ПЕРВАЯ НАЦИОНАЛЬНАЯ КОНФЕРЕНЦИЯ
«ОЦЕНКА СОЦИАЛЬНЫХ ПРОЕКТОВ И ПРОГРАММ В РЕСПУБЛИКЕ КАЗАХСТАН»

Опыт проведения Оценки воздействия (Impact Evaluation) проекта Бюджетная грамотность в трех регионах Российской Федерации (Russia Budget Literacy Project)»

Тренер: Сергей Гуляев

г.Астана, 19 ноября 2018г.

DECENTA®

Что такое оценка воздействия/влияния (Impact Evaluation)?

Оценка воздействия:

Это особый тип программной оценки, который стремится ответить на конкретный вопрос причинно-следственного характера: каково воздействие (или причинный эффект) программы на результат?

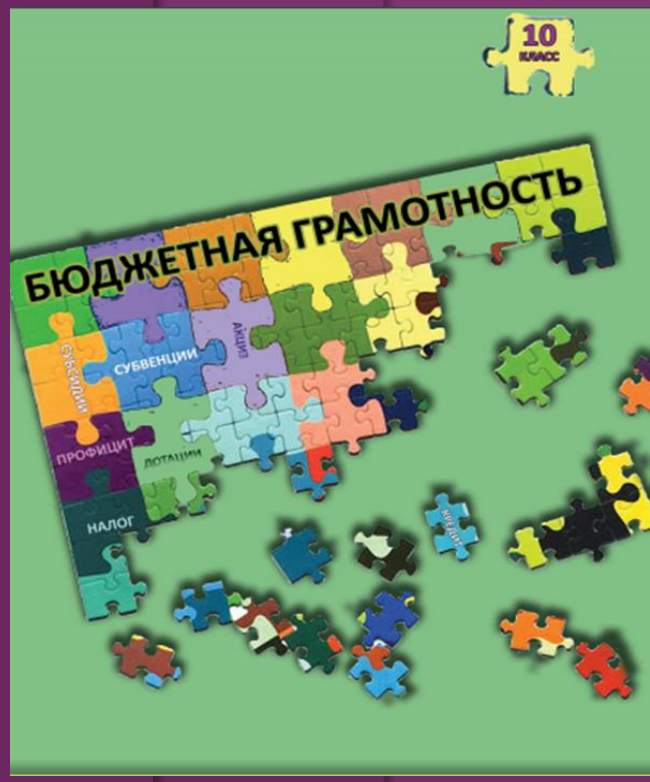
Оценка воздействия помогает нам преодолеть проблему установления причинно-следственной связи путем эмпирического определения того, в какой степени конкретная программа - и только эта программа - внесла вклад и оказала воздействие на ключевые компоненты, определяет степень, в которой программа - и одна только программа - вызывает изменение результата.

Цели данной оценки:

Оценка результатов обучения по Бюджетной грамотности, пилотируемой в трех регионах России, и выработка рекомендаций по ее улучшению.

Результат оценки должен был показать степень, до которой оцениваемый курс смог оказать влияние на учащихся, на их знания и навыки о бюджете, а также достижения двух основных образовательных целей Курса «Бюджетная грамотность»:

- понимать систему государственного бюджета;
- понимать свою личную роль в бюджетном процессе.



Основные этапы оценки воздействия

- ОВ состоит, как правило, из двух этапов или двух замеров состояния предмета оценки «до» вмешательства программы, а также «после».
- Целью исследования на 1-м этапе было выявить начальный уровень знаний у школьников, их осведомленность (на тот момент) о бюджете до изучения курса «Бюджетная грамотность».
- На 2-м этапе цель состояла в выявлении уровня знаний у контрольных групп после участия в курсе, а у сравнительных групп – на момент проведения тестирования.

Задачи исследования:

- Собрать количественные данные в виде заполненных тестовых заданий.
- По количеству правильных ответов определить степень сложности тестовых заданий, а также умение респондентов проводить анализ.
- Собрать и подытожить качественные данные посредством фокус-групп.
- Проанализировать базовый уровень знаний в части бюджетной грамотности у учащихся пилотных регионов до начала изучения курса.
- Сопоставить, насколько схожи знания и навыки о бюджетной грамотности у экспериментальных и сравнительных групп до начала изучения курса.
- Провести итоговое тестирование для тех же школ и учащихся, что и на начальном уровне, оценить воздействие курса «Бюджетная грамотность» в школах на основе результатов тестирования, сравнив изменения результатов у экспериментальных и сравнительных групп.

Информационная база исследования

- Период проведения: сентябрь 2016 – февраль 2017.
- Выборка: N (первый этап) = 474, N (второй этап) = 650; 12 школ по 4 из каждого пилотного региона.
- Регионы: Алтайский край, Пермский край, Республика Башкортостан.

Методология. Дизайн оценки.

Квази-эксперимент

ИЛИ

Чистый/истинный эксперимент

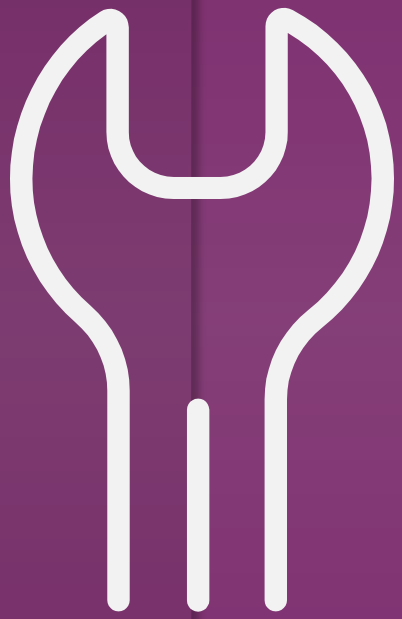
?

Был выбран квази-экспериментальный дизайн оценки

- Невозможно было случайным образом выбрать отдельных учащихся;
- Случайным образом выбирались целые классы в одной школе, и, если в этой школе был только один 10 класс, необходимо было подобрать десятиклассников в другой школе, которые бы выступили в качестве сравнительного класса.
- Выбирая целые классы, была сильная вероятность того, что будет разное количество учащихся в классах, уровень, а также гендерный состав будут отличаться.

Переменные, которые можно было контролировать:

- Показатели по успеваемости;
- Географическое положение (городская, сельская или пригородная местность);
- Экспериментальные и сравнительные классы имели аналогичную специализацию (тип образовательного учреждения) – общеобразовательная школа, лицей или гимназия.



Гипотезы оценки

Сравнивались две гипотезы: Нулевая гипотеза (H0), согласно которой различие между полученными результатами является недостаточно существенным, соответственно, обучение в рамках курса не оказало никакого влияния на полученный результат; Альтернативная гипотеза (HA), которая заключалась в том, что различия в результатах будут значительными, что обусловлено влиянием курса.

Выборка

Была сформирована выборка из 12 школ, по 4 школы в каждом регионе. Далее в каждом из трех пилотных регионов (Алтайский край, Пермский край, Республика Башкортостан) были отобраны группы экспериментальные и группы сравнения с максимально схожими признаками, такими как тип населенного пункта, тип школы, профиль класса, успеваемость.

Методы сбора информации:

- Работа с документами (обзор данных административно-управленческого характера, имеющих отношение к проекту, таких как: проектная заявка, заявки участников проекта, финансовая информация, описание школ, отчеты о реализации проекта и т.д.);
- Тестирование учащихся (входное и выходное);
- Проведение фокус-групп.

Работа с инструментарием

- Необходимо разработать тестовое задание
- Согласовать количество вопросов (уложиться в часть (!) урочного времени)
- Разработать вопросы теста и дистракторы
- Вопросы теста должны: проверить уровень знаний, умение аргументировано (!) отстаивать свою точку зрения, оценить возможности применения знаний – умения и навыки (т.е. включить в тест практические задачи)
- Разработать гайд для проведения фокус-группы

Апробация инструментария

Перед тем, как приступить к полномасштабному исследованию, была проведена пилотная апробация инструментария в Алтайском крае для проверки подхода, примененного для выборки, а также обоснованности инструментов исследования.

В качестве школы для проведения апробации выступил КГБУ «Алтайский краевой педагогический лицей».

Обработка данных

Для проверки гипотезы статистического исследования была составлена дезагрегированная форма, в которой отражены все респонденты (полный список по всем регионам и образовательным учреждениям). Данная форма (список) или выборочная совокупность была трансформирована в интервальный ряд распределения, где респонденты были разбиты на интервальные группы. Для подтверждения нормальности распределения респондентов по группам (интервалам) была использована статистическая компьютерная программа SPSS. Далее выполнялся расчет средних показателей, ошибки и отклонения.

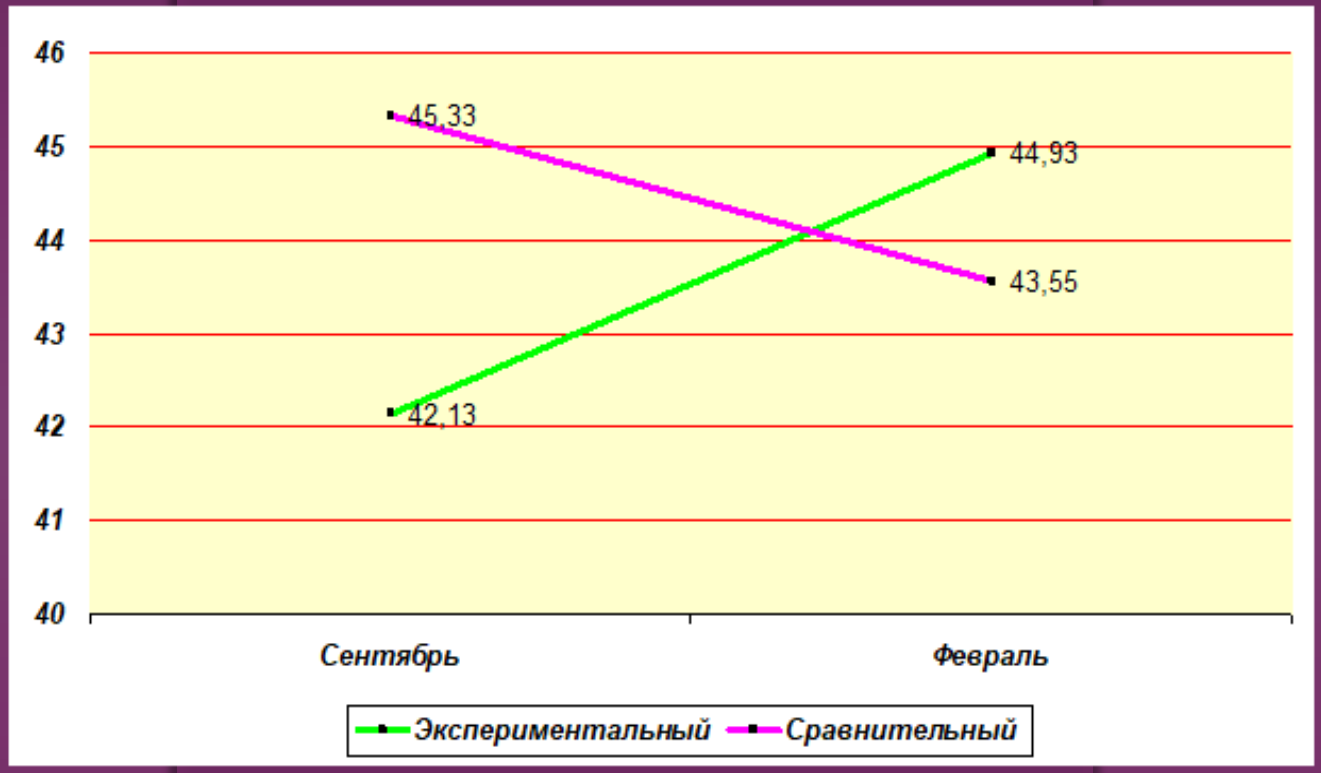
Стратегия оценки

- Первоначально планировалось использовать метод парного t-критерия, либо ANOVA. Но от использования этих методов пришлось отказаться. Применение дисперсионного анализа в данном случае для подтверждения гипотезы нецелесообразно вследствие значительного количества факторов и небольшой длины интервалов. Также в данном случае показатели стандартных ошибок и отклонений незначительны.
- Использование парного t-критерия также оказалось затруднительным вследствие невозможности обеспечить сопоставимую парность.
- В качестве методов оценки, прежде всего, использовался метод простого сравнения «до изучения курса» и «после изучения», хотя его, конечно же, было недостаточно, поэтому был применен метод Разности в Различиях, или Двойная разница.

Применение метода двойной разницы

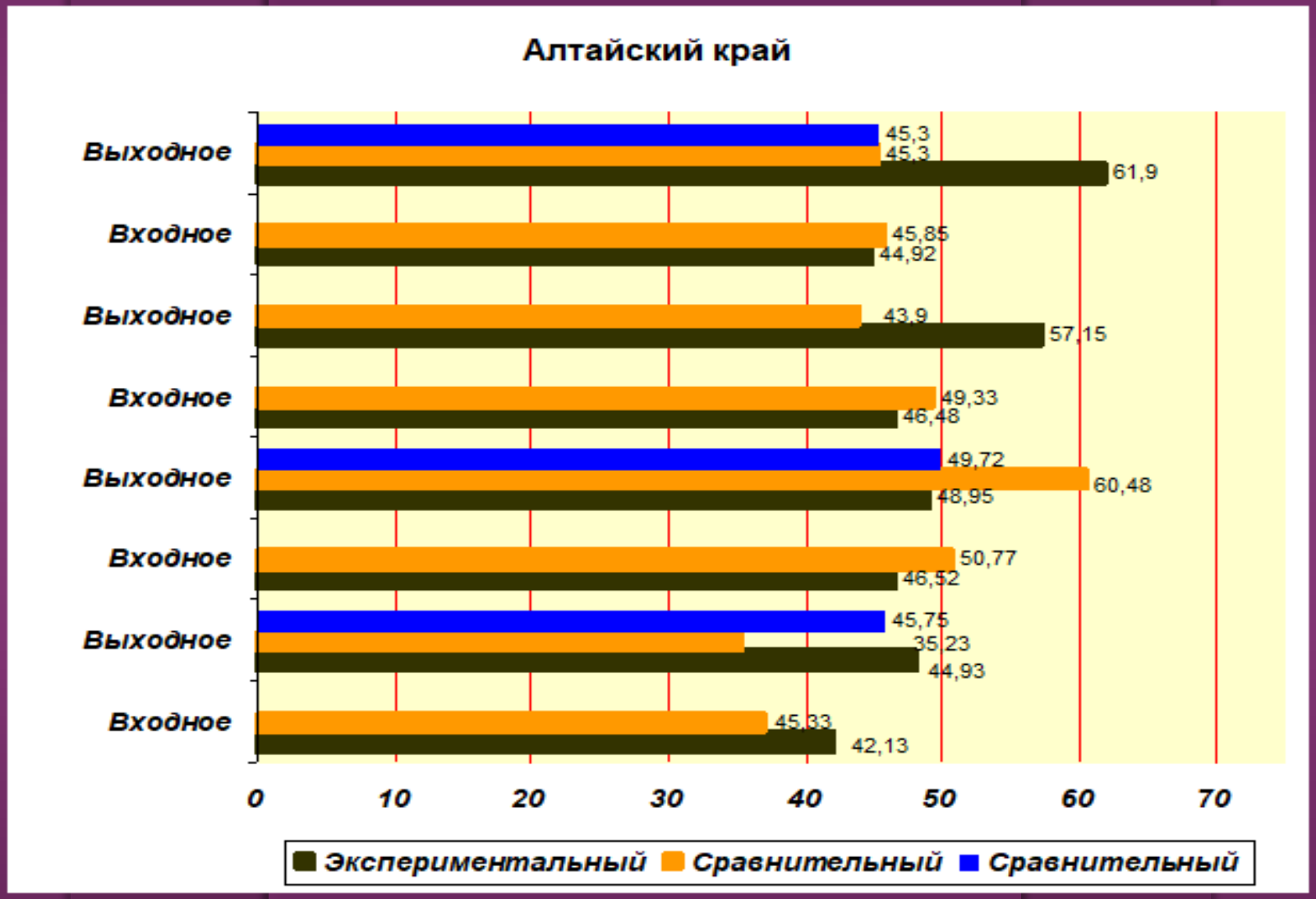
- В данном случае первая разница – это отличие в результатах экспериментальной группы (разница результатов: балл выходного тестирования минус балл входного тестирования).
- Вторая разница – это различие результатов у сравнительной группы.
- Оценка воздействия с применением метода двойной разницы вычисляется следующим образом:
- 1) Рассчитывается разница в результатах между ситуацией до и после программы для экспериментальной группы.
- 2) Рассчитывается разница в результатах между ситуацией до и после для сравнительной группы.
- 3) Рассчитывается разница между разницей в результатах экспериментальной группы и разницей у сравнительной группы, это и есть двойная разница, или разность разностей. Эта двойная разница и есть оценка воздействия в нашем случае.

Пример: Результаты входного и выходного тестирования в экспериментальных и сравнительных группах, Алтайский край



Таким образом, эффект при применении двойной разницы равняется: (44,93 – 43,55) – (42,13 – 45,33) = 1,38 – (-3,2) = 4,58.

Результаты входного и выходного тестирования в экспериментальных и сравнительных группах, Алтайский край.



Результаты

Результаты чётко показывают, что, во-первых, статистически значимой разницы между группами во время предварительного тестирования не наблюдалось. Во-вторых, существует статистически значимая разница между результатами теста экспериментальной и сравнительной групп после прохождения курса, а также между результатами предварительного теста и теста после прохождения курса экспериментальной группы.

Статистика

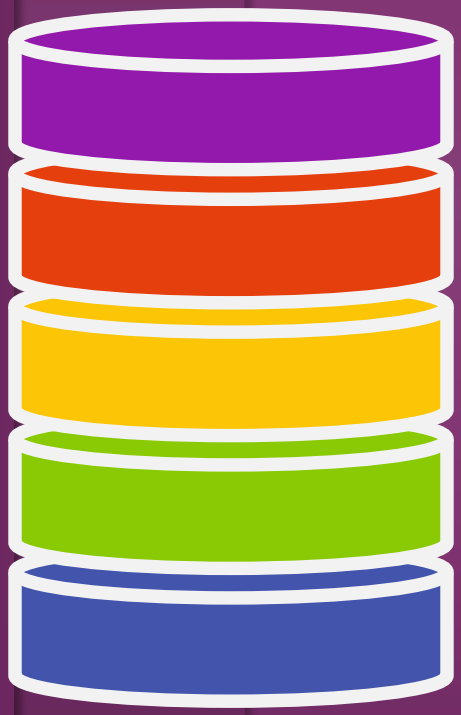
- Р-Значение и статистическая значимость:

Двустороннее Р-значение равно 0,4177, что статистически неважно.

- Среднее значение разницы - сравнительная группа минус экспериментальная, равно

-0.7244

- Доверительный интервал 95% для этой разницы: от -2,4794 до 1,0306
- t = 0.8111
- df = 472
- Стандартная ошибка разницы = 0.893



Результаты t-теста. Предварительное тестирование: Сравнительная группа – Экспериментальная группа.

Группа	**Сравнительная**	**Экспериментальная**
Среднее	50.1131	50.8375
SD	9.4775	9.9151
SEM	0.6318	0.6283
Значение	225	249

Степень эффекта

- d Коэна = (50.8375 - 50.1131)/9.698768 = 0.07469 (эффект отсутствует)
- delta Гейтса = (50.8375 - 50.1131)/9.4775 = 0.076434 (эффект отсутствует)
- g Хеджеса= (50.8375 - 50.1131)/9.709885 = 0.074604 (эффект отсутствует)

Подходы при обработке данных

D Коэна – Степень эффекта

Степень эффекта - это простой способ количественной оценки разницы между двумя группами. Степень эффекта подчеркивает именно степень различия, а не подменяет это размером выборки. D Коэна – это подходящий способ измерить степень воздействия/эффекта, если две группы имеют схожие стандартные отклонения и имеют схожий размер.

Delta Гейтса, которая использует только стандартные отклонения контрольной группы, является альтернативной мерой, если каждая группа имеет различные стандартные отклонения.

G Хеджеса, которая обеспечивает измерение степени эффекта, взвешенную в соответствии с относительным размером каждой из групп/выборок, является альтернативой для случаев, в которых имеются различные размеры выборки.

SEM – стандартная ошибка среднего

SD – стандартное отклонение

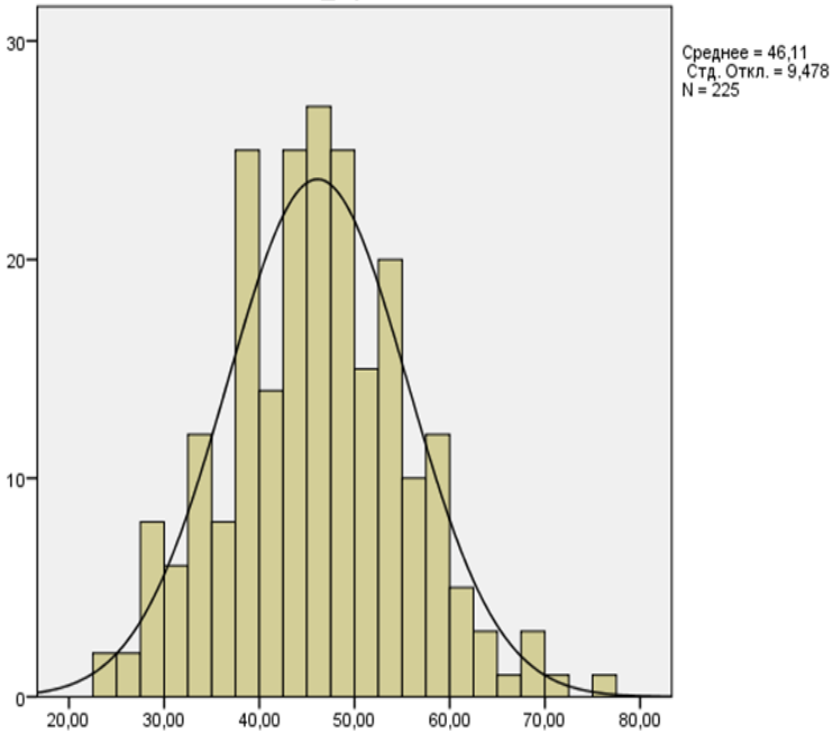
Регрессионный анализ

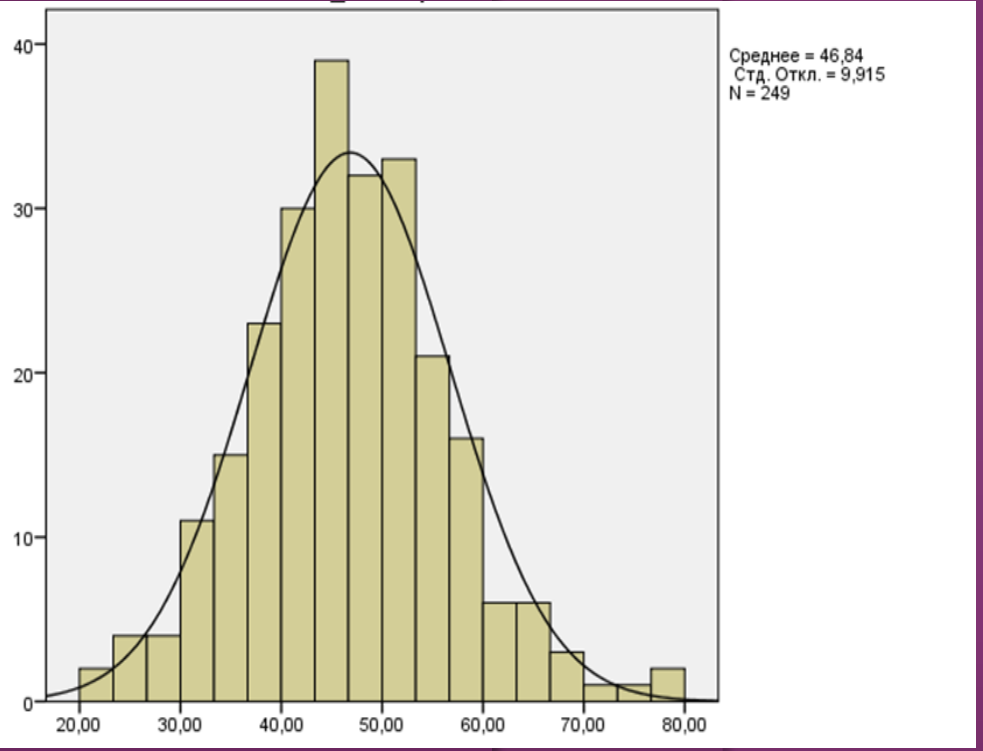
	Входная экспериментальная	Выходная экспериментальная
Экспериментальный класс - группа	0,049 (коэффициент бета) (0,427) (значимость)	-0,036 (0,569)
Экспериментальный класс - Регион	-0,037 (коэффициент бета) (0,635) (значимость)	0,032 (0,679)
Экспериментальный класс – Классификация населенного пункта	-0,198 (коэффициент бета) (0,009) (значимость)	-0,191 (0,012)
Экспериментальный класс – Профиль класса	-0,102 (коэффициент бета) (0,174) (значимость)	-0,195 (0,003)
R- квадрат	0,101	0,063
Количество учеников	474	650
Средняя зависимая переменная сравнительный класс	46,1131	45,6921
Стандартное отклонение зависимой переменной сравнительный класс	9,4775	10,1344

Наблюдения и выводы оценки:

Базовый уровень знаний по бюджетной грамотности оказался практически одинаковым.

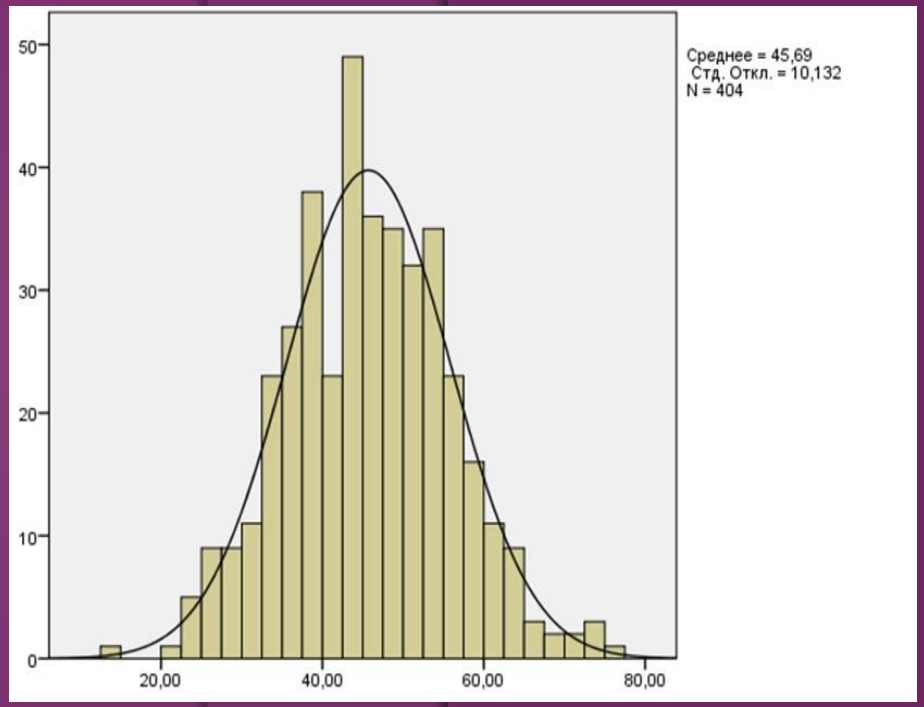
Результаты входного тестирования: Сравнительная группа - 46,11; Экспериментальная – 46,84.

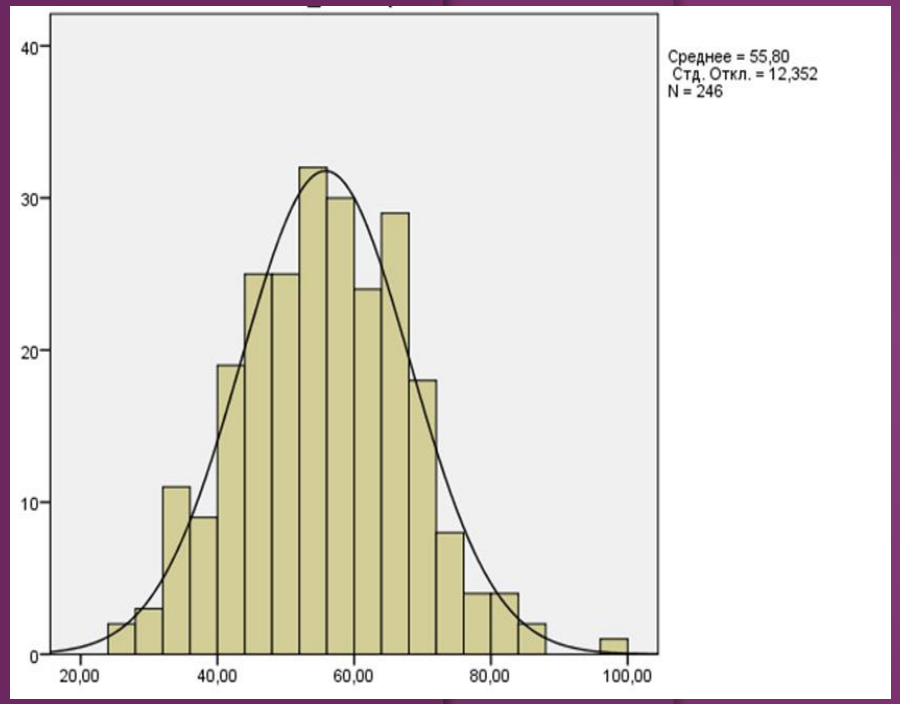




По окончании курса «Бюджетная грамотность» в экспериментальных группах наблюдается значительный эффект. Результаты выходного тестирования: Сравнительная группа – 45,69; Экспериментальная – 55,80.

.





Все экспериментальные классы (12 классов во всех 3 пилотных регионах) при итоговом тестировании показали более высокий результат в сравнении с предварительным тестированием.

При обработке данных были получены следующие результаты при заключительном тестировании:

Группа	Сравнительная	Экспериментальная
Среднее	45.6921	55.7967
SD	10.1344	12.3521
SEM	0.5042	0.7875
Значение	404	246

При сравнении результатов экспериментальных групп до и после изучения курса наблюдается значительный эффект. Данный эффект с большой долей вероятности объясняется воздействием курса «Бюджетная грамотность».

Группа	Экспериментальная до	Экспериментальная после
Среднее	46.8375	55.7967
SD	9.9151	12.3521
SEM	0.6283	0.7875
Значение	249	246

Результаты регрессионного анализа могут говорить о том, что влияние других факторов было не существенным.

	Входная экспериментальная	Выходная экспериментальная
Экспериментальный класс - группа	0,049 (коэффициент бета) (0,427) (значимость)	-0,036 (0,569)
Экспериментальный класс - Регион	-0,037 (коэффициент бета) (0,635) (значимость)	0,032 (0,679)
Экспериментальный класс – Классификация населенного пункта	-0,198 (коэффициент бета) (0,009) (значимость)	-0,191 (0,012)
Экспериментальный класс – Профиль класса	-0,102 (коэффициент бета) (0,174) (значимость)	-0,195 (0,003)
R- квадрат	0,101	0,063
Количество учеников	474	650
Средняя зависимая переменная сравнительный класс	46,1131	45,6921
Стандартное отклонение зависимой переменной сравнительный класс	9,4775	10,1344

Статистическая значимость воздействия является значительной (двустороннее Р-значение равно менее 0.0001, что статистически важно. Среднее значение разницы между сравнительной группой и экспериментальной группой равно -10.1046. Доверительный интервал 95% для этой разницы: от -11.8564 до -8.3528).

Выводы (продолжение)

- Экспериментальные классы при проведении выходного тестирования чаще пользовались своими знаниями, нежели просто угадывали правильный ответ.
- Умение применять знания из сферы бюджетной грамотности, а также аргументированно отстаивать свою позицию у сравнительных и экспериментальных групп находится примерно на одном уровне.
- Однозначного мнения о том, насколько необходимо преподавать «Бюджетную грамотность» в виде отдельного курса, либо сделать его интегрированной частью других предметов, не сложилось.
- Внедрение курса в пилотном режиме состоялось. Учебно-методические материалы курса получили положительную оценку.

Рекомендации

1. Данный Курс готов к внедрению с учетом замечаний, собранных из разных источников во время пилотирования (координаторы, учителя, школы, учащиеся). Курс мог бы стать интегрированной частью какого-то другого курса или предмета, например, Экономики или Финансовой грамотности.
2. В случае, если курс останется самостоятельным, то его рекомендовалось бы сделать более продолжительным, рассчитанным на весь учебный год.
3. Важно решить вопрос с мотивацией преподавателей.
4. Использовать в учебно-методических материалах активные ссылки на Интернет-ресурсы, риски частых обновлений, модернизаций, реконструкций которых будут минимальными.
5. Использовать данный курс в качестве элективного для непрофильных классов.
6. Возможно, стоит повторно сопоставить содержание курса «Бюджетная грамотность» с учебной программой, с целью избежать дублирования с содержанием других предметов.

Еще больше интересной информации на

www.decenta.org