

# СТАТИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ ХАРАКТЕРИСТИК СРЕДСТВ ИЗМЕРЕНИЯ (ТЕСТОВ), ИСПОЛЬЗУЕМЫХ В АСИ<sup>1</sup>

© **Бояшова С. А.**, доктор технических наук, научный руководитель сетевого эксперимента Приморского района Санкт-Петербурга: «Разработка районной системы независимой оценки качества образования на основе широкого использования современных информационно-коммуникационных технологий».

© Директор Государственного бюджетного образовательного учреждения гимназии № XXX Приморского района Санкт-Петербурга<sup>2</sup>.

Обобщенные рекомендации к разработке тестовых заданий и композиции целостного теста в соответствии с современной теорией и практикой педагогических измерений приводятся в работах В. С. Аванесова А. Н. Майорова<sup>3</sup>.

Основываясь на перечисленных выше авторских работах, определим основные статистические характеристики тестовых заданий, которые используются при тестировании обученности школьников по русскому языку и математике по завершению их обучения в начальной школе с использованием АСИ (статистическая трудность и дифференцирующая способность).

Исходя из особенностей построения АСИ, вначале проводится исследование названных статистических характеристик применительно ко всем отдельным структурным элементам системы теста с использованием данных тестирования различных выборок испытуемых.

В границах первоначального исследования структурный элемент системы теста рассматривается как одно тестовое задание, условно выполняемое всеми испытуемыми. В реальных условиях тестирования каждый испытуемый выполняет свое задание, относящееся к одному элементу системы теста, которое выбрано автоматически из базы АСИ в момент начала тестирования.

Если структурный элемент теста, рассматриваемый как одно тестовое задания, не соответствует статистическим требованиям, то пересматриваются все задания в базе АСИ, которые соответствуют данному элементу.

База заданий корректируется экспертами, и далее, проводятся последующие исследования качества структурных элементов системы теста.

Каждое новое измерение позволяет автоматически улучшать качество тестов в АСИ.

---

<sup>1</sup> АСИ – автоматизированная система измерений

<sup>2</sup> ФИО директора и номер гимназии не указан исходя из принципа соблюдения конфиденциальности информации, предоставляемой пользователям автоматизированной системы измерений (АСИ).

<sup>3</sup> Аванесов В. С. Теоретические основы разработки заданий в тестовой форме : учебное пособие для профессорско-преподавательского состава высшей школы. / В. С. Аванесов.– М. : Изд-во МГТА, 1995. – 198 с.

Аванесов, В. С. Композиция тестовых заданий : учебное пособие. / В. С. Аванесов. – М. : Изд-во Центр тестирования, 2002. – 217 с.

Майоров А. Н. Теория и практика создания тестов для системы образования. / А. Н. Майоров. – М. : Изд-во Народное образование, 2000.

Майоров А. Н. Тесты школьных достижений. Конструирование, проведение, использование. / А. Н. Майоров. – СПб. : Изд-во Образование и культура, 1996.

Далее, исходя из выше сказанного, понятие «статистическая трудность задания» будем рассматривать как аналогичное понятию «статистическая трудность элемента теста» (СТЭТ), а понятие «дифференцирующая способность задания» как аналогичное понятию «дифференцирующая способность элемента теста» (ДСЭТ).

Основные статистические характеристики элементов теста (СТЭТ и ДСЭТ) рассчитываются по данным, полученным при апробации АСИ в режиме реального времени на двух различных по объему и качественному составу статистических выборках испытуемых.

Минимальная статистическая выборка отбирается по заявке образовательного учреждения в количестве тридцати испытуемых из общего количества тестируемых в следующем соотношении:

- 15 испытуемых – учащиеся, имеющие оценки в исследуемой предметной области «4» или «5» по пятибалльной шкале;
- 15 испытуемых – учащиеся, имеющие оценки в исследуемой предметной области «3» или «2» по пятибалльной шкале.

Полная статистическая выборка в качественном соотношении формируется произвольно по заявке образовательного учреждения (все желающие пройти тестирование). Количественные характеристики выборок, на которых проводилось исследование основных статистических характеристик элементов моделей тестов по русскому языку и математике, представлены в табл. 1.

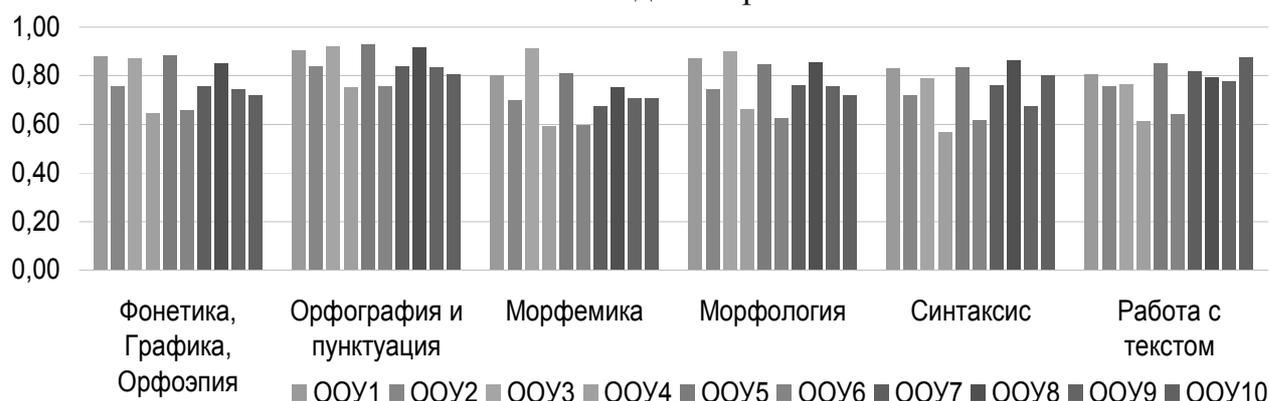
**Таблица 1.** Характеристики объемов выборок испытуемых

	2013 год	2014 год
Общая выборка испытуемых из числа школьников 4х классов (русский язык, математика)	1665	1163
• из них выборка по русскому языку	784	590
• из них СОШ	471	276
• из них гимназий	228	297
• из них лицеев	50	17
• из них коррекционных школ	35	0
• из них выборка по математике	881	573
• из них СОШ	561	293
• из них гимназий	235	263
• из них лицеев	49	17
• из них коррекционных школ	36	0

Из диаграмм справляемости испытуемых<sup>4</sup> с блоками заданий, теста, например, по русскому языку (рис. 1) видно, что СТЭТ зависит от качества статистических выборок. Поэтому далее при оценке СТЭТ и ДСЭТ будут указаны как качественные, так и количественные характеристики выборок испытуемых, на которых данное исследование проводилось.

<sup>4</sup> Понятие «справляемость с тестовым заданием» противоположно понятию «трудность выполнения тестового задания». Статистическая справляемость показывает какая часть испытуемых в исследуемой выборке выполнила задание правильно.

## Справляемость испытуемых с блоками элементов теста по русскому языку 2014 год эксперимента



**Рис. 1.** Выборки испытуемых из групп: средняя общеобразовательная школа, гимназия, лицей.

### Статистическая трудность элемента теста

СТЭТ будем рассчитывать по формуле статистической трудности тестового задания. Статистическая трудность задания или СТЭТ равна:

$$P = \frac{N_1}{N},$$

где:  $P$  – статистическая трудность задания,  $N_1$  – число испытуемых, правильно выполнивших задание,  $N$  – общее число испытуемых, выполнявших задание.

Согласно тестовой теории задания, имеющие статистическую трудность более 80% и менее 20%, не считаются тестовыми и выбраковываются. В таблицах 2 и 3 приведены результаты расчета СТЭТ на минимальной и полной статистических выборках испытуемых.

**Таблица 2.** Русский язык. Статистическая трудность элементов теста. Минимальная выборка испытуемых (Гимназия ХХХ). Полная статистическая выборка.<sup>5</sup>

	Номер элемента теста																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$P$	7	7	17	7	7	17	7	3	3	10	10	7	10	17	32	3	10	3	7	3
$P_o$	19	51	43	24	25	38	29	37	21	18	25	27	22	28	39	22	25	31	30	29
	Номер элемента теста																			
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
$P$	13	7	3	17	17	20	17	40	10	17	3	13	10	13	3	3	17	33	10	27
$P_o$	26	38	25	28	31	57	29	64	30	37	39	35	28	43	41	38	71	43	31	62
	Номер элемента теста																			
	41	42	43	44	45	46														
$P$	20	23	7	10	17	30														
$P_o$	47	51	28	39	34	42														

<sup>5</sup> Статистическая трудность элемента теста –  $P$  (определена по минимальной статистической гимназической выборке испытуемых). Статистическая трудность элемента теста –  $P_o$  (определена по общей статистической выборке испытуемых 2014 год эксперимента).

**Таблица 3.** Математика. Статистическая трудность элементов теста. Минимальная выборка испытуемых (Гимназия ХХХ). Полная статистическая выборка

	Номер элемента теста																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
P	22	30	7	7	11	7	26	15	7	11	7	30	0	15	30	26	11	41	26	30
P <sub>o</sub>	42	27	23	31	17	29	32	41	28	40	29	46	15	24	51	43	38	26	30	28
	Номер элемента теста																			
	21	22	23																	
P	11	11	30																	
P <sub>o</sub>	12	34	54																	

Из данных таблиц 2 и 3 следует, что при увеличении объема выборки СТЭТ увеличивается. В конечном счете, требуется повышение сложности заданий, относящихся к элементам моделей теста:

- элементы модели теста по русскому языку – 1, 10;
- элементы модели теста по математике – 5, 13, 21.

### Дифференцирующая способность элемента теста

Дифференцирующую способность элемента теста будем оценивать аналогично дифференцирующей способности задания, как меру соответствия между успешностью выполнения одного элемента теста в заданной выборке испытуемых и всего теста. Пример исследования ДСЭТ в АСИ представим на полной выборке испытуемых 2014 года (табл. 4, 5).

Количественной характеристикой ДСЭТ будем считать коэффициент дискриминации. Коэффициент дискриминации ДСЭТ равен:

$$\kappa_{\delta} = \frac{\bar{X}_i - \bar{X}}{\delta} \sqrt{\frac{N_n}{N - N_n}},$$

где:  $\bar{X}_i$  – среднее арифметическое значение баллов, полученных испытуемыми, правильно выполнившими задание  $I$  (элемент теста),  $\bar{X}$  – среднее арифметическое значение баллов, полученных испытуемыми по всему тесту,  $\delta$  – среднеквадратическое отклонение баллов, полученных испытуемыми по всему тесту,  $N_n$  – число испытуемых в данной выборке, правильно выполнивших задание (элемент теста)  $i$ .  $N$  – число всех испытуемых.

Коэффициент дискриминации тестового задания может принимать значения, находящиеся в интервале от  $-1$  до  $+1$ .

Если коэффициент дискриминации равен  $+1$ , то элемент теста имеет высокую дифференцирующую способность и позволяет различать испытуемых с высоким и низким уровнем подготовки.

Если коэффициент дискриминации равен  $-1$ , то элемент теста непригоден для теста. Если коэффициент дискриминации равен  $0$ , то элемент теста сформулирован некорректно.

Если тест содержит небольшое количество элементов или отдельные блоки, то дифференцирующую способность элемент теста оценивается по коэффициенту корреляции.

**Таблица 4.** Русский язык. Дифференцирующая способность элементов теста. Полная выборка испытуемых 2014 год<sup>6</sup>

	Номер элемента теста											
	1	2	3	4	5	6	7	8	9	10	11	12
$\kappa_{\partial}$	0,86	0,41	0	0,68	0,54	0,82	0,61	0,1	0,77	0,72	0,55	0,48
	Номер элемента теста											
	13	14	15	16	17	18	19	20	21	22	23	24
$\kappa_{\partial}$	0,11	-0,64	0,29	0,77	0,44	0,77	0,67	0	0,48	-0,07	0,58	0,56
	Номер элемента теста											
	25	26	27	28	29	30	31	32	33	34	35	36
$\kappa_{\partial}$	0,47	0	0,52	0,39	0	0	0,96	0,38	0,06	-0,1	0,1	0,54
	Номер элемента теста											
	37	38	39	40	41	42	43	44	45	46		
$\kappa_{\partial}$	0,09	0,36	0,67	0,34	0,55	0,4	0,88	0,5	0,21	-0,03		

**Таблица 5.** Математика. Дифференцирующая способность элементов теста. Полная выборка испытуемых 2014

	Номер элемента теста											
	1	2	3	4	5	6	7	8	9	10	11	12
$\kappa_{\partial}$	0,28	0,39	0,23	0,6	0,32	0,39	0,41	0,25	0,44	0,19	0,52	0,91
	Номер элемента теста											
	13	14	15	16	17	18	19	20	21	22	23	
$\kappa_{\partial}$	0,1	0,23	0,68	0,44	0,71	0,62	0,57	0,28	0,39	0,61	0,19	0,39

Анализ данных таблиц 4 и 5 показывает, что:

- в тесте по русскому следует обратить внимание на формулировки заданий, относящиеся к элементам модели теста: 3,20,26, 29,30;
- тестовые задания, требующие полной переработки, относятся к элементам: 14, 22, 34, 46;
- тестовые задания, относящиеся ко всем элементам модели теста по математике, соответствуют требованиям к их дифференцирующей способности.

В целом, следует отметить, что в настоящий момент АСИ обученности школьников в области русского языка и математики (степень начального общего образования) содержит необходимую и достаточную базу тестовых заданий, имеющих статистические характеристики, соответствующие требованиям тестовой теории. Русский язык – более 1380 заданий. Математика – более 690 заданий. Задания, из которых формируются элементы теста в АСИ, исследуются на предмет оценки их СТЭТ и ДСЭТ с 2011 года, и на основании полученных результатов можно утверждать, что их можно использовать для формирования метрологически пригодных средств педагогического измерения.

<sup>6</sup> Русский язык. Полная выборка испытуемых – 590 учащихся.  
Математика. Полная выборка испытуемых – 573 учащихся

## Исследование системных свойств тестов

Исследование системных свойств тестов, используемых в АСИ проводится на основе данных сравнительного анализ результатов независимого тестирования и результатов экспертной оценки грамотности школьников в предметных областях. К системным свойствам относится надежность шкалы средства измерения, которая оценивается по трем критериям<sup>7</sup>: обоснованность, устойчивость, точность.

Обоснованность (валидность) – это статистическая характеристика средства измерения, понимаемая как его способность измерять определенное заданное свойство или признак, не смешивая его с другими.

Контроль средств измерения на обоснованность его оценочной шкалы проводится в четыре шага.

*Шаг первый* – экспертиза кодификатора на его соответствие содержанию учебных программ и ФГОС в выделенной предметной области.

*Шаг второй* – экспертиза тестовых заданий на соответствие структурной единице кодификатора.

*Шаг третий* – проведение пробного тестирования минимальной статистической выборки испытуемых (30 человек из числа выпускников начальной школы), и сопоставление полученных данных с результатами итоговой аттестации за курс начальной школы в соответствующей предметной области.

*Шаг четвертый* – расчет коэффициента корреляции между двумя рядами случайных величин:  $X_i$  (ряд оценок, полученных в выборке испытуемых за тест) и  $Y_i$  (ряд итоговых оценок, полученных в той же выборке испытуемых).

Исследуем два ряда случайных величин. Первый – данные итоговой успеваемости учащихся начальной школы Гимназии № XXX в предметных областях «Русский язык» и «Математика» 2014 год эксперимента. Второй – результаты тестирования испытуемых в АСИ этих же учащихся (табл. 6).

Определим, существует ли корреляционная зависимость между переменными величинами в случае парной зависимости, используя коэффициент корреляции Пирсона<sup>8</sup>

$$r_{xy} = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\sqrt{\overline{X^2} - (\bar{X})^2} \cdot \sqrt{\overline{Y^2} - (\bar{Y})^2}} \quad 3$$

Значение  $r$  всегда заключено в пределах  $-1 < r < +1$ .

Если  $r = \mp 1$ , то между рассматриваемыми величинами существует прямая связь в первом случае и обратная – во втором. Если  $r = 0$ , то это значит, что между рассматриваемыми величинами не существует ни функциональной, ни корреляционной обратной связи, но криволинейная корреляционная связь может существовать. Чем ближе  $r \rightarrow +1$  или  $r \rightarrow -1$ , тем точнее и теснее корреляционная связь между величинами  $X$  и  $Y$ .

<sup>7</sup> Методы системного педагогического исследования. / Сб. ст. – М. : Изд-во Народное образование, 2002.

<sup>8</sup> Г.В. Горелова, И.А. Кацко. Теория вероятности и математическая статистика в примерах и задачах с применением Excel – Изд.4-е, – Ростов н/Д : Феникс, 2006. – 475, [1] с.: ил. – (Высшее образование).

При:  $r = 0 \div 0,2$  - связь незначительная,  $r = \pm 0,2 \div \pm 0,40$  - низкая степень связи,  $r = \pm 0,40 \div 0,70$  - ясно выраженная корреляция,  $r = \pm 0,70 \div 1,0$  - высокая или очень высокая степень корреляции.

**Таблица 1.** Сравнение итоговой оценки и результатов тестирования. Гимназия № XXX

Испытуемый	Русский язык					Математика				
	Итоговая		Тест		$X_i Y_i$	Итоговая		Тест		$X_i Y_i$
	$X_i$	$X_i^2$	$Y_i$	$Y_i^2$		$X_i$	$X_i^2$	$Y_i$	$Y_i^2$	
Школьник 1	4	16	5	25	20	5	25	5	25	25
Школьник 2	5	25	5	25	25	5	25	5	25	25
Школьник 3	4	16	5	25	20	4	16	5	25	20
Школьник 4	4	16	3	9	12	4	16	4	16	16
Школьник 5	4	16	5	25	20	4	16	5	25	20
Школьник 6	4	16	4	16	16	4	16	5	25	20
Школьник 7	4	16	5	25	20	4	16	4	16	16
Школьник 8	4	16	5	25	20	4	16	5	25	20
Школьник 9	4	16	5	25	20	5	25	5	25	25
Школьник 10	5	25	5	25	25	4	16	3	9	12
Школьник 11	5	25	5	25	25	4	16	5	25	20
Школьник 12	5	25	5	25	25	5	25	5	25	25
Школьник 13	4	16	3	9	12	4	16	5	25	20
Школьник 14	5	25	5	25	25	5	25	5	25	25
Школьник 15	4	16	5	25	20	5	25	4	16	20
Школьник 16	5	25	5	25	25	5	25	5	25	25
Школьник 17	4	16	3	9	12	4	16	4	16	16
Школьник 18	4	16	5	25	20	4	16	5	25	20
Школьник 19	4	16	5	25	20	4	16	5	25	20
Школьник 20	5	25	4	16	20	5	25	5	25	25
Школьник 21	4	16	4	16	16	5	25	5	25	25
Школьник 22	4	16	2	4	8	5	25	5	25	25
Школьник 23	4	16	5	25	20	4	16	4	16	16
Школьник 24	5	25	5	25	25	5	25	5	25	25
Школьник 25	4	16	5	25	20	4	16	4	16	16
Школьник 26	4	16	5	25	20	5	25	3	9	15
Школьник 27	4	16	5	25	20	5	25	5	25	25
Школьник 28	5	25	5	25	25	5	25	5	25	25
Школьник 29	5	25	5	25	25	5	25	4	16	20
Школьник 30	4	16	5	25	20	4	16	4	16	16
	$\bar{X}$	$\bar{X}^2$	$\bar{Y}$	$\bar{Y}^2$	$\overline{XY}$	$\bar{X}$	$\bar{X}^2$	$\bar{Y}$	$\bar{Y}^2$	$\overline{XY}$
	4,33	19	4,6	21,8	20,03	4,5	20,5	4,6	21,53	20,77
$r_{XY}$	<b>0,265165</b> (положительная корреляция)					<b>0,218218</b> (положительная корреляция)				

Анализ таблицы 6 показывает, что между выборочными значениями  $X_i$  и  $Y_i$  существует слабая взаимозависимость, следовательно, можно утверждать, что данные, полученные по шкалам экспертов, слабо сопоставляются с данными тестирования. Вместе с тем, слабая взаимозависимость не является демонстрацией слабой валидности шкалы в АСИ по отношению к исследуемому признаку, так как также можно предположить, что в итоговой оценке экспертов наряду с грамотностью содержатся другие признаки обученности выпускников начальной школы. В связи с выявленным фактом возникает новая задача исследования, связанная с выявлением влияния одного фактора (оценки за итоговый тест) на другой (итоговой экспертной оценки). Эта задача может быть решена методом дисперсионного анализа.

Устойчивость шкалы средства измерения определяется как однозначность данных, полученных при их использовании со значительным временным промежутком (корреляция между первой и второй серией измерения должна быть высокой – 0,9).

Исследование средств измерения в АСИ на устойчивость шкалы проводятся повторным их испытанием на одной и той же выборке испытуемых (табл. 7). Первое испытание – окончание обучения (4 класс). Второе испытание – начало обучения на новой ступени (5 класс).

**Таблица 7.** Сравнение результатов тестирования 4 класс и 5 класс. Гимназия № XXX

Испытуемый	Русский язык					Математика				
	Тест 5 класс		Тест 4 класс		$X_i Y_i$	Тест 5 класс		Тест 4 класс		$X_i Y_i$
	$X_i$	$X_i^2$	$Y_i$	$Y_i^2$		$X_i$	$X_i^2$	$Y_i$	$Y_i^2$	
Школьник 1	5	16	5	25	20	4	25	5	25	25
Школьник 2	5	25	5	25	25	5	25	5	25	25
Школьник 3	4	16	5	25	20	5	16	5	25	20
Школьник 4	3	16	3	9	12	4	16	4	16	16
Школьник 5	5	16	5	25	20	5	16	5	25	20
Школьник 6	4	16	4	16	16	5	16	5	25	20
Школьник 7	5	16	5	25	20	3	16	4	16	16
Школьник 8	4	16	5	25	20	5	16	5	25	20
Школьник 9	5	16	5	25	20	5	25	5	25	25
Школьник 10	5	25	5	25	25	4	16	3	9	12
Школьник 11	5	25	5	25	25	5	16	5	25	20
Школьник 12	5	25	5	25	25	5	25	5	25	25
Школьник 13	3	16	3	9	12	4	16	5	25	20
Школьник 14	5	25	5	25	25	5	25	5	25	25
Школьник 15	4	16	5	25	20	4	25	4	16	20
Школьник 16	5	25	5	25	25	5	25	5	25	25
Школьник 17	2	16	3	9	12	4	16	4	16	16
Школьник 18	4	16	5	25	20	5	16	5	25	20

Испытуемый	Русский язык					Математика				
	Тест 5 класс		Тест 4 класс		$X_i Y_i$	Тест 5 класс		Тест 4 класс		$X_i Y_i$
	$X_i$	$X_i^2$	$Y_i$	$Y_i^2$		$X_i$	$X_i^2$	$Y_i$	$Y_i^2$	
Школьник 19	5	16	5	25	20	5	16	5	25	20
Школьник 20	5	25	4	16	20	5	25	5	25	25
Школьник 21	4	16	4	16	16	4	25	5	25	25
Школьник 22	3	16	2	4	8	4	25	5	25	25
Школьник 23	5	16	5	25	20	4	16	4	16	16
Школьник 24	5	25	5	25	25	5	25	5	25	25
Школьник 25	5	16	5	25	20	4	16	4	16	16
Школьник 26	4	16	5	25	20	3	25	3	9	15
Школьник 27	5	16	5	25	20	4	25	5	25	25
Школьник 28	5	25	5	25	25	5	25	5	25	25
Школьник 29	5	25	5	25	25	3	25	4	16	20
Школьник 30	5	16	5	25	20	4	16	4	16	16
	$\bar{X}$	$\bar{X}^2$	$\bar{Y}$	$\bar{Y}^2$	$\overline{XY}$	$\bar{X}$	$\bar{X}^2$	$\bar{Y}$	$\bar{Y}^2$	$\overline{XY}$
	4,47	20,6	4,6	21,8	21,07	4,4	19,8	4,6	21,53	20,53
$r_{XY}$	<b>0,806916</b> (высокая положительная корреляция)					<b>0,723747</b> (высокая положительная корреляция)				

Анализ таблицы 7 показывает, что между выборочными значениями  $X_i$  и  $Y_i$  существует сильная взаимозависимость, следовательно, можно утверждать, что шкалы средств измерения, используемые в АСИ, характеризуются высокой устойчивостью.

Точность средства измерения определяется из оптимального соотношения между чувствительностью измеряемого объекта, и устойчивостью данных.

Оптимальность достигается экспериментально с помощью увеличения числа пунктов шкалы (числа элементов содержания в эталонном кодификаторе) и проверки шкалы на устойчивость.