

## УЛУЧШЕНИЕ КАЧЕСТВА КОНТРОЛЬНЫХ РАБОТ ПОСРЕДСТВОМ РАСЧЁТА СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК ЗАДАНИЙ

**Методология****Юлия Корешникова**

Краевое государственное бюджетное специализированное учреждение  
«Центр оценки качества образования» г. Красноярск  
KUN82@mail.ru

**Цель данной статьи — показать возможность применения научных методов педагогических измерений для улучшения качества традиционных контрольных работ, применяемых для проверки уровня подготовленности и аттестации учащихся.**

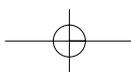
*Ключевые слова:* меры изменчивости, меры центральной тенденции, корреляция, мера трудности заданий, графические образы заданий, модель Раша.

### Введение

Одним из главных требований, предъявляемых современным обществом к оцениванию результатов обучения, является уменьшение субъективного фактора при контроле уровня подготовленности учащихся. Для достижения этого требования необходимо принципиально изменить подходы к созданию средств контроля. Методы, посредством которых можно получить максимально объективированную оценку результатов учебной деятельности, должны состоять из заданий известной трудности и отвечать ряду других статистических требований, используемых в педагогических измерениях. В противном случае мы рискуем получить данные, по которым весьма затруднительно будет сделать объективные выводы.

### Модель контрольной работы

Контрольные работы, результаты применения которых анализируются в данной статье, создавались с целью определения сформированных у учащихся предметных умений, рейтинга выпускников



ПЕД
измерения

начальной школы различных типов образовательных учреждений, а также для установления уровня трудности заданий, с которыми справляется каждый выпускник.

Контрольная работа состояла из 4 вариантов. Параллельность вариантов достигалась за счёт включения в работу взаимозаменяемых, однотипных, сопоставимых по трудности и по содержанию заданий, расположенных под одними и теми же порядковыми номерами во всех вариантах работы и проверяющих сформированность одних и тех же предметных умений.

Контрольные работы состояли из 2 частей, которые отличались формой и уровнем трудности включённых в них заданий. Часть А состояла из заданий в тестовой форме, часть Б — из заданий в традиционной, нетестовой форме. В качестве заданий в тестовой форме использовались задания с выбором одного правильного ответа, задания открытой формы и задания на установление правильной последовательности.

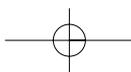
### **Методические принципы обработки результатов контрольных работ**

Обработка результатов контрольных работ проводилась при помощи пакета программ

автоматической обработки данных. В основу этого пакета заложен ряд принципов.

В качестве первого принципа было положено различие в количестве умений, что проверяется при помощи заданий в тестовой и нетестовой формах. Задание в тестовой форме соответствовало одному умению из кодификатора. С помощью заданий в нетестовой форме проверялось несколько умений. Для этого каждое такое задание делилось на части (действия), последовательное выполнение которых приводило учащегося к ответу на задание. Каждая часть оценивалась отдельно по дихотомической шкале: -1 за верное выполнение, 0 — за неверное выполнение или невыполнение. Таким образом, каждое действие в заданиях нетестовой формы, по замыслу разработчиков, рассматривалось как отдельное задание.

Второй принцип, заложенный в программе, основывался на различной степени значимости умений, оцениваемых при помощи контрольных работ. Все умения, перечисленные в кодификаторах к контрольной работе, проверялись с помощью различного количества действий. Для оценки сформированности умений выбрана дискретная шкала с тремя позициями: «сформировано», «частично сформировано» и «не сформировано».



Для дифференциации учащихся по уровню подготовленности все задания в контрольной работе разделены на три уровня трудности: базовый, повышенный и высокий. Это является третьим принципом, заложенным в программе.

После обработки результатов выполнения контрольных работ в образовательные учреждения, выпускники которых выполняли контрольные работы, была предоставлена информация по каждому ученику. Эта информация отражала результаты сформированности предметных умений; выводы о способности решать задания базового, повышенного и высокого уровня трудности; персональный рейтинг.

На основании полученных результатов на местах проводился рейтинг муниципалитетов, а так же школ внутри одного муниципалитета. В зависимости от рейтинга, относительно каждой школы принимались различные управленческие решения.

## Статистический анализ результатов контрольных работ

В 2010 году была проведена проверка данных, полученных в результате применения контрольных работ. Оценка качества заданий производилась с позиций теории педагогических измерений: определялись меры центральной тенденции и вариации данных, определялись значения коэффициентов корреляции между ответами на задания, проверяющими одно и то же умение. Оценивались также меры трудности каждого задания; строились по модели Раша графические образы заданий для каждой из трёх групп заданий.

Меры изменчивости и центральной тенденции (табл. 1) рассчитывались для того, чтобы проверить возможность проведения рейтинга по итогам контрольной работы. Максимально возможное число баллов в работе равнялось 40.

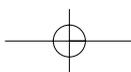
Меры центральной тенденции описывают типичного индивидуума из совокупности,

Методология

Таблица 1

Меры изменчивости и центральной тенденции

№ п\п	Мера	Значения
1.	Среднее арифметическое	32,6
2.	Мода	38
3.	Дисперсия	34,86
4.	Среднеквадратичное (стандартное) отклонение	5,9



ПЕД  
измерения

таким образом, по данным контрольной работы можно было сделать вывод, что типичный выпускник начальной школы очень хорошо подготовлен. Низкое значение стандартного отклонения свидетельствует о гомогенности изучаемой группы относительно исследуемого показателя.

График частотного распределения баллов испытуемых представлен на рис. 1.

$Z$  — это значения нормированных отклонений значений баллов испытуемых от средней арифметической. Эти значения вычислялись по формуле:  $Z_i = (X_i - M) / s_x$ , где  $X_i$  — исходный тестовый балл испытуемого  $i$ ;  $M$  — среднее значение тестового балла в группе испытуемых,  $s_x$  — стандартное отклонение тестовых результатов испытуемых исследуемой группы.

Графическое представление полученных данных свидетельствует о том, что распреде-

ление баллов является унимодальным, асимметричным, с максимумом частот правильных ответов, смещённым в область высоких баллов. Это подтверждают и меры центральной тенденции. Иначе говоря, для аттестации учащихся применялись в основном лёгкие задания. Что вполне согласуется с самим смыслом аттестации, главная цель которой — проверить соответствие уровня подготовленности учащихся минимальным требованиям, заложенным в государственных образовательных стандартах.

Стоит обратить внимание, что максимальные  $Z$ -оценки при помощи данной контрольной работы получить было невозможно, поскольку для проведения контроля применялись преимущественно лёгкие и очень лёгкие задания. То есть хорошо подготовленные учащиеся при применении данной

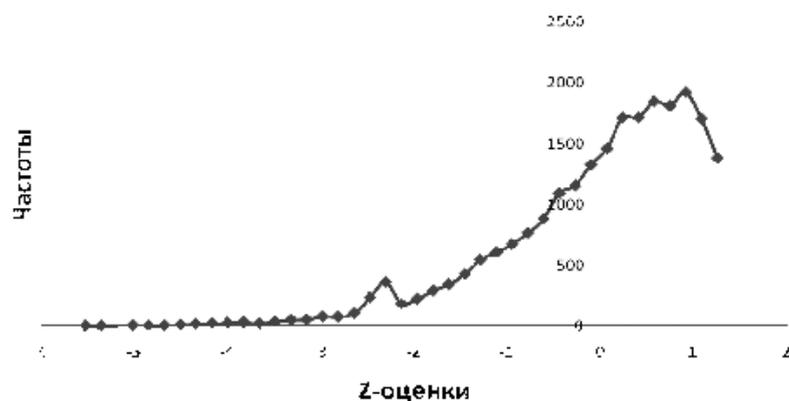


Рис. 1. Частотное распределение стандартизованных баллов

контрольной работы остались, образно говоря, «недомеренными», поскольку в контрольной работе не было подходящих для них заданий повышенной трудности.

Этот вывод о заниженной в целом мере трудности заданий опровергает возможность формулировки вывода о высокой подготовке выпускников начальной школы. Проблема заключается в ориентации разработчиков контрольных работ на подбор лёгких заданий, которые подходят для аттестации большинства учащихся, но бесполезны для действительного измерения уровня подготовленности хорошо и отлично подготовленных учащихся. Приближенные картины к действительному уровню подготовленности возможно в том случае, если в контрольную работу будут добавлены задания с более высоким уровнем трудности. Но тогда понадобится разрабатывать тест как систему заданий равномерно возрастающей трудности, что открывает возможность объективной оценки уровня подготовленности всех учащихся, с любым уровнем подготовленности.

Проведённая контрольная работа оказалась аттестационной, в принципе пригодной лишь для выявления слабо подготовленных учащихся, что важно для принятия педагогических решений относительно

возможности и направлений их дальнейшего обучения. Это и есть главная задача аттестации учащихся на соответствие обязательным для всех требованиям государственных образовательных стандартов. Но для объективного измерения уровня подготовленности учащихся среднего и высокого уровня подготовленности контрольная работа анализированного типа оказывается неподходящим методом.

Для проверки возможности обосновать выводы о сформированности предметных умений была построена корреляционная матрица<sup>1</sup>, так как многие умения проверялись при помощи совокупности заданий. Полученные коэффициенты корреляции представлены в табл. 2.

Как видно из данных таблицы, значения коэффициентов корреляции принимают преимущественно низкие значения, что свидетельствует о фактическом отсутствии связи заданий между собой и является первым признаком отсутствия у них метрических свойств. Следовательно, нельзя делать вывод о сформированности или несформированности умений, исходя из выбранной в данном случае совокупности облегчённых заданий. Так, если учащийся владеет умением, то он должен верно выполнять все задания, его проверяющие, тем более, если это задания одного

## Методология

1  
Использовался классический коэффициент корреляции Пирсона.

Таблица 2

## Значение коэффициентов корреляции Пирсона

№ п/п	Умения	Номера заданий	Коэффициенты корреляции Пирсона
1.	Определять количество букв и звуков	2,36	0,18
2.	Разбор слова по составу	4,31	0,14
3.	Однокоренные слова	9,28	0,16
4.	Число и падеж существительных	13,14	0,23
5.	Грамматическая основа предложения	26,32	0,19
6.	Безударные гласные	19,33	0,098
7.	Окончания существительных	5,22	0,072
8.	Писать «ъ» и «ь» знаки	6,23	0,11
9.	Знаки препинания в конце предложений	15,16	0,47

уровня трудности. Одна из причин низкой корреляции — высокая доля правильных ответов, что заметно снижает значение коэффициентов корреляции.

Данные табл. 3 дают представление о мере трудности использованных заданий. Из данных таблиц следует, что абсолютное большинство заданий, составляющих контрольную работу, оказались *слишком лёгкими*.

Этот вывод подтверждают графические образы заданий, построенных по модели Г. Раша (рис. 2–4). Все задания оказались примерно одного уровня трудности. Этот уровень слишком лёгкий, поэтому при помощи таких заданий невозможно делать выводы об уровне подготовленности учащихся решать задания повышенного и высокого уровней трудности.

Такова, по-видимому, реальная практика составления контрольных работ. И это можно понять, потому что, если задания окажутся трудными, то результаты проверок окажутся плохими, и за этим могут последовать оргвыводы, неприятные для учащихся, учителей и работников органов управления образованием. Но при этом надо также понимать, что облегчённые задания можно применять только для массовой аттестации учащихся, где главным критерий — овладение требуемым минимумом знаний и умений.

Полученные статистические данные свидетельствуют о том, что выводы, формулировавшиеся обычно по результатам применения контрольных работ, нельзя признать объективными по причине искус-

Таблица 3

Методология

**Мера трудности заданий**  
( $p$  – доля правильных ответов  $> 0,8$ , задания слишком лёгкие)

№ задания	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$p$	0,63	0,89	0,69	0,83	0,88	0,91	0,90	0,87	0,63	0,87	0,83	0,88	0,89	0,81
№ задания	15	16	17	18	19	20	21	22	23	24	25	26	27	28
$p$	0,88	0,61	0,68	0,83	0,97	0,92	0,95	0,90	0,94	0,98	0,98	0,93	0,86	0,85
№ задания	29	30	31	32	33	34	35	36	37	38	39	40		
$p$	0,66	0,87	0,64	0,64	0,93	0,91	0,57	0,61	0,81	0,77	0,83	0,77		

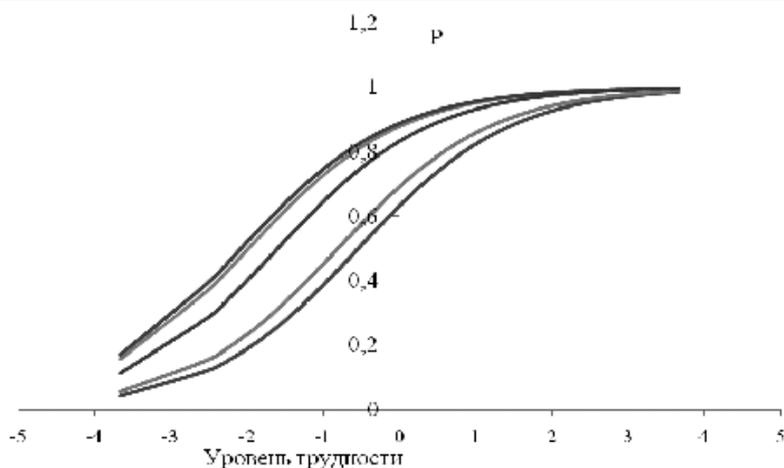


Рис. 2. Графические образы заданий базового уровня трудности

ственной облегченности большинства заданий.

Использовать данные результаты для принятия управленческих решений об уровне подготовленности всех обследованных учащихся, с точки зрения теории педагогических измерений, *не рекомендуется. По этим данным можно принимать только решения об аттестации или не аттестации слабо подготовленных учащихся.* Гранич-

ные значения баллов для аттестации принимаются обычно органами управления, примерно, как это делается при проведении ЕГЭ.

Данный пример подтверждает необходимость предварительного уточнения цели проведения контрольных работ. Потому что одним и тем же набором облегченных заданий контрольных работ достижение целей аттестации и объективного

ПЕД  
измерения

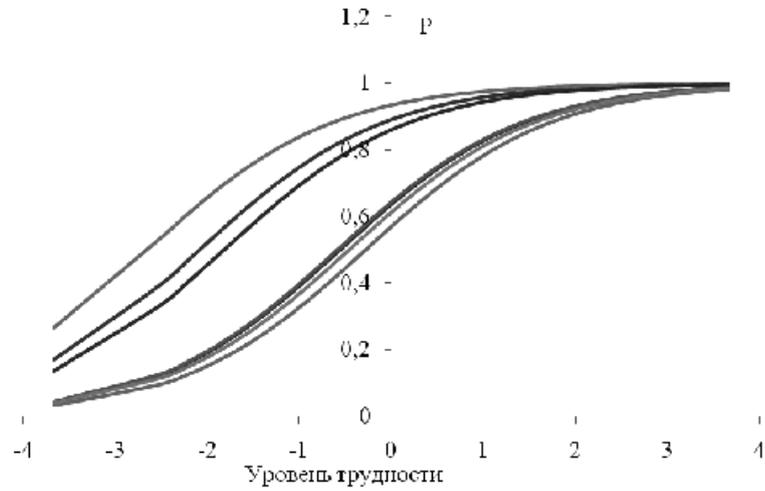


Рис. 3. Графические образы заданий повышенного уровня трудности

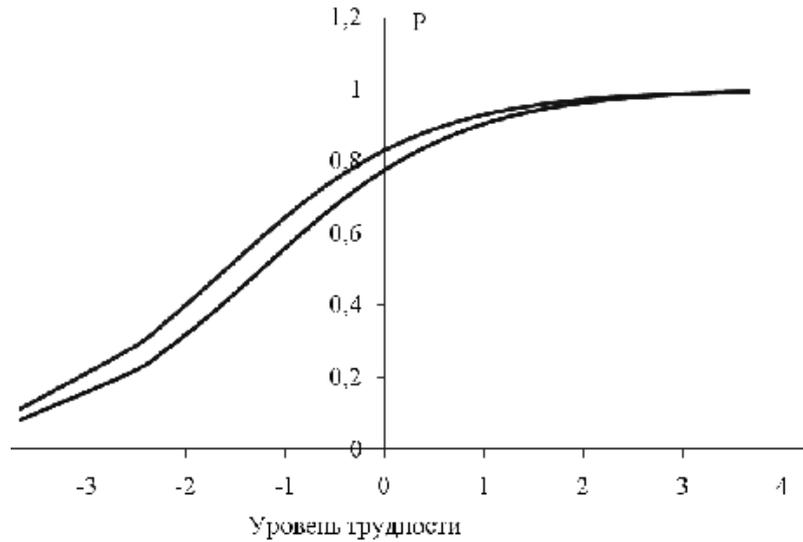


Рис. 4. Графические образы заданий высокого уровня трудности

измерения уровня учебных достижений всех учащихся, особенно высокого уровня подготовленности, оказывается делом невозможным.

Это и есть главный методический вывод, который можно сделать по итогам проведения статистического анализа результатов проведённых контрольных работ.