предполагается, что уровень мотивации к чтению является существенным при обучении ребенка математической грамотности, т.к. большинство заданий представлены в форме текстовых задач и требуют наличия развитых навыков работы с текстом. В свою очередь, интерес к математике как области знания так же играет немаловажную роль в движении учащегося к высоким академическим достижениям, т.к. может обусловливать присутствие у ребенка внутренних ресурсов для того, чтобы решать сложные задачи и разбираться с материалом повышенного уровня сложности. Ответы на вопросы представлены четырехбалльной шкалой («Никогда» – «Редко» – «Часто» – «Почти всегда»).

Оба блока состоят из 9 вопросов, дифференцированных на три тематических раздела, которые позволяют выявить уровень трех типов мотивации учащихся. Развитая внутренняя мотивация подразумевает, что учащийся принимает участие в освоении учебного предмета, руководствуясь внутренними причинами, основанными на положительных эмоциях: интересом к данной дисциплине, желанием получить удовольствие от процесса познания. Идентифицированная регуляция предполагает, что в течение обучения ученик действует сообразно намерению извлечь пользу из освоения образовательной дисциплины: расширить кругозор, получить практические навыки для дальнейшей жизни. Контролируемая регуляция строится вокруг стремления учащегося получить внешнее одобрение от значимых представителей его социального окружения: родителей, учителей, друзей. Примеры вопросов третьего и четвертого блока анкеты представлены в Таблице 4.

Таблица 4. Примеры вопросов, направленных на определение уровня мотивации респондентов к чтению и математике

|  |  |  |
| --- | --- | --- |
| Блок | Тип мотивации | Утверждение |
| Блок 3.  Мотивация к чтению | Внутренняя мотивация | 8.1. Мне нравится читать. |
| Идентифицированная регуляция | 8.4. Я могу узнать много полезного благодаря чтению. |
| Контролируемая регуляция | 8.7. Я читаю, чтобы получить какое-то вознаграждение. |
| Блок 4.  Мотивация к математике | Внутренняя мотивация | 9.1. Мне нравится математика. |
| Идентифицированная регуляция | 9.4. Я могу узнать много полезного благодаря занятиям математикой. |
| Контролируемая регуляция | 9.7. Я занимаюсь математикой, чтобы получить какое-то вознаграждение. |

**2.3. Содержательные и структурные особенности анкеты для учителей**

Анкета для учеников, разработанная для исследования, состоит из 23 вопросов. Вопросы условно дифференцированы на шесть разномасштабных разделов. Первый раздел направлен на сбор социально-демографической информации о респонденте. Второй раздел посвящен выяснению данных об уровне и направленности профессиональной подготовки испытуемого. Третий блок состоит из вопросов, ответы на которые позволяют сформировать представление о социально-демографическом благополучии класса посредством анализа таких характеристик учеников, как: уровень успеваемости, социальный статус семьи, состояние здоровье, социально-психологические и социально-культурные особенности. Четвертый блок имеет целью сбор информации о специфике педагогической практики, осуществляемой данным учителем и состоит из следующих тематизмов: распределение времени на уроке, субъективная оценка частоты взаимодействий определенного характера с учениками, структурно-содержательные особенности организации учебного процесса, частота и тематика внеклассных мероприятий. Пятый и шестой блоки вопросов направлены на изучение психологических особенностей респондентов, поэтому на их описании мы остановимся более подробно.

**Блок 5.**

**Удовлетворенность**

**работой**

**АНКЕТА УЧИТЕЛЯ**

**Блок 1. Социально-демографическая информация**

**Блок 2.**

**Профессиональная подготовка**

**Блок 3.**

**Благополучие**

**класса**

**Блок 4.**

**Особенности**

**педагогической**

**практики**

**Блок 6.**

**Профессиональное**

**выгорание**

*Рисунок 8. Содержательная структура анкеты учителя*

Выбор конструкта «Удовлетворенность работой» в качестве одного из содержательных блоков анкеты учителя обусловлен предположением, что высокий уровень удовлетворенности педагога местом работы и профессией может быть значимым фактором академической успешности учащихся. Мы можем подкрепить этот тезис, отметив, что если специалист в какой-либо сфере, в частности, образовательной, удовлетворен своей работой, то велика вероятность роста производительности труда. Удовлетворенность работой заключается в получении педагогом положительных эмоций от выполнения ежедневных функций. В контексте исследования факторов достижения высоких результатов в освоении математической грамотности удовлетворенность педагога работой может быть рассмотрена как предиктор качественного выстраивания образовательного процесса и осознанного, заинтересованного формирования учителем культурно-образовательной среды.

Так как вопросы анкеты представляют собой самоотчет учителя о своих психологических особенностях, во избежание проявления эффекта социальной желательности был введен еще один тематический блок – «Профессиональное выгорание». Профессиональное выгорание проявляется в эмоциональном истощении, обесценивании собственных профессиональных достижений и отчужденности в коммуникационном процессе. Специалист, который осуществляет свою деятельность в состоянии профессионального выгорания, отличается низким уровнем продуктивности и эффективности. В контексте исследования факторов достижения высоких результатов в освоении математической грамотности профессиональное выгорание педагога может стать предиктором формирования дефектов культурно-образовательной среды, что неизбежно скажется на успеваемости учащихся.

В тексте анкеты оба блока вопросов объединены в единый список и оцениваются по шкале по четырехбалльной шкале («Совершенно не согласен – Скорее не согласен – Скорее согласен – Полностью согласен»). Примеры вопросов представлены в Таблице 5.

Таблица 5. Примеры вопросов, направленных на выявление удовлетворенности педагога работой и его профессионального выгорания

|  |  |
| --- | --- |
| Утверждение | Фактор |
| 8.1. Я чувствую, что работа меня эмоционально выматывает. | Удовлетворенность работой |
| 8. 7. Я чувствую, что вкладываю слишком много усилий в свою работу. | Профессиональное выгорание |

# ГЛАВА 3. ПРОЦЕДУРА ИССЛЕДОВАНИЯ

**3.1. Выборка исследования**

Выборка исследования IRSA представлена учащимися выпускных классов начальной школы крупных городов стран: Армении, Беларуси, Кыргызстана, Таджикистана и муниципального образования РФ, сопоставимого с остальными городами по социально-экономическому уровню развития. В силу различия типов образовательных учреждений по статусу выборка является неоднородной, поэтому выбран стратифицированный путь формирования выборки, при котором в качестве страты выступает тип образовательного учреждения (школа/гимназия). В каждой страте отбор исследовательских единиц осуществлялся случайным образом.

**3.1.1. Выборка учеников**

Общая выборка учеников основного исследования, проведенного в 2019 году, составила 4862 человека из пяти стран. На нижеследующих диаграммах представлено распределение респондентов по странам (рисунок 9), по полу (рисунок 10), возрасту (рисунок 11), типам образовательных учреждений (рисунок 12).

*Рисунок 9. Распределение респондентов по странам*

*Рисунок 10. Распределение респондентов по полу*

*Рисунок 11. Распределение респондентов по возрасту*

*Рисунок 12. Распределение респондентов по типу образовательных учреждений*

Выборка респондентов из Армении составила 1025 человек. На нижеследующих диаграммах представлено распределение респондентов по полу (рисунок 13), возрасту (рисунок 14).

*Рисунок 13. Распределение респондентов из Армении по полу*

*Рисунок 14. Распределение респондентов из Армении по возрасту*

Выборка респондентов из Беларуси составила 1055 человек. На нижеследующих диаграммах представлено распределение респондентов по полу (рисунок 15), возрасту (рисунок 16), типам образовательных учреждений (рисунок 17).

*Рисунок 15. Распределение респондентов из Беларуси по полу*

*Рисунок 16. Распределение респондентов из Беларуси по возрасту*

*Рисунок 17. Распределение респондентов из Беларуси*

*по типу образовательного учреждения*

Выборка респондентов из Кыргызстана составила 1017 человек. На нижеследующих диаграммах представлено распределение респондентов по полу (рисунок 18), возрасту (рисунок 19), типам образовательных учреждений (рисунок 20).

*Рисунок 18. Распределение респондентов из Кыргызстана по полу*

*Рисунок 19. Распределение респондентов из Кыргызстана по возрасту*

*Рисунок 20. Распределение респондентов из Кыргызстана*

*по типу образовательного учреждения*

Выборка респондентов из России составила 752 человека. На нижеследующих диаграммах представлено распределение респондентов по полу (рисунок 21), возрасту (рисунок 22), типам образовательных учреждений (рисунок 23).

*Рисунок 21. Распределение респондентов из России по полу*

*Рисунок 22. Распределение респондентов из России по возрасту*

*Рисунок 23. Распределение респондентов из России*

*по типу образовательного учреждения*

Выборка респондентов из Таджикистана составила 1013 человек. На нижеследующих диаграммах представлено распределение респондентов по полу (рисунок 24), возрасту (рисунок 25), типам образовательных учреждений (рисунок 26).

*Рисунок 24. Распределение респондентов из Таджикистана по полу*

*Рисунок 25. Распределение респондентов из Таджикистана по возрасту*

*Рисунок 26. Распределение респондентов из Таджикистана*

*по типу образовательного учреждения*

**3.1.2. Выборка учителей**

Общая выборка учителей составила 147 человек из четырех стран. На нижеследующих диаграммах представлено распределение респондентов по странам (рисунок 27), по полу (рисунок 28), возрасту (рисунок 29), уровню образования (рисунок 30), стажу работы (рисунок 31).

*Рисунок 27. Распределение респондентов по странам*

*Рисунок 28. Распределение респондентов по полу*

*Рисунок 29. Распределение респондентов по возрасту*

*Рисунок 30. Распределение респондентов по уровню образования*

*Рисунок 31. Распределение респондентов по стажу работы*

Выборка респондентов из Беларуси составила 47 человек. 100% респондентов – женщины. На нижеследующих диаграммах представлено распределение респондентов по возрасту (рисунок 32), уровню образования (рисунок 33), стажу работы (рисунок 34).

*Рисунок 32. Распределение респондентов из Беларуси по возрасту*

*Рисунок 33. Распределение респондентов из Беларуси по уровню образования*

*Рисунок 34. Распределение респондентов из Беларуси по стажу работы*

Выборка респондентов из Кыргызстана составила 29 человек. 100% респондентов – женщины. На нижеследующих диаграммах представлено распределение респондентов по возрасту (рисунок 35), уровню образования (рисунок 36), стажу работы (рисунок 37).

*Рисунок 35. Распределение респондентов из Кыргызстана по возрасту*

*Рисунок 36. Распределение респондентов из Кыргызстана по уровню образования*

*Рисунок 37. Распределение респондентов из Кыргызстана по стажу работы*

Выборка респондентов из России составила 52 человека. 100% респондентов – женщины. На нижеследующих диаграммах представлено распределение респондентов по возрасту (рисунок 38), уровню образования (рисунок 39), стажу работы (рисунок 40).

*Рисунок 38. Распределение респондентов из России по возрасту*

*Рисунок 39. Распределение респондентов из России по уровню образования*

*Рисунок 40. Распределение респондентов из России по стажу работы*

Выборка респондентов из Таджикистана составила 29 человек. На нижеследующих диаграммах представлено распределение респондентов по полу (рисунок 41), возрасту (рисунок 42), уровню образования (рисунок 43), стажу работы (рисунок 44).

*Рисунок 41. Распределение респондентов из Таджикистана по полу*

*Рисунок 42. Распределение респондентов из Таджикистана по возрасту*

*Рисунок 43. Распределение респондентов из Таджикистана по уровню образования*

*Рисунок 44. Распределение респондентов из Таджикистана по стажу работы*

Примерно одинаковые количественные показатели по выделенным параметрам в каждой из стран-участниц исследования, а также схожий социо-культурный контекст, позволяют сделать вывод о том, что выборка в целом является 1) репрезентативной для крупных городов стран локализации и 2) сопоставимой между странами, взятыми в отдельности.

**3.2. Процедура локализации инструмента SAM**

Тест SAM по математике создавался на русском языке, и задача его перевода и адаптации / локализации на другие языки возникла позднее, после того как тест прошел все необходимые исследования по валидизации на родном языке. Попытки перевода и локализации теста SAM на другие языки уже предпринимались: тесты были переведены на национальные языки апробированы в Казахстане, Киргизии, Таджикистане. Следует отметить, что указанные страны и Россия имеют много общего в системе образования, что обуславливается общим историческим прошлым и близкими традициями школьного образования, что в итоге облегчило процесс адаптации. Однако ввиду того, что на данный момент была проведена только одна апробация, процесс создания национальных версий теста нельзя считать завершенным.

Тест SAM по математической грамотности изначально разрабатывался не как кросс-культурное исследование. В зависимости от цели создания кросс-культурных исследований можно выделить две категории подобных инструментов:

1. Инструменты, которые изначально создавались для того, чтобы применяться в различных культурах (к таковым относятся все международные сравнительные исследования, разрабатываемые с целью межстрановых сравнений, например, PISA (OECD, 2019), TIMSS (Mullis, I. V., & Martin, M. O. (2017)).
2. Инструменты, которые изначально создавались для применения в одной стране (на одном языке и в одной культуре), но после разработки методики по необходимости применялись в другом культурном контексте (Hambleton, Jon, 2003).

Под локализацией теста понимается процесс создания версии теста на другом языке, в основе которой лежит та же теоретическая модель, что и в оригинальной версии, но учтены культурные особенности страны локализации. Далее описаны основные этапы локализации теста SAM по математической грамотности на армянском языке. Аналогичная процедура локализации была проведена во всех странах.

**3.3.** **Этапы локализации SAM на армянском языке и результаты её выполнения**

Этап 1.Определение целевой аудитории локализованного инструмента SAM

В России тест рассчитан на выпускников начальной школы, а именно – на учащихся, находящихся в конце 4-го или начале 5-го года обучения. В России дети идут в школу в возрасте 6,5-7,5 лет, следовательно, к концу 4-го класса они достигают возраста 10-11 лет. Ученики возраста 10-11 лет, окончившие (или оканчивающие, если тестирование проходит в конце 4-го года обучения) начальную школу, составляют целевую аудиторию теста SAM в России.

Необходимо определить целевую аудиторию тестирования SAM в Армении: учащиеся какого возраста и какого года обучения будут принимать участие в тестировании?

**Результат**: целевой группой для теста SAM в Армении являются выпускники начальной школы.

Этап 2. Сопоставление программ обучения по математике

в начальной школе России и Армении

Тест SAM построен на математическом содержании, входящем в большинство существующих программ для начальной школы Российской Федерации. При отборе содержания теста SAM по математике в России используется Федеральный государственный стандарт начального общего образования второго поколения (ФГОС, 2010) Содержательная основа математического теста может быть представлена в виде матрицы, в которую включены (Таблица 1):

* разделы предметного содержания;
* математические средства (понятия, представления, принципы, правила, формулы, схемы и проч.), овладение которыми лежит в основе математической компетентности.

На основе анализа содержания образовательной программы в Армении необходимо определить предметное содержание версии теста SAM. В частности, необходимо выделить задания российской версии SAM, которые могут быть переведены и адаптированы для использования в Армении. В тесте SAM структурной единицей является блок из трех заданий. Каждый блок включает три задачи (1-го, 2-го и 3-го уровня), сконструированные из материала одного и того же раздела образовательной программы и образующие естественную иерархию по трудности. Решение о включении заданий в тест в стране локализации должно приниматься на уровне блока.

Необходимо определить, все ли блоки русскоязычной версии могут быть переведены и адаптированы. В случае принятия отрицательного решения должны быть выделены блоки, которые подлежат замене.

**Результат***:* проведен анализ содержания заданий теста SAM на предмет возможности их использования в Армении. Результаты экспертизы представлены в *Приложении 1*. По результатам проведенной экспертизы было выявлено, что все задания теста соответствуют программе 4-го класса в Армении, а также принято решение осуществить комплексные перевод и адаптацию заданий теста.

Таблица 6. Содержание теста по математике

|  |  |
| --- | --- |
| Разделы содержания | Средства ориентировки математических действий |
| Числа и вычисления | * последовательность натуральных чисел * числовая прямая * позиционный принцип * свойства арифметических действий * порядок действий |
| Измерение величин | * отношение между числом, величиной и единицей * отношение «целого и частей» * формула площади прямоугольника |
| Закономерности | * «индукционный шаг» * повторяемость (периодичность) |
| Зависимости | * отношения между однородными величинами (равенство, неравенство, кратности, разностное, «целого и частей») * прямая пропорциональная зависимость между величинами * производные величины: скорость, производительность труда * соотношения между единицами |
| Элементы геометрии | * форма и другие свойства фигур (основные виды геометрических фигур) * пространственные отношения между фигурами * симметрия |

Этап 3. Перевод и адаптация заданий SAM

Переведённые задания должны отражать одну и ту же содержательную область образовательной программы в обеих странах; слова, используемые в формулировках заданий, и картинки должны быть одинаково знакомы целевой аудитории; формулировки заданий должны быть корректными по отношению ко всем потенциальным участникам. В соответствии со стандартами AERA, APA&NCME (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. Standards for educational and psychological testing, 1999) наиболее подходящим способом перевода считается итерационный подход, при котором эксперты-билингвы осуществляют не дословный перевод, а подбирают наиболее близкое, эквивалентное (по смыслу, частоте употребления (распространённости в языке), числу слогов и т. д.) слово. На первом этапе работа экспертов-билингвов осуществляется индивидуально. На втором этапе происходит общее обсуждение созданных вариантов перевода и вырабатывается единое решение относительно формулировки каждого задания. После перевода необходимо провести экспертизу всех заданий на предмет сохранения их иерархичности в соответствии с используемой трехуровневой таксономией SAM.

**Результат***:* Все задания были переведены в соответствии с указанной процедурой.

Этап 4. Составление теста и обратный перевод

После перевода и адаптации все задачные блоки объединяются в тест, имеющий структуру тестов SAM, и производится обратный перевод всех заданий на язык оригинала (т. е. на русский язык). Российские эксперты проверяют задания. Главная цель экспертизы – контроль соответствия заданий уровням освоения материала в соответствии с трехуровневой таксономией SAM.

**Результат***:* Российские эксперты (разработчики российской версии теста) провели экспертизу заданий армянской версии. Экспертами было отмечено, что все задания армянской версии практически совпадают с русским вариантом. Все замечания по доработке заданий были учтены при разработке версии теста для апробации.

Этап 5. Доработка заданий и экспертиза теста

На основании заключения российских экспертов часть заданий была доработана. В результате создана версия на языке страны локализации, готовая к апробации. До этапа апробации проведена экспертиза – содержательная и тестологическая. Содержательная экспертиза теста проведена внешними экспертами – специалистами в области математического образования в начальной школе в стране локализации. Тестологическая экспертиза проведена специалистами-тестологами на предмет правильности формулировок заданий и соответствия формы заданий их содержанию.

**Результат:** Экспертиза теста была проведена специалистами в Армении.

Этап 6. Определение культурных эффектов, которые могут оказать влияние на функционирование заданий и теста в целом

Культурные эффекты ­– степень ознакомленности учащихся начальной школы в стране локализации с тестами, различными формами заданий, используемыми в SAM, стратегии ответов учащихся, а также возможные личностные характеристики учащихся и учителей (например, отсутствие интереса, мотивации у учащихся или учителей). Перечисленные особенности необходимо учитывать при подготовке материалов тестирования и при определении процедуры апробации.

**Результат**: Особых культурных эффектов, которые могли бы оказать влияние на результаты тестирования, выявлено не было.

Этап 7. Проведение качественного исследования по восприятию заданий школьниками в стране локализации

На данном этапе были поставлены две задачи:

- оценить «face validity» теста для учащихся данного возраста и возможные систематические ошибки, обусловленные непониманием заданий и/или инструкций к ним, на основе тестирования малой выборки учеников и наблюдения за их поведением и восприятием выполняемых заданий.

- определить время тестирования (замерить среднее время выполнения заданий различных уровней). В российском тесте общее время выполнения теста 90 минут (два урока по 45 минут с перерывом).

**Результат***:* В апробации приняли участие 11 учеников. За учениками велось наблюдение, все данные, в частности, время выполнения заданий, заносились в протокол. Время тестирования составило 90 минут. В среднем на задание 1-го уровня уходило 0.5-1 минута, 2-го уровня - 1-2 минуты, 3-го уровня - 2-3 минуты. Учащиеся справляются с заданиями разных уровней. Были проанализированы ошибки при решении заданий, которые классифицировались по адекватности. Наибольшее число пропусков и адекватных ошибок (т.е. ошибок в решении) наблюдается для заданий 3-го уровня. Неадекватные ошибки встречаются одинаково часто в заданиях 2-го и 3-го уровней. В целом, для учащихся в стране локализации тестирование оказалось труднее, чем для учеников из России. Рекомендована клиническая апробация заданий. Полный отчет об анализе данных качественного исследования представлен в *Приложении 2*.

Этап 8. Клиническая апробация

Цель клинической апробации – проверка функционирования отдельных заданий и теста в целом на малой выборке учеников. Выборка учеников для данного этапа является спланированной и сконструированной из учащихся школ различного типа и из семей с различным социально-экономическим статусом. Процедура проведения клинической апробации совпадает с разработанной процедурой проведения тестирования SAM. Все особенности поведения учеников и нештатные ситуации документировались для последующего обсуждения.

**Результат***:* Клиническая апробация проведена на выборке, состоящей из 58 учеников.

Этап 9. Обработка данных клинической апробации

Данные клинической апробации обрабатываются в рамках классической теории тестирования. Цель – получение психометрических характеристик заданий (трудности и дискриминативности), определение заданий, не удовлетворяющих критериям качества и экспертным ожиданиям. На данном этапе выявляются задания, которые оказались слишком трудными или слишком простыми для детей, а также задания, которые обладают низкой дискриминативностью (дифференцирующей способностью).

**Результат***:* Проведена обработка результатов апробации в рамках классической теории тестирования. Поведено сравнение показателей заданий между армянской версией и оригинальной российской (*Приложение 3*). Выработаны рекомендации для проведения доработки заданий с целью улучшения их психометрических свойств.

Выводы анализа функционирования заданий:

* Оба варианта теста имеют достаточно высокую надежность.
* Средняя трудность обоих вариантов имеет допустимое значение. Однако почти все задания 3-го уровня и некоторые задания 2-го уровня оказались очень трудными.
* Средний коэффициент дискриминативности имеет оптимальное значение для обоих вариантов. Однако ряд заданий имеет низкую (и даже отрицательную) дискриминативность. Это относится к некоторым заданиям 3-го уровня экстремальной трудности.
* Во всех блоках наблюдается иерархия по трудности заданий. Исключением является один блок, один и тот же в обоих вариантах – блок №7 (М-R-02, задания 19-21). Этот блок - самый проблемный в обоих вариантах, все 3 задания очень трудные.
* В целом, два варианта армянской версии по математике очень близки. По результатам предварительного исследования варианты можно признать параллельными.
* Армянская версия теста SAM оказалась труднее оригинальной российской версии. Это касается заданий всех уровней, однако, в первую очередь, заданий 3-го уровня.
* Некоторые задания 2-го уровня и практически все задания 3-го уровня нуждаются в экспертизе.

Полный отчет об анализе данных качественного исследования представлен в *Приложении 3*.

Этап 10. Определение выборки полномасштабной апробации

Выборка тестирования должна быть репрезентативной по отношению к генеральной совокупности учащихся, определенных как целевая аудитория исследования. Для стратификации выборки выбираются переменные, по которым осуществляется стратификация. При апробации теста SAM в России для стратификации выборки использовались две переменные: 1) тип школы (общеобразовательная школа / школа с особым статусом (гимназия, лицей) / школа с углубленным изучением некоторых предметов), и 2) местоположение школы (большой город / малый город / сельская местность). Количество учащихся не менее 200 человек на вариант.

Этап 11. Проведение полномасштабной апробации

Цель полномасштабной апробации – оценить качество теста как измерительного инструмента и решить специфические проблемы тестирования.

Основные задачи полномасштабной апробации:

* Оценить качество заданий теста.
* Получить статистические характеристики распределений тестовых баллов испытуемых (среднее значение, дисперсия, стандартное отклонение, вид распределения, гистограмма).
* Идентифицировать задания, имеющие проблемы функционирования (например, задания, демонстрирующие различное функционирование по отношению к различным группам испытуемых).
* Оценить надежность теста и ошибку измерения.

**Результат**: Апробация проводилась под контролем разработчиков с соблюдением всех инструкций и требований к выборке и процедуре тестирования, так как именно по результатам апробации будет сделан вывод о качестве инструмента. Выборка исследования составила 220 учеников.

Этап 12. Обработка данных полномасштабной апробации

Данные полномасштабной апробации обрабатываются в рамках классической теории тестирования и современной теории тестирования (IRT). По результатам обработки формулируется вывод о соответствии теста критериям качества.

**Результат**: В целом все задания теста имеют приемлемые психометрические характеристики. На основании результатов анализа функционирования заданий в рамках Классической и Современной теорий тестирования были определены задания, требующие дополнительной проверки на качество функционирования и перевода.

Общие выводы:

* Средний коэффициент решаемости имеет оптимальное значение: 0.4
* Средний индекс дискриминативности также обладает приемлемым значением: 0.38.
* Распределение первичных баллов близко к нормальному распределению.
* По результатам анализа теста на данной выборке надежность тестов высокая. Коэффициент альфа Кронбаха равен 0.87, следовательно, задания теста имеют высокую внутреннюю согласованность.

Полный отчет об анализе данных полномасштабной апробации представлен в *Приложении 4*.

**3.4. Результаты полномасштабной апробации**

В 2018 году была проведена полномасштабная апробация инструмента SAM в странах, участвующих в исследовании. Использовались варианты А1 иБ1. Выборка исследования составила более 1500 человек (таблица 7).

Таблица 7. Распределение участников исследования по вариантам тестирования

|  |  |  |
| --- | --- | --- |
| Страна | Вариант | Количество человек |
| Армения | А1 | 220 |
| Б1 | 177 |
| Беларусь | А1 | 262 |
| Б1 | 262 |
| Таджикистан | А1 | 203 |
| Б1 | 198 |
| Киргизия | А1 | 209 |
| Б1 | 216 |

По результатам полномасштабной апробации используемые варианты теста SAM показали приемлемые психометрические характеристики. Было определено, что варианты теста локализованы в каждой стране для использования в образовательных учреждениях.

Общие выводы о полномасштабной апробации в Армении:

* Средние коэффициенты решаемости по обоим вариантам имеют оптимальные значения: 0,4 и 0,38.
* Средний индекс дискриминативности по обоим вариантам также обладает оптимальными значениями: 0,38 и 0,36.
* Распределение первичных баллов во втором варианте скошено влево, что позволяет сделать вывод о том, что вариант Б1 оказался немного труднее для учащихся.
* По результатам анализа теста на данной выборке надежность тестов высокая. Коэффициент альфа Кронбаха равен 0,87 и 0,84.

Подробный отчет с результатами представлен в *Приложении 5*.

Общие выводы о полномасштабной апробации в республике Беларусь:

* + Средние коэффициенты решаемости по обоим вариантам достигают оптимальных значений: 0,45 и 0,42.
  + Средний индекс дискриминативности по обоим вариантам обладает оптимальным значением 0,41 и 0,39.
  + Распределение первичных баллов в варианте Б1 варианте немного скошено влево, что позволяет сделать вывод о том, что вариант Б1 оказался труднее для испытуемых, чем вариант А1.
  + По результатам анализа теста на данной выборке надежность тестов высокая. Коэффициент альфа Кронбаха равен 0,89 и 0,88.
* Подробный отчет с результатами представлен в *Приложении 6*.

Общие выводы о полномасштабной апробации в Киргизии:

* Средние коэффициенты решаемости по обоим вариантам чуть ниже оптимальных значений, 0,31 и 0,25.
* Средний индекс дискриминативности по обоим вариантам обладает оптимальным значением 0,32 и 0,3.
* Распределение первичных баллов во втором варианте скошено влево, что позволяет сделать вывод о том, что вариант Б1 оказался труднее для испытуемых.
* По результатам анализа теста на данной выборке надежность тестов высокая. Коэффициент альфа Кронбаха равен 0,85 и 0,83.

Подробный отчет с результатами представлен в *Приложении 7*.

Общие выводы о полномасштабной апробации в Таджикистане:

* Средние коэффициенты решаемости по обоим вариантам ниже оптимальных значений, 0,2 и 0,23.
* Средний индекс дискриминативности по обоим вариантам ниже оптимальных значений, 0,23 и 0,26.
* Распределение первичных баллов скошено влево, что позволяет сделать вывод о том, что варианты оказались достаточно трудными для испытуемых.
* По результатам анализа теста на данной выборке надежность тестов достаточно высокая. Коэффициент альфа Кронбаха равен 0,79 и 0,78.

Подробный отчет с результатами представлен в *Приложении 8*.

По итогам анализа в рамках современной и классической теории тестировании были выделены задачи, функционирование которых отличается от оптимального. Данные задачи требуют дополнительного внимания, так как на основе этой информации вносятся коррективы в варианты, которые будут использоваться в основном тестировании.

Варианты, прошедшие пилотную апробацию (А1 и Б1) являются параллельными по отношению к вариантам, которые будут использоваться в основном исследовании (А2 и Б2). Таким образом, после анализа функционирования заданий, использованных в пилотной апробации, были внесены коррективы в формулировки при подготовке тестовых заданий основного тестирования.

**3.5. Шкалирование результатов**

После полномасштабного исследования было проведено шкалирование вариантов тестирования, чтобы подтвердить качество инструмента в СНГ, а также иметь возможность сравнивать страны между собой, сравнивать учеников, решавших разные варианты

Процедура:

* + наложение шкалы результатов СНГ (варианты А1 и Б1) на шкалу России 2018 (варианты А1 и Б1) на основании использования узловых задач в количестве 6 задач по каждому варианту;
  + установление порогов перехода между уровнями для 5 стран участников на единой шкале вариантов;

Варианты теста параллельны в том смысле, что они отличаются лишь конкретным содержанием заданий, а количество заданий и их форма – одинаковы, и основные статистические характеристики достаточно близки.

**Задача**: зафиксировать трудности задач, используемых в тестировании 2018-2019 годов для:

* создания банка заданий;
* помещения всех вариантов тестирования на единую шкалу, что дает возможность сравнивать результаты.

Банк заданий – это множество откалиброванных заданий (заданий с известными параметрами, помещенных на единую общую шкалу), из которых могут быть построены различные варианты теста, порождающие эквивалентные меры. Наличие правильно созданного банка заданий убирает необходимость в дальнейшем выравнивании мер испытуемых. Параметры тестовых заданий можно вносить в банк заданий только при условии их сопоставимости, что достигается помещением их на единую шкалу в результате применения процедуры выравнивания.

Выравнивание - специальная процедура, позволяющая установить связь между баллами испытуемых по различным вариантам теста и затем поместить их на одну единую общую шкалу. В результате окончательный балл испытуемых с одинаковым уровнем подготовки будет одним и тем же вне зависимости от того, какой вариант теста им достался, и несмотря на то, что их первичные баллы могли отличаться.

Метод общих заданий представляет собой связывание двух вариантов осуществляется с помощью части заданий, общих для этих двух вариантов. Выравнивание вариантов А1, А2, Б1, Б2, которые имеют общие задания. Общие задания заложены заранее, в процессе разработки тестовых заданий и композиции вариантов в 2018 году. 15-20% - якорные (узловые, общие) задания. В нашем случае 6 из 45 заданий.

Основные этапы выравнивания:

• выбор метода и процедуры выравнивания;

• получение оценок всех параметров (испытуемых и заданий);

• оценка качества связующих звеньев между вариантами;

• преобразование параметров с различных вариантов на общую шкалу.

В логистических моделях Г. Раша первичные баллы являются достаточными статистиками. Это означает, что между первичными и окончательными баллами в пределах одного и того же варианта существует взаимно-однозначное соответствие. Это соответствие является, разумеется, нелинейным и устанавливается только после математической обработки всех исходных данных, но оно сохраняет упорядоченность. Это означает, что большему первичному баллу будет соответствовать и больший окончательный балл (в шкале логитов или в любой другой, производной от нее шкале, например в 100-балльной). Однако такое соответствие соблюдается только для испытуемых, выполнявших задания одного варианта.

Показатели общих заданий в двух вариантах очень близки и удовлетворяют критериям качества, а также демонстрируют хорошее согласие с моделью измерения.

В процессе анализа согласия экспериментальных данных тестирования с используемой моделью измерения не было выявлено заданий, которые не были достигнуты участниками (т. е. не были выполнены ввиду недостатка времени). Таким образом, практически все задания выравниваемых вариантов находятся в удовлетворительном согласии с используемой моделью измерения (*Приложение 9*).

**3.6. Установление пороговых баллов по данным пилотного исследования 2018**

В качестве модели тестирования используется одномерная дихотомическая модель Раша. В соответствии с этой моделью каждое задание теста характеризуется одним параметром – трудностью, каждый участник тестирования характеризуется одним параметром – уровнем подготовленности. Оценки всех параметров – и участников тестирования и заданий теста - находятся на общей метрической шкале (единица измерения на этой шкале называется логит) и снабжены характеристиками точности оценивания. Начало этой шкалы не определено и может быть выбрано произвольно. С целью фиксирования начала отсчета среднее значение оценок трудности всех заданий положено равным 0. Карта заданий представлена в *Приложении 10*.

Как можно заметить, распределение мер испытуемых близко к нормальному. Выборка испытуемых смещена вниз относительно выборки заданий, что означает, что тест оказался довольно трудным для данного контингента. Стоит обратить внимание на довольно большой размах мер испытуемых. Некоторые задания 1-го уровня оказались очень трудными, в то время как в верхней части карты достаточно много трудных заданий 3-го уровня.

Все задания 1-го уровня находятся в нижней части карты, далее следуют задания 2-го уровня, и далее – задания 3-го уровня, самые трудные в тесте.

Ниже первой ступени – не освоен даже первый уровень;

Первая ступень – освоен только первый уровень;

Вторая ступень – освоен второй уровень;

Третья ступень – освоен третий уровень.

Таблица 8. Установление пороговых баллов

|  |  |  |
| --- | --- | --- |
| Ступень | Трудность | Порог перехода |
| 1 | -1,7 | 430 |
| 2 | -0,014 | 500 |
| 3 | 1,67 | 570 |

По результатам шкалирования вариантов тестирования установлены пороговые баллы для стран-участниц исследования. Пороговые баллы получились идентичны баллам, установленным в России.

**3.7. Анализ различительного функционирования заданий**

Используется метод Mantel-Haenszel (MH DIF), который основан на методе логистической регрессии (Somes, 1986). Значения статистики имеют следующую классификацию:

* Категория А (| ln[αˆLA]|< 0,43) говорит об отсутствии DIF, следовательно, утверждение воспринимается одинаково респондентами из разных групп.
* Категория В (0.43≤ | ln[αˆLA]| < 0.64) говорит о наличии небольшого DIF, на который следует обратить внимание, но его наличие не критично. В данной работе он не оценивается.
* Категория С | ln[αˆLA]| ≥0.64 говорит о наличии существенного DIF, присутствие которого нельзя недооценивать. Именно этот DIF принимается во внимание, и утверждения с его наличием исключаются из дальнейшего исследования (Golia, 2012).

По результатам анализа заданий теста на различительное функционирование было определено, что большинство заданий функционируют одинаково во всех странах. При этом выявлены некоторые задания, демонстрирующие DIF. В основном эти задания относятся к заданиям 3 уровня (*Приложение 11*). Так, наличие DIF можно объяснить тем, что с заданиями именно 3 уровня справляется достаточно мало учеников.

Таким образом, локализованные версии инструмента SAM в странах можно характеризовать как качественный инструмент с хорошо функционирующими заданиями для измерения математической компетентности школьников.

# ГЛАВА 4. ПСИХОМЕТРИЧЕСКИЕ ХАРАКТЕРИСТИКИ ТЕСТОВ SAM

Анализ результатов теста по математике (варианты А2 и Б2, 2019 год)

Анализ результатов

Вариант А2

Анализ психометрических свойств варианта А2 проводился в рамках Современной теории тестирования (Item Response Theory) программе Winsteps в рамках модели 1PL (Rasch Model). Анализ психометрических свойств теста включал в себя: анализ надежности, анализ размерности, анализ заданий и анализ уровня подготовленности испытуемых и трудности заданий, расположенных на одной шкале. Далее все оценки приводятся в логитах.

Общая статистика

Тест включает 45 заданий. Количество испытуемых – 2437.

Вариант А2 демонстрирует высокую надежность – 0,89, что говорит о высокой согласованности заданий шкалы.

Шкала центрирована по заданиям, поэтому средний уровень трудности по заданиям положен равным 0, средняя ошибка по заданиям составляет 0,06. Стандартное отклонение составляет 1,47. Размах заданий по уровню трудности составляет от -2,89 до 3,4.

Средний уровень трудности по испытуемым составляет -1,31, ошибка измерения равна 0,44. Стандартное отклонение составляет 1,38. Размах по уровню подготовленности испытуемых составляет от -4,9 до 2,83.

В таблицах 9 и 10 представлены общие статистические показатели по заданиям теста и испытуемым.

Таблица 9. Общая статистика по заданиям, вариант А2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 0 | 0,06 | 0,99 | -0,7 | 1,32 | 0,9 |
| S.D. | 1,75 | 0,03 | 0,1 | 3,8 | 0,86 | 4,4 |
| MAX | 3,4 | 0,15 | 1,26 | 8,3 | 5,52 | 9,9 |
| MIN | -2,89 | 0,05 | 0,81 | -9,9 | 0,6 | -8,1 |

Общая статистика по испытуемым приводится без экстремальных значений, в нее включены 2429 человек.

Таблица 10. Общая статистика по испытуемым, вариант А2

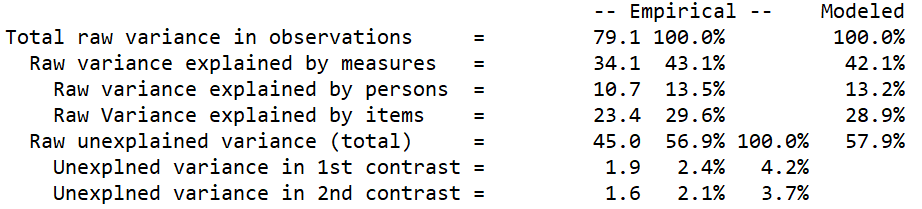
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Общий балл | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 14,8 | -1,31 | 0,44 | 0,99 | -0,1 | 1,25 | 0,2 |
| S.D. | 8 | 1,38 | 0,1 | 0,2 | 1 | 1,43 | 1,1 |
| MAX | 39 | 2,83 | 1,03 | 2,23 | 4,7 | 9,9 | 7,4 |
| MIN | 1 | -4,9 | 0,38 | 0,51 | -3,1 | 0,19 | -2,4 |

Размерность, вариант А2

Важным показателем психометрического качества теста является его одномерность, под которой подразумевается, что в результатах выполнения всех заданий теста преобладает один общий фактор.

Необъясненная дисперсия имеет значение 1,9, что ниже критического значения 2,0. Процент необъясненной дисперсии равен 2,4% (таблица 11). Разница между собственными значениями первого и второго контраста равна 0,3, а разница в общей необъясненной дисперсии составляет 0,3%. Таким образом, тест может быть признан существенно одномерным.

Таблица 11. Размерность теста, вариант А2



Подробный анализ заданий представлен в *Приложении 12*.

Вариант B2

Анализ психометрических свойств варианта В2 проводился в рамках Современной теории тестирования (Item Response Theory) программе Winsteps в рамках модели 1PL (Rasch Model). Анализ психометрических свойств теста включал в себя: анализ надежности, анализ размерности, анализ заданий и анализ уровня подготовленности испытуемых и трудности заданий, расположенных на одной шкале. Далее все оценки приводятся в логитах.

Общая статистика

Тест включает 45 заданий. Количество испытуемых – 2425.

Вариант В2 демонстрирует высокую надежность – 0,87, что говорит о высокой согласованности заданий шкалы.

Шкала центрирована по заданиям, поэтому средний уровень трудности по заданиям положен равным 0, средняя ошибка по заданиям составляет 0,06. Стандартное отклонение составляет 1,47. Размах заданий по уровню трудности составляет от -3,39 до 3,14.

Средний уровень трудности по испытуемым составляет -1,22, ошибка измерения равна 0,43. Стандартное отклонение составляет 1,31. Размах по уровню подготовленности испытуемых составляет от -4,76 до 4,69.

В таблицах 12 и 13 представлены общие статистические показатели по заданиям теста и испытуемым.

Таблица 12. Общая статистика по заданиям, вариант В2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 0 | 0,06 | 0,99 | -0,5 | 1,11 | 0,5 |
| S.D. | 1,47 | 0,02 | 0,1 | 3,9 | 0,42 | 4,3 |
| MAX | 3,14 | 0,13 | 1,28 | 7,7 | 2,72 | 9,9 |
| MIN | -3,39 | 0,05 | 0,82 | -9,9 | 0,74 | -7,9 |

Общая статистика по испытуемым приводится без экстремальных значений, в нее включены 2414 человек.

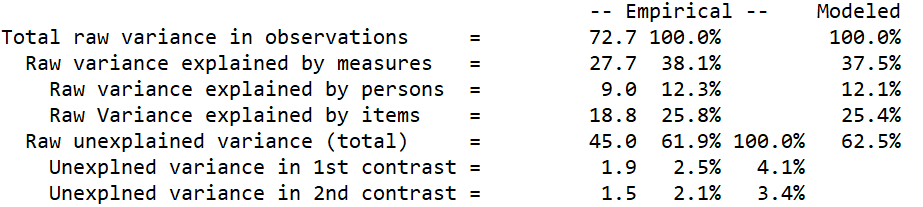
Таблица 13. Общая статистика по испытуемым, вариант В2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Общий балл | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 14,5 | -1,22 | 0,43 | 1 | 0 | 1,11 | 0,1 |
| S.D. | 7,9 | 1,31 | 0,12 | 0,19 | 0,9 | 0,89 | 1 |
| MAX | 44 | 4,69 | 1,04 | 1,99 | 4,5 | 9,9 | 4,8 |
| MIN | 1 | -4,76 | 0,36 | 0,53 | -3,5 | 0,11 | -2,5 |

Важным показателем психометрического качества теста является его одномерность, под которой подразумевается, что в результатах выполнения всех заданий теста преобладает один общий фактор. В случае, если тест не является одномерным, тестовый балл не может быть проинтерпретирован как оценка уровня выраженности измеряемой характеристики.

Необъясненная дисперсия имеет значение 1,9, что ниже критического значения 2,0. Процент необъясненной дисперсии равен 2,5% (таблица 1). Разница между собственными значениями первого и второго контраста равна 0,4, а разница в общей необъясненной дисперсии составляет 0,4%. Таким образом, тест может быть признан существенно одномерным.

Таблица 14. Размерность теста, вариант В2



Подробный анализ заданий представлен в *Приложении 13*.

# ГЛАВА 5. ПСИХОМЕТРИЧЕСКИЕ ХАРАКТЕРИСТИКИ ИССЛЕДУЕМЫХ ПСИХОЛОГИЧЕСКИХ КОНСТРУКТОВ, 2019 ГОД

**5.1. Психометрические характеристики конструктов из контекстной анкеты ученик**

Конструкт «Мотивация к чтению»

Анализ психометрических свойств шкалы «Мотивация к чтению» проводился в программе Winsteps в рамках модели Rating scale model (RSM, Rasch), которая относится к семейству моделей Раша и применяется для анализа данных, измеренных с помощью шкалы Ликерта. Прежде, чем анализировать связь результатов по математике с конструктом «Мотивация к чтению», необходимо проверить психометрическое качество шкалы, чтобы исключить значительные искажения в измерениях, вызванные параметрами самого инструмента. Анализ психометрических свойств шкалы эффективности включал в себя анализ надежности, анализ размерности, анализ заданий и анализ уровня выраженности установки респондентов и трудности утверждений, расположенных на одной шкале. Далее все оценки даны в логитах.

Общая статистика

Шкала «Мотивация к чтению» демонстрирует высокую надежность для психологических опросников – 0,69, что говорит о высокой согласованности заданий шкалы. Шкала центрирована по заданиям, поэтому средний уровень трудности по заданиям положен равным 0, средний уровень трудности по испытуемым составляет 1,69. Средняя ошибка по заданиям составляет 0,02, по испытуемым – 0,45. Средние значения статистик согласия (MNSQ и их стандартизованные версии ZSTD), характеризующих согласие экспериментальных данных с моделью измерения, находятся в пределах допустимых значений. В рамках модели Rating scale допустимыми границами для статистик MNSQ являются значения в пределах 0,6-1,4, для стандартизированных статистик – от -2 до 2. За пределы допустимых значений выходят максимальные и минимальные значения статистик, что указывает на то, что в тесте есть испытуемые и задания, которые не очень хорошо согласуются с моделью.

В таблицах 15 и 16 представлены общие статистические показатели по заданиям опросника и испытуемым.

Таблица 15. Общая статистика по заданиям.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 0 | 0,02 | 0,97 | -1,3 | 1,01 | -0,3 |
| S.D. | 0,63 | 0 | 0,26 | 8,4 | 0,25 | 8,0 |
| MAX | 1,05 | 0,02 | 1,37 | 9,9 | 1,39 | 9,9 |
| MIN | -0,75 | 0,02 | 0,66 | -9,9 | 0,69 | -9,9 |

Общая статистика по испытуемым приводится без экстремальных значений, в нее включены 4527 человек.

Таблица 16. Общая статистика по испытуемым.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Общий балл | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 26,8 | 1,69 | 0,45 | 1,05 | -0,1 | 1,01 | -0,1 |
| S.D. | 4,8 | 0,90 | 0,14 | 0,68 | 1,4 | 0,75 | 1,3 |
| MAX | 35 | 4,07 | 1,02 | 6,49 | 5,1 | 7,19 | 5,9 |
| MIN | 9 | -1,35 | 0,36 | 0,36 | -3,9 | 0,08 | -3,8 |

Анализ размерности

Важным показателем психометрического качества шкалы является ее одномерность, под которой подразумевается, что в результатах выполнения всех заданий теста преобладает один общий фактор. В случае, если шкала не является одномерной, во-первых, тестовый балл не может быть проинтерпретирован как оценка уровня выраженности измеряемой характеристики.

Необъясненная дисперсия имеет значение 2,5, что выше критического значения 2,0. Процент необъясненной дисперсии равен 15,6 %, что превышает принятый порог в 5%. (таблица 17).

Анализ размерности на симулированных данных продемонстрировал собственное значение первого контраста равное 1,2 и 6,6% необъясненной дисперсии (таблица 18). Учитывая, небольшую длину теста, в целом, мы можем признать данную шкалу существенно одномерной.

Таблица 17. Размерность шкалы

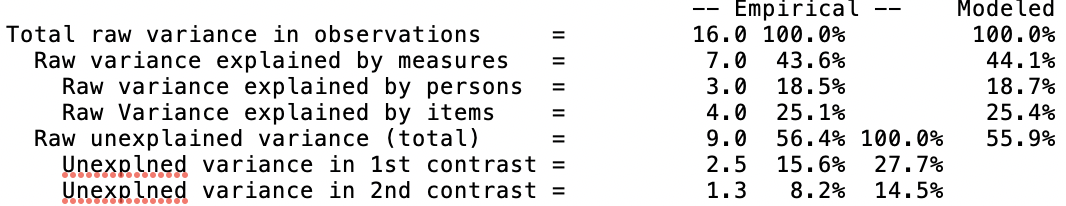
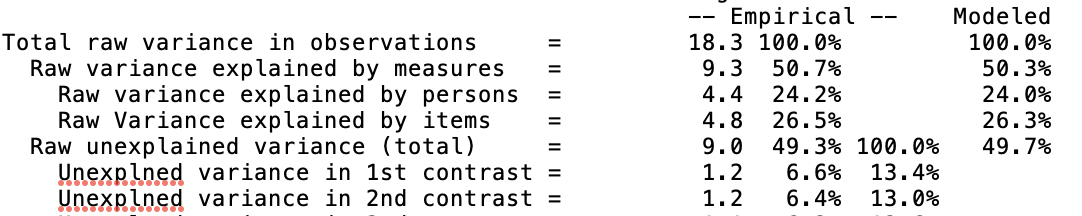


Таблица 18. Размерность шкалы после симуляции данных



Анализ заданий

Шкала «Мотивация к чтению» включает 9 заданий. Задания шкалы «Мотивация к чтению» обладают пониженной трудностью: легче всего испытуемым было согласиться с шестым утверждением («Для дальнейшей жизни важно научиться читать», уровень трудности -0,75), труднее всего – с седьмым утверждением («Я читаю, чтобы получить какое-то вознаграждение», уровень трудности 1,05).

Статистики INTFIT MNSQ и OUTFIT MNSQ по заданиям, которые являются показателями качества модели, находятся в пределах допустимых значений.

Все задания имеют удовлетворительную корреляцию с общим баллом по тесту – от 0,41 до 0,58.

Таблица 19. Статистические данные заданий теста, вариант В2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер задания | Вопрос | Трудность | Ошибка измерения | Корреляция | Статистики согласия | |
| INFIT MNSQ | OUTFIT MNSQ |
| 1 | Мне нравится читать | -0,32 | 0,02 | 0,58 | 0,66 | 0,72 |
| 2 | Мне очень интересно читать | -0,31 | 0,02 | 0,57 | 0,71 | 0,76 |
| 3 | Я читаю, даже если от меня этого не требуют | 0,18 | 0,02 | 0,57 | 0,93 | 1,02 |
| 4 | Я могу узнать много полезного благодаря чтению | -0,62 | 0,02 | 0,51 | 0,81 | 0,86 |
| 5 | Я читаю, чтобы многому научиться | -0,48 | 0,02 | 0,57 | 0,72 | 0,69 |
| 6 | Для дальнейшей жизни важно научиться читать | -0,75 | 0,02 | 0,41 | 1,06 | 1,16 |
| 7 | Я читаю, чтобы получить какое-то вознаграждение | 1,05 | 0,02 | 0,58 | 1,37 | 1,39 |
| 8 | Я читаю, чтобы порадовать родителей или моего учителя | 0,27 | 0,02 | 0,54 | 1,17 | 1,16 |
| 9 | Я читаю, чтобы показать другим, какой я молодец | 0,99 | 0,02 | 0,58 | 1,34 | 1,33 |

Карта переменных

Анализ уровня выраженности установки респондентов и трудности утверждений, расположенных на одной шкале, демонстрирует легкость заданий практически для всей выборки испытуемых, что не позволяет их оценивать с минимальной ошибкой измерения. Практически со всеми утверждениями шкалы испытуемым очень легко согласиться.



*Рисунок 45. Карта переменных*

Анализ функционирования категорий

Удовлетворительное функционирование категорий в тестах со шкалой Ликерта является также маркером качества теста (*Приложение 14*). К критериям удовлетворительного функционирования категорий в модели Rating scale относятся: равномерное распределение ответов респондентов по категориям задания, соответствие категорий модели, монотонное возрастание среднего уровня выраженности измеряемой характеристики с каждым пунктом шкалы и монотонное возрастание пороговых оценок по каждому пункту шкалы. В модели Rating scale предполагается одинаковое функционирование категорий у всех заданий шкалы.

В данной шкале использовались 4 категории в каждом задании: 1- никогда, 2- редко, 3- часто, 4- почти всегда.

Психометрический анализ показал, что, в целом, все категории являются наполненными, хотя более наполненными являются положительные категории.

Все задания можно признать соответствующими модели измерения, так как статистики согласия находятся в пределах допустимых значений.

Нарушение монотонного возрастания уровня выраженности характеристики наблюдается только в задании 8 («Я читаю, чтобы порадовать родителей или моего учителя»): трудность первой категории составляет 1,12 логита, второй – 1,11 логита. Разницу мы признаем незначительной. Обе категории являются достаточно наполненными.

Пороговые оценки для всех утверждений являются упорядоченными, среднее значение порога для перехода от категории 1 к категории 2 равняется 1,03, от второй к третьей – 1,58, от третьей к четвертой – 2,71 логита.

По результатам анализа мы признаем функционирование категорий удовлетворительным.

Конструкт «Мотивация к математике»

Анализ психометрических свойств шкалы «Мотивация к математике» проводился в программе Winsteps в рамках модели Rating scale model (RSM, Rasch), которая относится к семейству моделей Раша и применяется для анализа данных, измеренных с помощью шкалы Ликерта. Прежде, чем анализировать связь результатов по математике с конструктом «Мотивация к математике», необходимо проверить психометрическое качество шкалы, чтобы исключить значительные искажения в измерениях, вызванные параметрами самого инструмента. Анализ психометрических свойств шкалы эффективности включал в себя анализ надежности, анализ размерности, анализ заданий и анализ уровня выраженности установки респондентов и трудности утверждений, расположенных на одной шкале. Далее все оценки даны в логитах.

Общая статистика

Шкала «Мотивация к математике» демонстрирует высокую надежность для психологических опросников – 0,68, что говорит о высокой согласованности заданий шкалы. Шкала центрирована по заданиям, поэтому средний уровень трудности по заданиям положен равным 0, средний уровень трудности по испытуемым составляет 1,84. Средняя ошибка по заданиям составляет 0,02, по испытуемым – 0,46. Средние значения статистик согласия (MNSQ и их стандартизованные версии ZSTD), характеризующих согласие экспериментальных данных с моделью измерения, находятся в пределах допустимых значений. В рамках модели Rating scale допустимыми границами для статистик MNSQ являются значения в пределах 0,6-1,4, для стандартизированных статистик – от -2 до 2. За пределы допустимых значений выходят максимальные и минимальные значения статистик, что указывает на то, что в тесте есть испытуемые и задания, которые плохо согласуются с моделью.

В таблицах 20 и 21 представлены общие статистические показатели по заданиям опросника и испытуемым.

Таблица 20. Общая статистика по заданиям

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 0 | 0,02 | 0,96 | -2 | 1 | -0,4 |
| S.D. | 0,72 | 0 | 0,23 | 8,4 | 0,23 | 6,9 |
| MAX | 1,21 | 0,02 | 1,38 | 9,9 | 1,43 | 9,9 |
| MIN | -0,79 | 0,02 | 0,73 | -9,9 | 0,76 | -7,8 |

Общая статистика по испытуемым приводится без экстремальных значений, в нее включены 4449 человек.

Таблица 21. Общая статистика по испытуемым

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Общий балл | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 27,2 | 1,84 | 0,46 | 1,05 | 0 | 1 | -0,1 |
| S.D. | 4,8 | 0,91 | 0,14 | 0,69 | 1,3 | 0,81 | 1,2 |
| MAX | 35 | 4,17 | 1,02 | 8,77 | 6,6 | 9,9 | 6,9 |
| MIN | 9 | -1,37 | 0,36 | 0,06 | -4,2 | 0,06 | -3,6 |

Анализ размерности

Важным показателем психометрического качества шкалы является ее одномерность, под которой подразумевается, что в результатах выполнения всех заданий теста преобладает один общий фактор. В случае, если шкала не является одномерной, во-первых, тестовый балл не может быть проинтерпретирован как оценка уровня выраженности измеряемой характеристики.

Необъясненная дисперсия имеет значение 2,6, что выше критического значения 2,0. Процент необъясненной дисперсии равен 15,5 %, что превышает принятый порог в 5%. (таблица 22).

Анализ размерности на симулированных данных продемонстрировал собственное значение первого контраста равное 1,2 и 6,0% необъясненной дисперсии (таблица 23). Учитывая, небольшую длину теста, в целом, мы можем признать данную шкалу существенно одномерной.

Таблица 22. Размерность шкалы

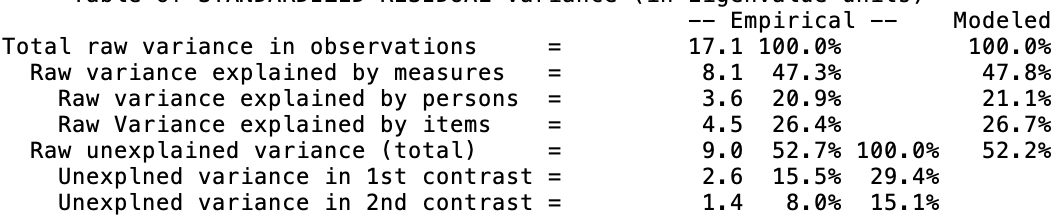
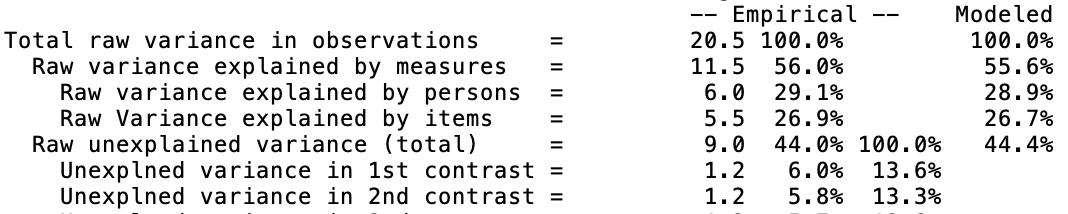


Таблица 23. Размерность шкалы после симуляции данных



Анализ заданий

Шкала «Мотивация к математике» включает 9 заданий. Задания шкалы «Мотивация к математике» обладают пониженной трудностью: легче всего испытуемым было согласиться с шестым утверждением («Для дальнейшей жизни важно научиться математике», уровень трудности -0,79), труднее всего – с седьмым утверждением («Я занимаюсь математикой, чтобы получить какое-то вознаграждение», уровень трудности 1,21).

Статистики INTFIT MNSQ и OUTFIT MNSQ по заданиям, которые являются показателями качества модели, находятся в пределах допустимых значений.

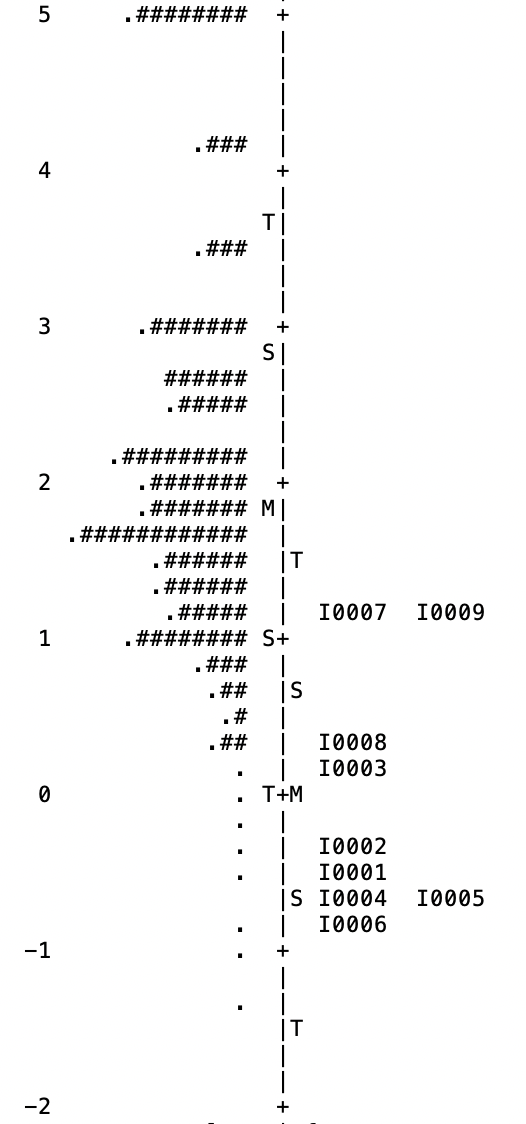
Все задания имеют удовлетворительную корреляцию с общим баллом по тесту – от 0,44 до 0,61.

Таблица 24. Статистические данные заданий теста, вариант В2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер задания | Вопрос | Трудность | Ошибка измерения | Корреляция | Статистики согласия | |
| INFIT MNSQ | OUTFIT MNSQ |
| 1 | Мне нравится математика | -0,42 | 0,02 | 0,55 | 0,78 | 0,84 |
| 2 | Мне очень интересно заниматься математикой | -0,32 | 0,02 | 0,57 | 0,73 | 0,8 |
| 3 | Я занимаюсь математикой, даже если от меня этого не требуют | 0,22 | 0,02 | 0,59 | 0,89 | 0,98 |
| 4 | Я могу узнать много полезного благодаря занятиям математикой | -0,69 | 0,02 | 0,52 | 0,73 | 0,76 |
| 5 | Я занимаюсь математикой, чтобы многому научиться | -0,66 | 0,02 | 0,53 | 0,76 | 0,77 |
| 6 | Для дальнейшей жизни важно научиться математике | -0,79 | 0,02 | 0,44 | 1 | 1,06 |
| 7 | Я занимаюсь математикой, чтобы получить какое-то вознаграждение | 1,21 | 0,02 | 0,59 | 1,38 | 1,43 |
| 8 | Я занимаюсь математикой, чтобы порадовать родителей или моего учителя | 0,37 | 0,02 | 0,57 | 1,14 | 1,11 |
| 9 | Я занимаюсь математикой, чтобы показать другим, какой я молодец | 1,09 | 0,02 | 0,61 | 1,27 | 1,29 |

Карта переменных

Анализ уровня выраженности установки респондентов и трудности утверждений, расположенных на одной шкале, демонстрирует легкость заданий практически для всей выборки испытуемых, что не позволяет их оценивать с минимальной ошибкой измерения. Практически со всеми утверждениями шкалы испытуемым очень легко согласиться.



*Рисунок 46. Карта переменных*

Анализ функционирования категорий

Удовлетворительное функционирование категорий в тестах со шкалой Ликерта является также маркером качества теста (Приложение 15). К критериям удовлетворительного функционирования категорий в модели Rating scale относятся: равномерное распределение ответов респондентов по категориям задания, соответствие категорий модели, монотонное возрастание среднего уровня выраженности измеряемой характеристики с каждым пунктом шкалы и монотонное возрастание пороговых оценок по каждому пункту шкалы. В модели Rating scale предполагается одинаковое функционирование категорий у всех заданий шкалы.

В данной шкале использовались 4 категории в каждом задании:1- никогда, 2- редко, 3- часто, 4- почти всегда.

Психометрический анализ показал, что большая часть категорий являются наполненными. Мало наполненной является категория 1 («никогда») в заданиях 4, 5, 6 («Я могу узнать много полезного благодаря занятиям математикой», «Я занимаюсь математикой, чтобы многому научиться», «Для дальнейшей жизни важно научиться математике»).

Все задания можно признать соответствующими модели измерения, так как статистики согласия находятся в пределах допустимых значений.

Уровень выраженности характеристики монотонно возрастает во всех заданиях опросника.

Пороговые оценки для всех утверждений являются упорядоченными, среднее значение порога для перехода от категории 1 к категории 2 равняется 1,17, от второй к третьей – 1,71, от третьей к четвертой – 2,91 логита.

По результатам анализа мы признаем функционирование категорий удовлетворительным.

Анализ конструкта «Субъективное благополучие школьника»

Конструкт «Субъективное благополучие школьника» представлен двумя шкалами многокомпонентного опросника удовлетворенности жизнью — шкал удовлетворенности школой и удовлетворенности друзьями. Инструмент был адаптирован и валидизирован на русском языке (Канонир, Угланова, Федерякин, 2018). Шкалы рекомендованы к использованию в исследовательских целях, в том числе и в мониторинговых исследованиях

Шкала удовлетворенности друзьями

Шкала удовлетворенности друзьями является частью Многокомпонентного опросника удовлетворенности жизнью для школьников. Шкала состоит из 6 утверждений и направлена на измерение удовлетворенности ребенка отношениями с друзьями. Каждую ситуацию-утверждение предлагалось оценить с помощью четырехбалльной шкалы Ликерта.

Анализ психометрических свойств шкалы удовлетворенности друзьями проводился в программе Winsteps в рамках модели Rating scale model (RSM, Rasch), которая относится к семейству моделей Раша и применяется для анализа данных, измеренных с помощью шкалы Ликерта. Анализ психометрических свойств шкалы эффективности включал в себя анализ надежности, анализ размерности, анализ заданий и анализ уровня выраженности установки респондентов и трудности утверждений, расположенных на одной шкале. Далее все оценки даны в логитах.

Общая статистика

Шкала центрирована по заданиям, поэтому средний уровень трудности по заданиям положен равным 0, средний уровень трудности по испытуемым составляет 1,73. Средняя ошибка по заданиям составляет 0,02, по испытуемым – 0,63. Средние значения статистик согласия (MNSQ и их стандартизованные версии ZSTD), характеризующих согласие экспериментальных данных с моделью измерения, находятся в пределах допустимых значений. В рамках модели Rating scale допустимыми границами для статистик MNSQ являются значения в пределах 0,6-1,4, для стандартизированных статистик – от -2 до 2. За пределы допустимых значений выходят максимальные и минимальные значения статистик, что указывает на то, что в тесте есть испытуемые и задания, которые плохо согласуются с моделью. Надежность шкалы – 0,71.

В таблицах 25 и 26 представлены общие статистические показатели по заданиям опросника и испытуемым.

Таблица 25. Общая статистика по заданиям

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 0 | 0,02 | 0,99 | -0,8 | 1,01 | 0 |
| S.D. | 0,12 | 0 | 0,16 | 5,9 | 0,16 | 5,5 |
| MAX | 0,16 | 0,02 | 1,31 | 9,9 | 1,33 | 9,9 |
| MIN | -0,21 | 0,02 | 0,8 | -8,9 | 0,8 | -8,1 |

Общая статистика по испытуемым приводится без экстремальных значений, в нее включены 3805 человек.

Таблица 26. Общая статистика по испытуемым.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Общий балл | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 18,9 | 1,73 | 0,63 | 1,02 | -0,2 | 1,01 | -0,2 |
| S.D. | 3,5 | 1,07 | 0,18 | 0,82 | 1,6 | 0,81 | 1,6 |
| MAX | 23 | 3,49 | 1,04 | 5,21 | 4,1 | 5,26 | 4,2 |
| MIN | 1 | -3,17 | 0,44 | 0,01 | -4,7 | 0,01 | -4,7 |

Анализ заданий

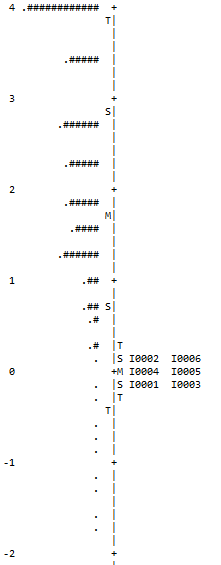
Шкала удовлетворенности друзьями включает 6 заданий. Все утверждения находятся в хорошем согласии с моделью (статистики согласия находятся в пределах допустимых значений). Можно увидеть, что для опросника характерен низкий уровень трудности утверждений — со всеми утверждениями опросника испытуемым довольно легко согласиться. Все задания имеют удовлетворительную корреляцию с общим баллом по тесту – от 0,63 до 0,71.

Таблица 27. Статистические данные заданий опросника

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер задания | Вопрос | Трудность | Ошибка измерения | Корреляция | Статистики согласия | |
| INFIT MNSQ | OUTFIT MNSQ |
| 7.2 | Мои друзья хорошо ко мне относятся | -0,21 | 0,02 | 0,66 | 0,89 | 0,93 |
| 7.5 | У меня достаточно друзей | 0,16 | 0,02 | 0,63 | 1,31 | 1,33 |
| 7.7 | У меня замечательные друзья | -0,08 | 0,02 | 0,67 | 1,04 | 0,98 |
| 7.10 | Я отлично провожу время с друзьями | 0,07 | 0,02 | 0,66 | 1,01 | 1,08 |
| 7.11 | Мои друзья доброжелательны со мной | -0,04 | 0,02 | 0,71 | 0,8 | 0,8 |
| 7.12 | Мои друзья помогут мне, если понадобится | 0,1 | 0,02 | 0,69 | 0,92 | 0,96 |

Карта переменных

Анализ уровня выраженности установки респондентов и трудности утверждений, расположенных на одной шкале, демонстрирует легкость всех утверждений опросника.



*Рисунок 47. Карта переменных*

Анализ функционирования категорий

Удовлетворительное функционирование категорий в тестах со шкалой Ликерта является также маркером качества теста (*Приложение 16*). К критериям удовлетворительного функционирования категорий в модели Rating scale относятся: равномерное распределение ответов респондентов по категориям задания, соответствие категорий модели, монотонное возрастание среднего уровня выраженности измеряемой характеристики с каждым пунктом шкалы и монотонное возрастание пороговых оценок по каждому пункту шкалы. В модели Rating scale предполагается одинаковое функционирование категорий у всех заданий шкалы.

В данной шкале использовались 4 категории в каждом задании:1- никогда, 2- редко, 3- часто, 4- почти всегда.

Психометрический анализ показал, что, в целом, все категории являются наполненными, хотя, в связи легкостью утверждений опросника, более наполненными являются положительные категории.

Все задания можно признать соответствующими модели измерения, так как статистики согласия находятся в пределах допустимых значений.

Уровень выраженности характеристики возрастает монотонно во всех утверждениях опросника. Пороговые оценки для всех утверждений являются упорядоченными, среднее значение порога для перехода от категории 1 к категории 2 равняется 0,83, от второй к третьей – 1,63, от третьей к четвертой – 3,28 логита.

Шкала удовлетворенности ребенка школой

Шкала удовлетворенности ребенка школой является частью Многокомпонентного опросника удовлетворенности жизнью для школьников. Шкала состоит из 7 утверждений и направлена на измерение удовлетворенности ребенка школой. Каждую ситуацию-утверждение предлагалось оценить с помощью четырехбалльной шкалы Ликерта.

Анализ психометрических свойств шкалы удовлетворенности друзьями проводился в программе Winsteps в рамках модели Rating scale model (RSM, Rasch), которая относится к семейству моделей Раша и применяется для анализа данных, измеренных с помощью шкалы Ликерта. Анализ психометрических свойств шкалы эффективности включал в себя анализ надежности, анализ размерности, анализ заданий и анализ уровня выраженности установки респондентов и трудности утверждений, расположенных на одной шкале. Далее все оценки даны в логитах.

Общая статистика

Шкала центрирована по заданиям, поэтому средний уровень трудности по заданиям положен равным 0, средний уровень трудности по испытуемым составляет 1,4. Средняя ошибка по заданиям составляет 0,02, по испытуемым – 0,46. Средние значения статистик согласия (MNSQ и их стандартизованные версии ZSTD), характеризующих согласие экспериментальных данных с моделью измерения, находятся в пределах допустимых значений. В рамках модели Rating scale допустимыми границами для статистик MNSQ являются значения в пределах 0,6-1,4, для стандартизированных статистик – от -2 до 2. За пределы допустимых значений выходят максимальные и минимальные значения статистик, что указывает на то, что в тесте есть испытуемые и задания, которые плохо согласуются с моделью. Надежность шкалы – 0,69.

В таблицах 18 и 29 представлены общие статистические показатели по заданиям опросника и испытуемым

Таблица 28. Общая статистика по заданиям

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 0 | 0,02 | 0,97 | -3,7 | 1,05 | -2 |
| S.D. | 0,83 | 0 | 0,27 | 7,7 | 0,35 | 8,2 |
| MAX | 1,46 | 0,02 | 1,55 | 9,9 | 1,79 | 9,9 |
| MIN | -0,77 | 0,02 | 0,75 | -9,9 | 0,77 | -9,9 |

Общая статистика по испытуемым приводится без экстремальных значений, в нее включены 4816 человек.

Таблица 29. Общая статистика по испытуемым.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Общий балл | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 19,7 | 1,4 | 0,46 | 0,99 | -0,2 | 1,05 | -0,1 |
| S.D. | 2,9 | 0,62 | 0,05 | 0,83 | 1,4 | 1,13 | 1,4 |
| MAX | 27 | 3,86 | 1,09 | 7,66 | 5,9 | 9,9 | 6,2 |
| MIN | 1 | -5,04 | 0,41 | 0,03 | -4,2 | 0,03 | -3,5 |

Анализ заданий

Шкала удовлетворенности друзьями включает 7 заданий. Шесть утверждений из семи находятся в хорошем согласии с моделью (статистики согласия находятся в пределах допустимых значений). В утверждении 7.6 «Я хотел бы, чтобы мне не нужно было ходить в школу» имеет место шум в данных (выше критических значений статистика OUTFIT MNSQ и обе стандартизированные статистики согласия).

Большая часть утверждений имеет низкий уровень трудности. С двумя утверждениями, 7.6 и 7.9, респондентам было согласиться на порядок сложнее («Я хотел бы, чтобы мне не нужно было ходить в школу» и «Мне плохо в школе» соответственно). Возможно, это связано с ярко негативной формулировкой задания.

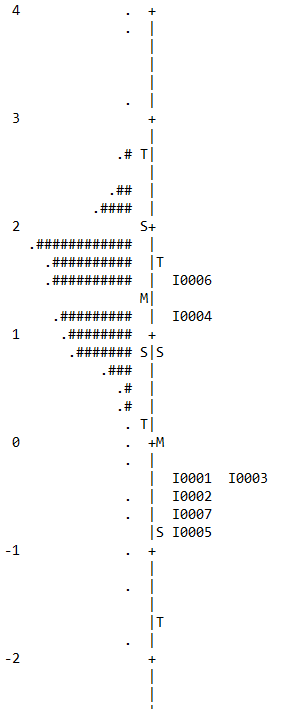
Все задания имеют удовлетворительную корреляцию с общим баллом по тесту – от 0,23 до 0,57.

Таблица 30. Статистические данные заданий опросника.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер задания | Вопрос | Трудность | Ошибка измерения | Корреляция | Статистики согласия | |
| INFIT MNSQ | OUTFIT MNSQ |
| 7.1 | Я с охотой хожу в школу | -0,27 | 0,02 | 0,53 | 0,88 | 0,93 |
| 7.3 | В школе интересно | -0,58 | 0,02 | 0,55 | 0,79 | 0,77 |
| 7.4 | Мне нравится то, что я делаю в школе | -0,35 | 0,02 | 0,54 | 0,75 | 0,78 |
| 7.6 | Я хотел бы, чтобы мне не нужно было ходить в школу | 1,1 | 0,02 | 0,23 | 1,55 | 1,79 |
| 7.8 | Я многому учусь в школе | -0,77 | 0,02 | 0,49 | 0,87 | 0,97 |
| 7.9 | Мне плохо в школе | 1,46 | 0,02 | 0,27 | 1,14 | 1,31 |
| 7.13 | Мне нравится находиться в школе | -0,6 | 0,02 | 0,57 | 0,79 | 0,77 |

Карта переменных

Анализ уровня выраженности установки респондентов и трудности утверждений, расположенных на одной шкале, демонстрирует легкость практически всех утверждений опросника, за исключением двух заданий, о которых было сказано выше (7.6 «Я хотел бы, чтобы мне не нужно было ходить в школу» и 7.9 «Мне плохо в школе»).



*Рисунок 48. Карта переменных*

Анализ функционирования категорий

Удовлетворительное функционирование категорий в тестах со шкалой Ликерта является также маркером качества теста (*Приложение 17*). К критериям удовлетворительного функционирования категорий в модели Rating scale относятся: равномерное распределение ответов респондентов по категориям задания, соответствие категорий модели, монотонное возрастание среднего уровня выраженности измеряемой характеристики с каждым пунктом шкалы и монотонное возрастание пороговых оценок по каждому пункту шкалы. В модели Rating scale предполагается одинаковое функционирование категорий у всех заданий шкалы.

В данной шкале использовались 4 категории в каждом задании:1- никогда, 2- редко, 3- часто, 4- почти всегда.

Психометрический анализ показал, что, в целом, все категории являются наполненными, хотя, в связи легкостью утверждений опросника, более наполненными являются положительные категории.

Большую часть заданий можно признать соответствующими модели измерения, так как статистики согласия ответных категорий находятся в пределах допустимых значений. В двух ответных категориях двух заданий статистики согласия незначительно выходят за пределы критических значений (7.6 «Я хотел бы, чтобы мне не нужно было ходить в школу» и 7.9 «Мне плохо в школе»).

В трех заданиях нарушена монотонность уровня выраженности характеристики. При этом, пороговые оценки для всех утверждений являются упорядоченными, среднее значение порога для перехода от категории 1 к категории 2 равняется 1,04, от второй к третьей – 1,3, от третьей к четвертой – 1,84 логита.

* 1. **Психометрические характеристики конструктов из контекстной анкеты учителя**

Опросник "Удовлетворенность работой и профессиональное выгорание"

Опросник "Удовлетворенность работой и профессиональное выгорание" состоит из 12 утверждений и подразумевает два фактора: удовлетворенность работой и профессиональное выгорание. На каждый фактор приходится по 6 утверждений. Учителям было предложено оценить утверждения опросника по четырехбалльной шкале без нейтральной ответной категории (Совершенно не согласен - Скорее не согласен - Скорее согласен - Полностью согласен).

Анализ психометрических свойств шкалы удовлетворенности работой и профессионального выгорания проводился в программе Winsteps в рамках модели Rating scale model (RSM, Rasch), которая относится к семейству моделей Раша и применяется для анализа данных, измеренных с помощью шкалы Ликерта. Анализ психометрических свойств шкалы эффективности включал в себя анализ надежности, анализ размерности, анализ заданий и анализ уровня выраженности установки респондентов и трудности утверждений, расположенных на одной шкале. Далее все оценки даны в логитах.

Общая статистика

Опросник демонстрирует высокий уровень демонстрирует высокую надежность для психологических опросников – 0,72. Шкала центрирована по заданиям, поэтому средний уровень трудности по заданиям положен равным 0, средний уровень трудности по испытуемым составляет 0,35. Средняя ошибка по заданиям составляет 0,12, по испытуемым – 0,41. Средние значения статистик согласия (MNSQ и их стандартизованные версии ZSTD), характеризующих согласие экспериментальных данных с моделью измерения, находятся в пределах допустимых значений. В рамках модели Rating scale допустимыми границами для статистик MNSQ являются значения в пределах 0,6-1,4, для стандартизированных статистик – от -2 до 2. За пределы допустимых значений выходят только максимальные и минимальные значения статистик.

В таблицах 31 и 32 представлены общие статистические показатели по заданиям опросника и испытуемым.

Таблица 31. Общая статистика по заданиям

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 0 | 0,12 | 0,99 | -0,1 | 0,98 | -0,2 |
| S.D. | 0,71 | 0,01 | 0,23 | 2 | 0,24 | 2,1 |
| MAX | 0,79 | 0,12 | 1,51 | 4,3 | 1,53 | 4,3 |
| MIN | -1,63 | 0,11 | 0,68 | -3 | 0,66 | -3,2 |

Общая статистика по испытуемым приводится без экстремальных значений, в нее включены 150 человек.

Таблица 32. Общая статистика по испытуемым

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Общий балл | Трудность | Ошибка измерения | Статистики согласия | | | |
| MNSQ | ZSTD | MNSQ | ZSTD |
| Среднее | 23,8 | 0,35 | 0,41 | 1 | -0,4 | 0,98 | -0,4 |
| S.D. | 5,3 | 0,9 | 0,04 | 0,91 | 1,9 | 0,93 | 1,9 |
| MAX | 39 | 2,6 | 0,52 | 5,56 | 5,6 | 6,22 | 5,7 |
| MIN | 12 | -2,07 | 0,37 | 0,05 | -4,6 | 0,05 | -4,6 |

Анализ размерности

Важным показателем психометрического качества шкалы является ее одномерность, под которой подразумевается, что в результатах выполнения всех заданий теста преобладает один общий фактор. В случае, если шкала не является одномерной, тестовый балл не может быть проинтерпретирован как оценка уровня выраженности измеряемой характеристики.

При анализе размерности собственное значение первого контраста имеет значение 2,8, что выше критического значения 2,0. Процент необъясненной дисперсии равен 13,6 %, что превышает принятый порог в 5%. (таблица 33).

Анализ размерности на симулированных данных продемонстрировал собственное значение первого контраста равное 1,7 и 7,6% необъясненной дисперсии (таблица 3.2). Учитывая небольшую длину опросника, мы можем признать данную шкалу существенно одномерной.

Таблица 33. Размерность шкалы

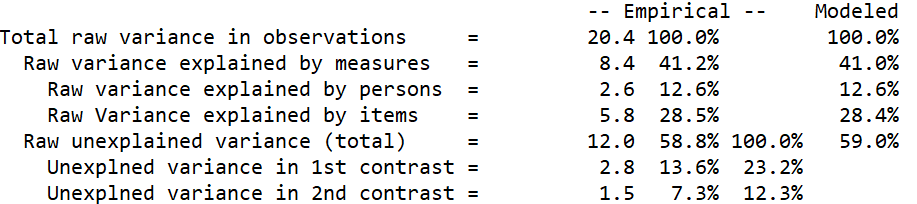
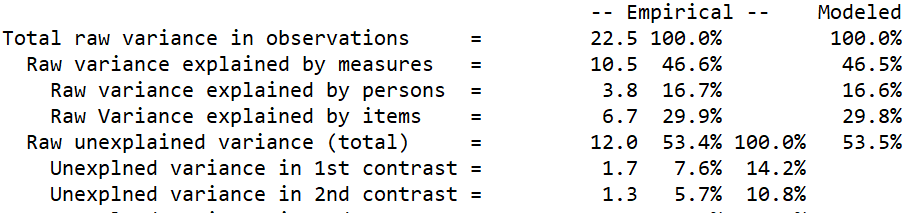


Таблица 34. Размерность шкалы после симуляции данных



Анализ заданий

Опросник включает 12 заданий. Все утверждения находятся в хорошем согласии с моделью (статистики согласия находятся в пределах допустимых значений). Можно увидеть, что для опросника характерен относительно низкий уровень трудности утверждений. Труднее всего респондентам было согласиться с утверждением 9.6 «Я чувствую разочарование в своем каждодневном труде» (0,79). Легче всего респондентам было согласиться с утверждением 9.7 «Я чувствую, что вкладываю слишком много усилий в свою работу» (-1,63). Все утверждения имеют удовлетворительную корреляцию с общим баллом – от 0,28 до 0,71.

Все задания опросника функционируют оптимально.

Таблица 35. Статистические данные заданий опросника

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Шкала | Номер задания | Вопрос | Трудность | Ошибка измерения | Корреляция | Статистики согласия | |
| INFIT MNSQ | OUTFIT MNSQ |
| Удовлетворенность работой | 9.1. | Я чувствую, что работа меня эмоционально выматывает | -0,64 | 0,11 | 0,71 | 0,89 | 0,92 |
| 9.2. | Я чувствую, что не получаю отдачу от своего труда | 0,31 | 0,12 | 0,63 | 1,01 | 0,95 |
| 9.3. | Я чувствую себя обессиленным в конце рабочего дня | -1,08 | 0,11 | 0,6 | 0,93 | 0,94 |
| 9.4. | Я чувствую себя усталым, когда просыпаюсь и собираюсь на работу | 0,22 | 0,12 | 0,69 | 0,71 | 0,66 |
| 9.5. | Я чувствую, что мне тяжело работать с учениками весь день | 0,18 | 0,12 | 0,59 | 0,96 | 0,95 |
| 9.6. | Я чувствую разочарование в своем каждодневном труде | 0,79 | 0,12 | 0,57 | 0,94 | 0,87 |
| Профессиональное выгорание | 9.7. | Я чувствую, что вкладываю слишком много усилий в свою работу | -1,63 | 0,11 | 0,31 | 1,51 | 1,53 |
| 9.8. | Я чувствую, что всё время нахожусь в стрессовой ситуации на работе | -0,04 | 0,11 | 0,66 | 0,84 | 0,86 |
| 9.9. | Я эффективно решаю проблемы своих учеников | 0,34 | 0,12 | 0,44 | 0,68 | 0,66 |
| 9.10. | Я чувствую, что позитивно влияю на жизнь своих учеников | 0,51 | 0,12 | 0,46 | 0,93 | 0,92 |
| 9.11. | Мне легко удается создать непринужденную, комфортную атмосферу в классе | 0,6 | 0,12 | 0,28 | 1,37 | 1,37 |
| 9.12. | Работа с учениками придаёт мне сил | 0,45 | 0,12 | 0,35 | 1,15 | 1,15 |

Карта переменных

Анализ уровня выраженности установки респондентов и трудности утверждений, расположенных на одной шкале, демонстрирует умеренную легкость утверждений опросника. Выборка респондентов смещена немного вверх относительно утверждений опросника.

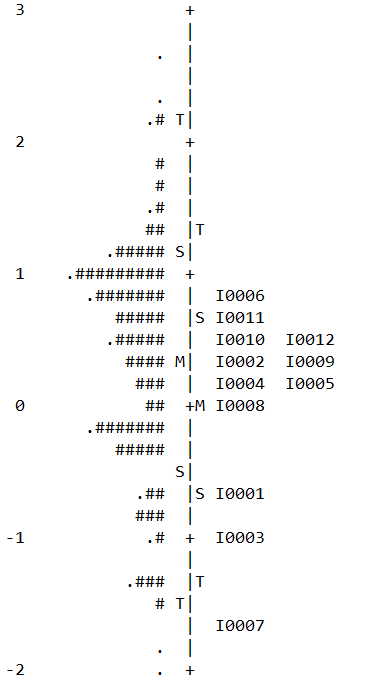


Рисунок 49. Карта переменных

Анализ функционирования категорий

Удовлетворительное функционирование категорий в тестах со шкалой Ликерта является также маркером качества теста (Приложение 18). К критериям удовлетворительного функционирования категорий в модели Rating scale относятся: равномерное распределение ответов респондентов по категориям задания, соответствие категорий модели, монотонное возрастание среднего уровня выраженности измеряемой характеристики с каждым пунктом шкалы и монотонное возрастание пороговых оценок по каждому пункту шкалы. В модели Rating scale предполагается одинаковое функционирование категорий у всех заданий шкалы.

В данной шкале использовались 4 категории в каждом задании: 1- Совершенно не согласен, 2 - Скорее не согласен, 3 - Скорее согласен, 4 - Полностью согласен. В опроснике было 4 обратных утверждения, которые были перекодированы для анализа.

Анализ наполненности ответных категорий опросника показывает, что все категории довольно часто выбираются испытуемыми. Есть незначительные нарушения монотонного возрастания среднего уровня выраженности измеряемой характеристики в нескольких утверждениях. Учитывая небольшой размер выборки и незначительный размер нарушений, этим можно пренебречь.

Трудность порогов и средний уровень выраженности конструкта у респондентов, выбравших каждую категорию, упорядочены монотонно, что говорит об оптимальном функционировании ответной шкалы. Среднее значение порога для перехода от категории 1 к категории 2 равняется 0,44, от второй к третьей – 1,02, от третьей к четвертой – 1,13 логита.

ОБЩИЕ ВЫВОДЫ ПО РЕЗУЛЬТАТАМ СТАТИСТИЧЕСКОГО АНАЛИЗА

Качество образования связано с многими факторами. Определение наиболее значимых предикторов академической успеваемости является достаточно сложной задачей. Ученики в школах принадлежат к различным категориям в зависимости от особенностей школы (тип школы и школьный климат), параметров учебного процесса (учительские практики), характеристик ученика (индивидуальные особенности). Это разнообразие очень обширное и комплексное. Принимая во внимание вышесказанное, проведено исследование, чтобы изучить и оценить связь различных факторов и уровень математической компетентности выпускников начальной школы.

В исследовании принимали участие выпускники начальной школы Армении, Беларуси, Киргизии, России, Таджикистана. Выборка стратифицирована в отношении типов образовательных учреждений. В каждой страте отбор исследовательских единиц осуществлялся случайным образом. В исследовании приняли участие 30-35 школ из крупных городов каждой страны (около 1000 учеников из каждой страны) в возрасте от 10 до 11 лет. По причине ограничений формирования выборки исследование носит описательно-сравнительный характер.

Student Achievement’s Monitoring (SAM) инструмент для измерения предметных образовательных результатов, дающий качественную характеристику знаний учащихся на основе соотнесения с таксономией педагогических целей. Трехуровневая таксономия освоения культурно-деятельностного содержания, разработанная Л.С. Выготским, вписана в возрастной контекст. Согласно концепции, выделено три качественных уровня овладения культурным способом действия: формальный, рефлексивный и функциональный. Каждый из этих уровней связан с определенным типом ориентировки в предметном содержании и проявляется в способности решать соответствующий тип задач.

Комплекс контекстных анкет позволяет оценить и сопоставить результаты академических предметных компетенций учеников, обучающихся в различных культурных и социально-экономических условиях, рассмотреть индивидуальные и межстрановые различия. Формат: анкета учителя и анкета ученика.

Анкета учителя позволяет определить основные педагогические особенности работы учителей, в частности, измеряет конструкт «Удовлетворенность работой и профессиональное выгорание» учителей начальных классов. Вопросы условно дифференцированы на шесть разномасштабных разделов. Первый раздел направлен на сбор социально-демографической информации о респонденте. Второй раздел посвящен выяснению данных об уровне и направленности профессиональной подготовки испытуемого. Третий блок состоит из вопросов, ответы на которые позволяют сформировать представление о социально-демографическом благополучии класса посредством анализа таких характеристик учеников, как: уровень успеваемости, социальный статус семьи, состояние здоровье, социально-психологические и социально-культурные особенности. Четвертый блок имеет целью сбор информации о специфике педагогической практики, осуществляемой учителем и состоит из следующих тематизмов: распределение времени на уроке, субъективная оценка частоты взаимодействий определенного характера с учениками, структурно-содержательные особенности организации учебного процесса, частота и тематика внеклассных мероприятий. Пятый и шестой блоки вопросов направлены на изучение психологических особенностей респондентов.

Анкета ученика раскрывает следующие характеристики:

- субъективное благополучие в школе (удовлетворенность школьным климатом);

- отношения с одноклассниками;

- субъективное физическое благополучие;

- мотивация к чтению;

- мотивация к математике.

Анкета ученика состоит из нескольких блоков вопросов. Вопросы из первого блока являются традиционными и направленными на сбор социально-демографической информации об учащемся. Второй блок вопросов представлен конструктом, который позволяет оценить уровень субъективного благополучия испытуемого. Данный блок включает в себя три тематических раздела, каждый из которых рассматривается в текущем исследовании как фактор, влияющий на уровень достижений в сфере математической грамотности: а) физическое благополучие; б) удовлетворенность школой; в) удовлетворенность друзьями. Третий и четвертый вопросные блоки направлены на оценку уровня мотивации к чтению и уровня мотивации к математике соответственно и так же содержат три маркированных рубрики: а) внутренняя мотивация; б) идентифицированная регуляция; в) контролируемая регуляция.

Инструмент SAM и контекстные анкеты были проанализированы в рамках современной теории тестирования (IRT): осуществлен анализ функционирования заданий, анализ учеников, DIF анализ, анализ ответных категорий.

Применение одинаковой методологии в каждой стране позволяет проанализировать содержание и дать корректную оценку образовательной ситуации в начальной школе.

В результате исследования:

* Локализованы 4 варианта тестов SAM по математике для начальной школы на армянском, киргизском и таджикском языках.
* Установлены пороговые баллы для определения ступени достижений учащихся.
* Проведен анализ различительного функционирования тесовых заданий в участвующих странах (DIF анализ).
* Получены данные о высоком качестве локализованных версий инструмента.
* Психологические конструкты, используемые в анкете ученика «Субъективное благополучие в школе», «Мотивация к чтению», «Мотивация к математике» и в анкете учителя «Удовлетворенность работой и профессиональное выгорание» имеют приемлемые психометрические характеристики, определены как надежные измерительные инструменты в контексте проведения данного исследования.

СПИСОК ЛИТЕРАТУРЫ

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999) Standards for educational and psychological testing. - Washington, DC: American Educational Research Association, 101 p.

Golia, S. (2012). A proposal for categorizing nonuniform DIF in polytomous items. Analysis and Modeling of Complex Data in Behavioural and Social Sciences, 39.

Hambleton, de Jon, (2003). Advances in translating and adapting educational and psychological tests.

Mullis, I. V., & Martin, M. O. (2017). TIMSS 2019 Assessment Frameworks. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

OECD (2019), PISA 2018 Assessment and Analytical Framework, PISA, OECD Publishing, Paris, https://doi.org/10.1787/b25efab8-en.

Somes, G. W. (1986). The generalized Mantel–Haenszel statistic. The American Statistician, 40(2), 106-108.

SAM: Инструмент мониторинга учебных достижений. Framework (2011), под ред. П. Г. Нежнова, Е. Ю. Кардановой, 106 с.

Диагностика учебной успешности в начальной школ (2009), под ред. П. Г. Нежнова, И. Фрумина, Б. И. Хасана, Б. Д. Эльконина, Открытый институт «Развивающее образование», 168с.

Канонир Т.Н., Угланова И.Л., Федерякин Д.А. Адаптация и валидизация шкал удовлетворенности школой и друзьями «Многокомпонентного опросника удовлетворенности жизнью для школьников» (2018), Современная зарубежная психология, Том 7, № 3, С. 64—74.

Федеральный государственный образовательный стандарт начального общего образования (2010), Министерство образования и науки Рос. Федерации, Просвещение, 31 с.

# ПРИЛОЖЕНИЕ 1

ЭКСПЕРТИЗА МАТРИЦЫ СОДЕРЖАНИЯ ТЕСТА SAM ПО МАТЕМАТИКЕ

Таблица 1. Экспертиза матрицы содержания теста SAM по математике

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Задание | № | Формат | Трудность | Раздел | Тема | Класс | Соответствие программе |
| М-С-01-1-1 | 1 | ОО | 0.58 | Натуральные числа | Деление натуральных чисел | 3,4 | Да |
| М-С-01-1-2 | 2 | ОО | 0.36 | Натуральные числа | Разряды натурального числа, умножение и деление | 3,4 | Да |
| М-С-01-1-3 | 3 | ОО | 0.2 | Натуральные числа | Разряды натурального числа, сумма | 3,4 | Да |
| М-С-03-1-1 | 4 | ОО | 0.85 | Натуральные числа | Сравнение натуральных чисел | 3,4 | Да |
| М-С-03-1-2 | 5 | ОО | 0.61 | Натуральные числа | Сравнение натуральных чисел | 3,4 | Да |
| М-С-03-1-3 | 6 | ОО | 0.25 | Натуральные числа | Разряды и сравнение натуральных чисел | 3,4 | Да |
| М-M-02-1-1 | 7 | ОО | 0.67 | Геометрия | Площадь прямоугольника | 3,4 | Да |
| М-М-02-1-2 | 8 | ОО | 0.49 | Геометрия | Площадь прямоугольника, деление на равные части | 3,4 | Да |
| М-М-02-1-3 | 9 | ОО | 0.15 | Геометрия | Площадь прямоугольника, деление на равные части | 3,4 | Да |
| М-М-03-1-1 | 10 | ОО | 0.73 | Геометрия | Длина отрезка | 3,4 | Да |
| М-М-03-1-2 | 11 | НВ | 0.65 | Геометрия | Длина отрезка, буквенное выражение | 3,4 | Да |
| М-М-03-1-3 | 12 | НВ | 0.3 | Геометрия | Длина отрезка, буквенное выражение | 3,4 | Да |
| М-М-06-1-1 | 13 | ОО | 0.72 | Геометрия | Измерение длины отрезка | 3,4 | Да |
| М-М-06-1-2 | 14 | ОО | 0.45 | Геометрия | Измерение длины отрезка | 4 | Да |
| М-М-06-1-3 | 15 | ОО | 0.23 | Геометрия | Измерение длины отрезка | 3,4 | Да |
| М-М-11-1-1 | 16 | ОО | 0.76 | Геометрия | Периметр прямоугольника | 4 | Да |
| М-М-11-1-2 | 17 | ОО | 0.41 | Геометрия | Периметр прямоугольника | 4 | Да |
| М-М-11-1-3 | 18 | ОО | 0.22 | Геометрия | Периметр прямоугольника | 4 | Да |
| М-R-02-1-1 | 19 | НВ | 0.7 | Геометрия | Прямоугольник | 3,4 | Да |
| М-R-02-1-2 | 20 | ОО | 0.41 | Закономерности | Продолжение по закономерности | 4 | Да |
| М-R-02-1-3 | 21 | НВ | 0.26 | Закономерности | Продолжение по закономерности | 4 | Да |
| М-R-05-1-1 | 22 | ОО | 0.81 | Закономерности | Продолжение по закономерности | 4 | Да |
| М-R-05-1-2 | 23 | ОО | 0.47 | Закономерности | Продолжение по закономерности | 4 | Да |
| М-R-05-1-3 | 24 | ОО | 0.16 | Закономерности | Продолжение по закономерности | 4 | Да |
| М-G-01-1-1 | 25 | CR | 0.82 | Геометрия | Геометрические фигуры | 4 | Да |
| М-G-01-1-2 | 26 | НВ | 0.63 | Геометрия | Геометрические фигуры | 3,4 | Да |
| М-G-01-1-3 | 27 | ОО | 0.18 | Геометрия | Геометрические фигуры | 3,4 | Да |
| М-D-03-1-1 | 28 | ОО | 0.7 | Геометрия | Единицы измерения длины, текстовая задача | 3,4 | Да |
| М-D-03-1-2 | 29 | ОО | 0.42 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-D-03-1-3 | 30 | ОО | 0.24 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-D-05-1-1 | 31 | ОО | 0.65 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-D-05-1-2 | 32 | ОО | 0.47 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-D-05-1-3 | 33 | ОО | 0.32 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-D-08-1-1 | 34 | ОО | 0.9 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-D-08-1-2 | 35 | ОО | 0.56 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-D-08-1-3 | 36 | ОО | 0.8 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-R-03-1-1 | 37 | ОО | 0.79 | Закономерности | Продолжение по закономерности | 4 | Да |
| М-R-03-1-2 | 38 | ОО | 0.65 | Закономерности | Продолжение по закономерности | 4 | Да |
| М-R-03-1-3 | 39 | ОО | 0.24 | Закономерности | Продолжение по закономерности | 4 | Да |
| М-С-05-1-1 | 40 | ОО | 0.7 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-С-05-1-2 | 41 | ОО | 0.44 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-С-05-1-3 | 42 | MC | 0.38 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-М-08-1-1 | 43 | ОО | 0.81 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-М-08-1-2 | 44 | ОО | 0.34 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |
| М-М-08-1-3 | 45 | ОО | 0.29 | Действия с натуральными числами | Текстовые задачи | 3,4 | Да |

# ПРИЛОЖЕНИЕ 2

АНАЛИЗ ДАННЫХ КАЧЕСТВЕННОГО ИССЛЕДОВАНИЯ

(АПРОБАЦИИ SAM НА МАЛОЙ ВЫБОРКЕ ДЕТЕЙ)

В апробации приняли участие 11 учеников. Для проведения апробации был разработан Протокол тестирования и Протокол наблюдения за ребенком во время тестирования, которые заполнялись организатором в аудитории. Ученикам понравилось выполнять задания, однако некоторые задачи показались им нелегкими. Все учащиеся, участвовавшие в апробации, выполнили задания теста за 90 минут. В среднем на задание 1-го уровня уходило 0.5-1 минута, 2-го уровня - 1-2 минуты, 3-го уровня - 2-3 минуты.

Далее была проанализирована трудность заданий по блокам. В таблице ниже указано количество учащихся, выполнивших задания 1-го, 2-го и 3-го уровней соответственно по каждому блоку. В последней строке таблицы указан средний процент учеников, выполнивших задания каждого уровня.

Таблица 1. Результаты качественного апробационного тестирования

|  |  |  |  |
| --- | --- | --- | --- |
| № блока | Уровень 1 | Уровень 2 | Уровень 3 |
| Вариант 1 | Вариант 1 | Вариант 1 |
| 1 | 8 | 5 | 1 |
| 2 | 10 | 7 | 2 |
| 3 | 9 | 5 | 0 |
| 4 | 7 | 3 | 1 |
| 5 | 8 | 5 | 1 |
| 6 | 10 | 5 | 0 |
| 7 | 5 | 2 | 0 |
| 8 | 10 | 6 | 0 |
| 9 | 6 | 3 | 2 |
| 10 | 5 | 2 | 2 |
| 11 | 8 | 4 | 1 |
| 12 | 8 | 5 | 1 |
| 13 | 9 | 7 | 1 |
| 14 | 5 | 2 | 1 |
| 15 | 5 | 3 | 0 |
| Среднее число | 7.5 | 4.3 | 0.87 |
| средний % | 0.68 | 0.39 | 0.09 |

Из таблицы видно, что задания 3-го уровня – трудные, в среднем их решают 7% детей. В блоках 3, 6, 7, 15 ни один учащийся не выполнил задачи 3-го уровня. Задания 1-го уровня - самые легкие (в среднем решаемость 68%), задания 2-го уровня - труднее (решаемость 39%). Иерархия по трудности наблюдается во всех блоках.

Дополнительно были проанализированы ошибки в решении заданий. Для этого для обозначения ошибок использовались следующие коды:

* 70 - адекватная ошибка (в решении);
* 79 - неадекватная ошибка (практически не относящаяся к решению);
* 99 - пустое задание, решения нет.

В таблице ниже показано распределение суммарного числа ошибок каждого типа в зависимости от уровня задания.

Таблица 2. Распределение ошибок по типу и уровню заданий

|  |  |  |  |
| --- | --- | --- | --- |
| Тип ошибки | Задания 1-го уровня | Задания 2-го уровня | Задания 3-го уровня |
| 99 (пропуск  задания) | 15 | 32 | 60 |
| 70 (адекватная ошибка) | 19 | 26 | 40 |
| 79 (неадекватная ошибка) | 20 | 43 | 41 |

Таким образом, наибольшее число пропусков и адекватных ошибок (т. е. ошибок в решении) наблюдается для заданий 3-го уровня. Неадекватные ошибки встречаются одинаково часто в заданиях 2-го и 3-го уровней.

**Общий вывод.** Учащиеся справляются с заданиями разных уровней. В целом, для учащихся в стране локализации тестирование оказалось труднее, чем для учеников из России.

# ПРИЛОЖЕНИЕ 3

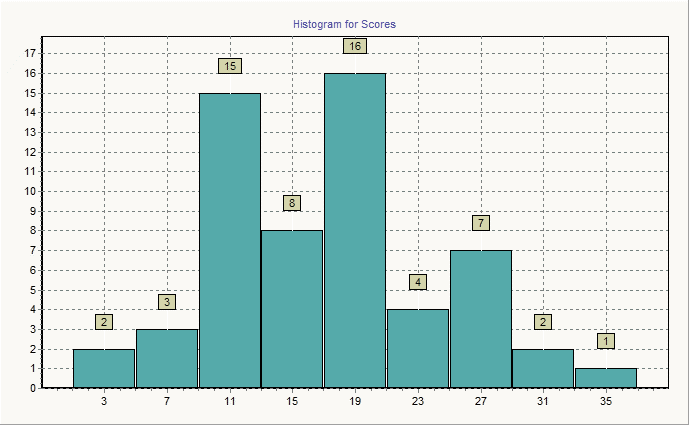
АНАЛИЗ КЛИНИЧЕСКОЙ АПРОБАЦИИ ТЕСТА

Тест SAM по математике в стране локазации имеет структуру, аналогичную оригинальной (российской) версии: 45 заданий, организованных в блоки из трех задач. В каждом блоке задания образовывают иерархию по трудности. Все задания оцениваются дихотомически: 1 балл – правильный ответ, 0 – неправильный. Выборка составила 58 учеников.

В таблице представлены общие статистические показатели. Тест имеет хорошую внутреннюю согласованность, коэффициент альфа Кронбаха равен 0.87.

Таблица 1. Общие результаты теста по математике

|  |  |
| --- | --- |
| Показатель | Значение |
| Число испытуемых | 58 |
| Средний балл | 17 |
| Стандартное отклонение | 6.9 |
| Коэффициент асимметрии | 0.14 |
| Коэффициент эксцесса | -0.49 |
| Средний коэффициент решаемости | 0.37 |
| Средний индекс дискриминативности | 0.34 |
| Средний скор. коэф. точ.-бис. корреляции | 0.32 |
| Показатель надежности (KR20) | 0.87 |
| Стандартная ошибка измерения | 2.5 |

****

*Рисунок 1. Гистограмма распределения первичных баллов участников тестирования*

Общие выводы

* Распределение баллов бимодальное. Однако это может объясняться малым объемом выборки и требует дальнейшей проверки.
* Средний индекс дискриминативности имеет приемлемое значение: 0.34
* Средний коэффициент решаемости составил 0.37, находится в пределах нормы.
* Тест обладают хорошей внутренней согласованностью (коэффициент альфа Кронбаха равен 0.87).

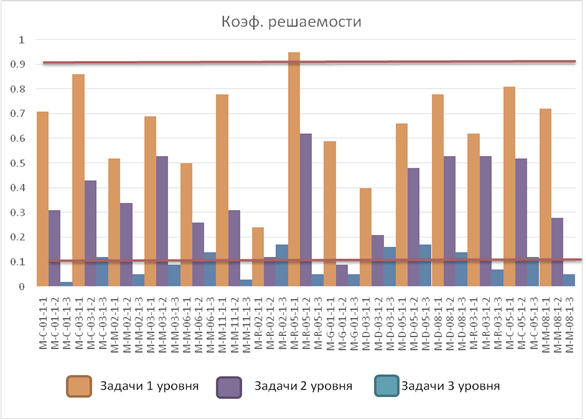
Сравнение результатов тестирования в стране локализации с оригинальной версией теста на русском языке

В таблице ниже цветом (оранжевый) отмечены задания с показателями, выходящими за критические значения (менее 0.1 и более 0.9 для коэффициента решаемости и менее 0.2 для показателей дискриминативности). Серым цветом выделены задания с неупорядоченностью по трудности внутри блока. Дополнительно в таблице приведены статистические показатели по заданиям этого же варианта в оригинальной версии на русском языке.

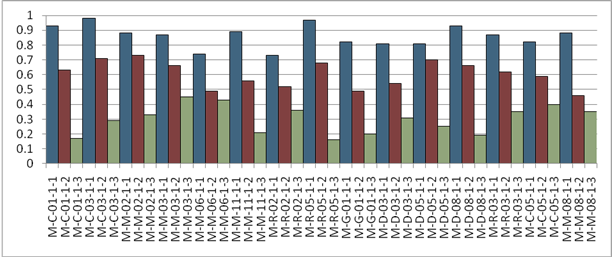
Таблица 2. Сравнительные показатели армянской и оригинальной (российской) версии тестирования SAM по математике

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | ID задания | Версия SAM на армянском языке | | Российская версия | |
| Коэффициент решаемости | Показатель дискриминативности | Коэффициент. решаемости | Показатель дискриминативности |
| 1 | M-C-01-1-1 | 0.71 | 0.34 | 0.93 | 0.13 |
| 2 | M-C-01-1-2 | 0.31 | 0.48 | 0.63 | 0.62 |
| 3 | M-C-01-1-3 | 0.02 | 0.05 | 0.17 | 0.37 |
| 4 | M-C-03-1-1 | 0.86 | 0.25 | 0.98 | 0.06 |
| 5 | M-C-03-1-2 | 0.43 | 0.69 | 0.71 | 0.47 |
| 6 | M-C-03-1-3 | 0.12 | 0.16 | 0.29 | 0.37 |
| 7 | M-M-02-1-1 | 0.52 | 0.54 | 0.88 | 0.31 |
| 8 | M-M-02-1-2 | 0.34 | 0.38 | 0.73 | 0.51 |
| 9 | M-M-02-1-3 | 0.05 | 0.11 | 0.33 | 0.53 |
| 10 | M-M-03-1-1 | 0.69 | 0.54 | 0.87 | 0.28 |
| 11 | M-M-03-1-2 | 0.53 | 0.69 | 0.66 | 0.63 |
| 12 | M-M-03-1-3 | 0.09 | 0.21 | 0.45 | 0.60 |
| 13 | M-M-06-1-1 | 0.5 | 0.44 | 0.74 | 0.49 |
| 14 | M-M-06-1-2 | 0.26 | 0.37 | 0.49 | 0.68 |
| 15 | M-M-06-1-3 | 0.14 | 0.21 | 0.43 | 0.54 |
| 16 | M-M-11-1-1 | 0.78 | 0.24 | 0.89 | 0.20 |
| 17 | M-M-11-1-2 | 0.31 | 0.53 | 0.56 | 0.72 |
| 18 | M-M-11-1-3 | 0.03 | 0.11 | 0.21 | 0.42 |
| 19 | M-R-02-1-1 | 0.24 | 0.38 | 0.73 | 0.60 |
| 20 | M-R-02-1-2 | 0.12 | 0.32 | 0.52 | 0.67 |
| 21 | M-R-02-1-3 | 0.17 | 0.11 | 0.36 | 0.51 |
| 22 | M-R-05-1-1 | 0.95 | 0.05 | 0.97 | 0.05 |
| 23 | M-R-05-1-2 | 0.62 | 0.29 | 0.68 | 0.49 |
| 24 | M-R-05-1-3 | 0.05 | 0 | 0.16 | 0.34 |
| 25 | M-G-01-1-1 | 0.59 | 0.54 | 0.82 | 0.27 |
| 26 | M-G-01-1-2 | 0.09 | -0.05 | 0.49 | 0.56 |
| 27 | M-G-01-1-3 | 0.05 | -0.05 | 0.20 | 0.32 |
| 28 | M-D-03-1-1 | 0.4 | 0.33 | 0.81 | 0.28 |
| 29 | M-D-03-1-2 | 0.21 | 0.42 | 0.54 | 0.61 |
| 30 | M-D-03-1-3 | 0.16 | 0.37 | 0.31 | 0.54 |
| 31 | M-D-05-1-1 | 0.66 | 0.39 | 0.81 | 0.40 |
| 32 | M-D-05-1-2 | 0.48 | 0.79 | 0.70 | 0.50 |
| 33 | M-D-05-1-3 | 0.17 | 0.11 | 0.25 | 0.42 |
| 34 | M-D-08-1-1 | 0.78 | 0.35 | 0.93 | 0.15 |
| 35 | M-D-08-1-2 | 0.53 | 0.59 | 0.66 | 0.54 |
| 36 | M-D-08-1-3 | 0.14 | 0.26 | 0.19 | 0.26 |
| 37 | M-R-03-1-1 | 0.62 | 0.49 | 0.87 | 0.28 |
| 38 | M-R-03-1-2 | 0.53 | 0.44 | 0.62 | 0.47 |
| 39 | M-R-03-1-3 | 0.07 | 0.16 | 0.35 | 0.49 |
| 40 | M-C-05-1-1 | 0.81 | 0.35 | 0.82 | 0.35 |
| 41 | M-C-05-1-2 | 0.52 | 0.64 | 0.59 | 0.61 |
| 42 | M-C-05-1-3 | 0.12 | 0.37 | 0.40 | 0.72 |
| 43 | M-M-08-1-1 | 0.72 | 0.65 | 0.88 | 0.29 |
| 44 | M-M-08-1-2 | 0.28 | 0.53 | 0.46 | 0.71 |
| 45 | M-M-08-1-3 | 0.05 | 0.16 | 0.35 | 0.67 |

* Несколько заданий в армянской версии имеют низкую дискриминативность. Все эти задания обладают экстремальной трудностью. Это задания 3-го уровня, однако одно задание (№26) – 2-го уровня. Все эти задания необходимо подвергнуть дополнительной экспертизе.
* Иерархия по трудности наблюдается во всех блоках, кроме блока №7 (задания 20 и 21). Этот блок необходимо проанализировать целиком (задания №19-21), т.к. все 3 задания в армянской версии намного сложнее, чем в российской (этот блок выделен цветом фиолетовый).
* В целом, почти все задания в армянской версии труднее, чем в оригинальной (российской).

****

*Рисунок 2. График распределения коэффициента решаемости по уровням заданий (армянская версия)*

****

*Рисунок 3. График распределения коэффициента решаемости по уровням заданий*

*(русская версия)*

Ниже представлены графики коэффициентов решаемости и показателей дискриминативности заданий армянской версии. Все показатели в норме. Все задания функционируют удовлетворительно. Часть заданий экстремальной трудности имеют низкую (и даже отрицательную) дискриминативность. Как отмечалось выше, эти задания необходимо проанализировать на предмет их содержания и формы.



*Рисунок 4. График распределения коэффициента решаемости*

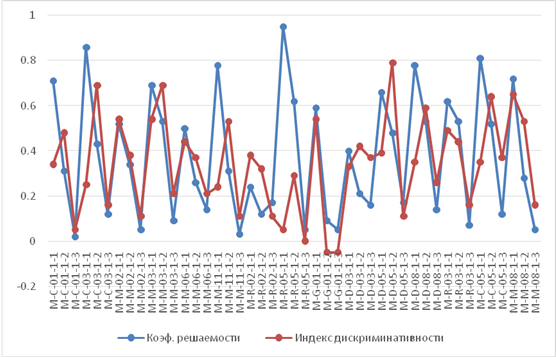
*(армянская версия)*

****

*Рисунок 5. График распределения показателя дифференцирующей*

*способности заданий*

*(армянская версия)*



*Рисунок 6. Сводная диаграмма распределения показателей коэффициентов решаемости и дифференцирующей способности*

*(армянская версия)*

# ПРИЛОЖЕНИЕ 4

АНАЛИЗ ПОЛНОМАСШТАБНОЙ АПРОБАЦИИ ТЕСТА

Тест SAM по математике состоит из 45 заданий, организованных в блоки из трех заданий. В каждом блоке задания должны образовывать иерархию по трудности. Все задания оцениваются дихотомически: 1 балл за правильный ответ, 0 – за неправильный. Выборка исследования составила 220 учеников.

В таблице 1 представлены общие статистические показатели. Тест имеет хорошую внутреннюю согласованность, коэффициент альфа Кронбаха равен 0.87.

Таблица 1. Общие результаты теста по математике

|  |  |
| --- | --- |
| Показатель | Вариант А1 |
| Число учащихся | 220 |
| Средний балл | 18 |
| Максимальный балл | 38 |
| Минимальный балл | 2 |
| Стандартное отклонение | 7,08 |
| Коэффициент асимметрии | 0,05 |
| Коэффициент эксцесса | -0,31 |
| Средний коэффициент решаемости | 0,4 |
| Средний индекс дискриминативности | 0,38 |
| Средний скор. коэф. точ.-бис. корреляции | 0,32 |
| Показатель надежности (KR20) | 0,87 |
| Стандартная ошибка измерения | 2,6 |

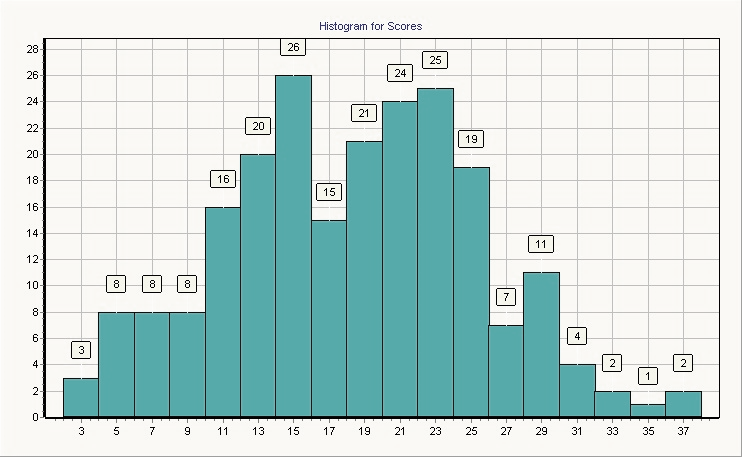


Рисунок 1. Гистограмма распределения первичных баллов участников

Выводы

* Средний коэффициент решаемости имеет оптимальное значение: 0.4
* Средний индекс дискриминативности обладает приемлемым значением: 0.38
* Распределение первичных баллов близко к нормальному.
* По результатам анализа теста на данной выборке надежность теста высокая, что говорит о достаточной внутренней согласованности заданий (Коэффициент альфа Кронбаха равен 0.87).

Общие результаты

Большинство заданий хорошо функционируют и демонстрируют удовлетворительные статистические показатели. Часть заданий имеют повышенный уровень трудности, который обуславливает их низкий уровень дифференцирующей способности. В обоих вариантах наблюдаются блоки с нарушением иерархии трудности. Более детальная информация по функционированию заданий представлена ниже, в таблице 2.

Анализ заданий проводился в рамках Классической теории тестирования (КТТ) и Современной теории тестирования (IRT). Анализ заданий в рамках Классической теории тестирования проводился с использованием показателей: трудность, дискриминативность и скорректированная точечно-бисериальная корреляция. В таблице ниже цветом отмечены задания с показателями, выходящими за критические значения (менее 0.2 для показателей дифференцирующей способности заданий (дискриминативности); для точечно-бисериальной корреляции менее 0.2). Голубым цветом выделены задания с неупорядоченностью по трудности внутри блока, зеленым – задания экстремальной трудности или легкости (выполнили менее 10% или более 90% участников тестирования).

В столбцах 7 и 8 таблицы представлены оценки трудности заданий и соответствующие ошибки измерения (в логитах), оцененные в рамках Современной теории тестирования с использованием дихотомической модели Раша. В последних четырех столбцах (9, 10, 11, 12) приведены значения статистик согласия, характеризующих согласие данных с используемой моделью измерения. В данном отчете используются следующие статистики согласия – простая статистика согласия (OUTFIT MNSQ), стандартизированная простая статистика согласия (OUTFIT ZSTD), взвешенная статистика согласия (INFIT MNSQ), стандартизированная взвешенная статистика согласия (INFIT ZSTD).

Простая статистика согласия более чувствительна к экстремально неожиданным ответам, когда подготовленный учащийся неожиданно неправильно отвечает на легкое задание, или наоборот, менее подготовленный учащийся неожиданно правильно отвечает на трудное задание. Взвешенная версия статистики согласия позволяет уменьшить влияние экстремально неожиданных ответов. Таким образом, эти две статистики относятся к различным частям распределения уровней подготовленности участников тестирования. Продуктивным для измерений называется интервал значений статистик согласия (0.8; 1.2) для простых статистик согласия и (-2; 2) – для стандартизированных статистик согласия. Показатели, выходящие за пределы критических значений, отмечены серым цветом.

Из анализа таблицы 2 следует, что у ряда заданий значения простой и взвешенных статистик согласия превышают правые критические значения. Это означает, что есть учащиеся, которые дали неожиданные ответы на эти задания.

Таблица 2. Статистические показатели функционирования заданий, вариант А1

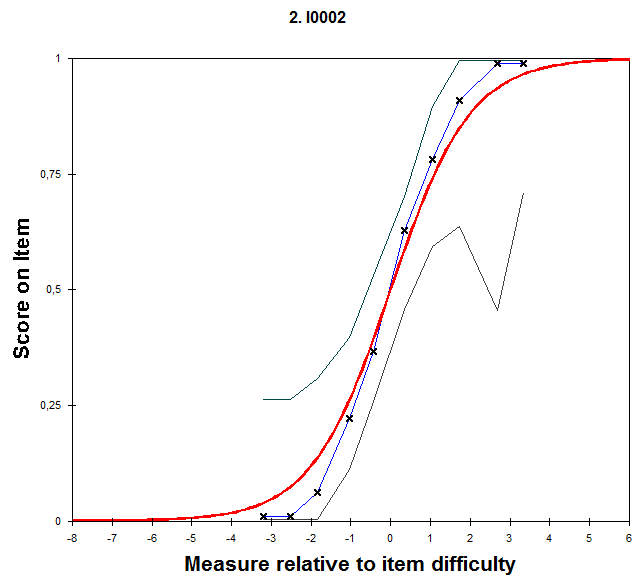
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | Количество выполнивших задание | КТТ | | | IRT | | | | | |
| Трудность | Дискр. | Скор. ТБК | Трудность | Ошибка измерения | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| 1 | М-С-01-1-1 | 178 | 0,81 | 0,28 | 0,24 | -2,61 | 0,19 | 1,1 | 0,9 | 1,24 | 1,1 |
| 2 | М-С-01-1-2 | 104 | 0,47 | 0,64 | 0,46 | -0,67 | 0,15 | 0,92 | -1,3 | 0,86 | -1,5 |
| 3 | М-С-0-1-3 | 16 | 0,07 | 0,13 | 0,25 | 2,22 | 0,27 | 0,91 | -0,4 | 1,12 | 0,4 |
| 4 | М-С-23-1-1 | 185 | 0,84 | 0,33 | 0,31 | -2,88 | 0,2 | 1 | 0,1 | 0,95 | -0,1 |
| 5 | M-C-23-1-2 | 68 | 0,31 | 0,56 | 0,44 | 0,19 | 0,16 | 0,92 | -1,1 | 0,84 | -1,2 |
| 6 | М-С-23-1-3 | 22 | 0,1 | 0,1 | 0,17 | 1,83 | 0,24 | 1,02 | 0,2 | 1,51 | 1,4 |
| 7 | М-С-05-1-1 | 178 | 0,81 | 0,46 | 0,43 | -2,61 | 0,19 | 0,9 | -0,9 | 0,74 | -1,3 |
| 8 | М-С-05-1-2 | 110 | 0,5 | 0,73 | 0,51 | -0,81 | 0,15 | 0,86 | -2,4 | 0,87 | -1,4 |
| 9 | М-С-05-1-3 | 46 | 0,21 | 0,4 | 0,4 | 0,83 | 0,18 | 0,91 | -0,9 | 0,83 | -0,8 |
| 10 | М-М-02-1-1 | 163 | 0,74 | 0,39 | 0,32 | -2,13 | 0,17 | 1,04 | 0,4 | 1,14 | 0,9 |
| 11 | М-М-02-1-2 | 114 | 0,52 | 0,45 | 0,35 | -0,9 | 0,15 | 1,04 | 0,7 | 1,08 | 0,8 |
| 12 | M-М-02-1-3 | 9 | 0,04 | 0,07 | 0,21 | 2,89 | 0,35 | 0,88 | -0,3 | 1,02 | 0,2 |
| 13 | М-М-03-1-1 | 120 | 0,55 | 0,45 | 0,29 | -1,04 | 0,15 | 1,11 | 1,8 | 1,17 | 1,7 |
| 14 | М-М-03-1-2 | 152 | 0,69 | 0,57 | 0,46 | -1,82 | 0,16 | 0,91 | -1,2 | 0,84 | -1,2 |
| 15 | М-М-03-1-3 | 30 | 0,14 | 0,12 | 0,05 | 1,43 | 0,21 | 1,28 | 2 | 1,45 | 1,5 |
| 16 | М-М-11-1-1 | 183 | 0,83 | 0,28 | 0,3 | -2,8 | 0,2 | 0,95 | -0,4 | 1,23 | 1 |
| 17 | М-М-11-1-2 | 84 | 0,38 | 0,66 | 0,49 | -0,2 | 0,15 | 0,89 | -1,9 | 0,78 | -2 |
| 18 | М-М-11-1-3 | 8 | 0,04 | 0,07 | 0,11 | 3,02 | 0,37 | 1,09 | 0,4 | 1,26 | 0,6 |
| 19 | М-М-08-1-1 | 181 | 0,82 | 0,32 | 0,31 | -2,72 | 0,19 | 0,99 | 0 | 1,27 | 1,2 |
| 20 | М-М-08-1-2 | 77 | 0,35 | 0,48 | 0,37 | -0,03 | 0,16 | 1,01 | 0,1 | 1 | 0,1 |
| 21 | М-М-08-1-3 | 19 | 0,09 | 0,13 | 0,22 | 2,01 | 0,25 | 0,98 | -0,1 | 1,24 | 0,7 |
| 22 | М-R-02-1-1 | 106 | 0,48 | 0,66 | 0,47 | -0,71 | 0,15 | 0,91 | -1,5 | 0,89 | -1,2 |
| 23 | М-R-02-1-2 | 61 | 0,28 | 0,62 | 0,51 | 0,38 | 0,17 | 0,84 | -2,2 | 0,71 | -2 |
| 24 | М-R-02-1-3 | 25 | 0,11 | 0,12 | 0,14 | 1,67 | 0,23 | 1,13 | 0,9 | 1,33 | 1 |
| 25 | М-R-05-1-1 | 151 | 0,69 | 0,44 | 0,32 | -1,79 | 0,16 | 1,1 | 1,2 | 1,06 | 0,5 |
| 26 | М-R-05-1-2 | 56 | 0,25 | 0,12 | 0,08 | 0,52 | 0,17 | 1,28 | 3,1 | 1,71 | 3,6 |
| 27 | М-R-05-1-3 | 6 | 0,03 | 0,02 | 0,11 | 3,34 | 0,43 | 0,9 | -0,1 | 2,28 | 1,8 |
| 28 | М-D-03-1-1 | 127 | 0,58 | 0,6 | 0,41 | -1,2 | 0,15 | 0,99 | -0,2 | 0,96 | -0,3 |
| 29 | М-D-03-1-2 | 76 | 0,35 | 0,59 | 0,44 | -0,01 | 0,16 | 0,92 | -1,2 | 0,91 | -0,7 |
| 30 | М-D-03-1-3 | 39 | 0,18 | 0,3 | 0,31 | 1,07 | 0,19 | 0,99 | -0,1 | 0,98 | 0 |
| 31 | М-D-20-1-1 | 150 | 0,68 | 0,6 | 0,46 | -1,77 | 0,16 | 0,88 | -1,6 | 0,89 | -0,8 |
| 32 | М-D-20-1-2 | 70 | 0,32 | 0,46 | 0,37 | 0,14 | 0,16 | 0,99 | -0,1 | 0,95 | -0,3 |
| 33 | М-D-20-1-3 | 7 | 0,03 | 0,05 | 0,14 | 3,17 | 0,4 | 1,03 | 0,2 | 1,1 | 0,4 |
| 34 | М-D-05-1-1 | 137 | 0,62 | 0,74 | 0,53 | -1,44 | 0,16 | 0,82 | -2,8 | 0,77 | -2,2 |
| 35 | М-D-05-1-2 | 114 | 0,52 | 0,52 | 0,34 | -0,9 | 0,15 | 1,06 | 1 | 1,2 | 2 |
| 36 | М-D-05-1-3 | 18 | 0,08 | 0,13 | 0,15 | 2,08 | 0,26 | 1,07 | 0,4 | 2,33 | 2,7 |
| 37 | М-D-08-1-1 | 152 | 0,69 | 0,51 | 0,39 | -1,82 | 0,16 | 1 | 0 | 0,9 | -0,7 |
| 38 | М-D-08-1-2 | 115 | 0,52 | 0,68 | 0,52 | -0,92 | 0,15 | 0,85 | -2,6 | 0,79 | -2,4 |
| 39 | М-D-08-1-3 | 41 | 0,19 | 0,35 | 0,35 | 1 | 0,19 | 0,96 | -0,4 | 0,87 | -0,5 |
| 40 | М-G-01-1-1 | 127 | 0,58 | 0,33 | 0,29 | -1,2 | 0,15 | 1,13 | 2,1 | 1,11 | 1,1 |
| 41 | М-G-01-1-2 | 68 | 0,31 | 0,37 | 0,26 | 0,19 | 0,16 | 1,1 | 1,5 | 1,26 | 1,8 |
| 42 | М-G-01-1-3 | 5 | 0,02 | 0,05 | 0,16 | 3,54 | 0,46 | 0,96 | 0,1 | 0,91 | 0,1 |
| 43 | М-G-03-1-1 | 147 | 0,67 | 0,43 | 0,3 | -1,69 | 0,16 | 1,12 | 1,6 | 1,13 | 1,1 |
| 44 | М-G-03-1-2 | 67 | 0,3 | 0,64 | 0,5 | 0,22 | 0,16 | 0,86 | -2 | 0,71 | -2,2 |
| 45 | М-G-03-1-3 | 9 | 0,04 | 0,07 | 0,08 | 2,89 | 0,35 | 1,13 | 0,5 | 1,26 | 0,6 |

На основании результатов анализа функционирования заданий в рамках Классической и Современной теориях тестирования были определены задания, требующие дополнительной проверки на качество функционирования и перевода.

Комментарии о функционировании заданий

В качестве примера хорошо функционирующего задания приводится характеристическая кривая задания №2 (М-С-01-1-2) (рисунок 2). Точки эмпирического распределения максимально приближены к вероятностной кривой, что означает, что задание хорошо дифференцирует учащихся с разным уровнем подготовленности.

Характеристическая кривая задания представляет собой вероятность правильного ответа на задание в зависимости от уровня подготовленности учащихся (красная линия). Крестиками на рисунке обозначены точки эмпирического распределения ответов учащихся на конкретное задание. Они представляют собой средний балл по данному заданию по группам учащихся (вся выборка была поделена на 10 частей в зависимости от балла по тесту). Также на рисунке показаны границы 95%-ного доверительного интервала для точек эмпирического распределения (серые линии).



*Рисунок 2.* Характеристическая кривая задания №2 (М-С-01-1-2)

ПРИЛОЖЕНИЕ 5

АНАЛИЗ РЕЗУЛЬТАТОВ ПОЛНОМАСШТАБНОЙ АПРОБАЦИИ ТЕСТА ПО МАТЕМАТИКЕ (ВАРИАНТЫ А1 И Б1)

АРМЕНИЯ, ВЕСНА 2018

Сводные результаты по двум вариантам

Тест состоял из двух вариантов по 45 заданий, объединенных в 15 блоков. В табл. 1.1 представлены общие статистические показатели.

Оба варианта теста имеют высокую внутреннюю согласованность, коэффициент альфа Кронбаха равен 0,87 и 0,84.

Таблица 1. Общие результаты теста по математике

|  |  |  |
| --- | --- | --- |
|  | Вариант А1 | Вариант Б1 |
| Число учащихся | 220 | 177 |
| Средний балл | 18 | 17 |
| Максимальный балл | 38 | 37 |
| Минимальный балл | 2 | 3 |
| Стандартное отклонение | 7,08 | 6,65 |
| Коэффициент асимметрии | 0,05 | 0,24 |
| Коэффициент эксцесса | -0,31 | -0,1 |
| Средний коэффициент решаемости | 0,4 | 0,38 |
| Средний индекс дискриминативности | 0,38 | 0,36 |
| Средний скор. коэф. точ.-бис. корреляции | 0,32 | 0,28 |
| Показатель надежности (KR20) | 0,87 | 0,84 |
| Стандартная ошибка измерения | 2,6 | 2,7 |

Вариант А1 Вариант Б1

|  |  |
| --- | --- |
|  |  |

*Рисунок 1. Гистограммы распределения первичных баллов участников (варианты А1 и Б1)*

Выводы:

* Средние коэффициенты решаемости по обоим вариантам имеют оптимальные значения: 0,4 и 0,38.
* Средний индекс дискриминативности по обоим вариантам также обладает оптимальными значениями: 0,38 и 0,36.
* Распределение первичных баллов во втором варианте скошено влево, что позволяет сделать вывод о том, что вариант Б1 оказался немного труднее для учащихся.
* По результатам анализа теста на данной выборке надежность тестов высокая. Коэффициент альфа Кронбаха равен 0,87 и 0,84.

Результаты по отдельным вариантам

Большинство заданий хорошо функционируют и демонстрируют удовлетворительные статистические показатели. Часть заданий имеют повышенный уровень трудности, который обуславливает их низкий уровень дифференцирующей способности. В обоих вариантах наблюдаются блоки с нарушением иерархии трудности. Более детальная информация по функционированию заданий представлена ниже, в таблицах 2.1 и 2.2.

Анализ заданий проводился в рамках Классической теории тестирования (КТТ) и Современной теории тестирования (IRT). Анализ заданий в рамках Классической теории тестирования проводился с использованием показателей: трудность, дискриминативность и скорректированная точечно-бисериальная корреляция. В таблице ниже цветом отмечены задания с показателями, выходящими за критические значения (менее 0,2 для показателей дифференцирующей способности заданий (дискриминативности); для точечно-бисериальной корреляции менее 0,2). Голубым цветом выделены задания с неупорядоченностью по трудности внутри блока, зеленым – задания экстремальной трудности или легкости (выполнили менее 10% или более 90% участников тестирования).

В столбцах 7 и 8 таблицы представлены оценки трудности заданий и соответствующие ошибки измерения (в логитах), оцененные в рамках Современной теории тестирования с использованием дихотомической модели Раша. В последних четырех столбцах (9, 10, 11, 12) приведены значения статистик согласия, характеризующих согласие данных с используемой моделью измерения. В данном отчете используются следующие статистики согласия – простая статистика согласия (OUTFIT MNSQ), стандартизированная простая статистика согласия (OUTFIT ZSTD), взвешенная статистика согласия (INFIT MNSQ), стандартизированная взвешенная статистика согласия (INFIT ZSTD).

Простая статистика согласия более чувствительна к экстремально неожиданным ответам, когда подготовленный учащийся неожиданно неправильно отвечает на легкое задание, или наоборот, менее подготовленный учащийся неожиданно правильно отвечает на трудное задание. Взвешенная версия статистики согласия позволяет уменьшить влияние экстремально неожиданных ответов. Таким образом, эти две статистики относятся к различным частям распределения уровней подготовленности участников тестирования. Продуктивным для измерений называется интервал значений статистик согласия (0,8; 1,2) для простых статистик согласия и (-2; 2) – для стандартизированных статистик согласия. Показатели, выходящие за пределы критических значений, отмечены серым цветом.

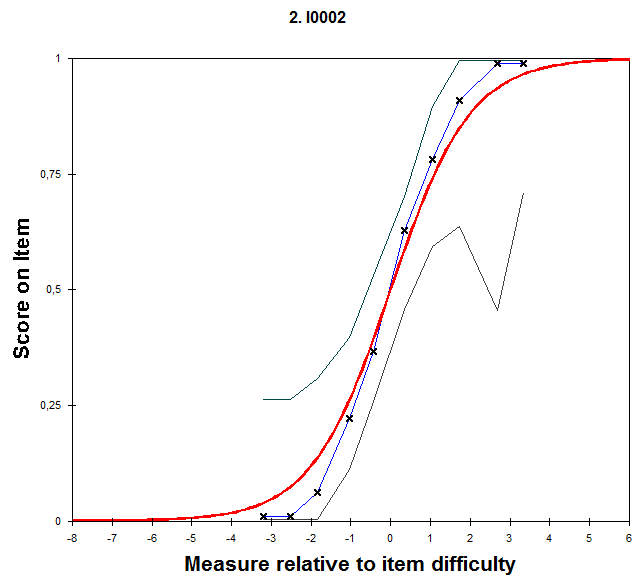
Из анализа таблицы 2.1. следует, что у ряда заданий значения простой и взвешенных статистик согласия превышают правые критические значения, что означает, что есть учащиеся, давшие неожиданные ответы на эти задания.

Таблица 2.1. Статистические показатели функционирования заданий, вариант А1

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | | Количество выполнивших задание | | КТТ | | | | | | IRT | | | | | | | | | | | |
| Трудность | | Дискр. | | Скор. ТБК | | Трудность | | Ошибка измерения | | INFIT MNSQ | | INFIT ZSTD | | OUTFIT MNSQ | | OUTFIT ZSTD | |
| **1** | **2** | **3** | | **4** | | **5** | | **6** | | **7** | | **8** | | **9** | | **10** | | **11** | | **12** | |
| 1 | М-С-01-1-1 | 178 | | 0,81 | | 0,28 | | 0,24 | | -2,61 | | 0,19 | | 1,1 | | 0,9 | | 1,24 | | 1,1 | |
| 2 | М-С-01-1-2 | 104 | | 0,47 | | 0,64 | | 0,46 | | -0,67 | | 0,15 | | 0,92 | | -1,3 | | 0,86 | | -1,5 | |
| 3 | М-С-0-1-3 | 16 | | 0,07 | | 0,13 | | 0,25 | | 2,22 | | 0,27 | | 0,91 | | -0,4 | | 1,12 | | 0,4 | |
| 4 | М-С-23-1-1 | 185 | | 0,84 | | 0,33 | | 0,31 | | -2,88 | | 0,2 | | 1 | | 0,1 | | 0,95 | | -0,1 | |
| 5 | M-C-23-1-2 | 68 | | 0,31 | | 0,56 | | 0,44 | | 0,19 | | 0,16 | | 0,92 | | -1,1 | | 0,84 | | -1,2 | |
| 6 | М-С-23-1-3 | 22 | | 0,1 | | 0,1 | | 0,17 | | 1,83 | | 0,24 | | 1,02 | | 0,2 | | 1,51 | | 1,4 | |
| 7 | М-С-05-1-1 | 178 | | 0,81 | | 0,46 | | 0,43 | | -2,61 | | 0,19 | | 0,9 | | -0,9 | | 0,74 | | -1,3 | |
| 8 | М-С-05-1-2 | 110 | | 0,5 | | 0,73 | | 0,51 | | -0,81 | | 0,15 | | 0,86 | | -2,4 | | 0,87 | | -1,4 | |
| 9 | М-С-05-1-3 | 46 | | 0,21 | | 0,4 | | 0,4 | | 0,83 | | 0,18 | | 0,91 | | -0,9 | | 0,83 | | -0,8 | |
| 10 | М-М-02-1-1 | 163 | | 0,74 | | 0,39 | | 0,32 | | -2,13 | | 0,17 | | 1,04 | | 0,4 | | 1,14 | | 0,9 | |
| 11 | М-М-02-1-2 | 114 | | 0,52 | | 0,45 | | 0,35 | | -0,9 | | 0,15 | | 1,04 | | 0,7 | | 1,08 | | 0,8 | |
| 12 | M-М-02-1-3 | 9 | | 0,04 | | 0,07 | | 0,21 | | 2,89 | | 0,35 | | 0,88 | | -0,3 | | 1,02 | | 0,2 | |
| 13 | М-М-03-1-1 | 120 | | 0,55 | | 0,45 | | 0,29 | | -1,04 | | 0,15 | | 1,11 | | 1,8 | | 1,17 | | 1,7 | |
| 14 | М-М-03-1-2 | 152 | | 0,69 | | 0,57 | | 0,46 | | -1,82 | | 0,16 | | 0,91 | | -1,2 | | 0,84 | | -1,2 | |
| 15 | М-М-03-1-3 | 30 | | 0,14 | | 0,12 | | 0,05 | | 1,43 | | 0,21 | | 1,28 | | 2 | | 1,45 | | 1,5 | |
| 16 | М-М-11-1-1 | 183 | | 0,83 | | 0,28 | | 0,3 | | -2,8 | | 0,2 | | 0,95 | | -0,4 | | 1,23 | | 1 | |
| 17 | М-М-11-1-2 | 84 | | 0,38 | | 0,66 | | 0,49 | | -0,2 | | 0,15 | | 0,89 | | -1,9 | | 0,78 | | -2 | |
| 18 | М-М-11-1-3 | 8 | | 0,04 | | 0,07 | | 0,11 | | 3,02 | | 0,37 | | 1,09 | | 0,4 | | 1,26 | | 0,6 | |
| 19 | М-М-08-1-1 | 181 | | 0,82 | | 0,32 | | 0,31 | | -2,72 | | 0,19 | | 0,99 | | 0 | | 1,27 | | 1,2 | |
| 20 | М-М-08-1-2 | 77 | | 0,35 | | 0,48 | | 0,37 | | -0,03 | | 0,16 | | 1,01 | | 0,1 | | 1 | | 0,1 | |
| 21 | М-М-08-1-3 | 19 | | 0,09 | | 0,13 | | 0,22 | | 2,01 | | 0,25 | | 0,98 | | -0,1 | | 1,24 | | 0,7 | |
| 22 | М-R-02-1-1 | 106 | | 0,48 | | 0,66 | | 0,47 | | -0,71 | | 0,15 | | 0,91 | | -1,5 | | 0,89 | | -1,2 | |
| 23 | М-R-02-1-2 | 61 | | 0,28 | | 0,62 | | 0,51 | | 0,38 | | 0,17 | | 0,84 | | -2,2 | | 0,71 | | -2 | |
| 24 | М-R-02-1-3 | 25 | | 0,11 | | 0,12 | | 0,14 | | 1,67 | | 0,23 | | 1,13 | | 0,9 | | 1,33 | | 1 | |
| 25 | М-R-05-1-1 | 151 | | 0,69 | | 0,44 | | 0,32 | | -1,79 | | 0,16 | | 1,1 | | 1,2 | | 1,06 | | 0,5 | |
| 26 | М-R-05-1-2 | 56 | | 0,25 | | 0,12 | | 0,08 | | 0,52 | | 0,17 | | 1,28 | | 3,1 | | 1,71 | | 3,6 | |
| 27 | М-R-05-1-3 | 6 | | 0,03 | | 0,02 | | 0,11 | | 3,34 | | 0,43 | | 0,9 | | -0,1 | | 2,28 | | 1,8 | |
| 28 | М-D-03-1-1 | 127 | | 0,58 | | 0,6 | | 0,41 | | -1,2 | | 0,15 | | 0,99 | | -0,2 | | 0,96 | | -0,3 | |
| 29 | М-D-03-1-2 | 76 | | 0,35 | | 0,59 | | 0,44 | | -0,01 | | 0,16 | | 0,92 | | -1,2 | | 0,91 | | -0,7 | |
| 30 | М-D-03-1-3 | 39 | | 0,18 | | 0,3 | | 0,31 | | 1,07 | | 0,19 | | 0,99 | | -0,1 | | 0,98 | | 0 | |
| 31 | М-D-20-1-1 | 150 | | 0,68 | | 0,6 | | 0,46 | | -1,77 | | 0,16 | | 0,88 | | -1,6 | | 0,89 | | -0,8 | |
| 32 | М-D-20-1-2 | 70 | | 0,32 | | 0,46 | | 0,37 | | 0,14 | | 0,16 | | 0,99 | | -0,1 | | 0,95 | | -0,3 | |
| 33 | М-D-20-1-3 | 7 | | 0,03 | | 0,05 | | 0,14 | | 3,17 | | 0,4 | | 1,03 | | 0,2 | | 1,1 | | 0,4 | |
| 34 | М-D-05-1-1 | 137 | | 0,62 | | 0,74 | | 0,53 | | -1,44 | | 0,16 | | 0,82 | | -2,8 | | 0,77 | | -2,2 | |
| 35 | М-D-05-1-2 | 114 | | 0,52 | | 0,52 | | 0,34 | | -0,9 | | 0,15 | | 1,06 | | 1 | | 1,2 | | 2 | |
| 36 | М-D-05-1-3 | 18 | | 0,08 | | 0,13 | | 0,15 | | 2,08 | | 0,26 | | 1,07 | | 0,4 | | 2,33 | | 2,7 | |
| 37 | М-D-08-1-1 | 152 | | 0,69 | | 0,51 | | 0,39 | | -1,82 | | 0,16 | | 1 | | 0 | | 0,9 | | -0,7 | |
| 38 | М-D-08-1-2 | 115 | | 0,52 | | 0,68 | | 0,52 | | -0,92 | | 0,15 | | 0,85 | | -2,6 | | 0,79 | | -2,4 | |
| 39 | М-D-08-1-3 | 41 | | 0,19 | | 0,35 | | 0,35 | | 1 | | 0,19 | | 0,96 | | -0,4 | | 0,87 | | -0,5 | |
| 40 | М-G-01-1-1 | 127 | | 0,58 | | 0,33 | | 0,29 | | -1,2 | | 0,15 | | 1,13 | | 2,1 | | 1,11 | | 1,1 | |
| 41 | М-G-01-1-2 | 68 | | 0,31 | | 0,37 | | 0,26 | | 0,19 | | 0,16 | | 1,1 | | 1,5 | | 1,26 | | 1,8 | |
| 42 | М-G-01-1-3 | 5 | | 0,02 | | 0,05 | | 0,16 | | 3,54 | | 0,46 | | 0,96 | | 0,1 | | 0,91 | | 0,1 | |
| 43 | М-G-03-1-1 | 147 | | 0,67 | | 0,43 | | 0,3 | | -1,69 | | 0,16 | | 1,12 | | 1,6 | | 1,13 | | 1,1 | |
| 44 | М-G-03-1-2 | 67 | | 0,3 | | 0,64 | | 0,5 | | 0,22 | | 0,16 | | 0,86 | | -2 | | 0,71 | | -2,2 | |
| 45 | М-G-03-1-3 | 9 | | 0,04 | | 0,07 | | 0,08 | | 2,89 | | 0,35 | | 1,13 | | 0,5 | | 1,26 | | 0,6 | |

Комментарии о функционировании заданий, вариант А1

В качестве примера хорошо функционирующего задания приводится характеристическая кривая задания №2 (М-С-01-1-2). Точки эмпирического распределения максимально приближены к вероятностной кривой, что означает, что задание хорошо дифференцирует учащихся с разным уровнем подготовленности.



*Рисунок 2.* Характеристическая кривая задания №2 (М-С-01-1-2)

Таблица 2.2. Статистические показатели функционирования заданий, вариант Б1

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | Количество выполнивших задание | КТТ | | | IRT | | | | | |
| Трудность | Дискр. | Скор. ТБК | Трудность | Ошибка измерения | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | М-С-03-1-1 | 158 | 0,89 | 0,12 | 0,12 | -3,26 | 0,26 | 1,02 | 0,2 | 1,67 | 1,9 |
| 2 | М-С-03-1-2 | 131 | 0,74 | 0,59 | 0,45 | -2,04 | 0,19 | 0,85 | -1,7 | 0,77 | -1,6 |
| 3 | М-С-03-1-3 | 29 | 0,16 | 0,17 | 0,16 | 1,1 | 0,22 | 1,07 | 0,5 | 1,32 | 1,4 |
| 4 | М-С-22-1-1 | 96 | 0,54 | 0,4 | 0,3 | -1,02 | 0,17 | 1,01 | 0,2 | 1,08 | 1 |
| 5 | М-С-22-1-2 | 74 | 0,42 | 0,22 | 0,13 | -0,4 | 0,17 | 1,18 | 2,9 | 1,32 | 3,1 |
| 6 | М-С-22-1-3 | 61 | 0,34 | 0,56 | 0,37 | -0,03 | 0,17 | 0,96 | -0,5 | 0,91 | -0,8 |
| 7 | М-С-25-1-1 | 137 | 0,77 | 0,41 | 0,3 | -2,25 | 0,19 | 0,98 | -0,2 | 1 | 0,1 |
| 8 | М-С-25-1-2 | 57 | 0,32 | 0,39 | 0,27 | 0,09 | 0,17 | 1,04 | 0,6 | 0,99 | 0 |
| 9 | М-С-25-1-3 | 5 | 0,03 | 0 | -0,03 | 3,17 | 0,46 | 1,1 | 0,4 | 1,65 | 1,1 |
| 10 | М-М-20-1-1 | 124 | 0,7 | 0,53 | 0,44 | -1,81 | 0,18 | 0,89 | -1,3 | 0,77 | -1,8 |
| 11 | М-М-20-1-2 | 54 | 0,31 | 0,39 | 0,32 | 0,18 | 0,18 | 0,98 | -0,2 | 1,09 | 0,7 |
| 12 | М-М-20-1-3 | 35 | 0,2 | 0,25 | 0,33 | 0,84 | 0,2 | 0,94 | -0,5 | 0,93 | -0,3 |
| 13 | М-М-23-1-1 | 117 | 0,66 | 0,34 | 0,27 | -1,59 | 0,17 | 1,04 | 0,5 | 1,08 | 0,7 |
| 14 | М-М-23-1-2 | 100 | 0,56 | 0,44 | 0,35 | -1,11 | 0,17 | 0,97 | -0,4 | 0,97 | -0,3 |
| 15 | М-М-23-1-3 | 45 | 0,25 | 0,41 | 0,29 | 0,47 | 0,19 | 1,02 | 0,2 | 0,91 | -0,5 |
| 16 | М-М-06-1-1 | 95 | 0,54 | 0,54 | 0,4 | -0,97 | 0,16 | 0,94 | -1 | 0,9 | -1,2 |
| 17 | М-М-06-1-2 | 60 | 0,34 | 0,51 | 0,37 | 0 | 0,17 | 0,96 | -0,6 | 0,9 | -0,9 |
| 18 | М-М-06-1-3 | 37 | 0,21 | 0,35 | 0,27 | 0,76 | 0,2 | 1,01 | 0,1 | 1,02 | 0,2 |
| 19 | М-M-24-1-1 | 139 | 0,79 | 0,59 | 0,5 | -2,33 | 0,2 | 0,8 | -1,9 | 0,64 | -2,3 |
| 20 | М-M-24-1-2 | 93 | 0,53 | 0,7 | 0,53 | -0,92 | 0,16 | 0,82 | -3,3 | 0,76 | -3,1 |
| 21 | М-M-24-1-3 | 29 | 0,16 | 0,37 | 0,37 | 1,1 | 0,22 | 0,91 | -0,7 | 0,77 | -1 |
| 22 | М-R-03-1-1 | 105 | 0,59 | 0,4 | 0,23 | -1,24 | 0,17 | 1,09 | 1,5 | 1,17 | 1,7 |
| 23 | М-R-03-1-2 | 87 | 0,49 | 0,4 | 0,25 | -0,75 | 0,16 | 1,08 | 1,3 | 1,12 | 1,3 |
| 24 | М-R-03-1-3 | 33 | 0,19 | 0,45 | 0,4 | 0,93 | 0,21 | 0,88 | -1 | 0,97 | -0,1 |
| 25 | М-R-22-1-1 | 146 | 0,82 | 0,41 | 0,4 | -2,62 | 0,21 | 0,84 | -1,3 | 0,82 | -0,8 |
| 26 | М-R-22-1-2 | 56 | 0,32 | 0,61 | 0,51 | 0,12 | 0,17 | 0,83 | -2,3 | 0,73 | -2,4 |
| 27 | М-R-22-1-3 | 33 | 0,19 | 0,37 | 0,3 | 0,93 | 0,21 | 1,01 | 0,2 | 0,83 | -0,8 |
| 28 | М-D-25-1-1 | 80 | 0,45 | 0,76 | 0,54 | -0,57 | 0,16 | 0,8 | -3,6 | 0,76 | -3 |
| 29 | М-D-25-1-2 | 73 | 0,41 | 0,54 | 0,39 | -0,37 | 0,17 | 0,95 | -0,8 | 0,9 | -1 |
| 30 | М-D-25-1-3 | 12 | 0,07 | 0,14 | 0,24 | 2,2 | 0,31 | 0,96 | -0,1 | 0,87 | -0,2 |
| 31 | М-D-21-1-1 | 33 | 0,19 | 0,35 | 0,3 | 0,93 | 0,21 | 0,97 | -0,2 | 1,11 | 0,6 |
| 32 | М-D-21-1-2 | 37 | 0,21 | 0,25 | 0,23 | 0,76 | 0,2 | 1,01 | 0,1 | 1,42 | 2,1 |
| 33 | М-D-21-1-3 | 9 | 0,05 | 0,02 | 0,04 | 2,52 | 0,35 | 1,06 | 0,3 | 1,4 | 0,9 |
| 34 | М-D-22-1-1 | 77 | 0,44 | 0,48 | 0,3 | -0,48 | 0,17 | 1,03 | 0,5 | 1,07 | 0,8 |
| 35 | М-D-22-1-2 | 51 | 0,29 | 0,41 | 0,33 | 0,27 | 0,18 | 0,98 | -0,2 | 0,99 | 0 |
| 36 | М-D-22-1-3 | 25 | 0,14 | 0,25 | 0,27 | 1,3 | 0,23 | 0,99 | 0 | 0,99 | 0,1 |
| 37 | М-D-26-1-1 | 104 | 0,59 | 0,41 | 0,3 | -1,22 | 0,17 | 1,04 | 0,6 | 0,98 | -0,2 |
| 38 | М-D-03-1-2 | 59 | 0,33 | 0,29 | 0,2 | 0,03 | 0,17 | 1,11 | 1,5 | 1,15 | 1,3 |
| 39 | М-D-26-1-3 | 6 | 0,03 | 0,02 | 0,07 | 2,97 | 0,42 | 0,96 | 0 | 4,74 | 3,6 |
| 40 | М-G-21-1-1 | 87 | 0,49 | 0,12 | 0,05 | -0,75 | 0,16 | 1,28 | 4,5 | 1,35 | 3,8 |
| 41 | М-G-21-1-2 | 87 | 0,49 | 0,44 | 0,3 | -0,75 | 0,16 | 1,03 | 0,5 | 1,02 | 0,2 |
| 42 | М-G-21-1-3 | 29 | 0,16 | 0,15 | 0,13 | 1,1 | 0,22 | 1,14 | 1,1 | 1,15 | 0,7 |
| 43 | М-G-24-1-1 | 110 | 0,62 | 0,35 | 0,22 | -1,39 | 0,17 | 1,09 | 1,3 | 1,27 | 2,5 |
| 44 | М-G-24-1-2 | 10 | 0,06 | 0,1 | 0,08 | 2,4 | 0,34 | 1,08 | 0,4 | 3,05 | 3,1 |
| 45 | М-G-24-1-3 | 3 | 0,02 | 0,04 | 0,07 | 3,71 | 0,59 | 1,04 | 0,3 | 0,85 | 0,1 |

# ПРИЛОЖЕНИЕ 6

АНАЛИЗ РЕЗУЛЬТАТОВ ПОЛНОМАСШТАБНОЙ АПРОБАЦИИ ТЕСТА ПО МАТЕМАТИКЕ (ВАРИАНТЫ А1 И Б1)

(БЕЛАРУСЬ, ВЕСНА 2018)

Сводные результаты по двум вариантам

Тест состоял из двух вариантов по 45 заданий, объединенных в 15 блоков. В таблице 1 представлены общие статистические показатели, а на Рисунке 1 - гистограммы распределения первичных баллов учащихся.

Оба варианта теста имеют высокую внутреннюю согласованность, коэффициент альфа Кронбаха равен 0,89 и 0,88.

Таблица 1. Общие результаты теста по математике

|  |  |  |
| --- | --- | --- |
|  | Вариант А1 | Вариант Б1 |
| Число учащихся | 262 | 262 |
| Средний балл | 20 | 18,7 |
| Максимальный балл | 41 | 40 |
| Минимальный балл | 0 | 1 |
| Стандартное отклонение | 7,5 | 7,443 |
| Коэффициент асимметрии | -0.139 | 0,357 |
| Коэффициент эксцесса | -0.229 | -0,209 |
| Средний коэффициент решаемости | 0,446 | 0,415 |
| Средний индекс дискриминативности | 0,407 | 0,387 |
| Средний скор. коэф. точ.-бис. корреляции | 0,404 | 0,395 |
| Показатель надежности (KR20) | 0,891 | 0,876 |
| Стандартная ошибка измерения | 2,478 | 2,621 |

Вариант А1 Вариант Б1

|  |  |
| --- | --- |
|  |  |

*Рисунок 1. Гистограммы распределения первичных баллов участников (варианты А1 и Б1)*

Выводы:

* Средние коэффициенты решаемости по обоим вариантам достигают оптимальных значений: 0,45 и 0,42.
* Средний индекс дискриминативности по обоим вариантам обладает оптимальным значением 0,41 и 0,39.
* Распределение первичных баллов в варианте Б1 варианте немного скошено влево, что позволяет сделать вывод о том, что вариант Б1 оказался труднее для испытуемых, чем вариант А1.
* По результатам анализа теста на данной выборке надежность тестов высокая. Коэффициент альфа Кронбаха равен 0,89 и 0,88.

Результаты по отдельным вариантам

Анализ варианта A1

Большинство заданий хорошо функционируют и демонстрируют удовлетворительные статистические показатели. Часть заданий имеют повышенный уровень трудности, который обуславливает их низкий уровень дифференцирующей способности. В обоих вариантах наблюдаются блоки с нарушением иерархии трудности. Более детальная информация по функционированию заданий представлена ниже, в таблицах 2.1. и 2.2.

Анализ заданий проводился в рамках Классической теории тестирования и Современной теории тестирования. Анализ заданий в рамках Классической теории тестирования проводился с использованием показателей: трудность, дискриминативность и скорректированная точечно-бисериальная корреляция. В таблице ниже цветом отмечены задания с показателями, выходящими за критические значения (менее 0,2 для показателей дифференцирующей способности заданий (дискриминативности); для точечно-бисериальной корреляции менее 0,2). Голубым цветом выделены задания с неупорядоченностью по трудности внутри блока, зеленым – задания экстремальной трудности или легкости (выполнили менее 10% или более 90% участников тестирования).

В столбцах 7 и 8 таблицы представлены оценки трудности заданий и соответствующие ошибки измерения (в логитах), оцененные в рамках Современной теории тестирования с использованием дихотомической модели Раша. В последних четырех столбцах (9, 10, 11, 12) приведены значения статистик согласия, характеризующих согласие данных с используемой моделью измерения. В данном отчете используются следующие статистики согласия – простая статистика согласия (OUTFIT MNSQ), стандартизированная простая статистика согласия (OUTFIT ZSTD), взвешенная статистика согласия (INFIT MNSQ), стандартизированная взвешенная статистика согласия (INFIT ZSTD). Простая статистика согласия более чувствительна к экстремально неожиданным ответам, когда сильный испытуемый неожиданно неправильно отвечает на легкое задание, или наоборот, слабый испытуемый неожиданно правильно отвечает на трудное задание. Взвешенная версия статистики согласия позволяет уменьшить влияние экстремально неожиданных ответов. Таким образом, эти две статистики относятся к различным частям распределения уровней подготовленности участников тестирования. Продуктивным для измерений называется интервал значений статистик согласия (0,8; 1,2) для простых статистик согласия и (-2; 2) – для стандартизированных статистик согласия. Показатели, выходящие за пределы критических значений, отмечены серым цветом.

Из анализа Таблицы 2.1. следует, что у ряда заданий значения простой и взвешенных статистик согласия превышают правые критические значения, что означает, что есть испытуемые, давшие неожиданные ответы на эти задания.

Таблица 2.1. Статистические показатели функционирования заданий, вариант А1

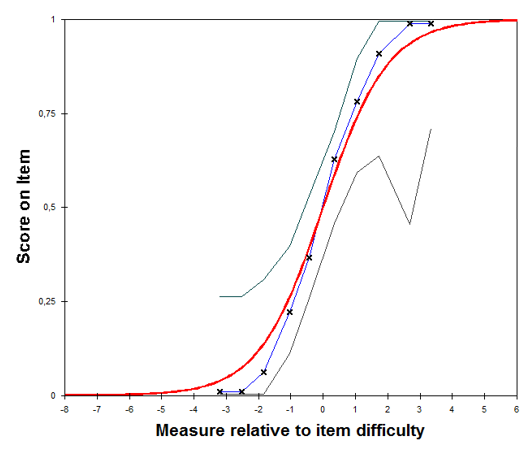
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | Количество выполнивших задание | КТТ | | | IRT | | | | | |
| Трудность | Дискр. | Скор. ТБК | Трудность | Ошибка измерения | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | М-С-01-1-1 | 241 | 0,92 | 0,33 | 0,28 | -3,52 | 0,24 | 0,98 | -0,04 | 0,98 | 0,06 |
| 2 | М-С-01-1-2 | 181 | 0,69 | 0,61 | 0,43 | -1,61 | 0,15 | 0,98 | -0,28 | 0,91 | -0,69 |
| 3 | М-С-0-1-3 | 69 | 0,26 | 0,29 | 0,39 | 0,68 | 0,16 | 1,07 | 0,85 | 0,94 | -0,33 |
| 4 | М-С-23-1-1 | 156 | 0,60 | 0,52 | 0,33 | -1,09 | 0,14 | 1,11 | 1,90 | 1,26 | 2,46 |
| 5 | M-C-23-1-2 | 132 | 0,50 | 0,64 | 0,31 | -0,63 | 0,14 | 1,15 | 2,67 | 1,29 | 3,03 |
| 6 | М-С-23-1-3 | 119 | 0,45 | 0,14 | 0,62 | -0,38 | 0,14 | 0,78 | -4,38 | 0,70 | -3,71 |
| 7 | М-С-05-1-1 | 186 | 0,71 | 0,24 | 0,44 | -1,72 | 0,15 | 0,94 | -0,80 | 0,98 | -0,12 |
| 8 | М-С-05-1-2 | 70 | 0,27 | 0,65 | 0,48 | 0,65 | 0,16 | 0,93 | -0,92 | 0,85 | -1,05 |
| 9 | М-С-05-1-3 | 11 | 0,04 | 0,57 | 0,33 | 3,13 | 0,32 | 0,95 | -0,10 | 0,38 | -1,52 |
| 10 | М-М-02-1-1 | 186 | 0,71 | 0,47 | 0,27 | -1,72 | 0,15 | 1,18 | 2,48 | 1,21 | 1,48 |
| 11 | М-М-02-1-2 | 70 | 0,27 | 0,71 | 0,36 | 0,65 | 0,16 | 1,03 | 0,46 | 1,27 | 1,73 |
| 12 | M-М-02-1-3 | 16 | 0,06 | 0,08 | 0,39 | 2,69 | 0,27 | 0,90 | -0,45 | 0,41 | -1,81 |
| 13 | М-М-03-1-1 | 194 | 0,74 | 0,62 | 0,39 | -1,90 | 0,15 | 1,00 | 0,08 | 0,96 | -0,20 |
| 14 | М-М-03-1-2 | 88 | 0,34 | 0,64 | 0,51 | 0,25 | 0,15 | 0,90 | -1,59 | 0,90 | -0,83 |
| 15 | М-М-03-1-3 | 23 | 0,09 | 0,29 | 0,48 | 2,25 | 0,23 | 0,78 | -1,34 | 0,46 | -1,97 |
| 16 | М-М-11-1-1 | 170 | 0,65 | 0,41 | 0,52 | -1,37 | 0,14 | 0,88 | -2,11 | 0,85 | -1,31 |
| 17 | М-М-11-1-2 | 94 | 0,36 | 0,61 | 0,44 | 0,12 | 0,14 | 0,98 | -0,25 | 1,02 | 0,22 |
| 18 | М-М-11-1-3 | 55 | 0,21 | 0,05 | 0,50 | 1,04 | 0,17 | 0,88 | -1,34 | 0,77 | -1,36 |
| 19 | М-М-08-1-1 | 179 | 0,68 | 0,41 | 0,34 | -1,56 | 0,15 | 1,09 | 1,36 | 1,21 | 1,62 |
| 20 | М-М-08-1-2 | 112 | 0,43 | 0,74 | 0,60 | -0,24 | 0,14 | 0,80 | -3,80 | 0,73 | -3,10 |
| 21 | М-М-08-1-3 | 53 | 0,20 | 0,22 | 0,47 | 1,10 | 0,17 | 0,94 | -0,63 | 0,74 | -1,51 |
| 22 | М-R-02-1-1 | 198 | 0,76 | 0,69 | 0,26 | -2,00 | 0,16 | 1,10 | 1,24 | 1,62 | 3,28 |
| 23 | М-R-02-1-2 | 146 | 0,56 | 0,64 | 0,33 | -0,90 | 0,14 | 1,13 | 2,43 | 1,16 | 1,67 |
| 24 | М-R-02-1-3 | 63 | 0,24 | 0,39 | 0,35 | 0,83 | 0,16 | 1,08 | 0,91 | 1,10 | 0,64 |
| 25 | М-R-05-1-1 | 232 | 0,89 | 0,37 | 0,43 | -3,07 | 0,21 | 0,87 | -0,87 | 0,58 | -1,63 |
| 26 | М-R-05-1-2 | 106 | 0,40 | 0,38 | 0,58 | -0,12 | 0,14 | 0,83 | -3,05 | 0,76 | -2,59 |
| 27 | М-R-05-1-3 | 86 | 0,33 | 0,16 | 0,48 | 0,29 | 0,15 | 0,96 | -0,65 | 0,86 | -1,20 |
| 28 | М-D-03-1-1 | 191 | 0,73 | 0,27 | 0,48 | -1,83 | 0,15 | 0,89 | -1,59 | 0,91 | -0,59 |
| 29 | М-D-03-1-2 | 133 | 0,51 | 0,76 | 0,50 | -0,65 | 0,14 | 0,92 | -1,55 | 0,90 | -1,12 |
| 30 | М-D-03-1-3 | 18 | 0,07 | 0,6 | 0,38 | 2,55 | 0,26 | 0,83 | -0,88 | 1,08 | 0,34 |
| 31 | М-D-20-1-1 | 126 | 0,48 | 0,62 | 0,45 | -0,51 | 0,14 | 1,00 | 0,01 | 0,97 | -0,31 |
| 32 | М-D-20-1-2 | 70 | 0,27 | 0,38 | 0,48 | 0,65 | 0,16 | 0,94 | -0,81 | 0,84 | -1,09 |
| 33 | М-D-20-1-3 | 19 | 0,07 | 0,12 | 0,27 | 2,49 | 0,25 | 0,97 | -0,10 | 1,63 | 1,58 |
| 34 | М-D-05-1-1 | 159 | 0,61 | 0,58 | 0,44 | -1,15 | 0,14 | 0,98 | -0,39 | 0,98 | -0,13 |
| 35 | М-D-05-1-2 | 98 | 0,37 | 0,53 | 0,23 | 0,04 | 0,14 | 1,26 | 3,94 | 1,36 | 3,12 |
| 36 | М-D-05-1-3 | 23 | 0,09 | 0,35 | 0,45 | 2,25 | 0,23 | 0,82 | -1,09 | 0,71 | -0,90 |
| 37 | М-D-08-1-1 | 211 | 0,81 | 0,18 | 0,39 | -2,35 | 0,17 | 0,97 | -0,26 | 0,89 | -0,50 |
| 38 | М-D-08-1-2 | 99 | 0,38 | 0,49 | 0,35 | 0,02 | 0,14 | 1,11 | 1,79 | 1,14 | 1,33 |
| 39 | М-D-08-1-3 | 13 | 0,05 | 0,22 | 0,31 | 2,94 | 0,30 | 0,96 | -0,10 | 0,81 | -0,31 |
| 40 | М-G-01-1-1 | 183 | 0,70 | 0,24 | 0,25 | -1,65 | 0,15 | 1,19 | 2,70 | 1,28 | 1,98 |
| 41 | М-G-01-1-2 | 46 | 0,18 | 0,29 | 0,19 | 1,31 | 0,18 | 1,22 | 1,93 | 1,64 | 2,72 |
| 42 | М-G-01-1-3 | 23 | 0,09 | 0,15 | 0,37 | 2,25 | 0,23 | 0,95 | -0,25 | 0,69 | -0,96 |
| 43 | М-G-03-1-1 | 176 | 0,67 | 0,31 | 0,41 | -1,50 | 0,15 | 1,00 | 0,07 | 1,02 | 0,22 |
| 44 | М-G-03-1-2 | 44 | 0,17 | 0,15 | 0,36 | 1,37 | 0,18 | 1,03 | 0,28 | 1,05 | 0,32 |
| 45 | М-G-03-1-3 | 29 | 0,11 | 0,18 | 0,02 | 1,95 | 0,21 | 1,35 | 2,22 | 2,00 | 2,87 |

На основании результатов анализа функционирования заданий в рамках Классической и Современной теориях тестирования были определены задания, требующие дополнительной проверки на качество функционирования и перевода.

В таблицах ниже приведены характеристические кривые наиболее проблемных заданий, представляющие собой вероятность правильного ответа на задание в зависимости от уровня подготовленности учащихся (красная линия). Крестиками на рисунке обозначены точки эмпирического распределения ответов учащихся на конкретное задание. Они представляют собой средний балл по данному заданию по группам учащихся (вся выборка была поделена на 10 частей в зависимости от балла по тесту). Также на рисунке показаны границы 95%-ного доверительного интервала для точек эмпирического распределения (серые линии). К каждому заданию дан более подробный комментарий по поводу его функционирования.

Комментарии о функционировании заданий, вариант А1

В качестве примера хорошо функционирующего задания приводится характеристическая кривая задания M-C-01-1-2. Точки эмпирического распределения максимально приближены к вероятностной кривой, что означает, что задание хорошо дифференцирует учащихся с разным уровнем подготовленности.



*Рисунок 2.* Характеристическая кривая задания №2 (М-С-01-1-2)

Карта переменных

На Рисунке 3 представлена карта переменных, на которой показано распределение испытуемых и заданий относительно друг друга на общей метрической шкале. Слева на рисунке находится шкала логитов. На карте испытуемые представлены слева, задания – справа. Более трудные задания и более подготовленные испытуемые находятся в верхней части карты, более легкие задания и менее подготовленные испытуемые находятся в нижней части карты. Карта позволяет проанализировать совместное распределение заданий относительно данного контингента испытуемых и диагностировать проблемы теста. Как можно заметить, распределение мер не является нормальным – распределение по испытуемым скошено вниз, т.е. данный тест оказался достаточно трудным и большинство учащихся находится в пределах уровня подготовленности от нуля до минус трех. Меры заданий распределены таким образом, что полностью покрывают как верхнюю часть, так и нижнюю часть карты, что говорит о достаточном количестве заданий разной трудности.



*Рисунок 3. Карта переменных (вариант А1)*

Вариант А1 продемонстрировал хорошие показатели в общей согласованности результатов. Почти все задачи обладают оптимальными показатели трудности и дискриминатиности. Следует уделить внимание заданию М-G-03-1-3, так как статистики согласия данного задания выходят за пределы оптимальных.

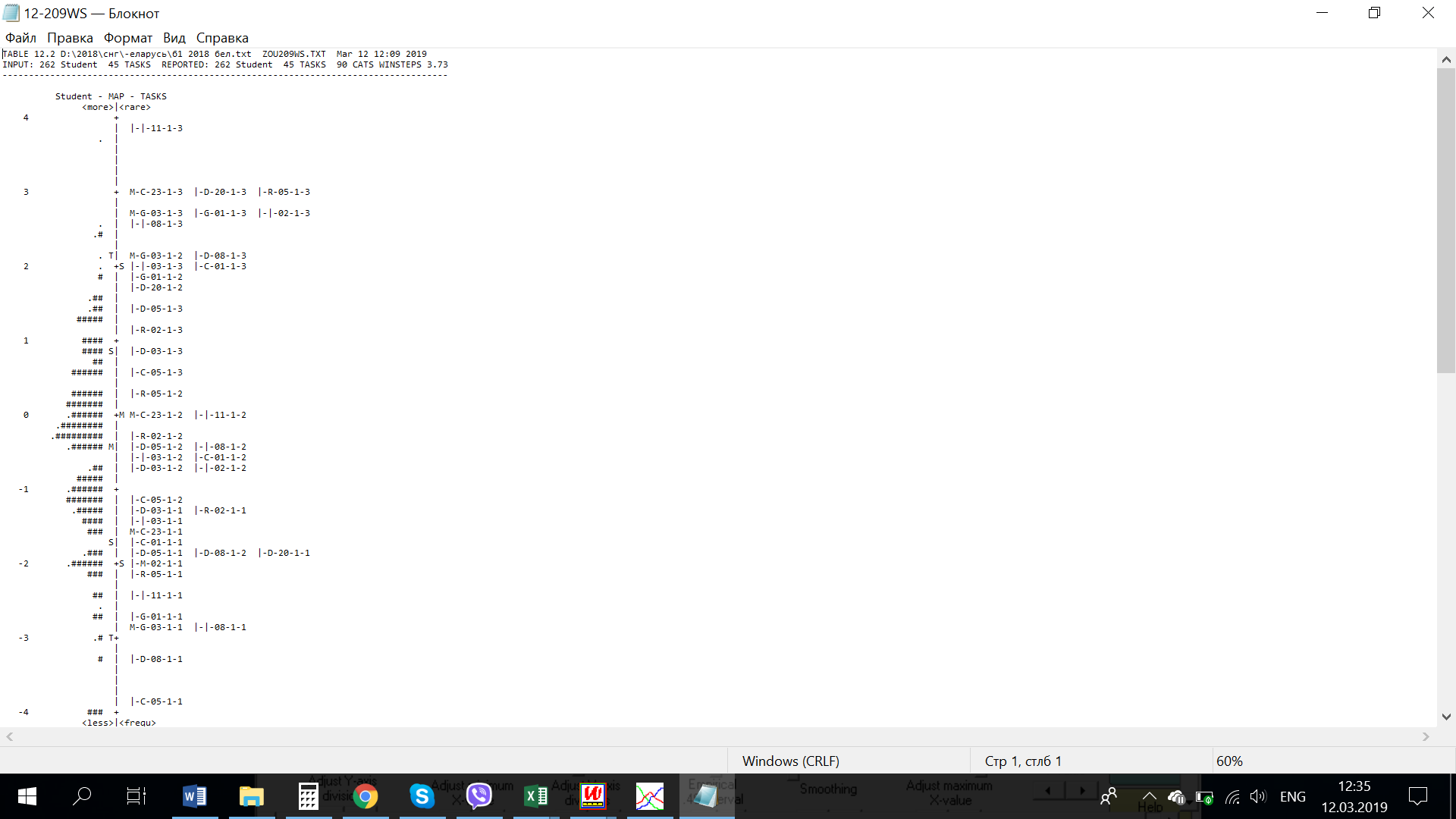
Анализ варианта Б1

Таблица 2.2. Статистические показатели функционирования заданий, вариант Б1

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | Количество выполнивших задание | КТТ | | | IRT | | | | | |
| Трудность | Дискр. | Скор. ТБК | Трудность | Ошибка измерения | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | М-С-03-1-1 | 191 | 0,73 | 0,18 | 0,33 | -1,77 | 0,16 | 1,19 | 2,29 | 1,40 | 2,20 |
| 2 | М-С-03-1-2 | 137 | 0,52 | 0,54 | 0,52 | -0,58 | 0,14 | 0,94 | -1,07 | 0,91 | -0,85 |
| 3 | М-С-03-1-3 | 33 | 0,13 | 0,43 | 0,32 | 1,99 | 0,2 | 1,05 | 0,39 | 0,93 | -0,15 |
| 4 | М-С-22-1-1 | 185 | 0,71 | 0,42 | 0,48 | -1,62 | 0,16 | 0,96 | -0,47 | 1,03 | 0,22 |
| 5 | М-С-22-1-2 | 111 | 0,42 | 0,38 | 0,51 | -0,05 | 0,14 | 0,93 | -1,14 | 0,87 | -1,19 |
| 6 | М-С-22-1-3 | 15 | 0,06 | 0,76 | 0,27 | 2,99 | 0,28 | 0,95 | -0,16 | 1,01 | 0,16 |
| 7 | М-С-25-1-1 | 244 | 0,93 | 0,45 | 0,44 | -3,85 | 0,28 | 0,93 | -0,26 | 0,52 | -1,16 |
| 8 | М-С-25-1-2 | 166 | 0,63 | 0,48 | 0,54 | -1,19 | 0,15 | 0,91 | -1,40 | 0,82 | -1,48 |
| 9 | М-С-25-1-3 | 82 | 0,31 | 0,15 | 0,48 | 0,57 | 0,15 | 0,94 | -0,88 | 0,87 | -0,91 |
| 10 | М-М-20-1-1 | 201 | 0,77 | 0,3 | 0,47 | -2,03 | 0,17 | 0,97 | -0,27 | 0,94 | -0,23 |
| 11 | М-М-20-1-2 | 146 | 0,56 | 0,4 | 0,55 | -0,76 | 0,14 | 0,89 | -1,81 | 0,80 | -2,05 |
| 12 | М-М-20-1-3 | 18 | 0,07 | 0,22 | 0,13 | 2,77 | 0,26 | 1,16 | 0,87 | 2,49 | 2,70 |
| 13 | М-М-23-1-1 | 176 | 0,67 | 0,43 | 0,51 | -1,41 | 0,15 | 0,94 | -0,81 | 0,88 | -0,88 |
| 14 | М-М-23-1-2 | 134 | 0,51 | 0,59 | 0,50 | -0,52 | 0,14 | 0,96 | -0,63 | 0,94 | -0,56 |
| 15 | М-М-23-1-3 | 33 | 0,13 | 0,27 | 0,38 | 1,99 | 0,2 | 0,90 | -0,76 | 0,96 | -0,03 |
| 16 | М-М-06-1-1 | 214 | 0,82 | 0,64 | 0,50 | -2,43 | 0,18 | 0,89 | -1,03 | 0,84 | -0,65 |
| 17 | М-М-06-1-2 | 106 | 0,40 | 0,56 | 0,49 | 0,05 | 0,14 | 0,98 | -0,34 | 0,89 | -0,98 |
| 18 | М-М-06-1-3 | 7 | 0,03 | 0,44 | 0,15 | 3,86 | 0,4 | 0,99 | 0,07 | 1,87 | 1,35 |
| 19 | М-M-24-1-1 | 224 | 0,86 | 0,42 | 0,50 | -2,79 | 0,2 | 0,91 | -0,73 | 0,69 | -1,16 |
| 20 | М-M-24-1-2 | 129 | 0,49 | 0,77 | 0,56 | -0,42 | 0,14 | 0,87 | -2,31 | 0,79 | -2,17 |
| 21 | М-M-24-1-3 | 20 | 0,08 | 0,49 | 0,34 | 2,64 | 0,25 | 0,93 | -0,33 | 0,68 | -0,80 |
| 22 | М-R-03-1-1 | 168 | 0,64 | 0,2 | 0,54 | -1,23 | 0,15 | 0,91 | -1,35 | 0,82 | -1,48 |
| 23 | М-R-03-1-2 | 123 | 0,47 | 0,36 | 0,46 | -0,3 | 0,14 | 1,03 | 0,46 | 1,02 | 0,26 |
| 24 | М-R-03-1-3 | 61 | 0,23 | 0,34 | 0,39 | 1,09 | 0,16 | 0,99 | -0,04 | 1,21 | 1,07 |
| 25 | М-R-22-1-1 | 203 | 0,77 | 0,31 | 0,34 | -2,09 | 0,17 | 1,18 | 1,88 | 1,29 | 1,42 |
| 26 | М-R-22-1-2 | 93 | 0,36 | 0,7 | 0,29 | 0,33 | 0,15 | 1,21 | 3,10 | 1,71 | 4,65 |
| 27 | М-R-22-1-3 | 15 | 0,06 | 0,5 | 0,23 | 2,99 | 0,28 | 1,09 | 0,48 | 0,76 | -0,44 |
| 28 | М-D-25-1-1 | 173 | 0,66 | 0,5 | 0,31 | -1,34 | 0,15 | 1,24 | 3,28 | 1,41 | 2,80 |
| 29 | М-D-25-1-2 | 142 | 0,54 | 0,58 | 0,55 | -0,68 | 0,14 | 0,88 | -2,05 | 0,84 | -1,60 |
| 30 | М-D-25-1-3 | 72 | 0,27 | 0,16 | 0,47 | 0,8 | 0,16 | 0,92 | -1,05 | 0,86 | -0,82 |
| 31 | М-D-21-1-1 | 193 | 0,74 | 0,54 | 0,55 | -1,82 | 0,16 | 0,86 | -1,78 | 0,80 | -1,18 |
| 32 | М-D-21-1-2 | 41 | 0,16 | 0,53 | 0,44 | 1,69 | 0,19 | 0,88 | -1,09 | 0,67 | -1,31 |
| 33 | М-D-21-1-3 | 14 | 0,05 | 0,14 | 0,23 | 3,07 | 0,29 | 0,99 | 0,04 | 0,97 | 0,09 |
| 34 | М-D-22-1-1 | 195 | 0,74 | 0,51 | 0,55 | -1,87 | 0,16 | 0,86 | -1,70 | 0,81 | -1,11 |
| 35 | М-D-22-1-2 | 132 | 0,50 | 0,22 | 0,40 | -0,48 | 0,14 | 1,12 | 1,97 | 1,24 | 2,19 |
| 36 | М-D-22-1-3 | 50 | 0,19 | 0,23 | 0,38 | 1,4 | 0,17 | 0,97 | -0,33 | 1,16 | 0,73 |
| 37 | М-D-26-1-1 | 236 | 0,90 | 0,35 | 0,35 | -3,34 | 0,23 | 1,04 | 0,32 | 1,16 | 0,55 |
| 38 | М-D-03-1-2 | 192 | 0,73 | 0,42 | 0,45 | -1,79 | 0,16 | 1,05 | 0,61 | 0,97 | -0,12 |
| 39 | М-D-26-1-3 | 30 | 0,11 | 0,14 | 0,29 | 2,12 | 0,21 | 1,03 | 0,29 | 1,24 | 0,80 |
| 40 | М-G-21-1-1 | 221 | 0,84 | 0,28 | 0,34 | -2,67 | 0,19 | 1,14 | 1,14 | 1,26 | 1,02 |
| 41 | М-G-21-1-2 | 37 | 0,14 | 0,15 | 0,29 | 1,84 | 0,19 | 1,07 | 0,60 | 1,42 | 1,39 |
| 42 | М-G-21-1-3 | 18 | 0,07 | 0,24 | 0,25 | 2,77 | 0,26 | 0,99 | 0,00 | 1,08 | 0,33 |
| 43 | М-G-24-1-1 | 225 | 0,86 | 0,42 | 0,41 | -2,83 | 0,2 | 1,04 | 0,36 | 0,91 | -0,23 |
| 44 | М-G-24-1-2 | 30 | 0,11 | 0,32 | 0,22 | 2,12 | 0,21 | 1,14 | 0,98 | 1,69 | 1,93 |
| 45 | М-G-24-1-3 | 18 | 0,07 | -0,01 | 0,27 | 2,77 | 0,26 | 1,00 | 0,09 | 0,80 | -0,37 |

Карта переменных

Распределение мер варианта Б1 близко к нормальному (Рисунок 4). Размах трудности заданий в пределах от минус четырех до четырех логитов. Все задания условно сгруппированы в зависимости от их уровня по шкале, но часть заданий второго уровня попали в верхнюю часть карты, что говорит об относительно высокой трудности нескольких заданий второго уровня. Для данной группы учащихся четыре задания третьего уровня слишком трудны, так как их уровень трудности находится выше уровня подготовленности самых подготовленных учащихся.



*Рисунок 4. Карта переменных (вариант Б1)*

Выводы:

Вариант Б1 продемонстрировал хорошие показатели надежности, а также большинство задач обладает высокой дискриминативностью и оптимальной трудностью в соответствии с теоретической моделью инструмента. Вариант Б1 оказался немного труднее по своим характеристикам, чем вариант А1. Более детально в Варианте Б1 было рассмотрено три задания теста. Наиболее проблемными являются задачи третьего уровня М-М-06-1-3 и М-М-20-1-3 – с этими заданиями справилось слишком малое количество учащихся и как итог – по данным заданиям невозможно получить достаточной информации об их функционировании.

# ПРИЛОЖЕНИЕ 7

АНАЛИЗ РЕЗУЛЬТАТОВ ТЕСТА ПО МАТЕМАТИКЕ (ВАРИАНТЫ А1 И Б1)

(КИРГИЗСКАЯ РЕСПУБЛИКА, ВЕСНА 2018)

Сводные результаты по двум вариантам

Тест состоял из двух вариантов по 45 заданий, объединенных в 15 блоков. В таблице 1 представлены общие статистические показатели.

Оба варианта теста имеют высокую внутреннюю согласованность, коэффициент альфа Кронбаха равен 0,85 и 0,83.

Таблица 1. Общие результаты теста по математике

|  |  |  |
| --- | --- | --- |
|  | Вариант А1 | Вариант Б1 |
| Число учащихся | 209 | 216 |
| Средний балл | 14 | 11 |
| Максимальный балл | 29 | 31 |
| Минимальный балл | 0 | 0 |
| Стандартное отклонение | 5,95 | 5,92 |
| Коэффициент асимметрии | 0,32 | 0,28 |
| Коэффициент эксцесса | -0,53 | -0,07 |
| Средний коэффициент решаемости | 0,31 | 0,25 |
| Средний индекс дискриминативности | 0,32 | 0,3 |
| Средний скор. коэф. точ.-бис. корреляции | 0,28 | 0,27 |
| Показатель надежности (KR20) | 0,85 | 0,83 |
| Стандартная ошибка измерения | 2,34 | 2,45 |

Вариант А1 Вариант Б1

|  |  |
| --- | --- |
|  |  |

*Рисунок 1. Гистограммы распределения первичных баллов участников (варианты А1 и Б1)*

Выводы:

* Средние коэффициенты решаемости по обоим вариантам чуть ниже оптимальных значений, 0,31 и 0,25.
* Средний индекс дискриминативности по обоим вариантам обладает оптимальным значением 0,32 и 0,3.
* Распределение первичных баллов во втором варианте скошено влево, что позволяет сделать вывод о том, что вариант Б1 оказался труднее для испытуемых.
* По результатам анализа теста на данной выборке надежность тестов высокая. Коэффициент альфа Кронбаха равен 0,85 и 0,83.

Результаты по отдельным вариантам

Большинство заданий хорошо функционируют и демонстрируют удовлетворительные статистические показатели. Часть заданий имеют повышенный уровень трудности, который обуславливает их низкий уровень дифференцирующей способности. В обоих вариантах наблюдаются блоки с нарушением иерархии трудности. Более детальная информация по функционированию заданий представлена ниже, в таблицах 2.1 и 2.2.

Анализ заданий проводился в рамках Классической теории тестирования и Современной теории тестирования. Анализ заданий в рамках Классической теории тестирования проводился с использованием показателей: трудность, дискриминативность и скорректированная точечно-бисериальная корреляция. В таблице ниже цветом отмечены задания с показателями, выходящими за критические значения (менее 0,2 для показателей дифференцирующей способности заданий (дискриминативности); для точечно-бисериальной корреляции менее 0,2). Голубым цветом выделены задания с неупорядоченностью по трудности внутри блока, зеленым – задания экстремальной трудности или легкости (выполнили менее 10% или более 90% участников тестирования).

В столбцах 7 и 8 таблицы представлены оценки трудности заданий и соответствующие ошибки измерения (в логитах), оцененные в рамках Современной теории тестирования с использованием дихотомической модели Раша. В последних четырех столбцах (9, 10, 11, 12) приведены значения статистик согласия, характеризующих согласие данных с используемой моделью измерения. В данном отчете используются следующие статистики согласия – простая статистика согласия (OUTFIT MNSQ), стандартизированная простая статистика согласия (OUTFIT ZSTD), взвешенная статистика согласия (INFIT MNSQ), стандартизированная взвешенная статистика согласия (INFIT ZSTD). Простая статистика согласия более чувствительна к экстремально неожиданным ответам, когда сильный испытуемый неожиданно неправильно отвечает на легкое задание, или наоборот, слабый испытуемый неожиданно правильно отвечает на трудное задание. Взвешенная версия статистики согласия позволяет уменьшить влияние экстремально неожиданных ответов. Таким образом, эти две статистики относятся к различным частям распределения уровней подготовленности участников тестирования. Продуктивным для измерений называется интервал значений статистик согласия (0,8; 1,2) для простых статистик согласия и (-2; 2) – для стандартизированных статистик согласия. Показатели, выходящие за пределы критических значений, отмечены серым цветом.

Из анализа табл. 2.1 следует, что у ряда заданий значения простой и взвешенных статистик согласия превышают правые критические значения, что означает, что есть испытуемые, давшие неожиданные ответы на эти задания.

Таблица 2.1 Статистические показатели функционирования заданий, вариант А1

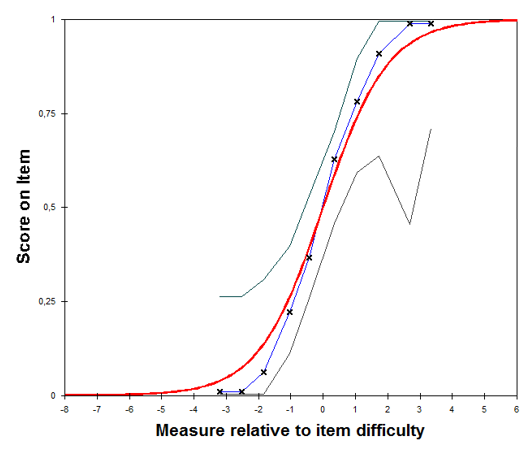
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | Количество выполнивших задание | КТТ | | | IRT | | | | | |
| Трудность | Дискр. | Скор. ТБК | Трудность | Ошибка измерения | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | М-С-01-1-1 | 183 | 0,88 | 0,17 | 0,23 | -3,86 | 0,23 | 0,97 | -0,10 | 1,30 | 1,00 |
| 2 | М-С-01-1-2 | 47 | 0,22 | 0,45 | 0,35 | 0,12 | 0,18 | 1,01 | 0,10 | 0,90 | -0,50 |
| 3 | М-С-0-1-3 | 0 | 0 | 0 | 0 | 5,73 | 1,83 |  |  |  |  |
| 4 | М-С-23-1-1 | 140 | 0,67 | 0,49 | 0,39 | -2,34 | 0,17 | 0,96 | -0,50 | 0,93 | -0,50 |
| 5 | M-C-23-1-2 | 54 | 0,26 | 0,53 | 0,45 | -0,10 | 0,18 | 0,91 | -1,00 | 0,76 | -1,50 |
| 6 | М-С-23-1-3 | 3 | 0,01 | 0 | 0,01 | 3,40 | 0,59 | 1,05 | 0,30 | 1,26 | 0,60 |
| 7 | М-С-05-1-1 | 167 | 0,8 | 0,44 | 0,39 | -3,17 | 0,19 | 0,94 | -0,60 | 0,74 | -1,30 |
| 8 | М-С-05-1-2 | 91 | 0,44 | 0,56 | 0,43 | -1,10 | 0,16 | 0,96 | -0,50 | 0,94 | -0,50 |
| 9 | М-С-05-1-3 | 38 | 0,18 | 0,48 | 0,49 | 0,44 | 0,20 | 0,84 | -1,50 | 0,61 | -1,90 |
| 10 | М-М-02-1-1 | 153 | 0,73 | 0,49 | 0,38 | -2,71 | 0,17 | 0,97 | -0,30 | 0,93 | -0,40 |
| 11 | М-М-02-1-2 | 53 | 0,25 | 0,47 | 0,44 | -0,07 | 0,18 | 0,90 | -1,10 | 0,89 | -0,60 |
| 12 | M-М-02-1-3 | 9 | 0,04 | 0,03 | 0,02 | 2,23 | 0,35 | 1,11 | 0,50 | 1,57 | 1,10 |
| 13 | М-М-03-1-1 | 109 | 0,52 | 0,48 | 0,34 | -1,55 | 0,16 | 1,07 | 1,10 | 1,07 | 0,80 |
| 14 | М-М-03-1-2 | 41 | 0,2 | 0,45 | 0,43 | 0,33 | 0,19 | 0,89 | -1,00 | 0,85 | -0,70 |
| 15 | М-М-03-1-3 | 1 | 0 | 0,02 | 0,15 | 4,52 | 1,01 | 0,96 | 0,30 | 0,16 | -1,40 |
| 16 | М-М-11-1-1 | 184 | 0,88 | 0,27 | 0,26 | -3,92 | 0,23 | 0,99 | 0,00 | 1,68 | 1,90 |
| 17 | М-М-11-1-2 | 70 | 0,33 | 0,62 | 0,44 | -0,56 | 0,16 | 0,95 | -0,70 | 0,89 | -0,80 |
| 18 | М-М-11-1-3 | 9 | 0,04 | 0,08 | 0,15 | 2,23 | 0,35 | 1,03 | 0,20 | 0,99 | 0,20 |
| 19 | М-М-08-1-1 | 116 | 0,56 | 0,58 | 0,43 | -1,72 | 0,16 | 0,97 | -0,50 | 0,88 | -1,30 |
| 20 | М-М-08-1-2 | 45 | 0,22 | 0,48 | 0,4 | 0,19 | 0,19 | 0,95 | -0,50 | 0,81 | -1,00 |
| 21 | М-М-08-1-3 | 10 | 0,05 | 0,08 | 0,16 | 2,11 | 0,33 | 1,00 | 0,10 | 1,27 | 0,70 |
| 22 | М-R-02-1-1 | 30 | 0,14 | 0,37 | 0,37 | 0,77 | 0,21 | 0,94 | -0,40 | 0,79 | -0,70 |
| 23 | М-R-02-1-2 | 16 | 0,08 | 0,25 | 0,37 | 1,57 | 0,27 | 0,91 | -0,40 | 0,45 | -1,60 |
| 24 | М-R-02-1-3 | 37 | 0,18 | 0,28 | 0,24 | 0,48 | 0,20 | 1,07 | 0,70 | 1,86 | 3,00 |
| 25 | М-R-05-1-1 | 129 | 0,62 | 0,33 | 0,18 | -2,05 | 0,16 | 1,21 | 3,00 | 1,38 | 3,10 |
| 26 | М-R-05-1-2 | 37 | 0,18 | 0,18 | 0,21 | 0,48 | 0,20 | 1,08 | 0,70 | 1,48 | 1,90 |
| 27 | М-R-05-1-3 | 0 | 0 | 0 | 0 | 5,73 | 1,83 |  |  |  |  |
| 28 | М-D-03-1-1 | 158 | 0,76 | 0,39 | 0,31 | -2,87 | 0,18 | 1,03 | 0,40 | 1,07 | 0,40 |
| 29 | М-D-03-1-2 | 76 | 0,36 | 0,51 | 0,37 | -0,72 | 0,16 | 1,01 | 0,20 | 1,02 | 0,20 |
| 30 | М-D-03-1-3 | 9 | 0,04 | 0,02 | 0,06 | 2,23 | 0,35 | 1,05 | 0,30 | 1,98 | 1,60 |
| 31 | М-D-20-1-1 | 86 | 0,41 | 0,61 | 0,41 | -0,98 | 0,16 | 0,98 | -0,30 | 0,96 | -0,40 |
| 32 | М-D-20-1-2 | 14 | 0,07 | 0,18 | 0,32 | 1,73 | 0,29 | 0,91 | -0,30 | 0,56 | -1,10 |
| 33 | М-D-20-1-3 | 3 | 0,01 | 0,05 | 0,14 | 3,40 | 0,59 | 1,00 | 0,20 | 0,45 | -0,60 |
| 34 | М-D-05-1-1 | 116 | 0,56 | 0,83 | 0,57 | -1,72 | 0,16 | 0,80 | -3,30 | 0,71 | -3,30 |
| 35 | М-D-05-1-2 | 134 | 0,64 | 0,49 | 0,32 | -2,18 | 0,16 | 1,09 | 1,20 | 1,16 | 1,40 |
| 36 | М-D-05-1-3 | 27 | 0,13 | 0,28 | 0,34 | 0,92 | 0,22 | 0,93 | -0,50 | 0,92 | -0,20 |
| 37 | М-D-08-1-1 | 179 | 0,86 | 0,36 | 0,34 | -3,67 | 0,22 | 0,93 | -0,50 | 0,76 | -0,90 |
| 38 | М-D-08-1-2 | 97 | 0,46 | 0,71 | 0,5 | -1,25 | 0,16 | 0,89 | -1,80 | 0,84 | -1,70 |
| 39 | М-D-08-1-3 | 6 | 0,03 | 0,02 | -0,03 | 2,67 | 0,42 | 1,10 | 0,40 | 2,18 | 1,60 |
| 40 | М-G-01-1-1 | 33 | 0,16 | 0,35 | 0,39 | 0,64 | 0,21 | 0,92 | -0,60 | 0,78 | -0,90 |
| 41 | М-G-01-1-2 | 17 | 0,08 | 0,12 | 0,19 | 1,50 | 0,27 | 0,99 | 0,00 | 1,26 | 0,70 |
| 42 | М-G-01-1-3 | 1 | 0 | 0 | 0,01 | 4,52 | 1,01 | 1,01 | 0,30 | 1,14 | 0,40 |
| 43 | М-G-03-1-1 | 88 | 0,42 | 0,45 | 0,26 | -1,03 | 0,16 | 1,16 | 2,40 | 1,23 | 2,10 |
| 44 | М-G-03-1-2 | 31 | 0,15 | 0,23 | 0,21 | 0,73 | 0,21 | 1,07 | 0,60 | 1,89 | 2,70 |
| 45 | М-G-03-1-3 | 41 | 0,2 | 0,2 | 0,1 | 0,33 | 0,19 | 1,26 | 2,40 | 1,44 | 1,90 |

Таблица 2.2 Статистические показатели функционирования заданий, вариант Б1

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | Количество выполнивших задание | КТТ | | | IRT | | | | | |
| Трудность | Дискр. | Скор. ТБК | Трудность | Ошибка измерения | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | М-С-03-1-1 | 144 | 0,67 | 0,38 | 0,24 | -2,65 | 0,16 | 1,16 | 2,00 | 1,44 | 3,40 |
| 2 | М-С-03-1-2 | 60 | 0,28 | 0,34 | 0,24 | -0,60 | 0,17 | 1,09 | 1,20 | 0,95 | -0,30 |
| 3 | М-С-03-1-3 | 23 | 0,11 | 0,08 | 0,05 | 0,73 | 0,23 | 1,15 | 0,90 | 1,67 | 1,90 |
| 4 | М-С-22-1-1 | 64 | 0,3 | 0,47 | 0,41 | -0,71 | 0,16 | 0,92 | -1,10 | 0,84 | -1,20 |
| 5 | М-С-22-1-2 | 63 | 0,29 | 0,46 | 0,33 | -0,68 | 0,16 | 0,99 | -0,10 | 1,31 | 2,20 |
| 6 | М-С-22-1-3 | 38 | 0,18 | 0,33 | 0,38 | 0,08 | 0,19 | 0,90 | -0,90 | 0,95 | -0,20 |
| 7 | М-С-25-1-1 | 131 | 0,61 | 0,61 | 0,46 | -2,32 | 0,16 | 0,88 | -1,80 | 0,89 | -1,20 |
| 8 | М-С-25-1-2 | 49 | 0,23 | 0,39 | 0,35 | -0,29 | 0,18 | 0,96 | -0,50 | 0,87 | -0,80 |
| 9 | М-С-25-1-3 | 0 | 0 | 0 | 0 | 5,31 | 1,83 |  |  |  |  |
| 10 | М-М-20-1-1 | 105 | 0,49 | 0,34 | 0,23 | -1,70 | 0,15 | 1,13 | 2,30 | 1,15 | 1,80 |
| 11 | М-М-20-1-2 | 46 | 0,21 | 0,12 | 0,09 | -0,19 | 0,18 | 1,19 | 2,00 | 1,31 | 1,70 |
| 12 | М-М-20-1-3 | 1 | 0 | 0,02 | 0,22 | 4,10 | 1,01 | 0,92 | 0,20 | 0,09 | -1,60 |
| 13 | М-М-23-1-1 | 108 | 0,5 | 0,38 | 0,19 | -1,77 | 0,15 | 1,20 | 3,30 | 1,33 | 3,60 |
| 14 | М-М-23-1-2 | 20 | 0,09 | 0,07 | 0,05 | 0,90 | 0,25 | 1,15 | 0,80 | 1,64 | 1,70 |
| 15 | М-М-23-1-3 | 11 | 0,05 | 0,08 | 0,11 | 1,59 | 0,32 | 1,05 | 0,30 | 1,50 | 1,10 |
| 16 | М-М-06-1-1 | 112 | 0,52 | 0,65 | 0,43 | -1,86 | 0,15 | 0,93 | -1,30 | 0,90 | -1,20 |
| 17 | М-М-06-1-2 | 103 | 0,48 | 0,52 | 0,33 | -1,65 | 0,15 | 1,03 | 0,50 | 1,02 | 0,20 |
| 18 | М-М-06-1-3 | 30 | 0,14 | 0,38 | 0,44 | 0,39 | 0,21 | 0,85 | -1,20 | 0,61 | -1,70 |
| 19 | М-M-24-1-1 | 161 | 0,75 | 0,48 | 0,38 | -3,14 | 0,18 | 1,03 | 0,30 | 0,89 | -0,70 |
| 20 | М-M-24-1-2 | 92 | 0,43 | 0,58 | 0,45 | -1,40 | 0,15 | 0,91 | -1,70 | 0,83 | -1,90 |
| 21 | М-M-24-1-3 | 22 | 0,1 | 0,29 | 0,44 | 0,78 | 0,24 | 0,82 | -1,10 | 0,53 | -1,70 |
| 22 | М-R-03-1-1 | 76 | 0,35 | 0,44 | 0,34 | -1,02 | 0,16 | 1,00 | 0,10 | 1,06 | 0,60 |
| 23 | М-R-03-1-2 | 90 | 0,42 | 0,56 | 0,43 | -1,35 | 0,15 | 0,93 | -1,30 | 0,84 | -1,80 |
| 24 | М-R-03-1-3 | 16 | 0,07 | 0,18 | 0,28 | 1,16 | 0,27 | 0,96 | -0,10 | 0,68 | -0,80 |
| 25 | М-R-22-1-1 | 127 | 0,59 | 0,49 | 0,37 | -2,22 | 0,16 | 1,01 | 0,20 | 1,00 | 0,00 |
| 26 | М-R-22-1-2 | 29 | 0,13 | 0,35 | 0,41 | 0,44 | 0,21 | 0,87 | -0,90 | 0,69 | -1,30 |
| 27 | М-R-22-1-3 | 32 | 0,15 | 0,27 | 0,3 | 0,31 | 0,20 | 0,98 | -0,10 | 0,89 | -0,40 |
| 28 | М-D-25-1-1 | 93 | 0,43 | 0,64 | 0,49 | -1,42 | 0,15 | 0,87 | -2,50 | 0,80 | -2,30 |
| 29 | М-D-25-1-2 | 55 | 0,25 | 0,45 | 0,4 | -0,46 | 0,17 | 0,92 | -0,90 | 0,78 | -1,50 |
| 30 | М-D-25-1-3 | 2 | 0,01 | 0,02 | 0,15 | 3,40 | 0,72 | 0,94 | 0,10 | 0,71 | -0,10 |
| 31 | М-D-21-1-1 | 32 | 0,15 | 0,3 | 0,33 | 0,31 | 0,20 | 0,96 | -0,30 | 0,75 | -1,10 |
| 32 | М-D-21-1-2 | 12 | 0,06 | 0,13 | 0,27 | 1,49 | 0,31 | 0,91 | -0,30 | 0,89 | -0,10 |
| 33 | М-D-21-1-3 | 4 | 0,02 | 0,03 | 0,12 | 2,68 | 0,51 | 0,96 | 0,10 | 1,19 | 0,50 |
| 34 | М-D-22-1-1 | 74 | 0,34 | 0,61 | 0,47 | -0,97 | 0,16 | 0,88 | -2,00 | 0,77 | -2,10 |
| 35 | М-D-22-1-2 | 53 | 0,25 | 0,14 | 0,1 | -0,41 | 0,17 | 1,19 | 2,20 | 1,38 | 2,20 |
| 36 | М-D-22-1-3 | 5 | 0,02 | 0,02 | 0,03 | 2,44 | 0,46 | 1,06 | 0,30 | 1,68 | 1,10 |
| 37 | М-D-26-1-1 | 112 | 0,52 | 0,49 | 0,38 | -1,86 | 0,15 | 0,99 | -0,20 | 0,95 | -0,50 |
| 38 | М-D-03-1-2 | 42 | 0,19 | 0,25 | 0,24 | -0,06 | 0,18 | 1,05 | 0,60 | 0,98 | 0,00 |
| 39 | М-D-26-1-3 | 1 | 0 | -0,01 | -0,05 | 4,10 | 1,01 | 1,02 | 0,40 | 2,67 | 1,60 |
| 40 | М-G-21-1-1 | 99 | 0,46 | 0,45 | 0,34 | -1,56 | 0,15 | 1,02 | 0,50 | 0,98 | -0,20 |
| 41 | М-G-21-1-2 | 23 | 0,11 | 0,18 | 0,23 | 0,73 | 0,23 | 0,98 | 0,00 | 1,24 | 0,80 |
| 42 | М-G-21-1-3 | 6 | 0,03 | -0,02 | -0,08 | 2,25 | 0,42 | 1,09 | 0,40 | 3,00 | 2,40 |
| 43 | М-G-24-1-1 | 61 | 0,28 | 0,5 | 0,39 | -0,63 | 0,16 | 0,94 | -0,80 | 0,91 | -0,60 |
| 44 | М-G-24-1-2 | 6 | 0,03 | 0,02 | 0,07 | 2,25 | 0,42 | 1,03 | 0,20 | 1,31 | 0,70 |
| 45 | М-G-24-1-3 | 22 | 0,1 | 0,04 | 0,06 | 0,78 | 0,24 | 1,11 | 0,70 | 1,98 | 2,60 |

На основании результатов анализа функционирования заданий в рамках Классической и Современной теориях тестирования были определены задания, требующие дополнительной проверки на качество функционирования и перевода.

В качестве примера хорошо функционирующего задания приводится характеристическая кривая задания М-С-01-1-2. Точки эмпирического распределения максимально приближены к вероятностной кривой, что означает, что задание хорошо дифференцирует учащихся с разным уровнем подготовленности.



*Рисунок 2.* Характеристическая кривая задания №2 (М-С-01-1-2)

Характеристическая кривая задания представляет собой вероятность правильного ответа на задание в зависимости от уровня подготовленности учащихся (красная линия). Крестиками на рисунке обозначены точки эмпирического распределения ответов учащихся на конкретное задание. Они представляют собой средний балл по данному заданию по группам учащихся (вся выборка была поделена на 10 частей в зависимости от балла по тесту). Также на рисунке показан границы 95%-ного доверительного интервала для точек эмпирического распределения (серые линии).

ПРИЛОЖЕНИЕ 8

АНАЛИЗ РЕЗУЛЬТАТОВ ПОЛНОМАСШТАБНОЙ АПРОБАЦИИ ТЕСТА ПО МАТЕМАТИКЕ (ВАРИАНТЫ А1 И Б1)

(РЕСПУБЛИКА ТАДЖИКИСТАН, ВЕСНА 2018)

Сводные результаты по двум вариантам

Тест состоял из двух вариантов по 45 заданий, объединенных в 15 блоков. В таблице 1 представлены общие статистические показатели.

Оба варианта теста имеют достаточно высокую внутреннюю согласованность, коэффициент альфа Кронбаха равен 0,79 и 0,78.

Таблица 1. Общие результаты теста по математике

|  |  |  |
| --- | --- | --- |
|  | Вариант А1 | Вариант Б1 |
| Число учащихся | 203 | 198 |
| Средний балл | 9 | 10 |
| Максимальный балл | 26 | 27 |
| Минимальный балл | 1 | 1 |
| Стандартное отклонение | 4,6 | 4,9 |
| Коэффициент асимметрии | 0,84 | 0,56 |
| Коэффициент эксцесса | 0,57 | 0,37 |
| Средний коэффициент решаемости | 0,2 | 0,23 |
| Средний индекс дискриминативности | 0,23 | 0,26 |
| Средний скор. коэф. точ.-бис. корреляции | 0,22 | 0,22 |
| Показатель надежности (KR20) | 0,79 | 0,78 |
| Стандартная ошибка измерения | 2,16 | 2,35 |

Вариант А1 Вариант Б1

|  |  |
| --- | --- |
|  |  |

*Рисунок 1. Гистограммы распределения первичных баллов участников (варианты А1 и Б1)*

Выводы:

* Средние коэффициенты решаемости по обоим вариантам ниже оптимальных значений, 0,2 и 0,23.
* Средний индекс дискриминативности по обоим вариантам ниже оптимальных значений, 0,23 и 0,26.
* Распределение первичных баллов скошено влево, что позволяет сделать вывод о том, что варианты оказались достаточно трудными для испытуемых.
* По результатам анализа теста на данной выборке надежность тестов достаточно высокая. Коэффициент альфа Кронбаха равен 0,79 и 0,78.

Результаты по отдельным вариантам

Большинство заданий хорошо функционируют и демонстрируют удовлетворительные статистические показатели. Часть заданий имеют повышенный уровень трудности, который обуславливает их низкий уровень дифференцирующей способности. В обоих вариантах наблюдаются блоки с нарушением иерархии трудности. Более детальная информация по функционированию заданий представлена ниже, в таблицах 2.1 и 2.2.

Анализ заданий проводился в рамках Классической теории тестирования и Современной теории тестирования. Анализ заданий в рамках Классической теории тестирования проводился с использованием показателей: трудность, дискриминативность и скорректированная точечно-бисериальная корреляция. В таблице ниже цветом отмечены задания с показателями, выходящими за критические значения (менее 0,2 для показателей дифференцирующей способности заданий (дискриминативности); для точечно-бисериальной корреляции менее 0,2). Голубым цветом выделены задания с неупорядоченностью по трудности внутри блока, зеленым – задания экстремальной трудности или легкости (выполнили менее 10% или более 90% участников тестирования).

В столбцах 7 и 8 таблицы представлены оценки трудности заданий и соответствующие ошибки измерения (в логитах), оцененные в рамках Современной теории тестирования с использованием дихотомической модели Раша. В последних четырех столбцах (9, 10, 11, 12) приведены значения статистик согласия, характеризующих согласие данных с используемой моделью измерения. В данном отчете используются следующие статистики согласия – простая статистика согласия (*OUTFIT MNSQ*), стандартизированная простая статистика согласия (*OUTFIT ZSTD*), взвешенная статистика согласия (*INFIT MNSQ*), стандартизированная взвешенная статистика согласия (*INFIT ZSTD*). Простая статистика согласия более чувствительна к экстремально неожиданным ответам, когда сильный испытуемый неожиданно неправильно отвечает на легкое задание, или наоборот, слабый испытуемый неожиданно правильно отвечает на трудное задание. Взвешенная версия статистики согласия позволяет уменьшить влияние экстремально неожиданных ответов. Таким образом, эти две статистики относятся к различным частям распределения уровней подготовленности участников тестирования. Продуктивным для измерений называется интервал значений статистик согласия (0,8; 1,2) для простых статистик согласия и (-2; 2) – для стандартизированных статистик согласия. Показатели, выходящие за пределы критических значений, отмечены серым цветом.

Из анализа табл. 2.1 следует, что у ряда заданий значения простой и взвешенных статистик согласия превышают правые критические значения, что означает, что есть испытуемые, давшие неожиданные ответы на эти задания.

Таблица 2.1 Статистические показатели функционирования заданий, вариант А1

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | Количество выполнивших задание | КТТ | | | IRT | | | | | |
| Трудность | Дискр. | Скор. ТБК | Трудность | Ошибка измерения | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | М-С-01-1-1 | 172 | 0,85 | 0,15 | 0,15 | -4,06 | 0,21 | 1,06 | 0,50 | 1,23 | 1,00 |
| 2 | М-С-01-1-2 | 30 | 0,15 | 0,34 | 0,36 | 0,10 | 0,21 | 0,94 | -0,40 | 0,77 | -0,90 |
| 3 | М-С-0-1-3 | 0 | 0 | 0 | 0 | 5,08 | 1,83 |  |  |  |  |
| 4 | М-С-23-1-1 | 116 | 0,57 | 0,27 | 0,16 | -2,37 | 0,16 | 1,17 | 2,70 | 1,20 | 2,00 |
| 5 | M-C-23-1-2 | 17 | 0,08 | 0,18 | 0,25 | 0,82 | 0,27 | 0,99 | 0,00 | 1,10 | 0,40 |
| 6 | М-С-23-1-3 | 7 | 0,03 | -0,06 | -0,17 | 1,84 | 0,39 | 1,16 | 0,60 | 3,81 | 3,10 |
| 7 | М-С-05-1-1 | 82 | 0,4 | 0,63 | 0,42 | -1,53 | 0,16 | 0,92 | -1,30 | 0,89 | -1,10 |
| 8 | М-С-05-1-2 | 39 | 0,19 | 0,34 | 0,34 | -0,27 | 0,19 | 0,98 | -0,20 | 0,86 | -0,70 |
| 9 | М-С-05-1-3 | 21 | 0,1 | 0,25 | 0,36 | 0,56 | 0,24 | 0,90 | -0,50 | 0,78 | -0,60 |
| 10 | М-М-02-1-1 | 106 | 0,52 | 0,64 | 0,41 | -2,12 | 0,16 | 0,90 | -1,70 | 0,89 | -1,20 |
| 11 | М-М-02-1-2 | 38 | 0,19 | 0,34 | 0,31 | -0,23 | 0,19 | 0,99 | -0,10 | 0,94 | -0,20 |
| 12 | M-М-02-1-3 | 3 | 0,01 | 0,02 | 0,02 | 2,74 | 0,59 | 1,05 | 0,30 | 1,76 | 1,10 |
| 13 | М-М-03-1-1 | 82 | 0,4 | 0,35 | 0,19 | -1,53 | 0,16 | 1,15 | 2,30 | 1,26 | 2,50 |
| 14 | М-М-03-1-2 | 14 | 0,07 | 0,09 | 0,08 | 1,05 | 0,29 | 1,11 | 0,60 | 1,47 | 1,20 |
| 15 | М-М-03-1-3 | 0 | 0 | 0 | 0 | 5,08 | 1,83 |  |  |  |  |
| 16 | М-М-11-1-1 | 143 | 0,7 | 0,52 | 0,35 | -3,07 | 0,17 | 0,95 | -0,60 | 0,86 | -1,00 |
| 17 | М-М-11-1-2 | 24 | 0,12 | 0,24 | 0,36 | 0,39 | 0,23 | 0,91 | -0,50 | 0,79 | -0,70 |
| 18 | М-М-11-1-3 | 8 | 0,04 | 0,1 | 0,22 | 1,69 | 0,37 | 0,99 | 0,10 | 0,70 | -0,40 |
| 19 | М-М-08-1-1 | 76 | 0,37 | 0,57 | 0,35 | -1,38 | 0,16 | 0,98 | -0,30 | 1,03 | 0,30 |
| 20 | М-М-08-1-2 | 12 | 0,06 | 0,17 | 0,29 | 1,23 | 0,31 | 0,96 | -0,10 | 0,58 | -1,00 |
| 21 | М-М-08-1-3 | 0 | 0 | 0 | 0 | 5,08 | 1,83 |  |  |  |  |
| 22 | М-R-02-1-1 | 13 | 0,06 | 0,17 | 0,36 | 1,14 | 0,30 | 0,90 | -0,40 | 0,53 | -1,20 |
| 23 | М-R-02-1-2 | 8 | 0,04 | 0,13 | 0,37 | 1,69 | 0,37 | 0,87 | -0,30 | 0,33 | -1,60 |
| 24 | М-R-02-1-3 | 10 | 0,05 | -0,02 | -0,06 | 1,44 | 0,34 | 1,17 | 0,70 | 2,11 | 1,90 |
| 25 | М-R-05-1-1 | 60 | 0,3 | 0,33 | 0,22 | -0,95 | 0,17 | 1,11 | 1,40 | 1,13 | 1,00 |
| 26 | М-R-05-1-2 | 10 | 0,05 | 0,1 | 0,22 | 1,44 | 0,34 | 0,95 | -0,10 | 0,92 | 0,00 |
| 27 | М-R-05-1-3 | 0 | 0 | 0 | 0 | 5,08 | 1,83 |  |  |  |  |
| 28 | М-D-03-1-1 | 79 | 0,39 | 0,43 | 0,35 | -1,46 | 0,16 | 1,00 | 0,00 | 0,98 | -0,10 |
| 29 | М-D-03-1-2 | 18 | 0,09 | 0,2 | 0,31 | 0,75 | 0,26 | 0,95 | -0,20 | 0,84 | -0,30 |
| 30 | М-D-03-1-3 | 9 | 0,04 | 0,08 | 0,24 | 1,56 | 0,35 | 0,93 | -0,10 | 0,77 | -0,30 |
| 31 | М-D-20-1-1 | 80 | 0,39 | 0,65 | 0,5 | -1,48 | 0,16 | 0,84 | -2,60 | 0,78 | -2,30 |
| 32 | М-D-20-1-2 | 6 | 0,03 | 0,1 | 0,32 | 2,00 | 0,42 | 0,90 | -0,10 | 0,32 | -1,30 |
| 33 | М-D-20-1-3 | 0 | 0 | 0 | 0 | 5,08 | 1,83 |  |  |  |  |
| 34 | М-D-05-1-1 | 51 | 0,25 | 0,49 | 0,47 | -0,68 | 0,18 | 0,87 | -1,50 | 0,76 | -1,60 |
| 35 | М-D-05-1-2 | 37 | 0,18 | 0,13 | 0,1 | -0,19 | 0,20 | 1,18 | 1,50 | 1,49 | 2,10 |
| 36 | М-D-05-1-3 | 17 | 0,08 | 0,03 | 0 | 0,82 | 0,27 | 1,18 | 1,00 | 1,88 | 2,00 |
| 37 | М-D-08-1-1 | 127 | 0,63 | 0,59 | 0,42 | -2,64 | 0,16 | 0,89 | -1,80 | 0,83 | -1,70 |
| 38 | М-D-08-1-2 | 76 | 0,37 | 0,7 | 0,48 | -1,38 | 0,16 | 0,85 | -2,30 | 0,80 | -1,90 |
| 39 | М-D-08-1-3 | 5 | 0,02 | 0,02 | 0,04 | 2,20 | 0,46 | 1,06 | 0,30 | 1,86 | 1,30 |
| 40 | М-G-01-1-1 | 80 | 0,39 | 0,59 | 0,42 | -1,48 | 0,16 | 0,92 | -1,30 | 0,83 | -1,70 |
| 41 | М-G-01-1-2 | 6 | 0,03 | 0,07 | 0,22 | 2,00 | 0,42 | 0,95 | 0,00 | 0,94 | 0,10 |
| 42 | М-G-01-1-3 | 0 | 0 | 0 | 0 | 5,08 | 1,83 |  |  |  |  |
| 43 | М-G-03-1-1 | 104 | 0,51 | 0,31 | 0,17 | -2,07 | 0,16 | 1,15 | 2,60 | 1,19 | 2,00 |
| 44 | М-G-03-1-2 | 26 | 0,13 | 0,18 | 0,18 | 0,29 | 0,22 | 1,09 | 0,70 | 1,04 | 0,20 |
| 45 | М-G-03-1-3 | 2 | 0,01 | 0,02 | -0,01 | 3,16 | 0,72 | 1,04 | 0,30 | 1,41 | 0,70 |

Выводы:

* С шестью заданиями в варианте А1 не справился ни один испытуемый, в связи с чем отсутствуют статистические показатели функционирования этих заданий. Все шесть заданий являются заданиями третьего уровня, в связи с чем от испытуемых не ожидается, что они справятся с этими заданиями.
* Все остальные задания варианта А1 можно признать функционирующими удовлетворительно.

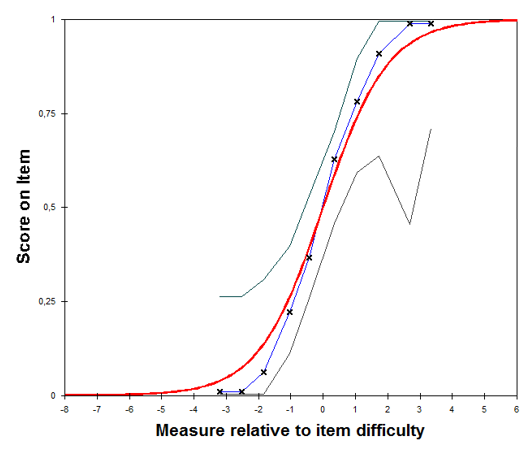
Таблица 2.2 Статистические показатели функционирования заданий, вариант Б1

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | Количество выполнивших задание | КТТ | | | IRT | | | | | |
| Трудность | Дискриминативность | Скор. ТБК | Трудность | Ошибка измерения | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | М-С-03-1-1 | 139 | 0,7 | 0,31 | 0,2 | -3,05 | 0,17 | 1,08 | 1 | 1,18 | 1,4 |
| 2 | М-С-03-1-2 | 81 | 0,41 | 0,52 | 0,33 | -1,59 | 0,16 | 0,98 | -0,4 | 0,99 | 0 |
| 3 | М-С-03-1-3 | 12 | 0,06 | 0,12 | 0,19 | 1,09 | 0,31 | 0,99 | 0,1 | 0,82 | -0,3 |
| 4 | М-С-22-1-1 | 7 | 0,04 | 0,02 | -0,03 | 1,69 | 0,39 | 1,09 | 0,4 | 5,28 | 4,3 |
| 5 | М-С-22-1-2 | 63 | 0,32 | 0,4 | 0,23 | -1,12 | 0,17 | 1,06 | 0,9 | 1,04 | 0,4 |
| 6 | М-С-22-1-3 | 1 | 0,01 | 0,02 | 0,12 | 3,71 | 1,01 | 0,99 | 0,3 | 0,31 | -0,7 |
| 7 | М-С-25-1-1 | 152 | 0,77 | 0,43 | 0,29 | -3,46 | 0,18 | 0,99 | -0,1 | 0,93 | -0,4 |
| 8 | М-С-25-1-2 | 96 | 0,48 | 0,59 | 0,4 | -1,95 | 0,16 | 0,93 | -1,4 | 0,92 | -1 |
| 9 | М-С-25-1-3 | 0 | 0 | 0 | 0 | 4,92 | 1,83 |  |  |  |  |
| 10 | М-М-20-1-1 | 79 | 0,4 | 0,39 | 0,25 | -1,54 | 0,16 | 1,05 | 0,8 | 1,03 | 0,4 |
| 11 | М-М-20-1-2 | 31 | 0,16 | 0,34 | 0,23 | -0,06 | 0,21 | 1,04 | 0,4 | 0,85 | -0,6 |
| 12 | М-М-20-1-3 | 1 | 0,01 | 0,02 | 0,2 | 3,71 | 1,01 | 0,96 | 0,3 | 0,15 | -1,2 |
| 13 | М-М-23-1-1 | 90 | 0,45 | 0,39 | 0,24 | -1,81 | 0,16 | 1,08 | 1,5 | 1,09 | 1,1 |
| 14 | М-М-23-1-2 | 31 | 0,16 | 0,29 | 0,33 | -0,06 | 0,21 | 0,94 | -0,4 | 0,85 | -0,6 |
| 15 | М-М-23-1-3 | 2 | 0,01 | 0,02 | 0,11 | 3 | 0,72 | 0,98 | 0,2 | 0,96 | 0,2 |
| 16 | М-М-06-1-1 | 62 | 0,31 | 0,55 | 0,44 | -1,09 | 0,17 | 0,88 | -1,7 | 0,76 | -2,1 |
| 17 | М-М-06-1-2 | 67 | 0,34 | 0,5 | 0,39 | -1,23 | 0,16 | 0,93 | -1,1 | 0,87 | -1,2 |
| 18 | М-М-06-1-3 | 17 | 0,09 | 0,21 | 0,24 | 0,69 | 0,26 | 1 | 0,1 | 0,76 | -0,6 |
| 19 | М-M-24-1-1 | 145 | 0,73 | 0,45 | 0,32 | -3,23 | 0,18 | 0,95 | -0,5 | 0,91 | -0,6 |
| 20 | М-M-24-1-2 | 44 | 0,22 | 0,4 | 0,38 | -0,55 | 0,18 | 0,92 | -0,8 | 0,86 | -0,8 |
| 21 | М-M-24-1-3 | 18 | 0,09 | 0,21 | 0,34 | 0,62 | 0,26 | 0,9 | -0,5 | 0,72 | -0,8 |
| 22 | М-R-03-1-1 | 88 | 0,44 | 0,39 | 0,25 | -1,77 | 0,16 | 1,05 | 0,9 | 1,05 | 0,6 |
| 23 | М-R-03-1-2 | 60 | 0,3 | 0,45 | 0,32 | -1,04 | 0,17 | 0,98 | -0,3 | 0,98 | -0,1 |
| 24 | М-R-03-1-3 | 11 | 0,06 | 0,05 | 0,07 | 1,19 | 0,32 | 1,07 | 0,4 | 1,33 | 0,8 |
| 25 | М-R-22-1-1 | 102 | 0,52 | 0,46 | 0,25 | -2,1 | 0,16 | 1,04 | 0,8 | 1,03 | 0,4 |
| 26 | М-R-22-1-2 | 13 | 0,07 | 0,14 | 0,26 | 1 | 0,3 | 0,95 | -0,2 | 0,7 | -0,7 |
| 27 | М-R-22-1-3 | 8 | 0,04 | 0,12 | 0,33 | 1,54 | 0,37 | 0,91 | -0,2 | 0,43 | -1,2 |
| 28 | М-D-25-1-1 | 88 | 0,44 | 0,58 | 0,43 | -1,77 | 0,16 | 0,9 | -1,9 | 0,84 | -1,9 |
| 29 | М-D-25-1-2 | 39 | 0,2 | 0,31 | 0,36 | -0,38 | 0,19 | 0,91 | -0,8 | 0,93 | -0,3 |
| 30 | М-D-25-1-3 | 3 | 0,02 | 0 | 0,08 | 2,58 | 0,59 | 0,97 | 0,1 | 1,57 | 0,9 |
| 31 | М-D-21-1-1 | 40 | 0,2 | 0,36 | 0,32 | -0,42 | 0,19 | 0,97 | -0,3 | 0,83 | -0,9 |
| 32 | М-D-21-1-2 | 23 | 0,12 | 0,17 | 0,17 | 0,32 | 0,23 | 1,06 | 0,4 | 0,99 | 0,1 |
| 33 | М-D-21-1-3 | 6 | 0,03 | -0,02 | -0,02 | 1,85 | 0,42 | 1,09 | 0,4 | 1,53 | 1 |
| 34 | М-D-22-1-1 | 51 | 0,26 | 0,26 | 0,23 | -0,77 | 0,17 | 1,05 | 0,6 | 1,44 | 2,6 |
| 35 | М-D-22-1-2 | 66 | 0,33 | 0,13 | 0,04 | -1,2 | 0,16 | 1,23 | 3,2 | 1,33 | 2,6 |
| 36 | М-D-22-1-3 | 2 | 0,01 | 0,03 | 0,25 | 3 | 0,72 | 0,94 | 0,1 | 0,19 | -1,2 |
| 37 | М-D-26-1-1 | 114 | 0,58 | 0,63 | 0,42 | -2,39 | 0,16 | 0,89 | -2 | 0,82 | -2,1 |
| 38 | М-D-03-1-2 | 42 | 0,21 | 0,38 | 0,31 | -0,49 | 0,19 | 0,97 | -0,3 | 0,85 | -0,8 |
| 39 | М-D-26-1-3 | 4 | 0,02 | 0,02 | -0,02 | 2,29 | 0,51 | 1,06 | 0,3 | 1,42 | 0,8 |
| 40 | М-G-21-1-1 | 66 | 0,33 | 0,18 | 0,05 | -1,2 | 0,16 | 1,22 | 3,1 | 1,5 | 3,8 |
| 41 | М-G-21-1-2 | 28 | 0,14 | 0,33 | 0,4 | 0,07 | 0,22 | 0,88 | -0,8 | 0,74 | -1 |
| 42 | М-G-21-1-3 | 6 | 0,03 | 0,02 | 0,03 | 1,86 | 0,42 | 1,06 | 0,3 | 1,7 | 1,2 |
| 43 | М-G-24-1-1 | 42 | 0,21 | 0,35 | 0,34 | -0,48 | 0,19 | 0,94 | -0,6 | 1,11 | 0,7 |
| 44 | М-G-24-1-2 | 2 | 0,01 | -0,02 | -0,09 | 3 | 0,72 | 1,04 | 0,3 | 3,75 | 2,3 |
| 45 | М-G-24-1-3 | 8 | 0,04 | 0,07 | 0,07 | 1,55 | 0,37 | 1,06 | 0,3 | 1,42 | 0,9 |

На основании результатов анализа функционирования заданий в рамках Классической и Современной теориях тестирования были определены задания, требующие дополнительной проверки на качество функционирования и перевода.

В качестве примера хорошо функционирующего задания приводится характеристическая кривая задания М-С-01-1-2. Точки эмпирического распределения максимально приближены к вероятностной кривой, что означает, что задание хорошо дифференцирует учащихся с разным уровнем подготовленности.

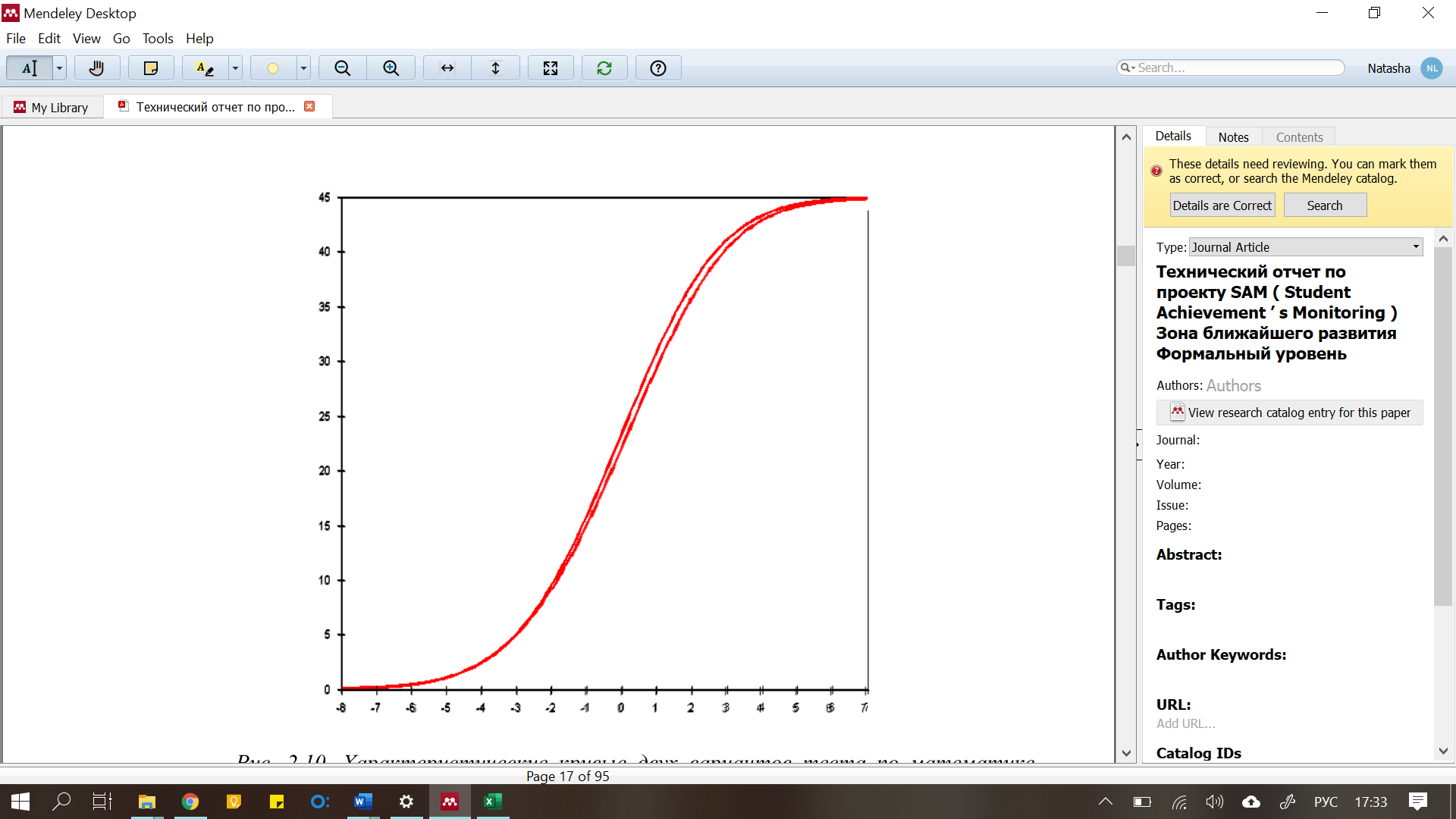
Характеристическая кривая представляет собой вероятность правильного ответа на задание в зависимости от уровня подготовленности учащихся (красная линия). Крестиками на рисунке обозначены точки эмпирического распределения ответов учащихся на конкретное задание. Они представляют собой средний балл по данному заданию по группам учащихся (вся выборка была поделена на 10 частей в зависимости от балла по тесту). Также на рисунке показаны границы 95%-ного доверительного интервала для точек эмпирического распределения (серые линии).



*Рисунок 2.* Характеристическая кривая задания №2 (М-С-01-1-2)

# ПРИЛОЖЕНИЕ 9

РЕЗУЛЬТАТЫ ВЫРАВНИВАНИЯ



*Рисунок 1. Характеристические кривые двух вариантов теста по математике после процедуры выравнивания*

На рисунке выше показаны в качестве примера характеристические кривые двух вариантов теста по математике. Характеристические кривые практически совпадают, следовательно, меры испытуемых, полученных в обоих вариантах, эквивалентны. В таком случае процедура выравнивания считается успешной.

Таблица 1. Сравнение показателей задач-клонов после выравнивания (вариант А1)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № задачи | А1 логит | А2 логит | Разница логит | А1 КТТ трудность | А2 КТТ трудность | Разница КТТ |
| 1 | -1,6 | -1,81 | 0,21 | 0,78 | 0,74 | 0,04 |
| 2 | 0,4 | 0,45 | -0,05 | 0,42 | 0,33 | 0,09 |
| 3 | 3,19 | 2,37 | 0,82 | 0,06 | 0,09 | -0,03 |
| 4 | -1,23 | -1,81 | 0,58 | 0,73 | 0,74 | -0,01 |
| 5 | 0,99 | 0,45 | 0,54 | 0,31 | 0,33 | -0,02 |
| 6 | 3,68 | 1,49 | 2,19 | 0,04 | 0,17 | -0,14 |
| 7 | -1,91 | -2,35 | 0,44 | 0,82 | 0,81 | 0,01 |
| 8 | 0,5 | -0,16 | 0,66 | 0,4 | 0,44 | -0,04 |
| 9 | 1,65 | 0,45 | 1,2 | 0,21 | 0,33 | -0,12 |
| 10 | -1,2 | -2,39 | 1,19 | 0,72 | 0,81 | -0,10 |
| 11 | -0,06 | 0,07 | -0,13 | 0,51 | 0,40 | 0,11 |
| 12 | 2,65 | 2,77 | -0,12 | 0,1 | 0,06 | 0,04 |
| 13 | -1,87 | -1,65 | -0,22 | 0,82 | 0,71 | 0,10 |
| 14 | -0,51 | -1,02 | 0,51 | 0,6 | 0,60 | -0,01 |
| 15 | 4,31 | 2,37 | 1,94 | 0,02 | 0,09 | -0,07 |
| 16 | -1,8 | -2,54 | 0,74 | 0,81 | 0,83 | -0,03 |
| 17 | 1,29 | 0,05 | 1,24 | 0,26 | 0,40 | -0,15 |
| 18 | 2,83 | 3,48 | -0,65 | 0,09 | 0,03 | 0,06 |
| 19 | -2,11 | -2,93 | 0,82 | 0,85 | 0,87 | -0,03 |
| 20 | 0,62 | -0,25 | 0,87 | 0,38 | 0,46 | -0,08 |
| 21 | 2,83 | 1,96 | 0,87 | 0,09 | 0,12 | -0,04 |
| 22 | -0,39 | -0,83 | 0,44 | 0,58 | 0,57 | 0,01 |
| 23 | 0,43 | 0,14 | 0,29 | 0,42 | 0,38 | 0,03 |
| 24 | 0,96 | 0,36 | 0,6 | 0,32 | 0,34 | -0,03 |
| 25 | -1,2 | -2,15 | 0,95 | 0,72 | 0,78 | -0,07 |
| 26 | 0,67 | 0,45 | 0,22 | 0,37 | 0,33 | 0,04 |
| 27 | 4,12 | 3,48 | 0,64 | 0,03 | 0,03 | 0,00 |
| 28 | -0,46 | -3,25 | 2,79 | 0,59 | 0,90 | -0,31 |
| 29 | 0,15 | -1,02 | 1,17 | 0,47 | 0,60 | -0,14 |
| 30 | 2,01 | 0,87 | 1,14 | 0,16 | 0,26 | -0,10 |
| 31 | -1,17 | -1,96 | 0,79 | 0,72 | 0,77 | -0,06 |
| 32 | 2,71 | 1,92 | 0,79 | 0,1 | 0,09 | 0,01 |
| 33 | 4,12 | 2,82 | 1,3 | 0,03 | 0,12 | -0,09 |
| 34 | -1,2 | -1,71 | 0,51 | 0,72 | 0,68 | 0,03 |
| 35 | -0,18 | -0,87 | 0,69 | 0,54 | 0,49 | 0,05 |
| 36 | 2,21 | 2,41 | -0,2 | 0,14 | 0,12 | 0,02 |
| 37 | -3,04 | -3,25 | 0,21 | 0,9 | 0,90 | 0,00 |
| 38 | -1,13 | -1,02 | -0,11 | 0,68 | 0,60 | 0,07 |
| 39 | 2,35 | 2,05 | 0,3 | 0,08 | 0,11 | -0,04 |
| 40 | -1,84 | -1,71 | -0,13 | 0,81 | 0,75 | 0,06 |
| 41 | 2,21 | 1,58 | 0,63 | 0,14 | 0,12 | 0,02 |
| 42 | 3,46 | 4,45 | -0,99 | 0,05 | 0,01 | 0,03 |
| 43 | -2,05 | -2,46 | 0,41 | 0,8 | 0,82 | -0,03 |
| 44 | -0,47 | -0,77 | 0,3 | 0,57 | 0,56 | 0,01 |
| 45 | 2,56 | 2,31 | 0,25 | 0,28 | 0,09 | 0,19 |
| ср знач | 0,61 | 0,02 | 0,59 | 0,43 | 0,44 | -0,02 |
| ст отклон | 2,02 | 2,04 | 0,69 | 0,29 | 0,29 | 0,08 |

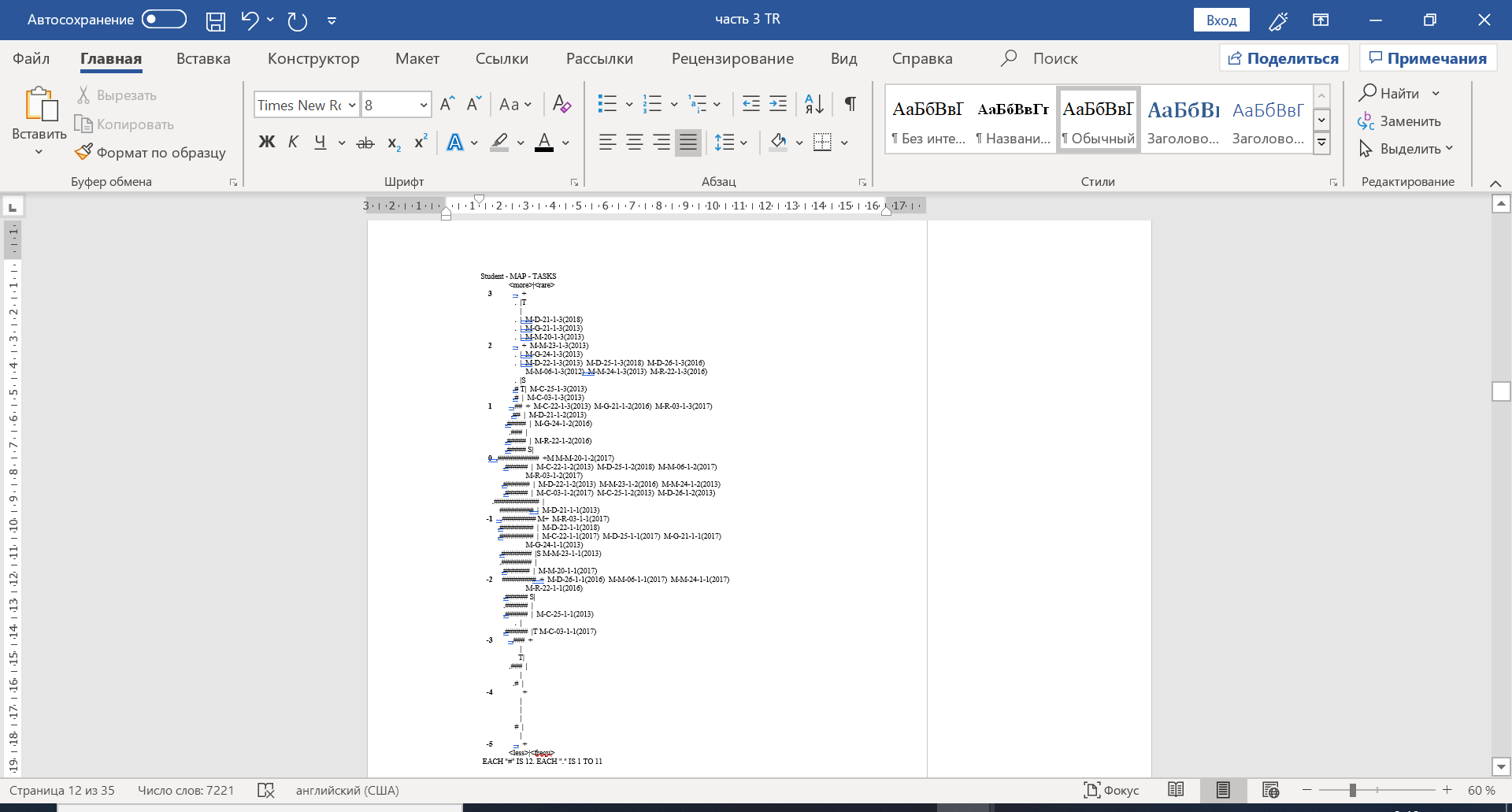
Таблица 2. Сравнение показателей задач-клонов после выравнивания (вариант B1)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № задачи | Б1 логит | Б2 логит | Разница логит | Б1 КТТ трудность | Б2 КТТ трудность | Разница КТТ |
| 1 | -4,15 | -3,76 | -0,39 | 0,85 | 0,93 | -0,08 |
| 2 | -2,24 | -2,12 | -0,12 | 0,64 | 0,77 | -0,13 |
| 3 | 0,32 | 0,81 | -0,49 | 0,24 | 0,26 | -0,02 |
| 4 | -0,85 | 0,07 | -0,92 | 0,42 | 0,38 | 0,04 |
| 5 | -1,54 | 0,52 | -2,06 | 0,55 | 0,3 | 0,25 |
| 6 | 0,77 | 1,06 | -0,29 | 0,17 | 0,22 | -0,05 |
| 7 | -2,7 | -1,94 | -0,76 | 0,74 | 0,74 | 0 |
| 8 | -0,17 | -0,03 | -0,14 | 0,31 | 0,4 | -0,09 |
| 9 | 4,92 | 2,59 | 2,33 | 0 | 0,07 | -0,07 |
| 10 | -2,28 | -1,41 | -0,87 | 0,68 | 0,72 | -0,04 |
| 11 | -0,09 | 0,7 | -0,79 | 0,29 | 0,3 | -0,01 |
| 12 | 2,26 | 2,73 | -0,47 | 0,05 | 0,11 | -0,06 |
| 13 | -2,19 | -1,75 | -0,44 | 0,67 | 0,6 | 0,07 |
| 14 | -0,68 | 0,19 | -0,87 | 0,39 | 0,37 | 0,02 |
| 15 | 1,77 | 2,35 | -0,58 | 0,08 | 0,04 | 0,04 |
| 16 | -1,72 | -1,05 | -0,67 | 0,58 | 0,59 | -0,01 |
| 17 | -1,17 | -0,18 | -0,99 | 0,48 | 0,43 | 0,05 |
| 18 | -0,23 | 0,6 | -0,83 | 0,32 | 0,29 | 0,03 |
| 19 | -1,96 | -1,48 | -0,48 | 0,67 | 0,67 | 0 |
| 20 | -0,56 | -0,42 | -0,14 | 0,43 | 0,47 | -0,04 |
| 21 | 0,91 | 1,16 | -0,25 | 0,14 | 0,21 | -0,07 |
| 22 | -2,76 | -2,03 | -0,73 | 0,75 | 0,75 | 0 |
| 23 | -1,42 | -0,42 | -1 | 0,53 | 0,47 | 0,06 |
| 24 | 0,47 | 0,57 | -0,1 | 0,21 | 0,3 | -0,09 |
| 25 | -2,86 | -2,24 | -0,62 | 0,77 | 0,74 | 0,03 |
| 26 | -0,88 | 0,35 | -1,23 | 0,43 | 0,34 | 0,09 |
| 27 | 0,21 | 1,63 | -1,42 | 0,25 | 0,2 | 0,05 |
| 28 | -2,09 | -1,25 | -0,84 | 0,65 | 0,63 | 0,02 |
| 29 | -1,22 | -0,1 | -1,12 | 0,49 | 0,41 | 0,08 |
| 30 | 1,7 | 2,44 | -0,74 | 0,09 | 0,08 | 0,01 |
| 31 | -0,8 | -0,59 | -0,21 | 0,42 | 0,5 | -0,08 |
| 32 | 0,84 | 0,66 | 0,18 | 0,16 | 0,28 | -0,12 |
| 33 | 2,73 | 1,74 | 0,99 | 0,04 | 0,14 | -0,1 |
| 34 | -3 | -1,54 | -1,46 | 0,79 | 0,68 | 0,11 |
| 35 | 0,31 | 1,19 | -0,88 | 0,23 | 0,2 | 0,03 |
| 36 | 2,87 | 2,59 | 0,28 | 0,03 | 0,07 | -0,04 |
| 37 | -2,73 | -2,12 | -0,61 | 0,75 | 0,77 | -0,02 |
| 38 | -0,98 | -0,37 | -0,61 | 0,45 | 0,46 | -0,01 |
| 39 | 2,59 | 2,67 | -0,08 | 0,04 | 0,07 | -0,03 |
| 40 | -1,98 | -0,75 | -1,23 | 0,63 | 0,54 | 0,09 |
| 41 | -2,25 | -1,25 | -1 | 0,68 | 0,63 | 0,05 |
| 42 | 0,84 | 1,12 | -0,28 | 0,16 | 0,21 | -0,05 |
| 43 | -2,09 | -1,56 | -0,53 | 0,65 | 0,68 | -0,03 |
| 44 | 0,87 | 0,6 | 0,27 | 0,16 | 0,29 | -0,13 |
| 45 | 1,99 | 1,37 | 0,62 | 0,07 | 0,18 | -0,11 |
| ср знач | -0,47 | 0,03 | -0,50 | 0,40 | 0,41 | -0,01 |
| ст отклон | 1,93 | 1,57 | 0,70 | 0,26 | 0,24 | 0,07 |

ПРИЛОЖЕНИЕ 10

КАРТА ЗАДАНИЙ

На карте переменных показано распределение испытуемых и заданий относительно друг друга на общей метрической шкале. Слева на рисунке 1 находится шкала логитов. На карте испытуемые представлены слева, задания – справа. Более трудные задания и более подготовленные испытуемые находятся в верхней части карты, более легкие задания и менее подготовленные испытуемые находятся в нижней части карты. Карта позволяет проанализировать совместное распределение заданий относительно данного контингента испытуемых и диагностировать проблемы теста.



*Рисунок 1. Карта заданий пилотного исследования*

ПРИЛОЖЕНИЕ 11

РЕЗУЛЬТАТЫ DIF АНАЛИЗА

Таблица 1. Результаты DIF анализа

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Номер** | **Задание** | **Армения** | **Беларусь** | **Киргизия** | **Россия** | **Таджикистан** |
| 1 | M-C-03-1-1(2017) | 0 | 0,42 | -0,48 | 0,84 | -0,64 |
| 2 | M-C-03-1-2(2017) | -0,63 | -0,45 | 0,44 | 0,09 | -0,31 |
| 3 | M-C-03-1-3(2013) | 0,45 | 0,04 | -1,48 | 0,92 | 0,64 |
| 4 | M-C-22-1-1(2017) | -0,4 | 0,18 | -0,04 | -0,19 | 0,39 |
| 5 | M-C-22-1-2(2013) | 0,19 | -0,07 | -0,23 | 0,22 | -0,16 |
| 6 | M-C-22-1-3(2013) | -0,18 | -0,19 | 0 | 0,14 | 1,17 |
| 7 | M-C-25-1-1(2013) | 0,16 | 0,18 | -0,07 | -0,53 | 0,14 |
| 8 | M-C-25-1-2(2013) | 0,11 | 0,24 | -0,21 | 0,37 | -0,62 |
| 9 | M-C-25-1-3(2013) | 0,19 | 0,04 | -0,34 | 0,04 | -0,09 |
| 10 | M-M-20-1-1(2017) | -0,08 | 0,07 | -0,21 | 0,21 | 0 |
| 11 | M-M-20-1-2(2017) | -0,32 | 0 | 0,13 | 0,32 | -0,14 |
| 12 | M-M-20-1-3(2013) | -0,38 | 0,59 | 0,31 | -0,53 | 0,89 |
| 13 | M-M-23-1-1(2013) | 0,31 | -0,03 | 0,36 | -0,59 | -0,12 |
| 14 | M-M-23-1-2(2016) | -0,05 | 0,26 | 0,57 | -0,36 | 0,62 |
| 15 | M-M-23-1-3(2013) | -0,92 | 0,16 | 0,4 | 0,57 | 1,49 |
| 16 | M-M-06-1-1(2017) | 0,31 | 0,36 | -0,34 | -0,4 | 0,04 |
| 17 | M-M-06-1-2(2017) | 0,1 | 0,29 | -0,91 | 0,7 | -0,39 |
| 18 | M-M-06-1-3(2012) | 0,08 | 0,34 | -0,59 | 0,12 | -0,53 |
| 19 | M-M-24-1-1(2017) | -0,45 | 0,04 | 0,02 | 0,54 | -0,23 |
| 20 | M-M-24-1-2(2013) | -0,07 | 0,04 | -0,2 | -0,06 | 0,42 |
| 21 | M-M-24-1-3(2013) | -0,12 | 0 | -0,4 | 0,53 | -0,27 |
| 22 | M-R-03-1-1(2017) | 0,13 | -0,49 | 0,97 | -0,68 | 0,29 |
| 23 | M-R-03-1-2(2017) | 0,12 | -0,31 | 0,06 | 0,19 | 0,08 |
| 24 | M-R-03-1-3(2017) | 0,44 | 0,06 | -0,37 | -0,29 | 0,36 |
| 25 | M-R-22-1-1(2016) | 0 | -0,38 | 0,06 | -0,32 | 0,5 |
| 26 | M-R-22-1-2(2016) | 0 | -0,24 | 0,06 | 0 | 0,76 |
| 27 | M-R-22-1-3(2016) | 0,25 | -0,33 | -0,41 | 0,43 | 0,68 |
| 28 | M-D-25-1-1(2017) | 0,58 | -0,08 | -0,48 | 0,21 | -0,18 |
| 29 | M-D-25-1-2(2018) | 0,11 | -0,36 | -0,29 | 0,43 | 0,31 |
| 30 | M-D-25-1-3(2018) | -0,08 | -0,15 | 0,29 | 0,1 | 0,28 |
| 31 | M-D-21-1-1(2013) | 0,39 | -0,09 | 0,67 | -0,84 | -0,07 |
| 32 | M-D-21-1-2(2013) | -0,42 | 0,24 | 1,01 | -0,3 | 0 |
| 33 | M-D-21-1-3(2018) | 0,36 | 0,19 | -0,3 | -0,35 | 0,08 |
| 34 | M-D-22-1-1(2018) | 0,21 | -0,24 | -0,13 | -0,14 | 0,41 |
| 35 | M-D-22-1-2(2013) | 0,06 | 0,22 | -0,76 | 0,66 | -0,36 |
| 36 | M-D-22-1-3(2013) | 0,09 | 0,08 | -0,45 | 0,21 | -0,27 |
| 37 | M-D-26-1-1(2016) | 0,5 | -0,65 | 1,02 | -0,59 | -0,51 |
| 38 | M-D-26-1-2(2013) | 0,14 | -0,17 | 0,17 | 0,25 | -0,39 |
| 39 | M-D-26-1-3(2016) | 0,05 | 0,79 | -0,87 | -0,38 | 0,47 |
| 40 | M-G-21-1-1(2017) | 0,38 | -0,65 | 0,66 | -0,45 | 0 |
| 41 | M-G-21-1-2(2016) | -1,15 | 0,58 | 0,19 | 0,69 | -0,25 |
| 42 | M-G-21-1-3(2013) | -0,31 | 0,13 | 0 | 0,07 | 0,27 |
| 43 | M-G-24-1-1(2013) | -0,04 | -0,49 | 0,69 | -0,27 | 0,1 |
| 44 | M-G-24-1-2(2016) | 0,33 | 1,05 | 0,46 | -0,49 | 0,29 |
| 45 | M-G-24-1-3(2013) | 1,34 | 0,55 | -1,29 | -0,46 | 0,33 |

# ПРИЛОЖЕНИЕ 12

АНАЛИЗ ЗАДАНИЙ, ВАРИАНТ А2

Задания теста проанализированы по следующим показателям: трудность задания, ошибка измерения, корреляция с общим баллом по тесту и общие статистики согласия.

В столбцах 2 и 3 таблиц представлены оценки трудности заданий и соответствующие ошибки измерения. В процессе анализа данных тестирования не было выявлено заданий, которые не были достигнуты участниками. Во всех блоках теста соблюдается иерархия уровней трудности заданий в соответствии с теоретической моделью.

В 4 столбце приведен коэффициент корреляции между баллами по заданию и уровнем подготовленности испытуемых. Большая часть заданий имеют приемлемые значения корреляции. Наименьшая связь баллов по заданию с уровнем подготовленности испытуемых наблюдается в заданиях повышенной трудности 3-го уровня (М-M-02-2-3, М-М-08-2-3, М-R-05-2-3, М-G-01-2-3).

В столбцах 5 и 6 приведены значения нестандартизированных статистик согласия, характеризующих согласие данных с используемой моделью измерения. Используются две версии статистики согласия – простая и взвешенная. В данном тесте большая часть заданий находится в хорошем согласии с моделью, о чем свидетельствуют статистики согласия, находящиеся в пределах нормы (0,8 – 1,2).

В тесте присутствует несколько заданий, у которых есть выход за пределы критических значений статистик. Статистика OUTFIT MNSQ является индикатором наличия выбросов в данных, она чувствительна к неожиданным ответам испытуемых, что является свидетельством отклонения от одномерности в данных, но не в измеряемом конструкте. Выход за пределы критических значений статистики OUTFIT MNSQ наблюдается только в заданиях повышенной трудности 3-го уровня.

В целом все задания теста можно признать функционирующими удовлетворительно.

Таблица 1. Статистические данные заданий теста, вариант А2.

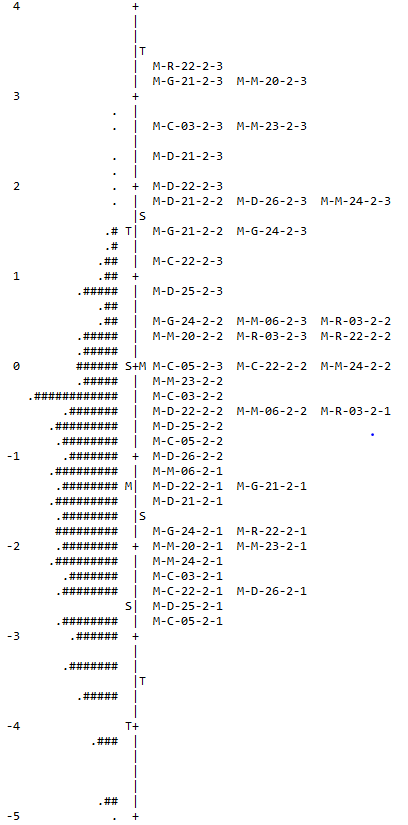
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | Трудность | Ошибка измерения | Корреляция | Статистики согласия | |
| INFIT MNSQ | OUTFIT MNSQ |
| 1 | М-С-01-2-1 | -2,36 | 0,05 | 0,35 | 1,21 | 1,42 |
| 2 | М-С-01-2-2 | -0,29 | 0,05 | 0,48 | 1,00 | 0,96 |
| 3 | М-С-01-2-3 | 2,7 | 0,11 | 0,3 | 0,9 | 0,6 |
| 4 | М-С-23-2-1 | -2,48 | 0,05 | 0,47 | 1,02 | 1,11 |
| 5 | М-С-23-2-2 | 0,08 | 0,05 | 0,49 | 0,95 | 0,88 |
| 6 | М-С-23-2-3 | 1,14 | 0,07 | 0,35 | 0,96 | 1,73 |
| 7 | М-С-05-2-1 | -2,89 | 0,05 | 0,49 | 0,97 | 0,93 |
| 8 | М-С-05-2-2 | -0,77 | 0,05 | 0,56 | 0,9 | 0,84 |
| 9 | М-С-05-2-3 | -0,07 | 0,05 | 0,4 | 1,09 | 1,27 |
| 10 | М-M-02-2-1 | -2,08 | 0,05 | 0,53 | 0,97 | 0,94 |
| 11 | М-M-02-2-2 | 0,25 | 0,05 | 0,44 | 0,99 | 1,04 |
| 12 | М-M-02-2-3 | 3,11 | 0,13 | 0,04 | 1,10 | 3,45 |
| 13 | М-М-03-2-1 | -1,94 | 0,05 | 0,46 | 1,08 | 1,12 |
| 14 | М-М-03-2-2 | -0,18 | 0,05 | 0,5 | 0,94 | 0,99 |
| 15 | М-М-03-2-3 | 2,67 | 0,11 | 0,26 | 0,94 | 0,92 |
| 16 | М-М-06-2-1 | -1,12 | 0,05 | 0,59 | 0,87 | 0,80 |
| 17 | М-М-06-2-2 | -0,48 | 0,05 | 0,43 | 1,10 | 1,14 |
| 18 | М-М-06-2-3 | 0,53 | 0,06 | 0,38 | 1,05 | 1,23 |
| 19 | М-М-08-2-1 | -2,17 | 0,05 | 0,6 | 0,85 | 0,73 |
| 20 | М-М-08-2-2 | -0,04 | 0,05 | 0,51 | 0,92 | 0,91 |
| 21 | М-М-08-2-3 | 1,9 | 0,08 | 0,16 | 1,16 | 1,75 |
| 22 | М-R-02-2-1 | -0,53 | 0,05 | 0,52 | 0,94 | 1,01 |
| 23 | М-R-02-2-2 | 0,53 | 0,06 | 0,45 | 0,93 | 0,98 |
| 24 | М-R-02-2-3 | 0,28 | 0,05 | 0,3 | 1,16 | 1,82 |
| 25 | М-R-05-2-1 | -1,8 | 0,05 | 0,43 | 1,13 | 1,17 |
| 26 | М-R-05-2-2 | 0,36 | 0,06 | 0,38 | 1,06 | 1,36 |
| 27 | М-R-05-2-3 | 3,4 | 0,15 | 0,08 | 1,05 | 3,26 |
| 28 | М-D-03-2-1 | -2,68 | 0,05 | 0,55 | 0,87 | 0,95 |
| 29 | М-D-03-2-2 | -0,73 | 0,05 | 0,57 | 0,88 | 0,82 |
| 30 | М-D-03-2-3 | 0,91 | 0,06 | 0,44 | 0,93 | 0,77 |
| 31 | М-D-20-2-1 | -1,56 | 0,05 | 0,63 | 0,81 | 0,74 |
| 32 | М-D-20-2-2 | 1,9 | 0,08 | 0,31 | 0,98 | 0,91 |
| 33 | М-D-20-2-3 | 2,31 | 0,1 | 0,25 | 0,99 | 1,06 |
| 34 | М-D-05-2-1 | -1,3 | 0,05 | 0,6 | 0,86 | 0,83 |
| 35 | М-D-05-2-2 | -0,44 | 0,05 | 0,54 | 0,92 | 0,87 |
| 36 | М-D-05-2-3 | 2,02 | 0,09 | 0,29 | 0,94 | 1,89 |
| 37 | М-D-08-2-1 | -2,58 | 0,05 | 0,52 | 0,94 | 1,01 |
| 38 | М-D-08-2-2 | -0,93 | 0,05 | 0,58 | 0,88 | 0,81 |
| 39 | М-D-08-2-3 | 1,75 | 0,08 | 0,19 | 1,12 | 2,05 |
| 40 | М-G-01-2-1 | -1,41 | 0,05 | 0,57 | 0,91 | 0,84 |
| 41 | М-G-01-2-2 | 1,55 | 0,07 | 0,28 | 1,04 | 1,44 |
| 42 | М-G-01-2-3 | 3,24 | 0,14 | 0,02 | 1,08 | 5,52 |
| 43 | M-G-03-2-1 | -1,8 | 0,05 | 0,48 | 1,05 | 1,03 |
| 44 | M-G-03-2-2 | 0,42 | 0,06 | 0,36 | 1,11 | 1,18 |
| 45 | M-G-03-2-3 | 1,57 | 0,07 | 0,13 | 1,26 | 2,17 |

Карта переменных, вариант А2

На карте переменных (рисунок 1) показано распределение испытуемых и заданий относительно друг друга на общей метрической шкале. Слева на рисунке находится шкала логитов. На карте испытуемые представлены слева, задания – справа. Более трудные задания и более подготовленные испытуемые находятся в верхней части карты, более легкие задания и менее подготовленные испытуемые находятся в нижней части карты. Карта позволяет проанализировать совместное распределение заданий относительно данного контингента испытуемых, а также диагностировать проблемы теста.

Распределение испытуемых на карте переменных смещено вниз относительно распределения заданий, что говорит о том, что тест оказался достаточно трудным для данной выборки.

Тест содержит достаточное количество заданий для испытуемых среднего уровня подготовленности. Уровень трудности некоторых заданий выше уровня подготовленности испытуемых (М-R-22-2-3, М-G-21-2-3, М-М-20-2-3), уровень трудности этих заданий выше 3 логитов. В тесте не хватает заданий для наименее подготовленных испытуемых, но стоит отметить, что уровень подготовленности тех испытуемых, для которых отсутствуют задания теста, располагается в диапазоне от -5 до -3 логитов.



*Рисунок 1. Карта переменных, вариант А2.*

ПРИЛОЖЕНИЕ 13

АНАЛИЗ РАЗМЕРНОСТИ, ВАРИАНТ В2

Анализ заданий, вариант В2

Задания теста проанализированы по следующим показателям: трудность задания, ошибка измерения, корреляция с общим баллом по тесту и общие статистики согласия.

В столбцах 2 и 3 таблиц представлены оценки трудности заданий и соответствующие ошибки измерения. В процессе анализа данных тестирования не было выявлено заданий, которые не были достигнуты участниками. Во всех блоках теста, кроме последнего, соблюдается иерархия уровней трудности заданий в соответствии с теоретической моделью.

В 4 столбце приведен коэффициент корреляции между баллами по заданию и уровнем подготовленности испытуемых. Большая часть заданий имеют приемлемые значения корреляции. Наименьшая связь баллов по заданию с уровнем подготовленности испытуемых наблюдается в заданиях повышенной трудности 3-го уровня (M-D-26-2-3, M-G-21-2-3).

В столбцах 5 и 6 приведены значения нестандартизированных статистик согласия, характеризующих согласие данных с используемой моделью измерения. Используются две версии статистики согласия – простая и взвешенная. В данном тесте большая часть заданий находится в хорошем согласии с моделью, о чем свидетельствуют статистики согласия, находящиеся в пределах нормы (0,8 – 1,2).

В тесте присутствует несколько заданий, у которых есть выход за пределы критических значений статистик. Выход за пределы критических значений наблюдается в четырех заданиях 2-го уровня трудности и в пяти заданиях 3-го уровня. Данную ситуацию можно объяснить тем, что задания относятся к блоку «Геометрия», с которым ученики трудно справляются. Наиболее проблемным выглядит задание 42 (M-G-21-2-3), так оно имеет низкую корреляцию с общим баллом по тесту, и, что менее типично для заданий 3-го уровня, имеет средний уровень трудности.

В целом задания теста можно признать функционирующими удовлетворительно.

Таблица 1. Статистические данные заданий теста, вариант В2.

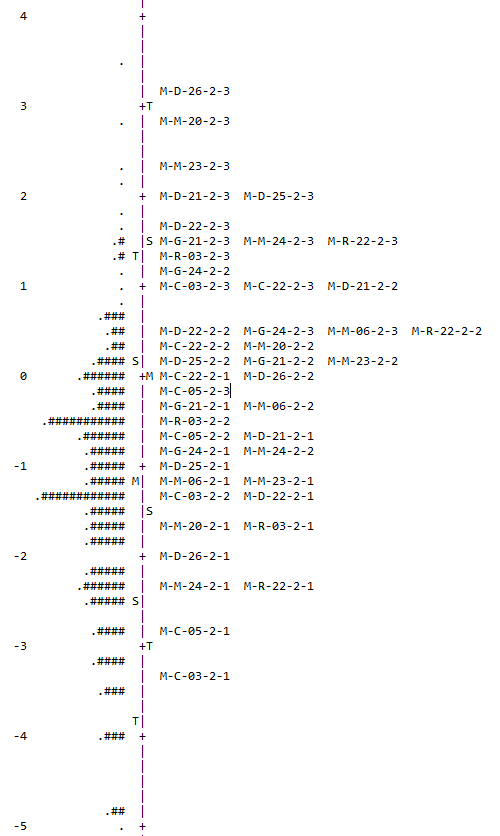
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер задания | | Трудность | Ошибка измерения | Корреляция | Статистики согласия | |
| INFIT MNSQ | OUTFIT MNSQ |
| 1 | M-C-03-2-1 | -3,39 | 0,06 | 0,47 | 0,93 | 0,92 |
| 2 | M-C-03-2-2 | -1,38 | 0,05 | 0,53 | 0,94 | 0,91 |
| 3 | M-C-03-2-3 | 1,03 | 0,06 | 0,37 | 0,98 | 0,86 |
| 4 | M-C-22-2-1 | 0,01 | 0,05 | 0,38 | 1,08 | 0,38 |
| 5 | M-C-22-2-2 | 0,28 | 0,05 | 0,4 | 0,98 | 1,39 |
| 6 | M-C-22-2-3 | 1 | 0,06 | 0,41 | 0,92 | 0,83 |
| 7 | M-C-05-2-1 | -2,87 | 0,06 | 0,53 | 0,89 | 0,78 |
| 8 | M-C-05-2-2 | -0,59 | 0,05 | 0,58 | 0,83 | 0,75 |
| 9 | M-C-05-2-3 | -0,19 | 0,05 | 0,42 | 1,03 | 1,11 |
| 10 | M-M-20-2-1 | -1,65 | 0,05 | 0,44 | 1,09 | 1,12 |
| 11 | M-M-20-2-2 | 0,37 | 0,05 | 0,25 | 1,17 | 1,69 |
| 12 | M-M-20-2-3 | 2,78 | 0,11 | 0,26 | 0,92 | 0,75 |
| 13 | M-M-23-2-1 | -1,13 | 0,05 | 0,45 | 1,05 | 1,1 |
| 14 | M-M-23-2-2 | 0,2 | 0,05 | 0,38 | 1,05 | 1,08 |
| 15 | M-M-23-2-3 | 2,27 | 0,09 | 0,3 | 0,91 | 0,76 |
| 16 | M-M-06-2-1 | -1,16 | 0,05 | 0,55 | 0,91 | 0,86 |
| 17 | M-M-06-2-2 | -0,35 | 0,05 | 0,41 | 1,07 | 1,09 |
| 18 | M-M-06-2-3 | 0,42 | 0,05 | 0,38 | 1,04 | 1,07 |
| 19 | M-M-24-2-1 | -2,36 | 0,05 | 0,53 | 0,92 | 0,87 |
| 20 | M-M-24-2-2 | -0,87 | 0,05 | 0,5 | 0,96 | 0,96 |
| 21 | M-M-24-2-3 | 1,53 | 0,07 | 0,38 | 0,89 | 0,74 |
| 22 | M-R-03-2-1 | -1,63 | 0,05 | 0,53 | 0,94 | 0,94 |
| 23 | M-R-03-2-2 | -0,56 | 0,05 | 0,44 | 1,04 | 1,02 |
| 24 | M-R-03-2-3 | 1,27 | 0,07 | 0,33 | 1 | 1,11 |
| 25 | M-R-22-2-1 | -2,3 | 0,05 | 0,53 | 0,94 | 0,86 |
| 26 | M-R-22-2-2 | 0,42 | 0,05 | 0,5 | 0,85 | 0,74 |
| 27 | M-R-22-2-3 | 1,47 | 0,07 | 0,35 | 0,95 | 0,91 |
| 28 | M-D-25-2-1 | -1,07 | 0,05 | 0,6 | 0,82 | 0,75 |
| 29 | M-D-25-2-2 | 0,15 | 0,05 | 0,5 | 0,89 | 0,75 |
| 30 | M-D-25-2-3 | 1,97 | 0,08 | 0,24 | 1 | 1,9 |
| 31 | M-D-21-2-1 | -0,64 | 0,05 | 0,37 | 1,13 | 1,27 |
| 32 | M-D-21-2-2 | 0,98 | 0,06 | 0,45 | 0,86 | 0,8 |
| 33 | M-D-21-2-3 | 2 | 0,08 | 0,22 | 1,07 | 1,17 |
| 34 | M-D-22-2-1 | -1,37 | 0,05 | 0,53 | 0,95 | 0,95 |
| 35 | M-D-22-2-2 | 0,43 | 0,05 | 0,21 | 1,23 | 1,87 |
| 36 | M-D-22-2-3 | 1,62 | 0,07 | 0,22 | 1,07 | 2,06 |
| 37 | M-D-26-2-1 | -2,01 | 0,05 | 0,59 | 0,85 | 0,8 |
| 38 | M-D-26-2-2 | 0,06 | 0,05 | 0,49 | 0,91 | 0,84 |
| 39 | M-D-26-2-3 | 3,14 | 0,13 | 0,15 | 1,01 | 1,76 |
| 40 | M-G-21-2-1 | -0,36 | 0,05 | 0,42 | 1,04 | 1,11 |
| 41 | M-G-21-2-2 | 0,23 | 0,05 | 0,39 | 1,04 | 1,03 |
| 42 | M-G-21-2-3 | 1,45 | 0,07 | 0,05 | 1,28 | 2,72 |
| 43 | M-G-24-2-1 | -0,9 | 0,05 | 0,48 | 0,99 | 1,01 |
| 44 | M-G-24-2-2 | 1,22 | 0,07 | 0,35 | 0,97 | 1,13 |
| 45 | M-G-24-2-3 | 0,53 | 0,06 | 0,22 | 1,22 | 1,83 |

Карта переменных, вариант В2

На карте переменных (рисунок 1) показано распределение испытуемых и заданий относительно друг друга на общей метрической шкале. Слева на рисунке находится шкала логитов. На карте испытуемые представлены слева, задания – справа. Более трудные задания и более подготовленные испытуемые находятся в верхней части карты, более легкие задания и менее подготовленные испытуемые находятся в нижней части карты. Карта позволяет проанализировать совместное распределение заданий относительно данного контингента испытуемых, а также диагностировать проблемы теста.

Распределение испытуемых на карте переменных смещено вниз относительно распределения заданий, что говорит о том, что тест оказался достаточно трудным для данной выборки.

Тест содержит достаточное количество заданий для испытуемых среднего уровня подготовленности. В тесте не хватает заданий для наименее подготовленных испытуемых, но стоит отметить, что уровень подготовленности тех испытуемых, для которых отсутствуют задания теста, располагается в диапазоне от -5 до -3 логитов. Также в тесте есть достаточно много заданий повышенной трудности, с которыми не справляется большая часть испытуемых.

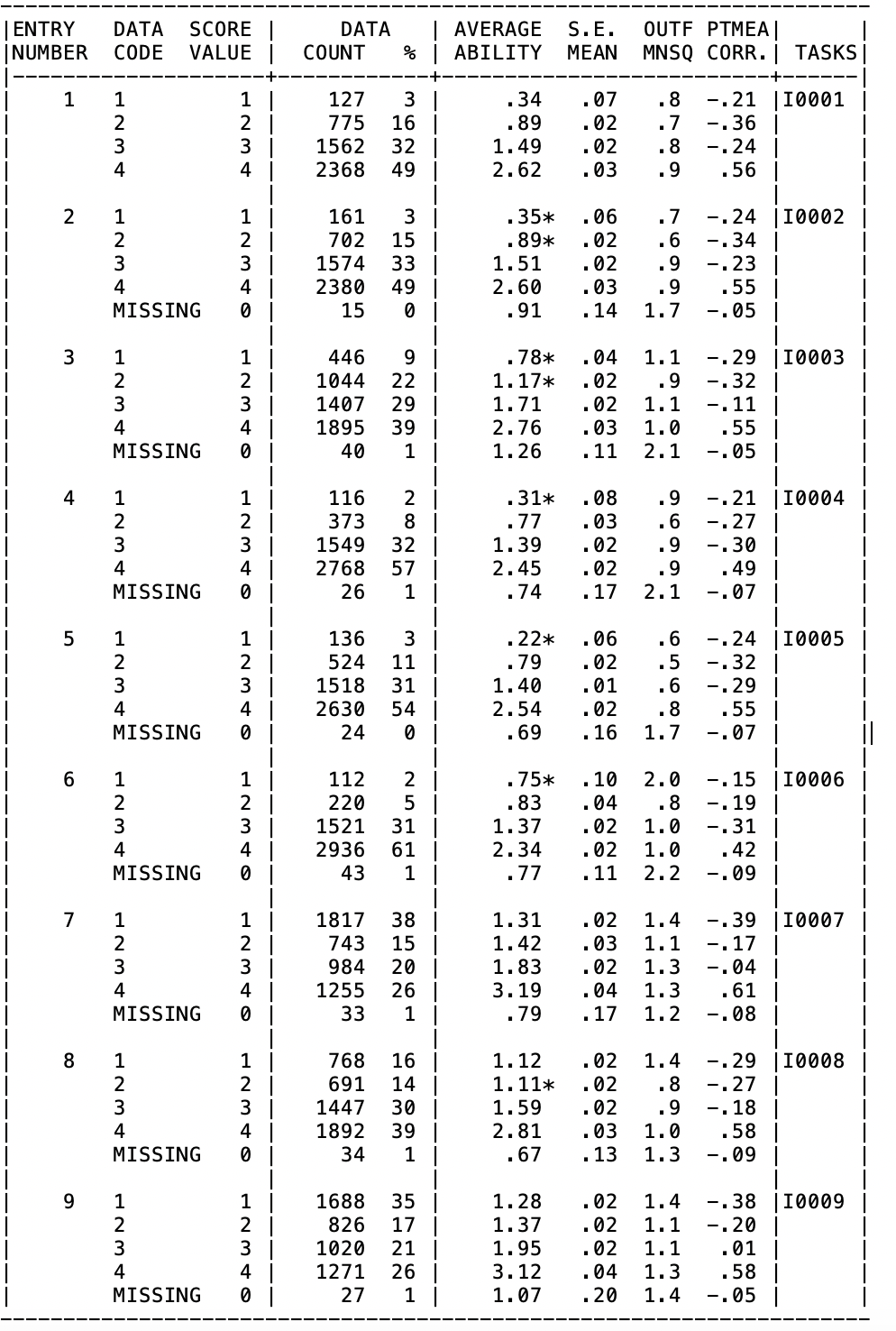
Практически все задания расположены в диапазоне от -3 до 3 логитов. У одного задания уровень трудности выше 3 логитов (M-D-26-2-3) и у одного задания уровень трудности ниже 3 логитов (M-C-03-2-1).

*Рисунок 1. Карта переменных, вариант В2*

# ПРИЛОЖЕНИЕ 14

АНАЛИЗ ФУНКЦИОНИРОВАНИЯ КАТЕГОРИЙ

(КОНСТРУКТ «МОТИВАЦИЯ К ЧТЕНИЮ»)



*Рисунок 1. Анализ категорий по Raiting scale*

ПРИЛОЖЕНИЕ 15

АНАЛИЗ ФУНКЦИОНИРОВАНИЯ КАТЕГОРИЙ

(КОНСТРУКТ «МОТИВАЦИЯ К МАТЕМАТИКЕ»)

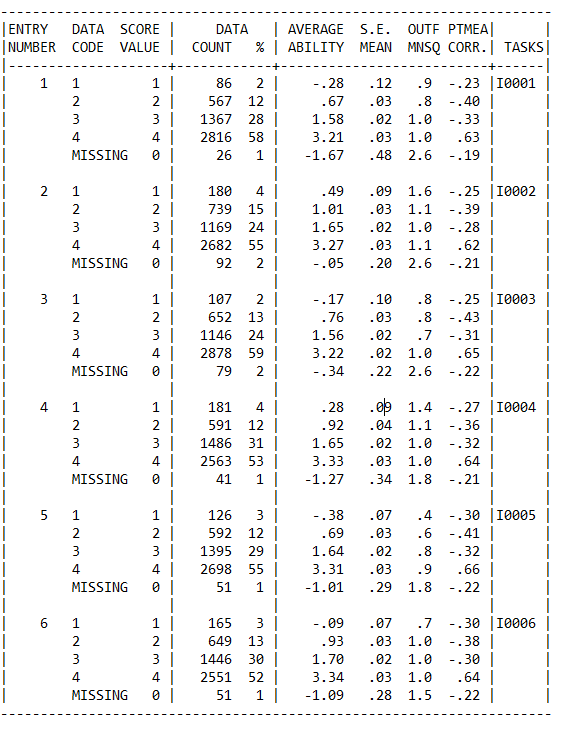


*Рисунок 1. Анализ категорий по Raiting scale*

ПРИЛОЖЕНИЕ 16

АНАЛИЗ ФУНКЦИОНИРОВАНИЯ КАТЕГОРИЙ

(КОНСТРУКТ «СУБЪЕКТИВНОЕ БЛАГОПОЛУЧИЕ ШКОЛЬНИКА. ШКАЛА УДОВЛЕТВОРЕННОСТИ ДРУЗЬЯМИ»)

*Рисунок 1. Анализ категорий по Raiting scale*

ПРИЛОЖЕНИЕ 17

АНАЛИЗ ФУНКЦИОНИРОВАНИЯ КАТЕГОРИЙ

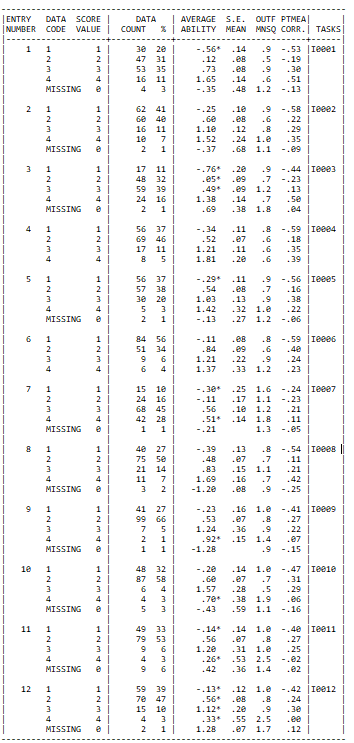
(КОНСТРУКТ «СУБЪЕКТИВНОЕ БЛАГОПОЛУЧИЕ ШКОЛЬНИКА. ШКАЛА УДОВЛЕТВОРЕННОСТИ ШКОЛОЙ»)

Таблица 1. Анализ категорий по Raiting Scale

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Утверждение | | Категория | Процент ответов (%) | Трудность | Ошибка измерения | OUTFIT MNSQ |
| 7.1 | Я с охотой хожу в школу | 1 | 8 | 0,78 | 0,03 | 1,1 |
| 2 | 16 | 1,02 | 0,02 | 0,8 |
| 3 | 32 | 1,32 | 0,01 | 0,8 |
| 4 | 42 | 1,8 | 0,01 | 0,9 |
| MISSING | 1 | -0,87 | 0,4 | 1,4 |
| 7.3 | В школе интересно | 1 | 3 | 0,28 | 0,07 | 0,6 |
| 2 | 13 | 0,92 | 0,02 | 0,7 |
| 3 | 33 | 1,25 | 0,01 | 0,7 |
| 4 | 51 | 1,75 | 0,01 | 0,9 |
| MISSING | 1 | -0,63 | 0,35 | 1,8 |
| 7.4 | Мне нравится то, что я делаю в школе | 1 | 4 | 0,52 | 0,06 | 0,8 |
| 2 | 19 | 1 | 0,02 | 0,8 |
| 3 | 34 | 1,33 | 0,01 | 0,9 |
| 4 | 42 | 1,79 | 0,01 | 0,9 |
| MISSING | 1 | -1,06 | 0,42 | 1,3 |
| 7.6 | Я хотел бы, чтобы мне не нужно было ходить в школу | 1 | 50 | 1,37 | 0,01 | 1,3 |
| 2 | 21 | 1,33 | 0,02 | 1,2 |
| 3 | 13 | 1,39 | 0,02 | 1,9 |
| 4 | 14 | 1,86 | 0,04 | 1,9 |
| MISSING | 2 | -0,22 | 0,29 | 1 |
| 7.8 | Я многому учусь в школе | 1 | 3 | 0,88 | 0,09 | 2 |
| 2 | 6 | 0,79 | 0,03 | 0,7 |
| 3 | 35 | 1,2 | 0,01 | 0,8 |
| 4 | 55 | 1,69 | 0,01 | 0,9 |
| MISSING | 1 | -0,89 | 0,34 | 2,1 |
| 7.9 | Мне плохо в школе | 1 | 56 | 1,38 | 0,01 | 1,2 |
| 2 | 30 | 1,37 | 0,01 | 1,4 |
| 3 | 7 | 1,41 | 0,04 | 1,9 |
| 4 | 7 | 2,24 | 0,07 | 1,7 |
| MISSING | 1 | -0,9 | 0,36 | 0,8 |
| 7.13 | Мне нравится находиться в школе | 1 | 4 | 0,43 | 0,05 | 0,7 |
| 2 | 13 | 0,9 | 0,02 | 0,7 |
| 3 | 30 | 1,23 | 0,01 | 0,6 |
| 4 | 53 | 1,75 | 0,01 | 0,9 |
| MISSING | 1 | -1,19 | 0,41 | 1,9 |

# ПРИЛОЖЕНИЕ 18

АНАЛИЗ ФУНКЦИОНИРОВАНИЯ КАТЕГОРИЙ

(КОНСТРУКТ «УДОВЛЕТВОРЕННОСТЬ РАБОТОЙ И ПРОФЕССИОНАЛЬНОЕ ВЫГОРАНИЕ»)

*Рисунок 1. Анализ категорий по Raiting scale*