

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ  
ПСИХОЛОГО-ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ»

М.Г. Сорокова, О.А. Ульянина, Г.В. Семья, О.И. Леонова,  
Д.В. Лубовский, Е.И. Исаев, Т.Г. Подушкина,  
Н.П. Бусыгина, Н.П. Радчикова, А.А. Шведовская

# **ДОКАЗАТЕЛЬНЫЙ ПОДХОД:** **Руководство по верификации** **программ, технологий, практик** **в образовании и социальной сфере**



Учебное пособие

Москва  
2024

**Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ  
ПСИХОЛОГО-ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ»**

**М.Г. Сорокова, О.А. Ульянина, Г.В. Семья,  
О.И. Леонова, Д.В. Лубовский, Е.И. Исаев, Т.Г. Подушкина,  
Н.П. Бусыгина, Н.П. Радчикова, А.А. Шведовская**

**Доказательный подход:  
Руководство по верификации  
программ, технологий, практик  
в образовании и социальной сфере**

**Учебное пособие**

**Москва  
2024**

УДК 378  
ББК 74  
Д63

Рецензенты:  
*Нечаев Н.Н., Микляева А.В.*

Под редакцией:  
*Марголис А.А., Сорокова М.Г., Семья Г.В.*

Авторы и разработчики:  
*Сорокова М.Г., Ульянина О.А., Семья Г.В., Леонова О.И., Лубовский Д.В.,  
Исаев Е.И., Подушкина Т.Г., Бусыгина Н.П., Радчикова Н.П., Шведовская А.А.*

Рекомендовано Ученым советом Московского государственного психолого-педагогического университета для студентов, обучающихся по направлениям и специальностям подготовки УГСН 37.00.00 «Психологические науки», УГСН 44.00.00 «Образование и педагогические науки».

**Сорокова М.Г., Ульянина О.А., Семья Г.В., Леонова О.И., Лубовский Д.В., Исаев Е.И., Подушкина Т.Г., Бусыгина Н.П., Радчикова Н.П., Шведовская А.А.**  
Д63 Доказательный подход: Руководство по верификации программ, технологий, практик в образовании и социальной сфере : учебное пособие / М.Г. Сорокова, О.А. Ульянина, Г.В. Семья, О.И. Леонова, Д.В. Лубовский, Е.И. Исаев, Т.Г. Подушкина, Н.П. Бусыгина, Н.П. Радчикова, А.А. Шведовская ; под ред. Марголиса А.А., Сороковой М.Г., Семья Г.В. — М. : ФГБОУ ВО МГППУ, 2024. — 130 с. ISBN 978-5-94051-333-9

*В учебном пособии представлены рамочные условия организации научной экспертизы и исследований по верификации психолого-педагогических программ и технологий в парадигме доказательного подхода. Раскрыты основные положения «Национального стандарта доказательности практик в сфере детства» как методологической основы для описания и анализа социальных практик на основе доказательного подхода. Предложена система критериев оценки качественного ассесмента — кандидатных программ и технологий с точки зрения их структуры, содержания и нормативной базы. Разработаны стандарты и общие требования к проведению эксперимента по верификации психолого-педагогических программ и технологий и критерии оценки его результатов на основе традиций отечественной науки с привлечением международного опыта. Показаны предпосылки возникновения и развития доказательного подхода в России и за рубежом. Представлены кейсы доказательных практик. Адресовано студентам направлений и специальностей подготовки УГСН 37.00.00 «Психологические науки», УГСН 44.00.00 «Образование и педагогические науки», а также широкому кругу авторов-разработчиков программ, технологий, практик, исследователей и практиков в сфере психологии и образования.*

УДК 378  
ББК 74

ISBN 978-5-94051-333-9

© ФГБОУ ВО «Московский государственный психолого-педагогический университет», 2024

## Оглавление

<b>Введение</b> .....	7
Актуальность и цели Руководства .....	7
Национальный стандарт доказательности практик в сфере детства: общая характеристика.....	8
Доказательный подход как основа научной верификации психолого-педагогических, образовательных и социальных программ и технологий .....	10
Стандартизированные психодиагностические инструменты, рекомендованные для измерения психолого-педагогических характеристик.....	14
Организация независимой оценки психолого-педагогических программ и стандартизации психодиагностических инструментов.....	15
<b>Глава 1. Теоретико-методологические основы и современные исследования в парадигме доказательного подхода в психологии и образовании</b> .....	17
1.1. Национальный стандарт доказательности как методологическая основа для описания и анализа социальных практик в парадигме доказательного подхода .....	17
1.2. Современные подходы к сбору и анализу данных при формировании доказательной базы социальных практик.....	21
1.3. Теоретико-методологические предпосылки доказательного подхода в психологии и образовании в трудах классиков российской психологии .....	24
1.4. Предпосылки возникновения и развития доказательного подхода в социальной сфере за рубежом .....	32
1.5. Современные прикладные экспериментальные исследования как примеры доказательного подхода в проектировании психолого-педагогических и образовательных программ, технологий, практик ....	36
<i>Кейс 1.</i> Верификация доказательной базы технологии индивидуализации обучения с классической схемой эксперимента и количественными методами.....	36
<i>Кейс 2.</i> Верификация доказательной базы просветительской программы с классической схемой эксперимента и количественными методами.....	41
<i>Кейс 3.</i> Верификация доказательной базы развивающей программы с классической схемой эксперимента при сочетании количественных и качественных методов .....	42

<i>Кейс 4.</i> Верификация доказательной базы деятельностной технологии при сочетании количественных и качественных методов .....	43
<i>Кейс 5.</i> Верификация доказательной базы рефлексивно- деятельностного подхода с акцентом на качественных методах .....	45
1.6. Ключевая формула доказательности в эксперименте: «количественные методы плюс опционально качественные методы» .....	47
<b>Глава 2. Рамочные условия организации научной экспертизы и исследований по верификации психолого-педагогических программ и технологий .....</b>	<b>50</b>
Этап 1. Паспорт верификации психолого педагогических, образовательных и социальных программ и технологий .....	50
Этап 2. Качественный ассесмент материалов психолого-педагогических, образовательных и социальных программ и технологий, представленных для верификации .....	53
Этап 3. Протокол верификации психолого-педагогических программ .....	60
Этап 4. Анализ данных эксперимента по верификации и оценка его результатов .....	63
4.1. Количественный анализ данных экспериментального исследования: общие рекомендации .....	63
4.2. Дисперсионный анализ и оценка взаимодействия .....	65
4.3. Анализ данных при использовании качественных методов исследования .....	66
4.4. Оценка результатов верификации: отчет о результатах, рейтинговая оценка и общий вывод .....	68
Этап 5. Публикация результатов эксперимента в рецензируемых научных журналах. Представление баз данных в Репозиториях открытого доступа .....	69
Этап 6. Включение верифицированной программы или технологии в «Реестр психолого-педагогических программ и технологий с доказанной эффективностью» и распространение опыта в профессиональном сообществе и СМИ .....	71
<b>Литература .....</b>	<b>73</b>
1.1. Документы, справочные издания и руководства-прототипы .....	73
1.2. Литература по методологии и методам психолого-педагогических исследований .....	74
1.3. Статьи по тематике эмпирических исследований верификации программ и технологий .....	78

**Приложение 1.**

Стандарты и общие требования к проведению эксперимента по верификации психолого-педагогических программ и технологий и оценке его результатов.....	81
1. Дизайн исследования: основные виды дизайна .....	81
1.1. Рандомизированные контролируемые испытания (РКИ) .....	81
1.2. Квази-экспериментальный дизайн (КЭД).....	83
1.3. Дизайн с регрессионным разрывом (ДРР).....	83
1.4. Дизайн с одним случаем (кейс-стади, ДОС) .....	84
2. Отсев выборки .....	85
2.1. Отсев в РКИ индивидуального уровня .....	85
2.2. Отсев выборки в кластерных РКИ.....	88
2.3. Потери выборки, которые не считаются отсевом .....	89
3. Базовая эквивалентность для групп в аналитической выборке .....	89
3.1. Проверка базовой эквивалентности: рекомендации МГППУ .....	89
3.2. Проверка базовой эквивалентности: рекомендации WWC.....	91
4. Представление результатов и отчетность.....	93
4.1. Стандарты и структура «Отчета о результатах верификации»: требования МГППУ .....	93
4.2. Стандарты и требования WWC к представлению и оценке результатов верификации .....	101
5. Смешивающие факторы .....	105
6. Выводы и оценка результатов верификации .....	106
6.1. Выводы экспертизы и оценка результатов эксперимента по верификации: требования МГППУ .....	106
6.2. Завершение обзора: требования WWC .....	107

**Приложение 2.**

Отчет о результатах верификации и размере эффекта.....	109
1. Требования и рекомендации МГППУ по оценке размера эффекта и статистической значимости результатов верификации программ и технологий .....	109
1.1. Методы количественного анализа данных, рекомендуемые для оценки результатов верификации .....	109
1.2. Примерный перечень показателей результатов анализа данных....	111
1.3. Протокол оценки результатов верификации: критерии экспертизы .....	112

2. Рекомендации WWC по оценке размера эффекта вмешательства....	114
2.1. Результат индивидуального анализа .....	115
2.1.1 Величина результата .....	115
2.1.2. Размеры эффекта для исследований с групповым дизайном .....	115
2.1.3. Размеры эффекта для исследований с дизайном с регрессионным разрывом (ДРР) .....	118
2.1.4. Размеры эффекта для исследований с дизайном отдельного случая .....	118
2.2. Индекс улучшения .....	120
2.3. Статистическая значимость результата .....	121
2.3.1. Расчет статистической значимости при анализе на индивидуальном уровне .....	121
2.3.2. Кластерная коррекция для «несоответствующих» случаев анализа .....	122
2.4. Выводы из нескольких вариантов анализа (множественного анализа) .....	123

### **Приложение 3.**

Требования к заполнению «Паспорта верификации психолого-педагогических программ и технологий для образования и социальной сферы» .....	125
--	-----

---

## Введение

### Актуальность и цели Руководства

В настоящее время в общем, профессиональном и дополнительном образовании, а также в социальной сфере реализуется целый ряд программ, технологий, практик и подходов, эффективность которых не является эмпирически подтвержденной. Отсутствуют сравнительные исследования эффективности различных программ и технологий в парадигме доказательного подхода, не определены даже параметры, в отношении которых ожидаемые эффекты могут проявиться. Вопрос о выборе технологии для реализации в образовательном учреждении зачастую решается либо только на основе мнений авторитетных экспертов, либо стихийно, без учета результатов научных эмпирических исследований.

Имеющиеся исследования носят, скорее, спорадический характер и связаны, главным образом, с работой над диссертациями. Однако что происходит с верифицированной в диссертации программой, технологией или методикой в последующий период — используется и совершенствуется ли она самим автором, распространяется ли она в профессиональном сообществе, исследуется ли ее эффективность при дальнейшей трансляции и т.д., — как правило, остается неизвестным. Кроме того, качество диссертационных исследований не всегда достаточно высокое, чтобы признать авторский подход действительно верифицированным. Предположительно, таким образом может быть растрочен потенциал разработки и реализации действительно эффективных программ и технологий в этой области.

Кроме того, отсутствует научно обоснованный с позиций доказательного подхода «Реестр психолого-педагогических программ и технологий с доказанной эффективностью для образования и социальной сферы», на основании которого Министерство просвещения РФ, Министерство науки и высшего образования РФ, образовательные учреждения общего, профессионального и дополнительного образования могли бы принимать решения о рекомендации для внедрения конкретной программы в школах, колледжах, центрах образовательной и социальной направленности и др., а также распространять этот опыт через систему повышения квалификации университетов, Федеральных координационных центров по обеспечению психологической службы в системе образования РФ и Научно-методических центров сопровождения педагогических работников. Вот почему организация исследований и систематическая работа по верификации психолого-педагогических, в том числе развивающих, коррекционно-развивающих, образовательных и социальных программ и технологий приобрела особую актуальность.

В данном контексте, социальные программы и технологии отвечают на потребности современного образования в воспитательных программах,



социальной адаптации учащихся, выстраивании отношений с одноклассниками, учителями, педагогами дополнительного образования, профилактике буллинга и употребления ПАВ в образовательных учреждениях, работе с семьей учащихся, учете социальных, медицинских и правовых статусов и различий обучающихся, проживающих на территориях с разной экологией и др. Важным условием внесения в Реестр социальных программ и технологий является применение психолого-педагогических методов при их верификации.

#### **Цели руководства:**

- Разработка рамочных требований и рекомендаций к исследованиям и процедурам научной верификации (т.е. проверки эффективности) психолого-педагогических программ и технологий в соответствии с положениями доказательного подхода в психологии и образовании и с учетом международного опыта;
- Операционализация процедур научной верификации с помощью 5-ти рамочных документов: Паспорта, Протокола качественного ассесмента, базового Протокола верификации психолого-педагогических, образовательных и социальных программ и технологий, Отчета о результатах верификации и Протокола оценки результатов верификации;
- Разработка рекомендаций по количественному анализу данных эксперимента по верификации и представлению его результатов;
- Создание на этой основе «Реестра психолого-педагогических программ и технологий с доказанной эффективностью для образования и социальной сферы» Министерства просвещения РФ.

## **Национальный стандарт доказательности практик в сфере детства: общая характеристика**

Постепенное накопление в социальной сфере и образовании существенного массива «лучших практик» в отсутствие системы профессиональной верификации качества, оценки эффективности и целесообразности применения имеющихся программ и технологий привело к необходимости разработки единого подхода к определению и измерению эффективности социальных практик в сфере детства. В 2018 году группой экспертов в рамках проекта «Развитие социально-ориентированных НКО: доказательность социальных практик», выполненного АНО «Эволюция и Филантропия» при поддержке межотраслевого профессионального объединения «Оценка программ в сфере детства», Фонда Тимченко и Центра доказательного социального проектирования МГППУ, была разработана первая версия стандарта доказательности практик, обобщающая зарубежный опыт применения и тиражирования доказательных практик в социальной сфере. Но уже к 2021 году в связи с апробацией разработанной методологии в сфере детства, а также задачей создания институцио-

нальной инфраструктуры для оценки и тиражирования практик возникла необходимость в переосмыслении некоторых положений документа. Основной задачей стало формулирование методологических принципов как ориентиров в деятельности экспертного и профессионального сообществ в РФ.

В рамках государственного задания Министерства просвещения Российской Федерации № 073–00041–21–03 от 11.06.2021 «Разработка методологических оснований для формирования единого национального подхода к определению социальных (в том числе образовательных) практик с доказанной эффективностью» в 2021 году была разработана обновленная версия Стандарта и его приложений — «Национальный стандарт доказательности практик в сфере детства». В 2022 году исследовательской группой МГППУ в рамках государственного задания Министерства просвещения Российской Федерации № 073–00110–22–01 от 21.01.2022 «Научно-методическое обеспечение единого подхода внедрения социальных (в том числе образовательных) практик с доказанной эффективностью в части реализации мероприятий Десятилетия Детства» была проведена его апробация и в 2023 году сформулирована итоговая версия [Национальный стандарт, 2023]. Документ носит общий характер и предназначен для широкой аудитории, в том числе: специалистов организаций сферы детства (некоммерческие организации, государственные учреждения и пр.), донорских организаций (крупные фонды — грантодателей; органов государственной власти и местного самоуправления; коммерческих организаций, реализующих социальные программы, программы корпоративной социальной ответственности, корпоративного волонтерства, частных доноров и др.), представителей экспертного и академического сообщества и благополучателей.

В структуру документа входят базовые понятия доказательного подхода в социальной сфере, описание практики в доказательном подходе, общие принципы сбора и анализа доказательной информации о практике. В разделе «Методологические аспекты применения доказательного подхода в практике специалистов сферы детства» рассмотрены основания для типологии социальных практик в контексте исследований влияния, охарактеризован принцип наилучших доступных данных как прагматическое основание для выбора исследовательской методологии и перечислены стратегии доказательного обоснования практик. В разделе «Компоненты доказательного анализа социальной практики» представлены общие подходы с точки зрения четырех аспектов: обоснованность практики, результативность практики, обоснованность данных о социальных результатах, регламентированность практики.

«Национальный стандарт доказательности практик в сфере детства» рассматривается как методологическая основа для оценки и отбора доказательных практик в социальной сфере и образовании. В нем отмечается необходимость профессиональной и научной верификации практики как основы для отбора социальных практик и создания реестров доказательных

практик, а также для принятия управленческих решений и реализации национальной модели доказательной социальной политики в сфере детства.

## **Доказательный подход как основа научной верификации психолого-педагогических, образовательных и социальных программ и технологий**

Проверка эффективности психолого-педагогических, образовательных и социальных программ и технологий независимо от их предметного содержания проводится специалистами и экспертами Московского государственного психолого-педагогического университета (МГППУ) и общероссийской общественной организации «Федерация психологов образования России» (ФПОР) в парадигме доказательного подхода, что предполагает:

- **Теоретическое обоснование программы, технологии.** Авторы и разработчики должны продемонстрировать понимание, сформулировать, какие представления отечественной психологической науки (культурно-историческая и деятельностьная научная школа — Л.С. Выготский, А.Н. Леонтьев, П.Я. Гальперин, А.В. Запорожец и др.) или теорий и подходов других научных школ, получивших международное признание, в том числе зарубежных, лежат в основе их программы, технологии. Важно также сослаться на результаты отечественных и международных исследований, современные публикации в рейтинговых научных журналах России и других стран по данной тематике. Чем более эрудированы авторы и разработчики, тем менее доморощенной выглядит их программа, технология и тем лучше она вписывается в современный профессиональный контекст. Примером технологии, основанной на идеях и положениях западных психологов, биологов, естествоиспытателей конца 19-го — первой трети 20-го в.в. (В. Штерн, Ш. Бюлер, Е. Джонс, А. Гезелл, Х. де Фриз и др.), но получившей распространение во всем мире, в том числе, в современной России, может служить педагогическая система М. Монтессори [Сорокова М.Г., 2020b], [Сорокова М.Г., Сороков Д.Г., 2018].
- **Корректный дизайн исследования.** Необходимо понять, в отношении каких психолого-педагогических характеристик (другие названия — конструкторов, переменных, признаков, параметров) должна быть эффективна методика, и провести проверку эффективности методики на основе естественно-научной методологии (номотетический подход) в соответствии с классической схемой психолого-педагогического эксперимента. Такой эксперимент предполагает проведение констатирующего, формирующего и контрольного этапов и наличие экспериментальной группы (группы вмешательства) и контрольной группы

(группы сравнения). Дополнением могут служить качественные методы исследования (идеографический подход), если их применение способствует лучшему пониманию статистически выявленных эффектов; позволяет прицельно работать с кейсами, для которых ожидаемый эффект не достигнут; на кейсе заметить феномен для последующей проверки на большой выборке и др.

- **Стандартизированные инструменты психолого-педагогических измерений.** Необходимо подобрать стандартизированные инструменты для измерения исследуемых переменных, валидность и надежность которых является доказанной в соответствии с международными требованиями и стандартами классической и/или современной теории тестов. Если таких инструментов недостаточно, можно либо собрать данные для стандартизации имеющихся отечественных или зарубежных методик, либо разработать и стандартизировать авторские методики. В последнем случае МГППУ может оказать научно-методическую поддержку при организации эксперимента, сборе и количественном анализе эмпирических данных для стандартизации. «Открытый реестр психодиагностических методик, вызывающих доверие профессионального сообщества» [Система функционирования психологических служб, 2020, С. 110], разработанный МГППУ, может быть взят за основу при выборе таких инструментов. Реестр систематически актуализируется.
- **Заслуживающие доверия процедуры сбора эмпирических данных.** Данные должны быть собраны квалифицированным персоналом в соответствии с инструкциями по проведению тестирования. Процедуры измерения и оценки исследуемых характеристик должны исключать манипулирование с отбором испытуемых, их данными, преднамеренным воздействием на испытуемых и их результаты и т.д., которые способны исказить результаты в сторону их повышения или понижения. Описание процедур сбора данных должно быть представлено в «Протоколе оценки результатов верификации».
- **Большие выборки,** желательно от 1000 до нескольких тысяч испытуемых. Данные больших выборок имеют следующие преимущества. Во-первых, эффектом формирующего воздействия (вмешательства) являются статистически значимые различия между экспериментальной и контрольной группами на констатирующем и контрольном этапах, т.е. до и после воздействия. В ходе количественного анализа данных любой статистический критерий учитывает объемы выборок, и различия с большей вероятностью выявляются именно на больших выборках, даже если размер эффекта невелик. Во-вторых, в любом психолого-педагогическом эксперименте всегда присутствует множество случайных факторов, которые создают «шум», ослабляют искомые закономерности. Чем меньше объем выборки, тем большую

роль играют случайные факторы и тем меньше уверенность, что эффект достигнут за счет формирующего воздействия, а не за счет уникальных свойств испытуемых, удачного стечения обстоятельств и т.д. Отметим, что большие выборки свыше 1000 испытуемых — это ориентир. Если психолого-педагогическая, образовательная, социальная программа или технология носит более локальный характер, то количество испытуемых может быть и меньше, например, порядка нескольких сотен.

При сборе эмпирических данных необходимы научные партнерства, методологическая, организационная, образовательная и, возможно, финансовая поддержка сотрудников, осуществляющих сбор данных. Научно-методическая, организационная и образовательная поддержка может быть оказана МГППУ. Автоматизация процессов сбора данных возможна за счет размещения стандартизованных методик на ресурсах Отдела мониторинга качества профессионального образования (ОМКПО) МГППУ.

- **Релевантные количественные методы анализа данных с применением стандартных статистических пакетов.** Арсенал современных количественных методов анализа данных и математического моделирования чрезвычайно богат и имеет тенденцию к расширению. Для анализа данных эксперимента по верификации психолого-педагогических, образовательных и социальных программ и технологий должны быть выбраны конкретные методы, необходимые оценки их эффективности. Рекомендации по выбору и применению таких методов см. в **Приложении 2**. Количественный анализ данных научного исследования должен осуществляться в стандартных статистических пакетах, например, в SPSS, Statistica, R, KNIME Analytics Platform и мн. др., поскольку реализованные там алгоритмы расчетов надежны, пользуются доверием международного научного сообщества, снижают вероятность ошибок, существенно ускоряют процесс расчетов и демонстрируют результаты анализа в требуемом виде, т.е. с указанием всех необходимых численных параметров доказательности. Качественные методы исследования и соответствующие методы анализа данных могут служить хорошим дополнением.
- **Правильная интерпретация результатов.** Это работа психологов, педагогов и других специалистов, участвующих в исследовательском проекте. Интерпретация результатов проводится на основании показателей количественного анализа данных (размер эффекта, эмпирические значения статистик критериев, уровни значимости и др.), рассчитанных в статистических пакетах, с учетом содержания измеренных психолого-педагогических характеристик и контекстных параметров. Если наряду с количественными методами применялись также качественные методы исследования, их анализ и интерпретация проводится в соответствии с инструкцией к ним.

- **Воспроизводимость результатов исследований.** Результаты верификации психолого-педагогических, образовательных и социальных программ и технологий должны быть представлены и описаны так, чтобы другие исследователи имели возможность повторить эксперимент и проверить результаты. Кроме того, при распространении программ и увеличении числа их пользователей иногда их качество снижается, поэтому здесь речь может идти о периодическом повторении исследований, мониторинге результатов, например, раз в 2–3 года.
- **Общедоступность баз данных исследований в репозиториях научных данных.** Данные исследований и краткое описание результатов должны быть представлены в репозиториях открытого доступа, что соответствует современным международным тенденциям открытой науки и требованиям рейтинговых научных журналов. Это повышает ответственность исследователей, дает возможность проведения мета-анализа данных, а также способствует формированию имиджа исследователей, соблюдающего этические нормы, и росту цитируемости их работ. В конечном счете, это содействует возрастанию доверия общества к научным открытиям. Пример — репозиторий RusPsyDATA <https://ruspsydata.mgppu.ru/>
- **Публикация результатов в рецензируемых научных журналах, обсуждение их на научных вебинарах, конференциях.** При подаче статьи в научный журнал она проходит процедуры двойного слепого рецензирования, когда ни автор, ни рецензенты не знают, чьи работы и кому поданы на экспертизу. Таким образом, статья получает независимую экспертную оценку и дорабатывается в соответствии с замечаниями экспертов. На основании опубликованных результатов верификации рекомендуется разработка и реализация программ дополнительного профессионального образования, масштабирование и продвижение программ и технологий с доказанной эффективностью, и др.
- **Создание «Реестра психолого-педагогических программ и технологий с доказанной эффективностью для образования и социальной сферы» Министерства просвещения РФ.** В Реестр будут включены профилактические, развивающие, коррекционно-развивающие, образовательные и социальные программы и технологии, прошедшие верификацию на основе эксперимента в соответствии с перечисленными выше положениями доказательного подхода в психологии и образовании.

## Стандартизированные психодиагностические инструменты, рекомендованные для измерения психолого-педагогических характеристик

«Открытый реестр психодиагностических методик, вызывающих доверие профессионального сообщества» представляет психодиагностический инструментарий, рекомендованный для использования психологами образовательных организаций. Это — составная часть документа «Система функционирования психологических служб в общеобразовательных организациях. Методические рекомендации» [Система функционирования психологических служб, 2020]. Документ подготовлен рабочей группой в составе: Рубцов В.В. (руководитель), Сергоманов П.А. (сооруководитель), Леонова О.И. (ответственный секретарь), Абушкин Б.М., Алехина С.В., Банников С.Г., Вихристюк О.В., Гаязова Л.А., Делибалт В.В., Драганова О.А., Дубровина И.В., Егоренко Т.А., Егорова М.А., Забродин Ю.М., Зарецкий В.К., Исаев Е.И., Ключева Т.Н., Лавриненко О.А., Лобанова А.В., Марголис А.А., Пахальян В.Э., Романова Е.С., Ослон В.Н., Сафронова М.А., Семья Г.В., Сорокова М.Г., Чиркина Р.В., Шарабарина О.Д., Шведовская А.А., Шумакова Н.Б.

**Структура Реестра** включает следующие позиции: Порядковый номер методики в Реестре; Автор методики, название методики; Измеряемый конструкт; Возрастная группа; Параметры стандартизации, доказательность: перечислены результаты стандартизации и проверенные параметры; Источник: даны ссылки на статьи в рецензируемых научных журналах, где подробно описаны все процедуры стандартизации и проверенные параметры; Год, наличие компьютерной версии методики; Статус: основная (т.е. методика, прошедшая процедуры стандартизации и валидации) или условно рекомендуемая (т.е. диагностические методики, в настоящее время широко используемые и/или вызывающие доверие профессионального сообщества, но пока не обладающие всеми необходимыми сведениями их доказательности); Целевая группа в соответствии с классификацией, представленной в документе «Система функционирования психологических служб в общеобразовательных организациях. Методические рекомендации».

В этом документе на основе анализа нормативных правовых актов, а также научно-методических документов в системе образования и социальной защиты населения представлена следующая **классификация целевых групп детей**, в отношении которых в общеобразовательных организациях реализуются программы адресной психологической помощи:

- норма (нормотипичные дети и подростки с нормативным кризисом взросления);
- дети, испытывающие трудности в обучении;



- социально уязвимые категории детей. К ним относятся: Дети, находящиеся в трудной жизненной ситуации (включая социально опасное положение), Одаренные дети. К числу Детей, находящиеся в трудной жизненной ситуации, относятся: Дети-сироты и дети, оставшиеся без попечения родителей (в приемных семьях и в организациях для детей-сирот); Обучающиеся с ОВЗ, дети-инвалиды; Дети-мигранты; Дети с отклоняющимся поведением (девиантное поведение детей и подростков, суицидальное поведение детей и подростков).

Кроме этого, в особо уязвимую категорию обучающихся входят дети, обучающиеся в малокомплектных сельских школах, кочевых школах для детей коренных народов Севера, школах-интернатах, проживающие на территориях, отнесенных к Арктической зоне, в экологически неблагоприятных районах, а также отдельные этнические группы (например, цыгане).

Взрослые представляют отдельную целевую группу. Диагностические методики, представленные в Реестре, сгруппированы в соответствии со следующими **разделами**: Развитие основных психических функций; Метапредметные компетенции и универсальные учебные действия; Социальное развитие и морально-ценностная сфера; Эмоционально-волевая сфера; Коммуникативная сфера; Поведенческая сфера; Профессиональная направленность, мотивация, характерологические особенности; Диагностика профессиональных и личностных проблем взрослых участников образовательного процесса. В последнем разделе представлен психодиагностический инструментарий для работы с персоналом образовательных организаций, учреждений социальной защиты населения, организаций для детей-сирот и детей, оставшихся без попечения родителей, и родителями.

## **Организация независимой оценки психолого-педагогических программ и стандартизации психодиагностических инструментов**

Жизненный цикл практики состоит из нескольких последовательных стадий, начиная с интуитивного поиска решений значимой социальной проблемы до регламентированной проверенной технологии, эффективность которой подтверждена экспериментально с применением стандартизированного инструментария, количественных и качественных методов. Для социальной сферы важно иметь возможность увидеть и поддержать развитие перспективных моделей социальных практик уже на начальном этапе, обеспечить их методическое сопровождение и экспертную оценку, направленные на «доразвивание» и дальнейшее совершенствование, чтобы со временем получить доказательную базу для перехода на более высокий уровень. Понимание этой ситуации привело профессиональное сообщество к необходимости развития системы **уровневой верификации**



практик с учетом их актуального этапа развития. Положения «Национального стандарта доказательности практик в сфере детства» носят скорее программный характер, настоящее же руководство ориентировано на научную верификацию и оценку доказательной базы именно зрелых практик, программ и технологий в образовании и социальной сфере.

С учетом современного уровня отечественной науки и международного опыта оценки эффективности был разработан четкий регламент независимой экспертизы и научной верификации доказательной базы развивающих, коррекционно-развивающих, профилактических, просветительских и других программ, в рамках которого действует секция «Доказательный подход в психологии и образовании» <https://mgppu.ru/project/538> Научно-экспертного совета МГППУ.

В статье [Сорокова М.Г., Леонова О.И., Ботова Ю.Б., Пятаков Е.О., 2023] представлена система проведения исследований по стандартизации психодиагностических методик и независимой оценке эффективности психолого-педагогических программ, разработанная и реализуемая на базе федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный психолого-педагогический университет» при участии общероссийской общественной организации «Федерация психологов образования России». Описаны особенности проведения экспертизы психолого- педагогических программ и технологий в парадигме доказательного подхода.

Доказательный подход рассматривается как ведущий принцип организации профессиональной деятельности педагога-психолога (психолога в системе образования), основной метод выявления и распространения лучших практик работы с проблемами и вызовами в социальной среде, формирования на этой базе реестра программ психологического обеспечения образовательного процесса для образовательных организаций. Обсуждается регламент верификации доказательной базы, экспертной оценки и отбора программ и технологий с доказанной эффективностью для образования и социальной сферы в соответствии с положениями доказательного подхода. Охарактеризована деятельность секции «Доказательный подход в психологии и образовании» в структуре Научно-экспертного совета на базе МГППУ, в рамках которой осуществляется экспертиза кандидатных программ, научно-методическая поддержка экспериментов по проверке эффективности психолого-педагогических программ, а также организация прикладных исследований по стандартизации психодиагностического инструментария.

# Глава 1.

## Теоретико-методологические основы и современные исследования в парадигме доказательного подхода в психологии и образовании

### 1.1. Национальный стандарт доказательности как методологическая основа для описания и анализа социальных практик в парадигме доказательного подхода

Постепенное накопление существенного массива «лучших практик» в отсутствие системы профессиональной верификации качества, оценки эффективности и целесообразности применения имеющихся программ и технологий в социальной сфере и образовании привело к необходимости создания единого стандарта в подходах к пониманию и измерению доказательности социальных практик в сфере детства. Проект Национального стандарта доказательности практик был подготовлен в рамках государственного задания Министерства Просвещения Российской Федерации № 073–00041–21–03 от 11.06.2021 «Разработка методологических оснований для формирования единого национального подхода к определению социальных (в том числе образовательных) практик с доказанной эффективностью».

Методология доказательного обоснования практик в социальной сфере и образовании закреплена в Национальном стандарте доказательности практик и призвана обеспечить научно-методическую преемственность на разных этапах поиска, отбора и оценки практик на основе доказательного подхода [Национальный стандарт, 2023]. Стандарт обеспечивает сетевым участникам отбора и верификации социальных практик общую методологическую основу для систематизации и анализа доказательной информации о практике: определяет систему сбора доказательств и обоснований, структурирует требования к данным, необходимым для оценки доказанности результатов, устанавливает формат описания практики и ее доказательной базы.

В качестве единицы доказательного обоснования и анализа в Стандарте рассматривается **«социальная практика»** — совокупность взаимосвязанных процедур и способов профессионального воздействия, направленных на достижение заранее запланированных позитивных изменений (социальных результатов) в жизни конкретных социальных групп или общества в целом. Понятие «социальная практика», предлагаемое в Стандарте, дает возможность учесть широкий спектр применяемых в социальной сфере и образовании форм работы с детьми, не ограничиваясь

каким-либо из существующих форматов организации профессиональной деятельности, но уделяя особое внимание анализу механизма происходящих изменений.

Границы понятия (критерии «социальной практики»):

- Деятельность ориентирована на решение конкретной проблемы определенных социально-уязвимых групп;
- Заранее заданы ожидаемые изменения в ситуации целевых групп;
- Может быть выделен необходимый и достаточный для изменений алгоритм (комплекс) действий;
- Известен или может быть описан механизм, за счет которого достигаются заявленные изменения.

В соответствии с методологией, изложенной в Стандарте, любая социальная практика может быть описана с опорой на принципы доказательного подхода с учетом актуального уровня ее зрелости, включая инновационные и пилотные практики, в отношении которых пока не получены надежные и достоверные сведения о достижении позитивных изменений в социальной ситуации благополучателей.

Сбор и анализ данных для разных типов практик будет иметь различный характер и предполагает различные стратегии при сборе доказательной базы. Анализ уникальных и универсальных аспектов реализации практики может рассматриваться как первый шаг к определению стратегии доказательного обоснования и точка отсчета для выбора исследовательской методологии. Четкое определение характера практики может существенно влиять на экспертную оценку ее доказательной базы, раскрывая некоторые специфические аспекты ее реализации и «объясняя» выбор тех или иных способов сбора и анализа данных. В Стандарте выделено несколько направлений, задающих основания для классификации практик и потенциально влияющих на выбор подходов к сбору и анализу данных:

1. **Стадия жизненного цикла практики** (инновационные, пилотные, устоявшиеся, масштабируемые практики). Применение исследовательских инструментов возможно и крайне желательно на всех этапах «жизненного цикла» практики, включая исследования по прояснению проблемы, раскрытию механизмов изменений, определению условий достижения желаемых результатов и др.;
2. **Направленность практики на разные стадии реализации социального риска**: профилактика возникновения риска, работа в ситуациях острого кризиса, сопровождение социально-уязвимых групп в трудной жизненной ситуации, минимизация вреда в ситуациях необратимых последствий уже реализовавшихся рисков;
3. **Степень сложности и уровень неопределенности причинно-следственных связей между деятельностью и результатом** (контроль над изменениями, подробнее см. ниже);

4. **Масштаб и формат реализации практики:** микроуровень — объединяет практики, направленные на оказание помощи в решение проблем индивида и семьи; мезоуровень — адресован отдельным социально-уязвимым группам; макроуровень — связывают с существенными переменами в обществе).

Подготовка социальной практики к процедуре базовой профессиональной верификации на основе Стандарта состоит из двух этапов:

1. Описание практики с опорой на принципы доказательного подхода
2. Подготовка доказательной базы практики

Описание практики с опорой на принципы доказательного подхода подразумевает форматирование социальной практики в соответствии с правилами доказательного проектирования с целью формулировки ожидаемых социально значимых изменений в ситуации конечных благополучателей. Основной задачей специалистов на данном этапе описания практики является разработка «логической модели» практики, которая описывает замысел специалистов в виде цепочки шагов, каждый из которых заканчивается достижением определенного результата. Логическая модель практики определяет и фиксирует содержание запланированных социальных изменений, отражая последовательную логику этих изменений с момента начала работы и заканчивая долгосрочными целями.

На этапе сбора доказательной базы специалисты систематизируют данные, свидетельствующие о наличии изменений в ситуации благополучателей, а также критически оценивают описанные в научной литературе теории и исследования, чтобы обосновать выбранный способ работы с проблемой. В основу сбора доказательной базы практики в соответствии со Стандартом положен принцип необходимости и достаточности данных как совокупности лучшего доступного на данный момент эмпирического базиса, экспертизы специалистов-практиков, анализа опыта и предпочтений благополучателей. Сбор доказательной базы практики рекомендуется осуществлять в нескольких направлениях:

1. *Научно-теоретическое и эмпирическое обоснование практики* (теоретическое и эмпирическое обоснование общих закономерностей и принципов, лежащих в основе деятельности специалиста, анализ фундаментальных теорий социальных наук и подходов, лежащих в основе осуществляемой деятельности).
2. *Анализ и систематизация имеющихся данных* (данные, накопленные в течение длительного времени и корректно систематизированные, могут продемонстрировать динамику изменений в ситуации благополучателей и быть использованы как эмпирический базис практики).
3. *Сбор и анализ эмпирических данных о влиянии практики* (для изучения эффектов и результатов социальных практик также используются данные первичных исследований: в Стандарте рекомендуется придерживаться прагматичного подхода к выбору методологии сбора данных

о социальных результатах, ориентируясь на получение данных высокого качества, но учитывая при этом целесообразность использования той или иной исследовательской методологии и реальные возможности получения данных).

Практика, описанная в соответствии с требованиями Стандарта, в обязательном порядке должна пройти независимую профессиональную и научную экспертизу для определения уровня ее доказательности. Решение о присвоении практике определенного статуса принимается на основании проведенной экспертизы, доверие к результатам верификации обеспечивается действием механизма распределенной ответственности за экспертизу. Методология, изложенная в Стандарте, задает четыре базовых направления доказательного анализа, позволяющих оценить актуальный уровень доказательности социальной практики.

#### *Регламентированность практики*

Данное направление сфокусировано на анализе стандартизованности процедур практики — насколько действия, осуществляемые в рамках практики, носят устойчивый, предсказуемый и воспроизводимый характер. Максимальная выраженность данного компонента означает высокую степень повторяемости действий в рамках практики, наличие регламентов по основным направлениям деятельности, формализованных в методических рекомендациях и сопровождающих документах (наличие и качество документов, в которых закреплены процедуры практики, формы и методы обеспечения качества работы специалистов, реализующих практику, соблюдение обязательных требований и процедур в процессе реализации практики).

#### *Научно-теоретическая и эмпирическая обоснованность практики*

Данное направление доказательного анализа сфокусировано на оценке замысла практики — внутренней логики, объясняющей как действия, осуществляемые в рамках практики позволяют достигать заявленных социальных результатов. Максимальная выраженность данного компонента предполагает развернутое описание и причинно-следственное обоснование механизма, раскрывающего каким образом положительные изменения в жизни благополучателей могут быть достигнуты в рамках предпринимаемых действий специалистов. Здесь анализируется непротиворечивость и убедительность причинно-следственных связей между деятельностью и ожидаемыми изменениями, обоснованность практики с точки зрения научных теорий и результатов исследований, обоснованность практики с точки зрения потребностей благополучателей (детей и их семей) и опыта специалистов-практиков.

#### *Результативность практики (достижение социальных результатов)*

В рамках данного направления анализируется, в какой мере достигаются заявленные изменения, как меняется ситуация благополучателей благодаря практике. Максимальная выраженность данного компонента

означает: наличие данных относительно достижения заявленных социальных результатов за счет реализации практики; устойчивости достигаемых социальных результатов в краткосрочной и долгосрочной перспективе; отсутствия (или анализа причин) негативного эффекта в краткосрочной и долгосрочной перспективе для благополучателей или сообщества в целом в ходе реализации практики.

*Обоснованность данных о социальных результатах практики*

Данный компонент направлен анализ корректности применения исследовательских инструментов в ходе сбора данных относительно результативности практики: в какой мере удачен выбор методов сбора и анализа данных, насколько строгой является методология оценочных исследований, достаточно ли данных и пр.

## **1.2. Современные подходы к сбору и анализу данных при формировании доказательной базы социальных практик**

Современные подходы к сбору и анализу данных при формировании доказательной базы практик вынуждены учитывать довольно широкий диапазон профессиональных интервенций: выбор исследовательской методологии должен осуществляться адекватно текущим задачам разработчиков, масштабу реализации и степени зрелости практики, сложности социальной проблемы, а также с учетом других факторов «реальности»: социальных и культурных контекстов, доступности данных, наличия исследовательских компетенций команды и т.п.

Стремление к использованию наиболее строгих методов для получения данных о доказательности (например, РКИ) должно быть пропорционально той цели, в соответствии с которой будут использованы результаты исследования. В процессе постепенного развития практики для повышения качества социальных результатов может допускаться использование менее строгих методов (при условии соблюдения правил), тем не менее измеряющих достижение результата с минимальной погрешностью. Таким образом, прагматический подход опирается на принцип использования лучших доступных данных: требования к строгости исследовательской методологии, применяемой для сбора и анализа данных относительно влияния социальных и образовательных практик ставятся в зависимость условий и целей проведения исследования (Рис 1).

К устоявшимся практикам, тиражирование которых поддерживается крупными организациями и ресурсами, предъявляются более высокие требования к качеству и методологии сбора данных, предполагающие строгий исследовательский дизайн (РКИ), тщательную проверку и независимую экспертизу результатов исследований. Однако, для эмпирического обоснования инновационных, пилотных практик микро- и мезо- уровней допускается применение различного дизайна исследований в зависимости

от целей и задач, стоящих перед разработчиками на разных этапах жизненного цикла практики.



Рис. 1. Иерархия дизайнов исследования для разных типов практик

Удачную концептуализацию условий реализации практик предлагает А. Эбрахим [Ebrahim A., 2019], [Ebrahim A., Rangan V.K., 2014]. Автор выделяет **два измерения**: первое — степень сложности и неопределенности причинно-следственных связей между деятельностью и результатом; второе — контроль над изменениями. В ряде практик связь между деятельностью и эффектом относительно отчетлива и линейна (например, практика, нацеленная на устройство в семье конкретных детей, оставшихся без попечения родителей, предполагает, что к эффекту — устройству в семью — приводит непосредственно деятельность специалистов). В других же практиках эта связь не очевидна, есть высокая вероятность действия большего числа иных факторов (сюда относятся, например, практики, нацеленные на работу с причинами, по которым дети остаются без попечения родителей: уменьшение количества детей, оставшихся без попечения родителей, может быть результатом действия множества факторов, вклад деятельности специалистов в такое изменение трудно отследить). Второе измерение — контроль над изменениями — также имеет две градации: низкий уровень и высокий уровень. Предполагается, что контроль более низкий, если специалисты сосредотачиваются на очень конкретной задаче (например, помощь в устройстве ребенка в семью), и более высокий, когда необходима координация специалистов, выполняющих разные функции (например, не только устройство в семью, но и оказание комплексной психологической и социальной поддержки семьям с приемными детьми).



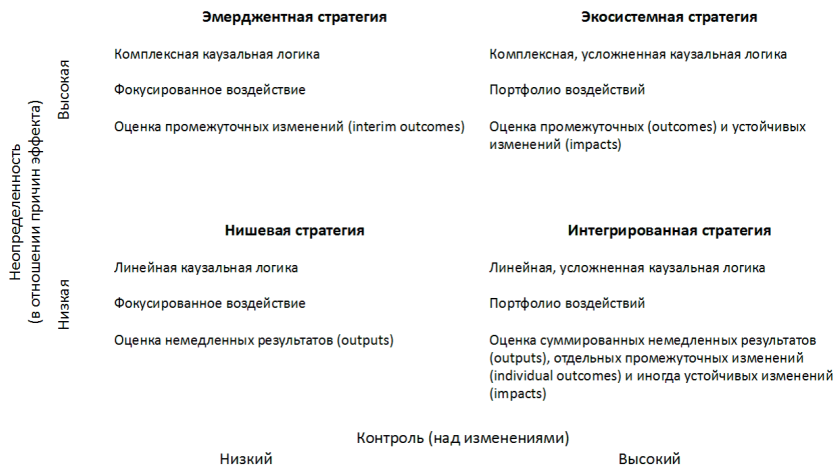


Рис. 2. Рамка условий социального действия

Четыре квадранта, полученные в результате пересечения описанных измерений, задают разные стратегии практической деятельности (Рис. 2). В квадранте нишевой стратегии располагаются практики с конкретным алгоритмом действия или практики по типу экстренной помощи, горячих линий и т.п., которые работают лишь на достижение немедленных результатов (сколько детей пристроено в семью, какое количество консультаций проведено и т.п.), соответственно, от них нельзя ожидать работы на устойчивые социальные изменения (например, нельзя оценивать работу телефона доверия по тому, уменьшается ли количество суицидов в регионе). В квадранте интегрированной стратегии — практики, занимающие несколько ниш, соответственно, их оценка требует измерения более мощных в социальном смысле результатов (например, требуется не просто фиксация количества устроенных в семью детей, но и измерение их благополучия). В квадрант эмерджентной стратегии попадают практики по защите прав, лоббированию интересов детей и т.п. Их деятельность по преобразованию общественных институтов и норм не вписывается в линейную логику: если, например, наблюдается изменение отношения в обществе к гражданским правам, и именно на это долгое время работали конкретные практики, можно ли сказать, что это непосредственно их заслуга, ведь в таких случаях мы всегда имеем дело с комплексными сдвигами? К квадранту экосистемной стратегии можно отнести практики, нацеленные на социальные преобразования и системные изменения, ставящие перед собой задачи влияния на национальную и международную политику (например, практики, связанные с поддержкой определенных инициатив в образовании или в сфере благополучия детей).



Очерченная А. Эбрахимом структурная рамка условий дает возможность более внимательно анализировать уровень эффектов, на которые рассчитывают разработчики тех или иных практик — работают ли они на достижение немедленных результатов, промежуточных или устойчивых социальных изменений. Такой анализ позволит специалистам делать более качественные обоснования своих практик, а экспертам — более реалистично их оценивать.

Можно ожидать, что в случае сложного характера отношений между деятельностью и эффектами (комплексной теории изменений) для оценки значимых социальных эффектов экспериментальные методы будут далеко не лучшим выбором из-за принципиальной невозможности контролировать множество факторов, влияющих на результат. Экспериментальные методы оценки эффектов хорошо работают в случае отчетливых причинно-следственных связей между деятельностью и эффектом (сфокусированной теории изменений). В случае же комплексной теории лучше выбирать другие методы, например, методы отслеживания (картирования) промежуточных изменений (outcome mapping) и определения наиболее значимых изменений, методы, позволяющие получать обратную связь от различных аудиторий, системный анализ и др. Как справедливо замечает А. Эбрахим [Ebrahim A., 2019], хотя подобные методы менее популярны, чем экспериментальные, они тем не менее представляются весьма перспективными, поскольку помогают преодолевать трудности, связанные с оценкой результатов в сложном и непредсказуемом мире.

### **1.3. Теоретико-методологические предпосылки доказательного подхода в психологии и образовании в трудах классиков российской психологии**

**Предпосылки доказательного подхода в научных психологических исследованиях.** Доказательство научной ценности и практической значимости полученных результатов составляет неотъемлемый атрибут методологии отечественных научно-психологических исследований. Доказательность достоверности и обоснованности результатов исследования обеспечивается опорой на фундаментальные психологические теории, обоснованными методологическими позициями, репрезентативной выборкой исследования и применением валидного научно-психологического инструментария, сопоставительным анализом процесса и результатов исследования в экспериментальных и контрольных группах, применением адекватных статистических методов. Методологическая компетентность является неотъемлемой составляющей исследовательской и практической деятельности психолога. Курс «Методологические основы психологии» входит в перечень обязательных учебных дисциплин профессиональной психологической подготовки [Василюк Ф.Е., Зинченко В.П., Мещеряков Б.Г.,

Петровский В.А., Пружинин Б.И., Щедрина Т.Г., 2012; Корнилова Т.В., Смирнов С.Д., 2024].

Доказательность результатов проведенного научного исследования нашла свое последовательное воплощение в психологическом эксперименте. В экспериментальной психологии тщательно прописаны основные характеристики психологического эксперимента: этапы исследования, экспериментальная выборка, экспериментальные переменные и способы их контроля. Описаны и оценены достоинства и ограничения основных планов исследования: доэкспериментального, квазиэкспериментального, экспериментального, корреляционного исследования. При планировании экспериментального (эмпирического) исследования и обработке его результатов в экспериментальной психологии применяется широкий арсенал методов математической статистики [Дружинин В.Н., 2019].

Теория и практика доказательного подхода в психологии и образовании берет свое начало в работах классиков культурно-исторической и деятельностной научной школы (Л.С. Выготский, А.Н. Леонтьев, П.Я. Гальперин, А.В. Запорожец, Д.Б. Эльконин, В.В. Давыдов и др.). А.А. Леонтьев подчеркивал, что в культурно-исторической теории важнейшей для педагогов и психологов является концепция зоны ближайшего развития как представления о том, «что ребенок может сейчас выполнять лишь в сотрудничестве, а завтра уже самостоятельно» [Выготский Л.С., 2005, С. 12]. При этом он акцентировал важность для практики трех теоретических выводов, каждый из которых крайне важен в наше время для проектирования психолого-педагогических и образовательных программ, технологий, социальных практик, эффективность которых предстоит доказать.

Один из этих выводов выражен в хорошо известном генетическом законе культурного развития, согласно которому «...всякая функция в культурном развитии ребенка появляется на сцену дважды, в двух планах, сперва — социальном, потом — психологическом, сперва между людьми, как категория интерпсихическая, затем внутри ребенка, как категория интрапсихическая. Это относится одинаково к произвольному вниманию, к логической памяти, к образованию понятий, к развитию воли» [Выготский Л.С., 1983, С. 145]. Примером из повседневной жизни может быть развитие произвольного внимания младших школьников в ходе обучения в начальной школе, когда в самом начале учитель для овладения вниманием учеников использует прием пошагового контроля. Необходимость в нем полностью отпадает к концу обучения в начальных классах, поскольку произвольное внимание, первоначально «появившееся на сцену» в виде совместной деятельности учителя и учеников, переходит из внешнего плана во внутренний и становится интерпсихической функцией. Распространенной ошибкой является слишком узкое понимание генетического закона, т.е. представление о том, что он относится только к высшим формам познавательных процессов. Между тем в широком смысле любое

овладение своим поведением при помощи интериоризированных культурных средств является высшей психической функцией.

Второй момент связан с коллективным характером процесса обучения и воспитания. А.А. Леонтьев подчеркивал, что в данном направлении шли многие педагогические поиски 1920–30-х годов, включая А.С. Макаренко. В данном направлении идет научная школа В.В. Давыдова, где за основу принято положение о том, что методом развивающего обучения является «...организация совместной учебной деятельности младших школьников (и прежде всего организация решения ими учебных задач)» [Давыдов В.В., 1996, С. 384]. В данной научной школе большое значение придается характеру детско-взрослого взаимодействия, в котором взрослый творит зону ближайшего развития, а между детьми складывается новая общность, делающая возможным формирование таких психологических новообразований, которые в индивидуальной деятельности детей (или в ситуации, где, как сказал бы Д.Б. Эльконин, дети учатся рядом, но не вместе) формируются значительно труднее или не формируются вовсе. Значение групповых форм учебной деятельности для развития школьников в обучении доказано многочисленными исследованиями, проведенными под руководством В.В. Рубцова [Совместная учебная деятельность, 2021].

Третий момент, по мнению А.А. Леонтьева, относится к деятельности учителя, смысл которой в том, «... чтобы направлять и регулировать деятельность учеников через коллективную деятельность, через организацию сотрудничества учеников друг с другом и учителя с учениками» [Выготский Л.С., 2005, С. 12]. Отсюда очевиден вывод о важности качества взаимодействий между детьми и подростками и детско-взрослых взаимодействий в психолого-педагогических, образовательных программах и технологиях, развивающих практиках, оцениваемых в рамках доказательного подхода. Очевидно и то, что одним из параметров оценки психолого-педагогических, образовательных программ и социальных практик должны быть характеристики детско-взрослого взаимодействия и их воспроизводимость при передаче программы или технологии в руки других людей.

Теоретико-методологические предпосылки доказательного подхода в образовании заложены в обоснованном Л.С. Выготским экспериментально-генетическом (каузально-генетическом) методе, позволившем изучать процесс становления психологических новообразований посредством их целенаправленного формирования. Полагалось, что если удастся при определенных условиях сформировать (изменить) некое психическое свойство, состояние, процесс, то сам ход формирования вполне адекватно отражает структуру и ход естественного, нормального развития этих психических особенностей. Л.С. Выготский писал: «...применяемый нами метод может быть назван методом экспериментально-генетическим в том смысле, что он искусственно вызывает и создает генетически процесс психического развития» [Выготский Л.С., 1983, С. 95].

Экспериментально-генетический метод был предложен Л.С. Выготским как исследовательское средство, дающее возможность проследить именно процесс развития, а не изменения готовой формы психического, которые могут изучаться при помощи классического психологического эксперимента. Экспериментальному методу старой (или, точнее, классической) психологии Л.С. Выготский противопоставлял экспериментальную методику двойной стимуляции, которая позволяла проследить процесс структурного преобразования изучаемой функции. Под руководством Л.С. Выготского были проведены экспериментальные исследования высших психических функций (памяти, внимания), ставшие классикой отечественной психологии.

В научной школе Л.С. Выготского были уточнены и детально проработаны представления о переменных, изучаемых посредством экспериментально-генетического метода. Примером такого уточнения и детализации может служить концепция поэтапного / планомерного формирования умственных действий и понятий. Л.Ф. Обухова [Обухова Л.Ф., 2022] в структуре экспериментального метода, предложенного П.Я. Гальпериним, выделяет три подструктуры условий:

1. Нужно наметить качества будущего действия и подобрать подсистему условий, которые обеспечивают формирование действия с заданными свойствами.
2. Должна быть такая подсистема условий, которая обеспечивает выполнение действия с первого раза и каждый раз правильно.
3. Должна быть подсистема условий, обеспечивающая перевод действия в умственный план» [Обухова Л.Ф., 2022, С. 180].

Л.Ф. Обухова подчеркивала, что в экспериментах научной школы П.Я. Гальперина продолжают играть важную роль понятия идеальной формы и двойной стимуляции. При этом под идеальной формой понимается, вслед за Л.С. Выготским, развитая форма, которая должна появиться в итоге развития, называемая «...идеальной в том смысле, что является образцом того, что должно появиться в конце развития. Или конечной — в смысле того, что должно в конце развития ребенка у него получиться» [Выготский Л.С., 2001, С. 88].

Л.Ф. Обухова раскрывала понятие двойной стимуляции также цитатой из работ Л.С. Выготского: «Желая изучить внутреннюю структуру высших психических процессов, мы не ограничиваемся обычно предъявлением испытуемому простых стимулов (все равно — элементарных раздражителей или сложных задач), на которые ждем непосредственного ответа. Мы одновременно предъявляем испытуемому и второй ряд стимулов, которые функционально должны играть особую роль — служить средством организации его собственного поведения» [Выготский Л.С., 1983, С. 78].

Эти уточнения экспериментального метода и связанные с ним понятия крайне важны для оценки эффективности психолого-педагогических,

образовательных и социальных программ, технологий и практик. Очевидно, что в основе проектирования должно быть отчетливое представление о конечном результате, достигаемом в образовательном или социальном проекте, например, о тех действиях, которые сформируются в итоге. Необходимо планирование условий, обеспечивающих правильное выполнение действий детьми или подростками, а также условий, благодаря которым действия, перейдя во внутренний план, станут интрапсихической функцией. Наконец, необходимо детальное планирование системы внешних средств, которые сначала предложены взрослым (учителем, педагогом-психологом и др.) для овладения детьми их поведением, а затем станут их внутриспсихологическими средствами управления собой. Материалы, представляемые для экспертизы, должны содержать описания конечного результата, условий его достижения, а также давать представление о тех средствах, с помощью которых сначала во внешнем, а затем во внутреннем плане дети и подростки организуют свою деятельность. Анализ эмпирических данных, полученных в ходе реализации проекта, должен давать представление о том, насколько первоначальный замысел реализован, что является одним из важнейших показателей возможности передачи программы или технологии другим специалистам (учителям, педагогам-психологам, социальным педагогам и др.).

А.А. Марголис провел анализ понятия «зона ближайшего развития» в сравнении с понятием «скаффолдинг» (Д. Брунер) [Марголис А.А., 2020a]. Рассматривая возможности теории учебной деятельности и практики развивающего обучения Д.Б. Эльконина — В.В. Давыдова, ученый показывает достижения этого подхода и выделяет проблемные области с точки зрения соотношения житейских и научных понятий. А.А. Марголис отмечает, что, по мнению Выготского, формирование научных понятий в процессе взаимодействия учащихся с учителем в ходе организованного школьного обучения «оказывается возможным только на основе спонтанных понятий, для которых формируемые научные понятия становятся могучим средством влияния, выступая в качестве зоны их ближайшего развития. ... Проще говоря, учебная деятельность, построенная по типу ЗБР, строится таким образом, при котором формируемые научные понятия становятся зоной ближайшего развития спонтанных понятий и всего целостного процесса умственного развития ребенка» [Марголис А.А., 2020b, С. 9–10].

Проектирование психолого-педагогических, образовательных и социальных программ, технологий, практик можно трактовать как создание детско-взрослых сообществ, где взаимодействие в детской группе при решении задач (учебных, творческих, социальных) и выступает движущей силой развития. В научной школе В.В. Давыдова накоплен огромный опыт проектирования взаимодействий учитель-школьники и взаимодействий между школьниками, а также опыт успешной реализации учебной

деятельности. В социальных практиках примером такого опыта было коммунаское движение 1960–80-х годов XX века (О.С. Газман, И.П. Иванов, С.Д. Поляков, Р.В. Соколов и др.).

Таким образом, в работах Л.С. Выготского, А.Н. Леонтьева, П.Я. Гальперина, А.В. Запорожца, Д.Б. Эльконина, В.В. Давыдова, В.В. Рубцова, А.А. Марголиса и др., по существу, заложены основополагающие представления о подлинной эффективности развивающей и формирующей работы. Во-первых, по-настоящему эффективная развивающая работа ведет к структурной перестройке функции, на развитие которой она направлена. Структурная перестройка выражается не только в количественных показателях, пусть даже имеющих статистически значимое увеличение по сравнению с исходными. Шаг в развитии ребенка или подростка виден, что называется, невооруженным глазом, так как дети или подростки не только показывают, что предложенные им в ходе развивающей работы средства ими освоены, но и начинают создавать свои средства на основе предложенных им на занятиях, значительно преобразуя то, чему их научили.

Во-вторых, критерием становится завершенность интериоризации как перехода формируемой функции из внешнего плана во внутренний. Внешним признаком этого становится применения тех приемов и средств, которыми ребенок овладел, без помощи внешних опор. Например, в хорошо известном экспериментальном исследовании П.Я. Гальперина и С.Л. Кабыльницкой школьники, нуждавшиеся в развивающей работе, направленной на формирование внимания, вначале пользовались в качестве ориентировочной основы списком действий, которые им необходимо выполнить при проверке своих письменных работ. Далее список был заменен карточками с порядковыми номерами этих действий, затем, по мере интериоризации действий контроля отпала необходимость и в такой ориентировочной основе [Гальперин П.Я., Кабыльницкая С.Л., 1974].

В-третьих, по-настоящему эффективная практика расширяет зону ближайшего развития, что становится заметно по возникновению у детей или подростков новых видов деятельности, которые имеют очевидную связь с содержанием развивающей работы, но не копируют ее. Например, в исследовании Е.В. Чудиновой показано, как школьники, овладев действиями моделирования условий задачи, преобразуют способ моделирования почти до неузнаваемости, индивидуализируя способ и проявляя тем самым инициативу в построении модели учебной задачи [Чудинова Е.В., 2021].

**Психологические исследования образования в парадигме доказательного подхода.** В конце 50-х годов прошлого века развернули свои исследования научные коллективы под руководством Л.В. Занкова и Д.Б. Эльконина — учеников Л.С. Выготского и последователей научных школ культурно-исторической психологии и деятельностного подхода в образовании. Исследовательские коллективы были нацелены на создание развивающей практики начального общего образования. В ходе

накопления опыта реализации широкомасштабных проектов экспериментального обучения и воспитания данными коллективами сложился особый тип формирующего эксперимента. В научной школе В.В. Давыдова он получил название генетико-моделирующего эксперимента. Исследователь теперь не ограничивался изучением сложившихся форм психики, регистрацией и объяснением достигнутого учащимися уровня психического развития, что характерно для констатирующего (лабораторного) эксперимента. На основе предварительного теоретического анализа закономерностей психического развития детей определенного возраста, психолого-педагогических условий формирования учебно-познавательных способностей обучающихся строилась модель образовательного процесса в специально создаваемых условиях, как правило, в экспериментальных классах или школах.

Реализация разработанной модели тщательно контролировалась и оценивалась специалистами разного профиля — педагогами, психологами, методистами, врачами и т.д. В ходе эксперимента выявлялись закономерности, механизмы, динамика, индивидуальные особенности психического развития школьников. Результаты эксперимента позволяли подтвердить, уточнить либо опровергнуть предварительно разработанную модель образовательного процесса. В психологическом экспериментальном исследовании осуществлялся поиск и построение новых форм учебно-воспитательного процесса, осуществлялось своеобразное «выращивание» продуктивных форм сотрудничества педагогов и школьников и одновременно происходило изучение образовательных технологий развития детей на определенном возрастном этапе.

Формирующий эксперимент стал основным методом реализации доказательного подхода в образовании. Его достоинствами являются: ориентация на развитие обучающихся в образовательном процессе, теоретическая обоснованность экспериментальной модели организации этого процесса, репрезентативность экспериментальной выборки, длительность исследования, гарантирующая обоснованность и надежность получаемых данных, использование валидного научно-психологического инструментария для оценки развивающих возможностей экспериментальной модели обучения. Основные теоретические результаты в отечественных психолого-педагогических исследованиях образования были получены в ходе реализации конкретных формирующих экспериментов.

Радикальные преобразования в экономической и общественной жизни России на рубеже XX–XXI вв. привели к серьезным изменениям в сфере образования, оформлению практики развивающего, лично-ориентированного образования. Новая практика требует и нового психолого-педагогического обеспечения, предполагает разработку принципиально новых методов работы не только для образования, но и с самим образованием. В настоящее время сложился принципиально



новый тип экспериментальной работы в образовании, и главное — с самим образованием. В системе наук об образовании складывается культура проектно-преобразующих экспериментальных разработок. Оформляется научность нового типа — наряду с научными знаниями складываются проектно-программные, инженерно-конструкторские знания, в наибольшей степени адекватные именно развивающемуся образованию.

Комплексный проект по реформированию педагогического образования, реализующий деятельностный подход, был проведен в 2014–2017 годах Московским государственным психолого-педагогическим университетом совместно с Национальным исследовательским университетом «Высшая школа экономики». В реализации комплексного проекта приняли участие 65 образовательных организаций высшего образования, расположенных в 51 субъекте. Данный проектно-преобразующий эксперимент строился в полном соответствии с требованиями доказательного подхода в образовании. Отметим основные достоинства проведенного эксперимента, его принципиальную научную новизну и практическую значимость.

В первую очередь выделим реализованный в эксперименте практико-ориентированный характер подготовки будущих педагогов. Практическая направленность эксперимента по реформированию педагогического образования проявилась в его целевых, содержательных, организационных, результирующих аспектах. В основу эксперимента была положена деятельностная модель педагогического образования, ставящая своей целью формирование у будущих педагогов способностей к проектированию и реализации своей профессиональной деятельности в соответствии с Профессиональным стандартом педагога. Освоение педагогической деятельности «предполагает овладение профессиональными (трудовыми) действиями, описанными в профессиональном стандарте, а также компетенциями и знаниями, необходимыми для их осуществления» [Марголис А.А., 2014. С. 106].

Деятельностный подход реализован в модульном принципе строения основной профессиональной образовательной программы подготовки педагогов. Переход от традиционного дисциплинарного принципа строения образовательной программы к модульному означал отказ от знаниевой парадигмы профессионального образования (знание начало и конец образования) и следование профессионально-деятельностной парадигме подготовки педагогов. В эксперименте отработаны новый принцип строения учебного модуля и этапы его изучения. Каждый модуль включает как теоретическое содержание, так и практики, решающие задачи «постановки проблемы осуществления профессионального действия и его отработки в специально организованной лабораторно-учебной среде (практикум) и на практической «клинической» базе (в условиях реальной образовательной организации)» [там же, с. 108]. При этом каждый модуль нацелен



на освоение конкретной трудовой функции педагогической деятельности в логике деятельностной образовательной технологии.

Особое внимание в образовательной программе подготовки педагогов, построенной на деятельностном подходе, уделяется практике. Реализация практико-ориентированной образовательной программы предполагает выделение на практику студентов бакалавриата до 60–80 зачетных единиц. При этом на выпускном курсе планируется долгосрочная производственная практика на одной из клинических баз, организованных на принципах школьно-университетского партнерства.

Модернизация педагогического образования на принципах деятельностного подхода предполагает внесение серьезных изменений в итоговую оценку качества подготовки выпускников. Предлагается переработать содержание государственной итоговой аттестации, включив в нее не только проверку знаний и компетенций, но и готовность выпускника к осуществлению профессиональных действий. Предлагается расширить тематику выпускных квалификационных работ, включив в нее темы, нацеленные на исследование и решение актуальных проблем профессиональной педагогической деятельности. Также предлагается проводить независимую оценку квалификации выпускников педагогических программ общественно-профессиональной ассоциацией или органами управления образованием [Марголис А.А., 2014; Марголис А.А., 2021; Марголис А.А., Сафронова М.А., 2018].

Итоги комплексного проекта по модернизации педагогического образования, построенного на принципах деятельностного подхода, дают основания утверждать, что он вобрал в себя все лучшие психолого-педагогические разработки по совершенствованию практики подготовки будущих педагогов. Разработана, обоснована и экспериментально подтверждена модель педагогического образования, отвечающая потребностям современного отечественного образования в квалифицированных педагогических кадрах. Перспектива дальнейшей работы — исследовательской и практической — в наполнении представленной модели деятельностным содержанием, отвечающим задачам подготовки педагогов к педагогической деятельности в дошкольном, начальном общем, основном общем, среднем общем образовании.

#### **1.4. Предпосылки возникновения и развития доказательного подхода в социальной сфере за рубежом**

В предыдущей главе были рассмотрены основания для возникновения и внедрения доказательного подхода в практическую деятельность специалистов в области психологии и образования в нашей стране. Но также важно рассмотреть, каким образом данное направление развивалось в глобальном мировом контексте, поскольку современный

доказательный подход представляет собой синтез отечественных теоретико-методологических подходов и зарубежного опыта внедрения доказательного подхода в области медицины и социальной помощи.

В настоящее время существует общепринятое мнение, что использование эмпирических данных в социальной практике является не только полезным, но и отвечает принципам этичности. Несмотря на то, что научное знание не может быть исчерпывающе полным из-за постоянно меняющихся социальных условий, поддержание и укрепление связи науки и практики необходимо для предоставления благополучателям качественной помощи и поддержки.

В 1917 году в своей работе «Социальные диагнозы» пионер в области образования американских социальных работников Мэри Эллен Ричмонд одной из первых начала пропагандировать идею использования результатов эмпирических исследований при принятии управленческих решений в социальной сфере [Rubin A., 2015]. В дальнейшем идея значимости научных знаний для практики социальной работы в противовес опоре на интуицию, опыт и экспертность отдельных специалистов неоднократно возникала в поле социальной помощи. Одной из наиболее значимых работ в этой области считается книга «Оценка социального воздействия» под редакцией Э. Мюллена и Дж. Дампсона, изданная в 1972 году [Mullen E.J., Dumpson J.R. (Eds.), 1972]. Изучив шестнадцать ключевых социальных проектов, реализованных в течение 1971 года, и собрав экспертные мнения относительно своего исследования, авторы сделали вывод о том, что данные программы далеко не всегда достигают заявленных социальных результатов. В результате экспертных обсуждений на конференции, собравшей 125 представителей ведущих американских организаций, занимавшихся разработкой основ социальной работы, были предложены рекомендации по изменению методологии социального воздействия и обучения специалистов социальной сферы, которые впоследствии были осмыслены как доказательный подход.

Попытки внедрения доказательного подхода в социальную сферу часто сталкивались с различными трудностями, например, отсутствием мотивации и знаний у специалистов, позволяющих использовать результаты научных исследований в своей практике или несовпадением результатов исследований с методиками, показавшими свою эффективность на практике [Manuel J.I. et al., 2009]; [Proctor E.K., Rosen A., 2008]. Специалисты социальной сферы предпочитали опираться на мнение своих более статусных коллег, супервизоров и экспертов, а также на традиции, принятые в их организации [Бусыгина Н.П., Подушкина Т.Г., Станилевский В.В., 2020]. Разрыв между эмпирическими исследованиями и практикой наблюдался также в сфере медицины и здравоохранения, что вызывало определенную тревогу в соответствующих социальных службах [Mullen E.J., Bledsoe S.E., & Bellamy, J.L., 2008].

Культура обращения к научной методологии сбора оснований, доказывающих эффективность социальных и образовательных практик начала складываться в западных странах благодаря доказательной медицине, возникшей в Канаде в 90-х годах прошлого века. Канадский врач Д. Сакетт с коллегами разработали методологию доказательной медицины как способ преодоления концептуального разрыва между наукой и практикой и улучшения качества оказания медицинской помощи. Они дали следующее определение доказательной медицины: «ответственное, корректное и разумное использование полученных в исследованиях фактических данных при принятии решения о подходе к лечению конкретных пациентов» [Sackett D.L. et al., 1996, С. 71]. Впоследствии авторы расширили понятие доказательной медицины, обратив внимание на то, что полученные в исследованиях данные должны быть интегрированы с клиническим опытом специалиста, а также уникальными ценностями и опытом пациента [Sackett D.L. et al., 2000]. Таким образом, доказательный подход появился как связующее звено между практикой и наукой, объединяя опыт академических исследований с клинической экспертизой и ожиданиями пациентов [Straus S.E. et al., 2005, С. 1]. Со временем доказательный подход в медицине начал распространяться в области социальной помощи, образования, психологии. Методология доказательного подхода в социальной сфере не является точной копией подходов, принятых в медицине, поскольку психосоциальные практики обладают собственной определенной спецификой. Ориентация на клинические исследования препаратов, проводимые в медицине, не может быть признана удовлетворительной для социальной сферы в силу ряда причин (отсутствие характерной для медицины очевидной связи между позитивистской наукой и практикой в психологии и социальных науках; гораздо меньшая определенность в понимании мишеней психологических и социальных практик и операционализации переменных; значительное разнообразие теоретико-методологических оснований психологических и социальных практик, теоретическая нагруженность ожидаемых эффектов и проблема возможности применения универсальной (независимой от фундирующей практику теории) методологии для их отслеживания и фиксации).

Основываясь на более широком определении доказательной медицины Д. Сакетта, К. Макнис и Б. Тайер описывают доказательный подход в социальной сфере как «интеграцию лучших научных данных с клиническим опытом и ценностями клиентов при принятии практических решений» [McNeece C.A., Thyer B.A., 2004, С. 9]. По мнению авторов, практика с доказанной эффективностью (evidence-based practice) должна быть основана на научных данных (соответствующих фундаментальных и прикладных научных исследованиях, исследованиях вмешательства в отношении результатов и методики оценки), профессиональном опыте специалиста (способности оценивать факторы окружающей среды, понимания ценно-

стей и предпочтений благополучателя), ценности благополучателя (его уникальных предпочтений, ожиданий, которые должны быть интегрированы в практические решения). Н.П. Бусыгина отмечает, что термин *evidence-based practice* в англоязычных источниках подразумевает практику как продукт, имеющий эмпирически подтвержденную эффективность и предлагает рассматривать термин «доказательная практика» в двух качествах: как ориентацию на применение практик с эмпирически подтвержденной эффективностью, чьи эффекты зафиксированы в соответствующих научных исследованиях, и как обоснованный процесс осуществления практики [Бусыгина Н.П., Подушкина Т.Г., Станилевский В.В., 2020]. Процесс доказательной практики включает в себя пять этапов: формулирование практического вопроса, на который можно ответить; поиск эмпирических данных в научных изданиях и практических руководствах; критическая оценка эмпирических данных; выбор типа интервенции с учетом интеграции эмпирических данных с клиническим опытом, предпочтениями и ценностями клиента и оценка эффективности выбранных практических решений [Straus S.E. et al., 2005].

С момента своего появления доказательный подход в социальной сфере вызывает оживленные дискуссии о его ценности для практиков. Сторонники применения доказательного подхода в социальной сфере аргументируют необходимость его внедрения в практику, во-первых, этической составляющей ответственности специалистов за качество и эффективность предоставляемых социальных услуг [Gambrell E., 2011]. Отсутствие эмпирически полученных доказательств могут приводить на практике к неэффективным воздействиям или воздействиям, приносящим больше вреда, чем пользы. Во-вторых, применение доказательного подхода повышает профессиональный уровень сотрудников социальных служб за счет развития исследовательской культуры и практики критического осмысления. Кроме того, использование научных данных повышает доверие благополучателей и грантодающих организаций к работе служб социальной помощи, в частности, их отчетной документации.

Критика доказательного подхода обычно основана на исключении критикующей стороной нескольких основополагающих элементов данной методологии и фокусировании внимания на том факте, что процесс принятия решения о социальном воздействии подразумевает лишь поиск наилучших доступных данных. При этом из поля зрения упускается профессиональный опыт специалиста, а также потребности и ценности благополучателя [Mullen E.J., Streiner D.L., 2004]. Некоторые исследователи ставят под вопрос методологию сбора данных, представленных для разработки практик, например, систематических обзоров, полученных при помощи мета-анализа [Pawson R., 2002].

Как показывает опыт внедрения доказательного подхода за рубежом, поиск его эффективных стратегий в практике социальной работы

осложняется недостаточным количеством эмпирических исследований эффективности практик, а также исследований, касающихся установок, ценностей и компетенций специалистов социальной сферы относительно его применения [Zwet R van der., 2018]. Несмотря на критику и продолжающиеся обсуждения профессиональным сообществом ценности и ограничений доказательного подхода, на рубеже тысячелетий он стал доминирующей парадигмой интеграции научного знания и практики, и программы психологической и социальной помощи, в основном, ориентируются на разработанные в его рамках принципы и стратегии работы.

### **1.5. Современные прикладные экспериментальные исследования как примеры доказательного подхода в проектировании психолого-педагогических и образовательных программ, технологий, практик**

В качестве иллюстраций доказательного подхода приведем несколько кейсов — примеров психолого-педагогических, образовательных и социальных программ, технологий, практик, описанных в литературе. На этих примерах можно проследить как реализацию в полном объеме положений, сформулированных выше, так и трудности, которые могут возникать при верификации подобных программ. Подобные трудности могут быть связаны с немногочисленностью выборки, ограниченными возможностями формализации и трудностью квантификации достигнутых результатов. В некоторых случаях принципиальная невозможность участия многочисленной выборки в реализации программы или технологии может создавать препятствия для статистической оценки результативности.

#### **Кейс 1. Верификация доказательной базы технологии индивидуализации обучения с классической схемой эксперимента и количественными методами**

Рассмотрим пример верификации доказательной базы 3-х образовательных программ по математическим методам в психологии и психолого-педагогических исследованиях на базе технологии индивидуализации обучения «перевернутый класс» и электронных курсов в парадигме доказательного подхода. В 2019–2021 г.г. в МГППУ был реализован научно-исследовательский проект «Цифровые технологии в высшем образовании: разработка технологии индивидуализации обучения средствами электронных учебных курсов (ЭУК)» (руководитель М.Г. Сорокова, исследовательская группа М.А. Одинцова, Н.П. Радчикова, С.Ю. Игнашев). Предметом исследования стали:

- возможности ЭУК как средства индивидуализации обучения на примере курсов «Математические методы в психологии» (ЭУК ММВП) для программ бакалавриата по направлению 37.03.01 «Психология» и специалитета по направлениям 37.05.01 «Клиническая психология»,

- «Педагогика и психология девиантного поведения»; «Статистические и математические методы в психолого-педагогических исследованиях» (ЭУК СММвППИ) для магистратуры направления 44.04.02 «Психолого-педагогическое образование»; «Математическая обработка данных психолого-педагогических исследований в программе SPSS» (ЭУК МОДППИвSPSS) для магистратуры направления 37.04.01 «Психология» (в настоящее время курс называется «Статистические методы в психологических исследованиях»).
- оценки студентами своего опыта обучения в цифровой образовательной среде (ЦОС) университета и их академические достижения на примере ЭУК ММвП для бакалавриата и специалитета, ЭУК СММвППИ и ЭУК МОДППИвSPSS для магистратуры.

Цели исследования:

- разработать базовый вариант технологии индивидуализированного обучения студентов средствами электронных курсов на примере ЭУК по математическим методам в психологии и в психолого-педагогических исследованиях и оценить ее возможности применительно к формированию положительного отношения студентов к электронному обучению, предмету, и достижению ими хороших образовательных результатов;
- сравнить отношение студентов разных уровней образования к обучению в цифровой образовательной среде (ЦОС) университета, их вовлеченность в учебный процесс (поведенческий компонент) и образовательные результаты; выявить сходство и различия;
- сравнить образовательные результаты студентов разных уровней высшего образования, завершивших обучение в электронных курсах в смешанном и онлайн-форматах;
- разработать и валидизировать диагностический инструмент «Шкала оценки цифровой образовательной среды (ЦОС) университета».

Кратко опишем дизайн исследования. Форматы изучения ЭУК были в 1-й год — смешанное обучение [Марголис А.А., 2018], во 2-й год — онлайн-обучение на основе модели «перевернутый класс» с применением элементов технологии скаффолдинга (Д. Брунер) [Марголис А.А., 2020а], основными характеристиками которой являются «угасающая помощь» и ориентация на зону ближайшего развития студента.

Методы и инструменты сбора данных: входное, выходное и текущее тестирование и оценка выполнения кейс-задания по ЭУК; рубежный контроль — тестирование через Отдел мониторинга качества (ОМКПО) МГППУ отдаленных результатов; анкета «Мнения студентов о ЭУК» в режиме обратной связи через LMS MOODLE; тестирование студентов и преподавателей через Гугл-форму при разработке диагностического инструмента «Шкала оценки ЦОС университета»; Опросник переживания

в деятельности (Леонтьев Д.А., Осин Е.Н. и др.) для изучения субъективной репрезентации соотношения текущей учебной деятельности с усилиями, удовольствием, смыслом/пустотой; Шкалы академической мотивации (Гордеева Т.О., Сычев О.А., Осин Е.Н.) для изучения внутренней и внешней мотивации учебной деятельности.

Применялись следующие методы анализа данных: количественные методы, в том числе, описательная статистика, критерий Манна — Уитни, критерий Уилкоксона, критерий Хи-квадрат, методы многомерной статистики: двухфакторный дисперсионный анализ для смешанной схемы, кластерный анализ, множественный регрессионный анализ, логистическая регрессия.

Общий объем выборки за 2 года составил  $N = 809$  студентов 45 академических групп МГППУ. Подробное описание выборок см. в перечисленных ниже статьях.

- Результаты исследования за 1-й год проекта представлены в статьях [Сорокова М.Г., 2020с], [Сорокова М.Г., 2020d], [Sorokova Marina G., 2020e]. Приведем основные результаты 1-го года:
- Разработаны и апробированы ЭУК по математическим методам в психологии и в психолого-педагогических исследованиях для бакалавриата (37.03.01), специалитета (37.05.01, 44.05.01) и магистратуры (44.04.02, 37.04.01) как цифровые образовательные ресурсы для реализации смешанного обучения и онлайн-обучения в университете.
- Эмпирически подтверждена эффективность ЭУК как инструмента формирования положительного отношения студентов к смешанному обучению в формате ЭУК и их позитивной самооценки вовлеченности в учебный процесс. В своих положительных отзывах студенты отметили значительно большую эффективность, чем при очном обучении; высокую информативность курса, структурированность материала и доступность его изложения в разных форматах с большим количеством примеров и аналогий; практическую пользу для научной работы. Выявлены достоверные связи между положительной оценкой объема, пользы и практической применимости полученных знаний с позитивным отношением к формату ЭУК и смешанному обучению.
- Показано, что формат ЭУК способствует статистически значимому росту академических достижений студентов по сравнению с исходным уровнем.
- Выявлены достоверно лучшие образовательные результаты студентов группы смешанного обучения в формате ЭУК (ЭГ,  $N = 78$ ) по сравнению со студентами традиционно-очного обучения («КГ»,  $N = 309$ ): на выходе знания студентов ЭГ достоверно выше, чем «КГ» ( $p < 0,001$ ), однако по абсолютной величине разность в средних невелика
- Гипотеза о том, что студенты магистратуры и студенты, обучающиеся на базе высшего образования (ВВ), будут испытывать больше трудностей при изучении ЭУК, чем студенты первого высшего образования (ПВ), а их отношение к формату электронных курсов будет более



скептическим, не подтвердилась. Для обеих категорий студентов была выявлена совокупность типичных мнений-предикторов об обучении в электронных курсах, характерных для каждой категории.

- В целом, у подавляющего большинства студентов обеих категорий существенных трудностей в изучении ЭУК в смешанном формате выявлено не было, тесты оказались полезны для лучшего усвоения материала, ЭУК действительно позволил им отслеживать свою индивидуальную траекторию, а личных контактов с преподавателем было достаточно. Это опровергает сложившиеся стереотипы.

Результаты 2-го года проекта представлены в публикациях: [Радчикова Н.П., Одинцова М.А., Сорокова М.Г., 2021], [Сорокова М.Г., 2021], [Сорокова М.Г., Одинцова М.А., Радчикова Н.П., 2021a], [Сорокова М.Г., Одинцова М.А., Радчикова Н.П., 2021b]. Приведем основные результаты 2-го года проекта:

- Подтвердилась гипотеза о том, что обучение в ЭУК в онлайн-формате по технологии индивидуализированного обучения не приведет к снижению академических достижений студентов по сравнению со смешанным обучением.
- Проведен сравнительный анализ образовательных результатов студентов, завершивших курсы по математическим методам в психологии при смешанном (N=404) и онлайн-обучении (N=405) (всего N = 809).
- Студенты онлайн-обучения показали в среднем более высокие результаты по сравнению с группой смешанного обучения по входному тесту, по итоговому тесту и по итоговой оценке за курс, однако разность в средних по всем трем показателям минимальна, а значимость различий обеспечивается большим объемом выборки. На меньшей выборке на этапе итогового тестирования и итоговой оценки за ЭУК значимых различий при смешанном и при онлайн обучении не выявлено
- Предметные результаты студентов магистратуры и ВВ, с одной стороны, и программ ПВ — с другой, также практически одинаковы: различия по итоговому тесту и итоговой оценке за ЭУК находятся на грани статистической значимости, причем разность в средних минимальна и составляет лишь около одного процентного пункта.
- Динамика изменения средних значений по показателям академических достижений в группах смешанного и онлайн обучения (Рис. 3.) демонстрирует весьма выраженный размер эффекта — около 50 процентных пунктов — рост показателей итогового теста по сравнению с входным, а затем менее выраженный спад результатов через 1–1,5 месяца, которые все же остаются значимо выше входных. При этом в группе онлайн-обучения спад очень незначительный, то есть динамика лучше. Последний результат требует дальнейшей проверки при более уравненных условиях внешнего тестирования.



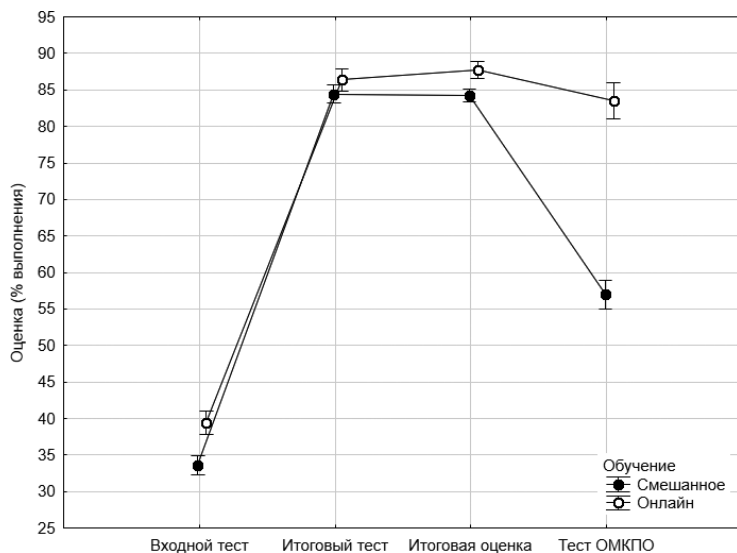


Рис. 3. Динамика изменения средних значений по показателям академических достижений в группах смешанного и онлайн обучения (по: Сорокова М.Г., Одинцова М.А., Радчикова Н.П., 2021а)

Разработана и стандартизована «Шкала оценки цифровой образовательной среды (ЦОС) университета» (N=406; 366 (90,1%) женщин; возраст от 19 до 72 лет, в среднем  $28,7 \pm 9,6$  лет (медиана 24 года)), позволяющие дать комплексную характеристику ЦОС на основании выделения шести индикаторов: удовлетворенность учебным процессом; удовлетворенность коммуникативным взаимодействием; стресснапряженность; необходимость поддержки; нечестные стратегии при контроле знаний; доступность. Доказана конструктивная валидность и внутренняя согласованность субшкал. Результаты конфирматорного факторного анализа подтверждают наличие шести субшкал (IFI=0,87;  $\chi^2/df=2,6$ ; RMSEA=0,06 [0,058; 0,066]; SRMR=0,06). Все субшкалы обладают приемлемой надежностью (альфа Кронбаха = 0,72–0,91, альфа Гутмана = 0,82–0,92) и демонстрируют предсказуемые взаимосвязи с показателями: переживания в учебной деятельности (усилия, удовольствие, смысл); познавательная мотивация, мотивация достижения, мотивация саморазвития, интроецированная мотивация, экстернальная мотивация, амотивация. Доказана критериальная валидность и расщепленная надежность. Рассчитаны станаины [Сорокова М.Г., Одинцова М.А., Радчикова Н.П., 2021b].

## **Кейс 2. Верификация доказательной базы просветительской программы с классической схемой эксперимента и количественными методами**

Рассмотрим пример верификации просветительской программы в парадигме доказательного подхода. В феврале — марте 2022 г. в Московском государственном психолого-педагогическом университете в рамках реализации программы воспитания проведен эксперимент по верификации программы просветительских мероприятий в области вакцинопрофилактики COVID-19 для студентов. Результаты эксперимента представлены в статье [Марголис А. и др., 2023]. База данных размещена в Репозитории психологических исследований и инструментов RusPsyDATA [Марголис А.А. и др., Датасет, 2023].

Цель исследования — выявление наиболее эффективных стратегий работы с установками студентов при выборе ими решения вакцинирования от коронавирусной инфекции.

Исследование проведено в три этапа дистанционно. На первом этапе — констатирующем — осуществлялся сбор данных: студенты, подписавшие «Информированное согласие», добровольно и анонимно заполняли гугл-формы «Анкеты отношения к вакцинации от COVID-19» и 6 опросников исследования. На втором этапе — формирующем — для анализа эффективности воздействия просветительских мероприятий в области вакцинопрофилактики студенты были поделены на четыре подгруппы: первая — контрольная, составили студенты, кто только слушал лекции специалистов в области медицины и вакцинопрофилактики; вторая — 1 экспериментальная — помимо лекций посещали семинары в традиционном формате; третья — 2 экспериментальная — вебинары с разбором кейсов на заданные темы (история человека, вакцинированного от COVID-19 и не заболевшего; история человека, вакцинированного от COVID-19 и переболевшего в легкой форме; история человека, не вакцинированного от COVID-19 и переболевшего в тяжелой форме; история, предложенная студентами во время занятий); четвертая — 3 экспериментальная — занятия с проведением студенческих дебатов. На третьем этапе — контрольном — студенты заполняли «Анкеты отношения к вакцинации от COVID-19» повторно. Объем аналитической выборки  $N = 780$ .

Методы количественного анализа данных: описательная статистика, однофакторный дисперсионный анализ ANOVA, парный t-критерий Стьюдента, критерий однородности Хи-квадрат, критерий Колмогорова — Смирнова, эксплораторный факторный анализ. Анализ выполнен в статистическом пакете SPSS V.23.

Кратко перечислим основные результаты эксперимента. Выделено 5 аспектов отношения к вакцинации от COVID-19 — измерительных шкал: польза вакцинации от коронавируса для человека и общества; отрицание опасности коронавируса и надежда на естественный иммунитет;

страх побочных эффектов вакцинации от коронавируса и недоверие к информации о безопасности вакцинирования; уверенность в серьезных негативных последствиях вакцины от коронавируса; неверие в доказанность эффективности российских вакцин на международном уровне. Проведен сравнительный анализ эффективности 4-х стратегий воздействия на отношение студентов к вакцинации от COVID-19: только лекции (КГ), лекции в сочетании с семинарами/вебинарами в традиционном формате (ЭГ1), с разбором кейсов из жизни студентов (ЭГ2), с проведением студенческих дебатов (ЭГ3). Различий между группами по полу нет. Различия по возрасту есть, но разность в средних не превышает 2-х лет.

На констатирующем этапе различий по шкалам нет, на контрольном этапе различия выявляются. По всем 5-ти шкалам положительный и отрицательный эффекты в целом по выборке почти равновероятны. Принципиально лучшую стратегию воспитательного воздействия выделить не удалось. В целом, эффективность просветительских мероприятий несколько больше при проведении лекций в сочетании с разного рода семинарскими занятиями по сравнению с только лекциями, однако эффективность везде невелика. Стандартные размеры эффекта  $d$  Коэна не превышают 0,44. Несколько выше эффективность при проведении лекций в сочетании с семинарами/вебинарами в традиционном формате или разбором кейсов. Стратегия, где помимо лекций проводились студенческие дебаты, себя не оправдала. Более подробно об этих и других результатах эксперимента см. в статье [Марголис А.А. и др., 2023].

### **Кейс 3. Верификация доказательной базы развивающей программы с классической схемой эксперимента при сочетании количественных и качественных методов**

Другой пример эффективной практики мы видим в экспериментальном исследовании А.З. Зака [Зак А.З., 2018], где для развития универсальных учебных действий (анализа и рефлексии) применена разработанная им развивающая программа «Интеллектика». В исследовании участвовали 117 учеников 4 класса (46 — экспериментальная группа, 71 — контрольная), 46 детей занимались по программе «Интеллектика» весь учебный год (32 занятия). Начальная и заключительная диагностика проводилась с учениками обеих групп на материале заданий авторской методики «Замещение», предназначенных для определения способа решения поисковых проблем (общий или частный) — задание 1 и рефлексии при решении задач (содержательная или формальная) — задание 2.

В данном случае реализовать принципы доказательности значительно проще, поскольку контрольная и экспериментальная группы достаточно многочисленны, а данные по диагностическим методикам легко представить в количественном виде (Табл. 1, по Зак А.З., 2018).

Таблица 1

**Результаты выполнения детьми контрольной и экспериментальной групп задания 1 в сентябре и мае, %**

Характеристика решения задач	Контрольная группа		Экспериментальная группа	
	Сентябрь	Май	Сентябрь	Май
Отсутствие решения	11,3	5,6	15,2	0,0
Решение частным способом	56,3	52,1*	54,5	36,9*
Решение общим способом	32,4	42,3**	30,4	63,1**

Примечание: \* –  $p < 0,05$ ; \*\* –  $p < 0,01$ .

Наряду с примененным в работе статистическим анализом данных, в подобном исследовании могут быть применены экспертные оценки учителей, а также повторная (через несколько месяцев) диагностика для оценки отсроченного эффекта, проявляющегося в дальнейшем развитии анализа и рефлексии, индикатором которого были бы статистически значимое повышение количественных показателей. В цитируемой публикации не приведены некоторые показатели, существенные для оценки доказательности, что, очевидно, обусловлено форматом журнальной статьи. Опыт применения такой развивающей программы может быть передан другим работникам образования (учителям, педагогам-психологам); для этого достаточно курсов повышения квалификации небольшой продолжительности.

**Кейс 4. Верификация доказательной базы деятельностной технологии при сочетании количественных и качественных методов**

В статье О.В. Рубцовой, Т.А. Посакаловой и А.Г. Соловьевой [Рубцова О.В., Посакалова Т.А., Соловьева А.Г., 2022]. представлены результаты проекта «Подростковый театр как деятельностная технология воспитания и формирования личностных образовательных результатов», реализованного Центром междисциплинарных исследований современного детства МГППУ в 2021–2022 гг. Авторы доказывают, что школьный театр эффективен как средство формирования личностных образовательных результатов и метапредметных компетенций. Большим достоинством программы является ее соответствие возрастным особенностям подростков, прежде всего ведущим типам деятельности, к которым относится ролевое экспериментирование. Данную практику трудно отнести однозначно к образовательным либо социальным, поскольку, наряду с образовательными эффектами, она имеет очевидную социальную значимость, наделяя подростков мощной системой средств овладения своим поведением и своим внутренним миром, прежде всего, эмоциональной сферой.

Программа, представленная в статье, включала серию занятий театральной деятельностью (30 уроков по 45 минут). Содержание данного проекта таково, что методы исследования (наблюдение, видеофиксацию и последующий анализ видеоматериалов занятий; анализ продуктов деятельности; регулярные интервью с подростками и учителями, участвовавшими в проекте) поддаются формализации с трудом, что делает довольно затруднительным соблюдение некоторых формальных требований, касающихся доказательности. Занятия театральной деятельностью были бы в принципе малоэффективны в многочисленной группе, поэтому участие только десяти человек вынуждает авторов ограничиваться описательной статистикой. Поэтому для доказательства эффективности программы авторы используют оценки самими участниками исследования их эмоционального состояния на занятии и после него (Рис. 4), а также метод кейсов. С точки зрения принципов культурно-исторической и деятельностной теорий, обсужденных в предыдущем разделе, реализация программы действительно имела заметные положительные эффекты, поскольку

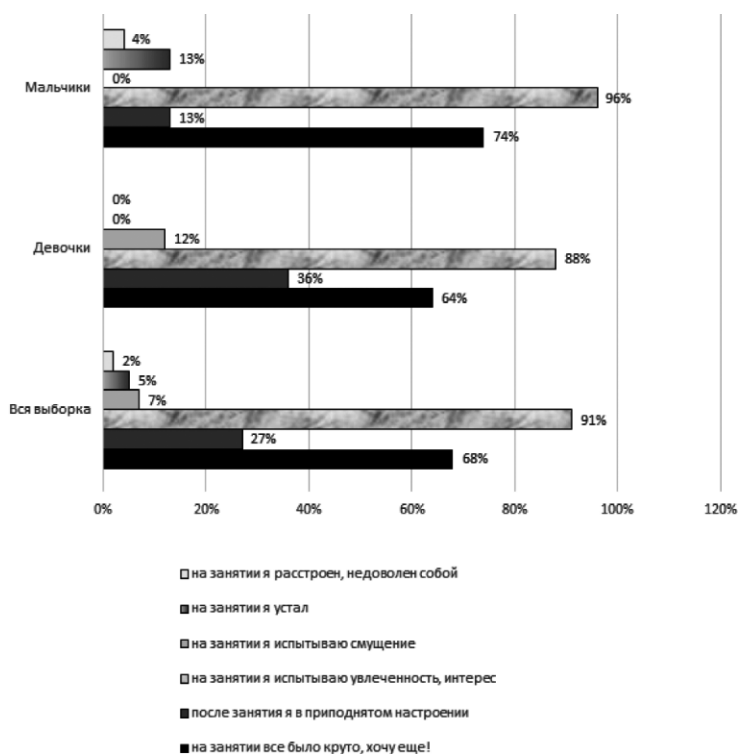


Рис. 4. Эмоциональное восприятие подростками занятий в рамках проекта (по: Рубцова О.В., Посакалова Т.А., Соловьева А.Г., 2022)

по приведенным авторами кейсам видно, как подростки-участники проекта проявляют в жизни компетенции, сформированные в ходе занятий театральной деятельностью. Однако формализация результатов, представленных в таком виде, ограничена и, следовательно, возможности применения математической статистики для оценки эффектов также ограничены. Передача такого опыта также имеет ограничения, поскольку программа может быть успешно реализована педагогом с театральным образованием, имеющим опыт работы с подростками. Подчеркнем однако, что авторы программы в 2023 г. провели исследование эффективности на большой выборке с применением стандартизированных тестов и опросников и количественных методов исследования в сочетании с качественными.

### **Кейс 5. Верификация доказательной базы рефлексивно-деятельностного подхода с акцентом на качественных методах**

Приведем пример технологии, которая давно зарекомендовала себя в общественном мнении как эффективное средство помощи детям и подросткам в преодолении учебных трудностей разного уровня сложности, вплоть до выученной беспомощности в учебной деятельности. В.К. Зарецкий первоначально предложил созданный им рефлексивно-деятельностный подход (далее — РДП) как средство помощи детям с ограниченными возможностями здоровья [Зарецкий В.К., 2013], однако в дальнейшем была доказана эффективность РДП в помощи детям и подросткам с нормотипическим развитием. Сущность РДП, по словам В.К. Зарецкого, в том, что «...процесс оказания помощи в преодолении трудностей в учебной деятельности — это процесс содействия выращиванию внутренних средств деятельности. А побочный продукт — избавление от тех внутренних препятствий, которые мешали осуществлению этой деятельности» [Зарецкий В.К., 2013, С. 15]. Между тем «...именно этот принцип — поддержка субъектного отношения ребенка к осуществляемой им деятельности — оказывается наиболее трудно реализуемым во взаимодействии ребенка и взрослого (учителя, консультанта, родителя)» [Зарецкий В.К., 2013, С. 15], что сдерживает возможности передачи данной технологии более широкому кругу людей, поскольку для того чтобы учитель или педагог-психолог был в полной мере готов к реализации подхода необходим не только опыт помощи детям, но и супервизия. В целом РДП представляет собой яркий пример педагогики сотрудничества — взаимодействие ребенка и взрослого основано на взаимных договоренностях и принципиальном равноправии взрослого и ребенка в процессе проработки учебных трудностей.

Для оценки эффектов РДП в процессе преодоления учебных трудностей В.К. Зарецкий предложена многовекторная модель зоны ближайшего развития (Рис. 5).

Затруднительность оценки эффективности РДП с помощью стандартных количественных процедур заключается в том, что движение в развитии ребенка или подростка оценивается при помощи содержательного анализа высказываний школьников во время занятий. Высказывания с некоторой долей условности можно отнести к тому или иному вектору развития (способности действовать в уме, плоскости отношения к трудностям и др.), однако некоторая доля субъективности в анализе высказываний неизбежна. Для оценки одного из важнейших результатов применения РДП, а именно формирования субъектной позиции школьников по отношению к учебной деятельности, Ю.В. Зарецким, В.К. Зарецким и И.Ю. Кулагиной был разработан опросник «Субъектная позиция» [Зарецкий Ю.В., Зарецкий В.К., Кулагина И.Ю., 2014], позволяющий оценить количественные показатели эффективности

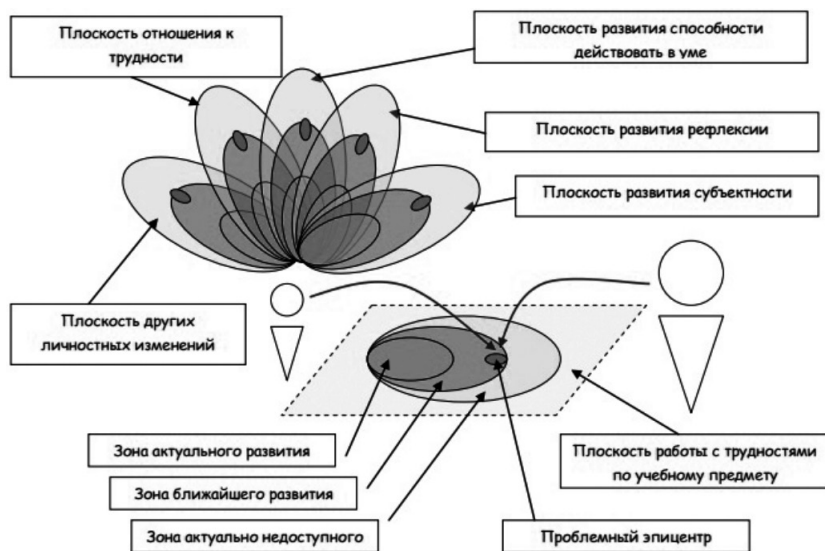


Рис. 5. Схема зоны ближайшего развития как совокупности векторов, по которым возможны «шаги» в развитии (по: Зарецкий В.К., 2013, Зарецкий В.К., 2016)

в этом направлении. Развитие отдельных видов рефлексии при реализации РДП также может быть оценено при помощи стандартизированных методик, например, предложенных А.З. Заком [Зак А.З., 2018]. Таким образом, рефлексивно-деятельностный подход также содержит хороший потенциал для проведения масштабированного эксперимента по верификации этой технологии в парадигме доказательного подхода с более активным привлечением стандартизированных опросников и количественных методов анализа данных.

В целом проведенный нами анализ кейсов успешных образовательных и социальных программ и технологий показывает, что доказательный подход в проектировании глубоко укоренен в культурно-исторической теории и деятельностном подходе, которые стали теоретико-методологической основой современного российского образования. Анализ теоретических положений и экспериментальных методов, созданный в данной научной школе, позволяет сделать ряд важнейших выводов относительно проектирования психолого-педагогических программ и технологий, а также выявить проблемы их качественного ассесмента и верификации доказательной базы в рамках доказательного подхода. В то же время, проведенный нами анализ позволяет и наметить пути преодоления этих проблем, для того чтобы действительно эффективные программы и технологии, содействующие развитию детей и подростков и расширяющие их зону ближайшего развития, получали бы заслуженно высокую оценку.

### **1.6. Ключевая формула доказательности в эксперименте: «количественные методы плюс опционально качественные методы»**

Проведенный нами анализ предпосылок доказательного подхода в проектировании психолого-педагогических, образовательных программ, технологий и социальных практик в отечественной психологической науке имеет не только теоретическое значение. Практический смысл заключается в том, что при оценке развивающих эффектов психолого-педагогической, образовательной или социальной практики при помощи одних лишь тестов и опросников (т.е. методик, направленных на оценку уровня актуального развития) порой ускользают именно те аспекты эффективности, которые определяются зоной ближайшего развития. Необходимы также наблюдение, архивный метод и другие качественные методы оценки результатов, которые позволят дать исчерпывающе полное представление об эффективности предложенной программы или технологии.

Вместе с тем, если применение количественных методов эмпирического исследования позволяет дать ясный ответ на вопрос, в чем выражается эффект применения психолого-педагогической, образовательной программы или технологии, а также на какие конкретно конструкты, психические функции и процессы, метапредметные компетенции и универсальные учебные действия, мягкие навыки и другие параметры оценивания она оказывает влияние и, что особенно важно, каков размер эффекта, то в чем выражается эффект при применении качественных методов оценки результатов сформулировать бывает достаточно трудно. Этого зачастую не могут сделать даже сами авторы-разработчики, так как здесь большую роль могут играть субъективная точка зрения «эксперта» и так называемые «смешивающие» или «искажающие факторы». И тогда мы рискуем



снова вернуться к ситуации, когда «эффективность» определяется авторитетом разработчиков, мнением влиятельных экспертов, «мнением благополучателей» или вообще стихийно.

Заметим, что дискуссии о том, что можно измерить при помощи тестов и опросников, а что — нет и каковы возможности и ограничения их применения ведутся практически с начала появления этих измерительных инструментов. История развития тестологии насчитывает уже более 100 лет. Первые тесты индивидуальных различий в сфере сенсомоторных, интеллектуальных функций и перцептивных процессов возникли в конце XIX века и связаны с работами Ф. Гальтона, Дж. Кеттела, Дж. Ястрова. В начале XX века стали широко известны интеллектуальные тесты А. Бине в особенности шкала Стэнфорд-Бине, были разработаны первые стандартизованные тесты для оценки результатов школьного обучения (Г. Эббингауз и др.). В дореволюционной России начала XX в. сторонники естественно-научного направления в развитии психологии — А.П. Нечаев, Н.Е. Румянцев, Г.И. Россолимо, Ф.Е. Рыбаков — способствовали распространению идей тестирования и разрабатывали методики измерения общих способностей, пространственного воображения и др. В советский период 1920–1930-х гг. практика тестирования получает широкое распространение в педологии и психотехнике. Плодотворно работают в области психологического тестирования М.Я. Басов, М.С. Бернштейн, П.П. Блонский, Г.И. Россолимо и др. [Сорокова М.Г., Карданова Е.Ю., Радчикова Н.П., Федоров В.В., 2024]

Мы не будем здесь подробно останавливаться на истории развития тестологии и психометрики как в России, так и за рубежом. В настоящее время психометрика — это междисциплинарная наука, имеющая международное признание и интегрирующая достижения в области психологии, когнитивных наук и образования, а также мощный аппарат многомерной математической статистики и других разделов математики. Она изучает теорию и методы психологических измерений, в том числе, основных психических функций, личностных качеств, характерологических особенностей, профессиональной направленности, академических достижений и мн. др. Психометрика занимается вопросами разработки и стандартизации валидного и надежного психодиагностического инструментария, отвечающего всем международным требованиям. С этой целью она использует широкий арсенал методов Классической теории тестов (КТТ) и Современной теории тестов (IRT). Для стандартизации тестов используются большие выборки вплоть до нескольких тысяч респондентов, а сбор данных облегчается появлением цифровых технологий тестирования. Анализ данных стандартизационной выборки осуществляется в стандартных статистических пакетах, таких как SPSS, Statistica, R, KNIME Analytics Platform, Jamovi и мн. др. В настоящее время психодиагностические тесты и опросники разрабатываются очень активно, а статьи об этих

измерительных инструментах, параметрах их надежности и валидности и показателях соответствия моделей эмпирическим данным публикуются в рейтинговых научных журналах. В образовании также широко используются тесты знаний и академических достижений.

Вот почему в экспериментах по оценке доказательности психолого-педагогических и образовательных программ, технологий, практик для сферы образования и социальной сферы в парадигме доказательного подхода на первый план выступают количественные методы исследования и анализа эмпирических данных, а качественные методы могут служить им полезным дополнением.

## **Глава 2.**

# **Рамочные условия организации научной экспертизы и исследований по верификации психолого-педагогических программ и технологий**

Согласно концепции Московского государственного психолого-педагогического университета (МГППУ) и общероссийской общественной организации «Федерация психологов образования России» (ФПОР), процесс верификации психолого-педагогических программ и технологий в соответствии с положениями доказательного подхода включает 6 этапов.

Если эксперимент по верификации уже проведен, то секцией «Доказательный подход в психологии и образовании» (секция ДП НЭС) МГППУ <https://mgppu.ru/project/538> Научно-экспертного совета может быть проведена научная экспертиза. Для научной экспертизы авторы и разработчики предоставляют 3 документа для оценивания — Паспорт программы, саму Программу и Отчет о результатах верификации. Экспертиза проводится коллегиально командой из 3-х экспертов на основе 2-х оценочных форм — Протокола качественного ассесмента, оценивающего представленную программу, технологию как документ, и Протокола оценки результатов верификации. По результатам экспертизы принимается решение о рекомендации программ, технологий, практик, получивших положительную нормативную оценку (заключение секции ДП НЭС) о степени соответствия стандартам доказательности, по включению в «Реестр психолого-педагогических программ и технологий с доказанной эффективностью для образования и социальной сферы» Министерства просвещения РФ и размещению их на цифровой платформе Министерства просвещения РФ.

### **Этап 1. Паспорт верификации психолого педагогических, образовательных и социальных программ и технологий**

При представлении программы или технологии в проект по ее верификации авторы или разработчики заполняют «Паспорт верификации психолого-педагогических программ и технологий для образования и социальной сферы», в котором описывают необходимую информацию для организации эксперимента. На основе этого паспорта в случае положительных результатов эксперимента составляется «Реестр психолого-педагогических программ и технологий с доказанной эффективностью для образования и социальной сферы» Министерства просвещения РФ.

Структура паспорта приведена ниже. Комментарии и требования по заполнению паспорта представлены в **Приложении 3**.

### **Паспорт верификации психолого-педагогических программ и технологий**

1. № п/п.
2. ФИО ответственных экспертов — инициаторов верификации.
3. Организация, на базе которой разработана и апробирована программа/технология.
4. Авторы — разработчики программы/технологии.
5. Наименование программы/технологии.
6. Уровень образования по № 273-ФЗ «Об образовании в РФ», ред. от 16.04.2022 (Общее образование: дошкольное образование, начальное общее образование, основное общее образование, среднее общее образование; Профессиональное образование: среднее профессиональное образование, высшее образование — бакалавриат, высшее образование — специалитет, магистратура, высшее образование — подготовка кадров высшей квалификации; Дополнительное образование: дополнительное образование детей и взрослых, дополнительное профессиональное образование) или Социальная сфера Вид программы в соответствии с профессиональным стандартом «Педагог-психолог (психолог в сфере образования)» (профилактическая, просветительская, программа коррекционно-развивающей работы и др.).
7. Целевая группа в соответствии с документом «Система функционирования психологических служб в общеобразовательных организациях. Методические рекомендации» [Система функционирования психологических служб, 2020] (1. Норма (нормотипичные дети и подростки с нормативным кризисом взросления); 2. Дети, испытывающие трудности в обучении; 3. Уязвимые категории детей: 3.1. Дети, находящиеся в трудной жизненной ситуации: 3.1.1. Дети-сироты и дети, оставшиеся без попечения родителей, 3.1.2. Обучающиеся с ОВЗ, детиинвалиды, 3.1.3 Дети с отклоняющимся поведением (девиантное поведение детей и подростков, суицидальное поведение детей и подростков), Дети-мигранты и 3.2. Одаренные дети; 3.3 Иные категории особо уязвимых детей (см. раздел 1.3); 4. Взрослые (с указанием основных характеристик, например, родители, педагоги и др.).
8. Целевая аудитория (конкретная более узкая группа, которой адресована программа/технология). Если целевая аудитория относится к нескольким категориям, то надо указать все, выделив первую как основную (например, дети с ОВЗ, проживающие в организации для детей-сирот, добровольно переданные родителями).
9. Сфера применения (сфера трудностей) в соответствии с документом «Система функционирования психологических служб в общеобразовательных организациях. Методические рекомендации» [Система

- функционирования психологических служб, 2020] и «Открытым реестром психодиагностических методик, вызывающих доверие профессионального сообщества» как его составной частью (Развитие основных психических функций, Метапредметные компетенции и универсальные учебные действия, Социальное развитие и морально-ценностная сфера, Эмоционально-волевая сфера, Коммуникативная сфера, Поведенческая сфера, Профессиональная направленность, мотивация, характерологические особенности, Профессиональные и личностные проблемы взрослых участников образовательного процесса, Профессиональное и/или дополнительное образование взрослых).
10. Укажите конкретно типы трудностей и проблемы, на решение которых направлена программа, а также ее социальную значимость.
  11. Проблемная ситуация, на решение которой направлена программа (краткая характеристика). Структура описания проблемной ситуации, например, ее проявления на когнитивном, эмоционально-волевом и поведенческом уровнях.
  12. Цели программы/технологии.
  13. Задачи программы/технологии.
  14. Краткая характеристика формирующего этапа эксперимента в соответствии с программой/технологией. Для социальных программ возможно описание эмпирического исследования.
  15. Измеряемые конструкты и критерии оценки планируемых результатов и характеристики (признаки, свойства, параметры и т.п.), на которые программа/технология должна оказывать эффект.
  16. Ожидаемые результаты реализации программы.
  17. Ограничения применения программы. Риски реализации программы (возможность нанесения вреда целевой группе или связанным с ней группами детей и взрослых).
  18. Психодиагностические методики, используемые для оценки достижения планируемых результатов (с опорой на «Открытый реестр психодиагностических методик, вызывающих доверие профессионального сообщества» [Система функционирования психологических служб, 2020, С. 110]).
  19. Длительность реализации программы/технологии (период времени и/или количество акад. часов).
  20. Гиперссылка на программу (если есть).
  21. Краткие сведения о проведении апробации.
  22. Исследования (статьи, диссертации и др.) эффективности программы/технологии (если есть).
  23. Ресурсные площадки — базы верификации программы/технологии (конкретные образовательные и социальные организации, учреждения, центры и т.д.): наименование организации, место нахождения, основные направления деятельности (по Уставу), численный состав обучающихся/воспитанников/педагогических работников и др.

24. Планируемые сроки проекта по верификации программы/технологии.
25. Результаты верификации (кратко).
26. Статьи по результатам верификации (библиографические описания, ссылки).
27. Полученные награды на всероссийских или межрегиональных конкурсах за представленную программу/технологию, благодарственные письма от органов законодательной и исполнительной власти регионов, положительные оценки профессионального сообщества.

## **Этап 2. Качественный ассесмент материалов психолого-педагогических, образовательных и социальных программ и технологий, представленных для верификации**

Ассесмент психолого-педагогических, образовательных и социальных программ с дополнительными материалами к ним, представленных для верификации, проводится первоначально на качественном уровне. Он осуществляется после заполнения авторами программы и/или экспертами-инициаторами «Паспорта верификации психолого-педагогических программ и технологий». Ассесмент проводится на основе метода экспертной оценки представленных разработок.

С этой целью, как правило, создается Экспертная группа на базе организации, выступающей организатором независимой оценки программ и технологий, соответствующих определенным видам профессиональной деятельности: в нашем случае, ФГБОУ ВО МГППУ. В состав Экспертной группы рекомендуется включать ведущих ученых и практиков в области, соответствующей профилю (направленности практики), имеющих опыт работы по проведению экспертизы практик. В состав Экспертной группы могут входить также представители общественных организаций, социально ориентированных некоммерческих организаций, автономных некоммерческих организаций (АНО), фондов, работающих в сфере науки, образования, здравоохранения и социальной сфере. Экспертная группа осуществляет оценку принятых для верификации программ и технологий на основании установленных критериев, для каждого критерия указывается максимальный балл. По результатам экспертной оценки каждой кандидатной программы и технологии составляется протокол экспертной оценки, в который вносятся результаты оценивания в баллах. Протокол качественного ассесмента является составной частью экспертного заключения.

Критериальный аппарат как система взаимосвязанных и взаимно дополняющих критериев качественной оценки психолого-педагогических программ и технологий приведен в форме Протокола (Табл. 2), разработанного с использованием конкурсной документации общероссийской

общественной организации «Федерация психологов образования России» (<https://rospsy.ru/>) и «Положения о Всероссийском конкурсе лучших психолого-педагогических программ и технологий в образовательной среде — 2022», утвержденного директором Департамента государственной политики в сфере защиты прав детей Минпросвещения России. Раздел 1 «Протокола качественного ассесмента» составляет чек-лист для проверки соответствия представленного описания программы или технологии «Паспорту верификации психолого-педагогических программ и технологий». Критерии экспертной оценки Раздела 2 «Ассесмент содержания программы, технологии и материалов к ней» соответствуют

«Протоколу экспертной оценки психолого-педагогической программы» и «Положению о Всероссийском конкурсе лучших психолого-педагогических программ и технологий в образовательной среде-2022», утвержденному Министерством просвещения РФ.

В качестве примера приведем комментарий для экспертов секции

«Доказательный подход в психологии и образовании» <https://mgppu.ru/project/538> Научно-экспертного совета МГППУ (секции ДП НЭС).

#### **Комментарий для экспертов:**

Настоящий Протокол качественного ассесмента состоит из 2-х разделов. В Разделе 1. Чек-лист соответствия «Паспорту верификации образовательных и социальных программ и технологий» эксперт отмечает, заполнена (балл = 1) или нет (балл = 0) соответствующая графа в «Паспорте верификации ...» по позициям 1–9, 19, 20 и 24. По остальным позициям 10–8 и 21–23 эксперт может выставить 2 балла (позиция раскрыта полностью), 1 балл (позиция раскрыта частично) или 0 баллов (позиция не раскрыта или вообще не заполнена). В таблице указан максимальный балл по каждой позиции, который может выставить эксперт. В таблице в графе **Итого по Разделу 1 (чек-лист)** указывается сумма баллов и какой процент она составляет от максимального балла, т.е. от 36 баллов.

Наиболее важным является Раздел 2. Ассесмент содержания программы, технологии и материалов к ней. Эксперт оценивает кандидатную программу и выставляет балл по каждому критерию. В таблице указан максимальный балл по каждому критерию, который может выставить эксперт. В таблице в графе **Итого по Разделу 2 (ассесмент содержания программы, технологии)** указывается сумма баллов и какой процент она составляет от максимального балла, т.е. от 72 баллов.

Окончательные выводы экспертизы по «Протоколу качественного ассесмента ...» эксперты согласовывают между собой, включают в коллегиальное Экспертное заключение и представляют на сессии секции «Доказательный подход в психологии и образовании» Научно-экспертного совета МГППУ (секция ДП НЭС) как результат консенсуса. **Вопросы, комментарии, рекомендации экспертов** указываются в таблице в нижней графе.

После представления программы или технологии и обсуждения результатов ее экспертизы на сессии секции ДП НЭС авторам и разработчикам программы предоставляется **Экспертное заключение секции ДП НЭС**. В Экспертном заключении указываются рейтинговые оценки (проценты от максимальных баллов), набранные по «Протоколу качественного ассесмента ...» и «Протоколу экспертной оценки результатов верификации ...», Заключение экспертизы (нормативная оценка), а также комментарии и рекомендации экспертов. Экспертное заключение секции ДП НЭС подписывают эксперты, руководитель и ученый секретарь секции ДП НЭС.

Таблица 2

**Протокол качественного ассесмента психолого-педагогических программ и технологий, представленных для верификации**

<b>Критерии</b>	<b>Максимальный балл</b>
<b>Раздел 1. Чек-лист соответствия «Паспорту верификации образовательных и социальных программ и технологий»</b>	<b>36</b>
1. ФИО ответственных экспертов — инициаторов верификации	1
2. Организация, на базе которой разработана и апробирована программа/технология	1
3. Авторы — разработчики программы/технологии	1
4. Наименование программы/технологии или специалисты, адаптировавшие уже разработанную программу/технологию с указанием оригинальных авторов	1
5. Уровень образования по № 273-ФЗ «Об образовании в РФ», ред. от 16.04.2022 (Общее образование: дошкольное образование, начальное общее образование, основное общее образование, среднее общее образование; Профессиональное образование: среднее профессиональное образование, высшее образование — бакалавриат, высшее образование — специалист, магистратура, высшее образование — подготовка кадров высшей квалификации; Дополнительное образование: дополнительное образование детей и взрослых, дополнительное профессиональное образование) или Социальная сфера	1
6. Вид программы в соответствии с профессиональным стандартом «Педагог-психолог (психолог в сфере образования)» (профилактическая, просветительская, программа коррекционно-развивающей работы и др.)	1
7. Целевая группа в соответствии с документом «Система функционирования психологических служб в общеобразовательных организациях. Методические рекомендации» [35]	1
8. Целевая аудитория (конкретная более узкая группа, которой адресована программа/технология)	1
9. Сфера применения (сфера трудностей) в соответствии с документом «Система функционирования психологических	1



Критерии	Максимальный балл
служб в общеобразовательных организациях. Методически рекомендации» [35] и «Открытым реестром психодиагностических методик, вызывающих доверие профессионального сообщества» как его составной частью (Развитие основных психических функций, Метапредметные компетенции и универсальные учебные действия, Социальное развитие и морально-ценностная сфера, Когнитивная сфера, Эмоционально-волевая сфера, Коммуникативная сфера, Поведенческая сфера, Профессиональная направленность, мотивация, характерологические особенности, Профессиональные и личностные проблемы взрослых участников образовательного процесса, Профессиональное и/или дополнительное образование взрослых)	1
10. Указаны конкретно типы трудностей и проблемы, на решение которых направлена программа, а также ее социальная значимость	2
11. Проблемная ситуация, на решение которой направлена программа (краткая характеристика). Структура описания проблемной ситуации, например, ее проявления на когнитивном, эмоционально-волевом и поведенческом уровнях, в социальной сфере	2
12. Цели программы/технологии	2
13. Задачи программы/технологии	2
14. Краткая характеристика формирующего этапа эксперимента в соответствии с программой/технологией	2
15. Измеряемые конструкты и критерии оценки планируемых результатов и характеристики (признаки, свойства, параметры и т.п.), на которые программа/технология должна оказывать эффект	2
16. Ожидаемые результаты реализации программы	2
17. Ограничения применения программы	2
18. Психодиагностические методики, используемые для оценки достижения планируемых результатов (с опорой на «Открытый реестр психодиагностических методик, вызывающих доверие профессионального сообщества» [35])	2
19. Длительность реализации программы/технологии (период времени и/или количество акад. часов)	1
20. Гиперссылка на программу (если есть)	1
21. Краткие сведения о проведении апробации	2
22. Исследования (статьи, диссертации и др.) эффективности программы/технологии (при наличии)	2
23. Ресурсные/пилотные/стажировочные площадки — базы верификации программы/технологии (конкретные образовательные и социальные организации, учреждения, центры и т.д.): наименование организации, место нахождения, основные	2

<b>Критерии</b>	<b>Максимальный балл</b>
направления деятельности (по Уставу), численный состав обучающихся/воспитанников/педагогических работников и др	2
24. Планируемые сроки проекта по верификации программы/технологии	1
<b>Итого по Разделу 1 (чек-лист):</b>	<b>Указать балл и процент от 36 баллов</b>
<b>Раздел 2. Ассесмент содержания программы, технологии и материалов к ней</b>	<b>Максимальный балл 72</b>
<b>1. Соответствие требованиям нормативных правовых документов, регламентирующих деятельность специалиста</b>	<b>8</b>
1.1 Соблюдение в программе (технологии) требований действующих нормативных и иных документов (профессиональный стандарт, ФГОС и т.п.), регламентирующих деятельность специалистов	3
1.2 Точность, конкретность и профессионализм аргументации в содержании требований к специалистам, реализующим программу (технологии)	2
1.3 Содержательность, реалистичность, логичность профессиональной аргументации и конкретность требований к материально-технической оснащённости организации для реализации программы (технологии) (помещение, оборудование, инструментарий и т.д.)	3
<b>2. Соответствие теме</b>	<b>8</b>
2.1 Соответствие содержания программы социально-психологическим особенностям тех, на помощь кому она направлена	2
2.2 Соответствие основного содержания программы (технологии) заявленным целям и задачам	2
2.3 Соответствие выбранных в программе (технологии) способов решения профессиональной задачи выделенным целям, задачам и особенностям контингента, на помощь которому направлено содержание	3
2.4 Четкость и последовательность в описании структуры программы (технологии), наличие внутренней логики построения, содержательная взаимосвязь основных модулей	1
<b>3. Результативность</b>	<b>18</b>
3.1 Реалистичность выделенных в программе целей, задач и сроков их достижения, решения	4
3.2 Технологичность и комплексность решения задач с помощью заявленных мероприятий	4
3.3 Конкретность и реалистичность в формулировке планируемых результатов реализации программы (технологии)	2

<b>Критерии</b>	<b>Максимальный балл</b>
3.4 Наличие адекватной целям и задачам Программы (технологии) системы контроля за ее реализацией и результатами использования в рамках апробации	4
3.5 Полнота и конкретность сведений о практической апробации программы (технологии)	4
<b>4. Содержательность и аргументированность</b>	<b>16</b>
4.1 Обоснованность выбора профиля программы (технологии) в соответствии с нормативными документами и проблемной ситуацией, на решение которой она направлена	2
4.2 Научная обоснованность выбора и согласованности используемого инструментария оценки конструкторов с описанной теоретической базой, целями и задачами программы. Профессиональная надежность и качество выбранных автором(ами) источников заимствования инструментария	4
4.3 Аргументированность выбора и точность описания требований к информационной обеспеченности для реализации программы (технологии)	2
4.4 Обоснованность алгоритма и последовательности реализации программы (технологии) в условиях указанного учреждения (образовательной среды) с указанными участниками	4
4.5 Содержательность и логичность критериев оценки достижений результатов (качественных и количественных) и способов/форм их использования	4
<b>5. Соответствие общим требованиям качественного ассесмента программ/технологии</b>	<b>8</b>
5.1 Полнота и точность в выполнении требований к оформлению, содержанию и структуре программы (технологии);	2
5.2 Соблюдение требований к представлению учебных и методических материалов, необходимых для реализации программы (технологии);	2
5.3 Конкретность, корректность и точность формулировки оснований программы/технологии как практической	2
5.4 Соответствие представленного текста требованиям к оформлению документа согласно ГОСТ;	1
5.5 Точность соблюдения требований организации-разработчика к оформлению текста Программы (титальный лист, библиографический список, таблицы, схемы, шрифт, интервал и т.п.)	1
<b>6. Оригинальность содержания</b>	<b>14</b>
6.1 Использование в программе цифровых инструментов и возможностей современных цифровых технологий	6
6.2 Уникальность идеи и технологии, следование принципу преемственности, развития традиций отечественных научных школ и опыту психолого-педагогической практики в России	4

<b>Критерии</b>	<b>Максимальный балл</b>
6.3 Соблюдение закона об авторских правах, требований к выполнению правил заимствованию/компиляции в профессиональных текстах и т.п.	4
<b>Итого по Разделу 2 (ассесмент содержания программы, технологии):</b>	<b>Указать балл и процент от 72-х баллов</b>
<b>Вопросы, комментарии, рекомендации экспертов</b>	

Из Табл. 2 видно, что часть критериев ассесмента представленной для верификации содержания программы/технологии и материалов к ней носит формальный характер (это перечень требований, определенных законами и постановлениями Правительства, юридическими документами, образовательными стандартами подготовки специалиста, стандартами профессиональной деятельности и т. п.). Они указывают на обязательное наличие тех или иных ее формальных показателей. Те критерии, которые отличаются от чисто формальных, отражают качество содержания самой программы или технологии и должны сопровождаться развернутым пояснением и обоснованием, а их удельный вес превышает значения первой категории критериев [Пахальян В.Э., 2021].

Опыт проведения экспертной оценки конкурсной документации ФПО России показывает, что у некоторых авторов программ и технологий имеются типовые трудности в определении ожидаемых результатов реализации программы и критериев эффективности, в оценке результатов апробации. В частности, выбор методического инструментария для оценки результатов не всегда отражает особенности по срокам, содержанию и сложности проекта. Особенно это касается программ, рассчитанных на длительный период времени, в которых трудно выделить развивающий эффект той или иной технологии.

Наибольшее количество замечаний экспертов связано со следующими оценочными экспертными показателями [Рубцов В.В., Романова Е.С., 2022]:

- 6.1. Использование в программе возможностей современных цифровых технологий (42%);
- 3.5. Полнота и конкретность сведений о практической апробации программы (технологии) (51,3%);
- 4.5. Содержательность и логичность критериев оценки достижения результатов (качественных и количественных) и способов / форм их использования (54%);
- 6.2. Уникальность идеи и технологии, следование принципу преемственности, развития традиций отечественных научных школ и опыту психолого-педагогической практики в России (60,3%).

### **Этап 3. Протокол верификации психолого-педагогических программ**

Настоящий «Протокол верификации психолого-педагогических программ и технологий» является рамочным типовым документом и создает основу для разработки протоколов аналогичных исследований в конкретных областях образования, психологии, социальной сферы. В случае необходимости он может быть дополнен и скорректирован с учетом специфики таких исследований. Структура Протокола верификации с краткими комментариями приведена ниже.

#### **Протокол верификации психолого-педагогических программ и технологий**

1. *Актуальность исследования и теоретико-методологическое обоснование программы*
  - 1.1. Название и авторы психолого-педагогической программы/технологии. Если название экспериментальной программы/технологии переводится с иностранного языка, приведите название на языке оригинала. Опишите, была ли проведена модификация или адаптация зарубежной методики для условий РФ;
  - 1.2. Ссылка на веб-сайт с информацией о программе/технологии (если есть);
  - 1.3. Теоретико-методологическое обоснование программы/технологии  
Непротиворечивость и убедительность причинно-следственных связей между деятельностью и ожидаемыми результатами/изменениями, обоснованность программы/технологии с точки зрения научных теорий и исследований, а также с точки зрения того, насколько практика отвечает на запросы благополучателей (детей и их семей) и представляется важной и продуктивной специалистам-практикам;
  - 1.4. Сформулировать, почему исследование эффективности этой программы актуально.
2. Краткий обзор предшествующих исследований (если есть).
3. Исследовательские вопросы.
4. Научное направление исследования в соответствии с государственным рубрикатором научно-технической информации (ГРНТИ) (по третьему уровню иерархии).
5. Цель и задачи исследования.
6. Дизайн исследования:
  - 6.1. Описание выборки: регион(ы) проведения исследования, учреждения — участники исследования, категория испытуемых, общий объем выборки (N).

- Экспериментальная группа (ЭГ): общая характеристика, объем выборки (N1), пол, возраст (среднее, стандартное отклонение, медиана, мода), другие контекстные характеристики, важные для исследования, например, социо-экономический статус, этнокультурная принадлежность, статус беженца, успеваемость в школе и др. Контрольная группа (КГ): общая характеристика, объем выборки (N2), пол, возраст (среднее, стандартное отклонение, медиана, мода), другие контекстные характеристики, важные для исследования.
- 6.2. Психолого-педагогические характеристики (конструкты, параметры, признаки, переменные и др.), в отношении которых программа, предположительно, должна быть эффективна: перечислить, дать четкое определение, что под ними понимается.
  - 6.3. Психодиагностический инструментарий:
    - стандартизованные инструменты (тесты, опросники, шкалы, диагностические методики и др.) для измерения исследуемых конструктов. Указать официальные названия инструментов, дать ссылки на статьи с результатами стандартизации в рецензируемых научных журналах. Ориентир — «Открытый реестр психодиагностических методик, вызывающих доверие профессионального сообщества» как составная часть документа «Система функционирования психологических служб в общеобразовательных организациях. Методические рекомендации» [Система функционирования психологических служб, 2020., С. 110];
    - другие инструменты (например, качественные методы) оценки исследуемых признаков (если есть).
  - 6.4. Этапы сбора эмпирических данных:
    - 6.4.1. констатирующий этап: планируемые сроки, длительность, содержание деятельности;
    - 6.4.2. формирующий этап: планируемые сроки, длительность, содержание формирующего воздействия, вмешательства (кратко); используемые цифровые устройства, приборы и др.;
    - 6.4.3. контрольный этап: планируемые сроки, длительность, содержание деятельности;
    - 6.4.4. контроль отдаленных результатов (если предусмотрен): планируемые сроки, длительность, содержание деятельности.
  - 6.5. Методы и процедуры сбора данных:
    - 6.5.1. Общая характеристика процедур сбора данных:
      - Тип дизайна планируемого эксперимента (РКИ индивидуального уровня, кластерное РКИ, квази-экспериментальный дизайн (КЭД), дизайн с регрессионным разрывом (ДРР), дизайн с одним случаем (кейс-стади, ДОС)) (классификацию типов дизайна см. в **Приложении 1, п. 1**);

- принцип группировки испытуемых при сборе данных (будет ли осуществляться воздействие на испытуемых индивидуально, в составе малой или большой группы, например, школьного класса?). Опишите наименьшую единицу воздействия в формирующем эксперименте (индивид, рабочая группа, класс), результаты который будут анализироваться с точки зрения величины эффекта;
  - метод отбора, если будет использован прием систематического отбора испытуемых;
  - условия и место, в которых будет происходить сбор данных;
  - все виды соглашений, которые будут заключены с участниками эксперимента, и выплаты, которые будут производиться;
  - соглашения с Наблюдательными Советами учреждений; информация о соблюдении этических норм.
- 6.5.2. Условия проведения формирующего этапа эксперимента (вмешательства):
- количестве воздействий и их длительность (т.е. сколько было экспериментальных сессий, эпизодов и событий, связанных с воздействием на испытуемых и какова была их продолжительность);
  - частота, с которой оказывалось воздействие на испытуемых (например, завершался ли эксперимент после однократной экспериментальной сессии или участники неоднократно подвергались воздействию, и в этом случае, каким был временной интервал между первой и последней сессией).
- 6.6. Информация о специалистах и персонале, осуществляющих экспериментальное воздействие (вмешательство):
- проводилось ли (будет ли проводиться) повышение квалификации и обучение специалистов для проведения формирующего эксперимента;
  - сколько и какие специалисты будут осуществлять формирующее воздействие;
  - конкретные действия и стимулы, с точки зрения их соответствия задачам исследования.
7. Методы анализа данных (планируемые):
- количественные методы, стандартные статистические пакеты для анализа данных научного исследования (SPSS, Statistica, R, KNIME Analytics Platform и др.);
  - качественные методы.

## **Этап 4. Анализ данных эксперимента по верификации и оценка его результатов**

### **4.1. Количественный анализ данных экспериментального исследования: общие рекомендации**

Корректное, объективное, полное и глубокое изложение результатов обработки данных (будь то количественный или качественный анализ) является одним из важнейших компонентов всех научно-исследовательских работ. Исследование, выполненное в естественно-научной методологии, предполагает количественный анализ больших массивов данных разного типа — номинативных (категориальных), порядковых, количественных, — в стандартных статистических пакетах: SPSS, Statistica, KNIME Analytics Platform и др. Методы количественного анализа должны быть релевантны исследовательским вопросам, типу и характеру данных, а также количеству выборок.

Поскольку аналитические приемы выбираются в соответствии с характером анализируемых данных и выдвинутых предположений, невозможно указать, что представляет собой «эффективный набор статистических процедур» для каждого типа анализа. Для количественных переменных обычно такой набор включает следующие показатели описательной статистики для каждой подгруппы: размер выборки ( $N$ ); меры центральной тенденции — выборочное среднее ( $M$ ), стандартная ошибка среднего, медиана ( $Me$ ), мода ( $Mod$ ); меры рассеяния — стандартное отклонение ( $SD$ ), асимметрия ( $A$ ), стандартная ошибка асимметрии, эксцесс ( $E$ ), стандартная ошибка эксцесса. Для категориальных переменных принято указывать частоты для каждой категории испытуемых и их процентные доли от общей выборки или соответствующих подвыборок.

Делать выводы только на основе описательной статистики некорректно, необходимо проверять статистические гипотезы. Исключение могут составлять социологические исследования, где основным результатом является описание ситуации с использованием выборочных характеристик, процентных долей и ранжирования ответов по частоте, а также разнообразных графиков, диаграмм, гистограмм. В настоящее время использование статистических пакетов сделало возможным широкое применение методов математического моделирования, Data Mining и многомерных статистических методов анализа эмпирических данных. Data Mining (другие термины — добыча данных, интеллектуальный анализ данных, глубинный анализ данных или просто майнинг данных) — это процесс извлечения полезной информации из необработанных больших данных эмпирического исследования. Одним из наиболее часто используемых методов количественной оценки эффективности формирующего воздействия



(вмешательства) является дисперсионный анализ (ANOVA), краткая характеристика которого представлена в п. 4.2.

Нужно стремиться проводить исследование на больших выборках — от нескольких сотен до нескольких тысяч испытуемых. Чем меньше выборка, тем большую роль играют случайные факторы, которые «создают шум» и могут ослабить искомые эффекты, связи и т.д. Все статистические критерии при расчете эмпирических значений статистик и уровней значимости учитывают объемы выборок. На больших выборках статистический критерий лучше выявляет различия, поэтому возрастает вероятность получения статистически значимого эффекта.

Отметим несколько позиций, которые нередко отражаются некорректно при представлении результатов. Использование параметрических методов требует проверки согласия с нормальным распределением исследуемых переменных. При проверке статистической значимости различий сопоставляемые группы не обязаны быть одинаковы по численности. Корреляционный или регрессионный анализ не доказывает причинно-следственную связь, а только сопутствующие изменения. Результаты анализа многих переменных нужно оформлять в виде таблиц, в которых указывать статистические критерии, эмпирические значения статистик, уровни значимости. Таблицы должны сопровождаться ясными комментариями, например: «Табл. 1 показывает ...», «Как видно из Табл. 2...».

Для оценки масштаба или важности выводов исследования необходимо сообщать о величине эффекта, например, с помощью доверительного интервала для каждой величины эффекта. Величины эффекта могут быть выражены в оригинальных единицах, например, среднее количество вопросов, на которые был получен правильный ответ, или кг/месяц, которые анализируются с помощью статистической оценки наклона линии регрессии. Они являются наиболее легко воспринимаемыми при описании реакций в разных подгруппах испытуемых. Иногда величину эффекта лучше представлять не в оригинальных единицах, а в стандартизированном виде (например, через  $d$  Коэна,  $g$  Хеджеса,  $\Delta$  Гласса или стандартизированный Бета-коэффициент регрессии). Методы расчета стандартных показателей размера эффекта для исследований с разными видами дизайна представлены в **Приложении 2**.

Результаты должны быть представлены максимально подробно, чтобы получить обоснованные выводы. Эмпирическое исследование проводится не с целью подтвердить гипотезу, а чтобы ее проверить. Во многих случаях гипотеза исследования подтверждается лишь частично или не подтверждается вообще: это нормально. Рекомендуется отметить все полученные результаты, включая те, которые противоречат исходным представлениям, а также упомянуть о малых величинах эффекта или статистически незначимых коэффициентах, в тех случаях, когда по теории предполагалась большая величина эффекта или статистически значимые показатели.

Результаты должны быть не только правильно представлены, но и правильно интерпретированы. Количественный анализ данных — не самоцель, мы проводим исследование в области психологии или образования, обсуждение результатов обязательно.

Эмпирические исследования, как правило, содержат ограничения.

О них нужно не бояться писать в разделе «Ограничения результатов».

## 4.2. Дисперсионный анализ и оценка взаимодействия

В исследованиях, которые проводятся для проверки эффективности воздействия некоторого фактора, как правило, используется сложная экспериментальная схема, в которой одна переменная является межгрупповой (контрольная группа — экспериментальная группа), а вторая — внутригрупповой (обе группы тестируются по интересующему исследователей показателю до воздействия и после воздействия). Для статистической проверки эффективности воздействия при этом часто используется двухфакторный дисперсионный анализ (ANOVA). Сущность дисперсионного анализа заключается в расчленении общей дисперсии изучаемого признака на отдельные компоненты, обусловленные влиянием конкретных факторов, и проверке гипотез о значимости влияния этих факторов на исследуемый признак. Данный метод позволяет оценить значимость не только эффектов одной независимой переменной (главных эффектов), но и взаимодействия независимых переменных. Взаимодействие — это количественный результат, обусловленный соотношением между действием двух или нескольких независимых переменных, выделенный в факторном эксперименте. Вычисляется как разность между различиями значений зависимой переменной, полученных при действии разных условий первой, второй и т. д. переменных, и иллюстрируется графически. Взаимодействие показывает, зависит ли величина воздействия фактора от значений других факторов (переменных). Как правило, в таких исследованиях предполагается, что взаимодействие будет статистически значимым. Следует помнить, что главные эффекты статистически независимы от эффектов взаимодействия, поэтому нельзя предсказать, какое будет взаимодействие, зная только главные эффекты. Предположения и ограничения многофакторного дисперсионного анализа: дисперсии в сравниваемых группах должны быть приблизительно равны, выборки (контрольная группа — экспериментальная группа) должны быть случайны и независимы. Зависимая переменная должна быть, по крайней мере, интервальной, и нормально распределена в каждой группе. Выполнение допущения о независимости выборок является обязательным в любом случае. Последствия нарушений остальных двух допущений требуют специального рассмотрения. Многочисленные исследования показали, что дисперсионный анализ очень устойчив к нарушению предположения о нормальности распределения (подробнее см. [Lindman H.R., 1974]), поэтому перед его

проведением нет особой необходимости в проверке соответствия выборочных распределений нормальному закону. Нарушение предположения о равенстве дисперсий имеет существенное значение в том случае, если сравниваемые выборки отличаются по численности [Наследов А.Д., 2004, С. 188–189]. Если эксцесс больше 0, то эмпирическое значение статистики дисперсионного анализа  $F$  может стать очень маленьким. Нулевая гипотеза при этом не может быть отвергнута, хотя она и не верна. Ситуация меняется на противоположную, если эксцесс меньше 0. Асимметрия распределения обычно незначительно влияет на  $F$  статистику. Если число наблюдений в ячейке достаточно большое, то отклонение от нормальности не имеет особого значения в силу *центральной предельной теоремы*, в соответствии с которой, распределение среднего значения при большом объеме выборки близко к нормальному, независимо от начального распределения. Подробное обсуждение устойчивости  $F$  статистики можно найти в [Box G.E.P., Anderson S.L., 1955] или [Lindman H.R., 1974]. Предполагается также, что дисперсии в разных группах одинаковы, однако, Линдман [Lindman H.R., 1974, С. 33] показывает, что  $F$  критерий вполне устойчив относительно нарушения предположений однородности дисперсии (см. также [Box G.E.P. 1954a], [Box G.E.P., 1954b]).

Для сравнения всех исследуемых условий между собой служат специальные апостериорные критерии, которые принимают во внимание множественность сделанных сравнений, и, следовательно, увеличивающийся риск допустить ошибку I рода, и контролируют его. Для дисперсионного анализа таких критериев достаточно много. Критерий наименьшей значимой разности (*least significant difference test* — *LSD test*) не принимает во внимание число сделанных сравнений и аналогичен, например, вычислению критерию Стьюдента несколько раз. Другие критерии, такие как критерий Дункана (*Duncan test*) или критерий Ньюмена-Кейлса (*Newman-Keuls test*), учитывают множественность сделанных сравнений. Еще более консервативными (и, значит, надежными) являются критерии Тьюки (*Tukey*) и Шеффе (*Sheffe*).

### **4.3. Анализ данных при использовании качественных методов исследования**

Качественные методы сбора и анализа данных являются неотъемлемой частью психологических исследований, исследований в сфере детства. В отличие от количественного анализа, основанного на сравнении и анализе количественных мер исследуемых объектов и явлений, качественный анализ основан на сравнении качеств. Качественная методология отражает междисциплинарный тренд развития современных социальных наук, который охватывает различные направления и стратегии исследований — этнографические, психоаналитические, феноменологические, дискурсивные, нарративные и ряд других [Melnikova O., Khoroshilov D., 2010].

Методология качественных исследований строится на четырех основных **принципах** — контекстуальной чувствительности; понимания; интерпретативной реконструкции и рефлексивности [Мельникова О.Т., Хорошилов Д.А., 2013]. К характерным чертам методологии качественных исследований относят направленность на описание случаев; предпочтение естественной обстановки для проведения исследования; применением «аналитических обобщений»; гибкости исследовательских процедур и учету контекста исследования; преимущественное использование качественных методов сбора и анализа данных [Бусыгина Н.П., 2010].

Целью качественных исследований является сбор качественных данных (любой тип вербальной (тексты) или визуальной (изображения) информации) и их последующий качественный анализ, при этом исследователи не ограничиваются только описанием феноменологии, полученной в опыте, но и объясняют причины происходящего [Бусыгина Н.П., 2019]. Эта двойная цель реализуется с помощью интерпретации — процесса приписывания или конструирования значений, основанного на теоретических концепциях. Таким образом, *качественные методы исследования* — это стратегии познания, направленные на определение значений и смыслов, которые отдельный человек и группа людей придают различным предметам, действиям и событиям окружающей жизни [Хорошилов Д.А., Балашова Е.Ю., 2018].

Можно выделить четыре **основных качественных метода исследования**: обоснованную теорию (grounded theory), интерпретативный феноменологический анализ (interpretative phenomenological analysis), дискурсивный анализ (discourse analysis) и нарративный анализ (narrative analysis. [Lyons E., Coyle A. (Eds.), 2007]. К качественным методам относится также анализ индивидуального случая (case studies). Анализ индивидуального случая характеризуется не методами, используемыми для сбора и анализа данных, а, скорее, тем, что оно сосредоточено на конкретной единице анализа: случае. Кейсом может быть организация, город, группа людей, сообщество, пациент, школа, вмешательство, ситуация, инцидент или опыт. Как и любые качественные методы он включает в себя глубокое, интенсивное и четко сфокусированное исследование изучаемого явления. Исследования индивидуального случая могут использовать **как качественные, так и количественные методы сбора и анализа данных** [Willig C., 2008]. К *методам сбора качественных данных* относят наблюдение, опрос в индивидуальной и групповой форме, изучение документов и другие.

Качественные методы сбора и анализа данных в психологии, образующие сегодня единый методологический подход, позволяют раскрыть эмоционально-смысловые отношения человека и мира, личности и общества, пациента и врача, клиента и терапевта [Хорошилов Д.А., Балашова Е.Ю., 2018].

В качественных исследованиях на первый план выступает субъективный смысл рассматриваемых феноменов с точки зрения испытуемого, наблюдаемого, респондента. Особенность подхода заключается в уникальности и индивидуальности исследовательского взгляда и, соответственно, встает вопрос о подтверждаемости получаемых данных. Результаты качественных исследований могут быть подтверждены прежде всего другими исследователями [Войскунский А.Е., Скрипкин С.В., 2001]. Для **валидации качественных исследований** используется включение в аналитический процесс специальных стратегий, позволяющих сделать его результаты прозрачными. Их использование не гарантирует автоматической валидности исследования, а является способом самопроверки исследователя и развитием критической рефлексивной позиции по отношению к изучаемой информации. К таким стратегиям относятся: ведение рефлексивного журнала (reflexive journal writing), аудит исследования (audit trail), проверка участниками (member checking), теоретическая выборка (theoretical sampling), анализ негативных случаев (negative case analysis), триангуляция (triangulation), а также опросы получателей услуг, мониторинг последствий, оценка устойчивости — долговременности полученных результатов (изменений), отзывы специалистов, применяющих данную практику, и ряд других.

Основная стратегия среди них — триангуляция, предполагающая одновременное обращение к нескольким источникам данных, методам, теориям и экспертным мнениям. Практическое применение триангуляции зависит от профессиональных компетенций и опыта исследователя, и результаты применения триангуляции отвечают общей специфике и основным критериям валидности в качественной методологии [Мельникова О.Т., Хорошилов Д.А., 2015].

#### **4.4. Оценка результатов верификации: отчет о результатах, рейтинговая оценка и общий вывод**

По результатам верификации с учетом количественного анализа данных и размера эффекта составляется **Отчет о результатах верификации**, общая структура которого представлена в **Приложении 1, п. 4.1**. Кроме того, по результатам верификации психолого-педагогическая программа или технология получает рейтинговую оценку и нормативную оценку в соответствии с **Протоколом оценки результатов верификации** (см. **Приложение 2, п. 1.3**). Нормативная оценка — это общий вывод о соответствии, соответствии с учетом дополнительных рекомендаций, соответствии на этапе пилотных проектов или несоответствии программы/технологии стандартам доказательности. Отчет используется для 3-х целей: 1. Как источник информации для заполнения экспертами Протокола оценки результатов верификации и выставления рейтинговой оценки;

2. Для написания Экспертного заключения; 3. Если программа или технология успешно прошла верификацию и получила высокую рейтинговую и нормативную оценки, информация Отчета используется для написания статьи в рецензируемом научном журнале.

## Этап 5. Публикация результатов эксперимента в рецензируемых научных журналах. Представление баз данных в Репозиториях открытого доступа

На основании статьи о результатах эксперимента по верификации программа/технология может быть включена в «Реестр психолого-педагогических программ и технологий с доказанной эффективностью для образования и социальной сферы» Министерства просвещения Российской Федерации. Результаты должны быть представлены в соответствии с рекомендациями и требованиями к **Отчету о результатах верификации**, общая структура которого представлена в **Приложении 1, п. 4.1**.

Базы данных исследований рекомендуется размещать в репозиториях открытого доступа, таких как Репозиторий психологических исследований и инструментов RusPsyDATA <https://ruspsydata.mgppu.ru/> МГППУ (Рис. 6). Разработчиком и оператором репозитория RusPsyDATA является Научно-практический центр по комплексному сопровождению психологических исследований PsyDATA <https://mgppu.ru/project/417> МГППУ (далее — Центр PsyDATA).



Рис. 6. Репозиторий психологических исследований и инструментов RusPsyDATA ФГБОУ ВО «Московский государственный психолого-педагогический университет»

Репозитории баз данных с каждым годом приобретают все большую распространенность. Это связано с укреплением позиций доказательного



подхода в науке, а также с растущей прозрачностью научных исследований. Все больше журналов заявляют обязательным при подаче статьи условие публикации базы данных исследования. Одним из требований многих научных грантов также является размещение базы данных в открытом доступе. Научные работы со ссылками на открытые базы данных выглядят более надежными в глазах мирового научного сообщества. По данным некоммерческого издательства PLOS <https://vimeo.com/865129965>, статьи, в которых есть ссылки на данные в репозиториях открытого доступа, имеют показатель цитируемости на 25% выше, чем статьи, авторы которых не делятся данными или делятся только файлами со вспомогательной информацией.

### **Семь преимуществ репозитория RusPsyDATA:**

1. **Базы данных хранятся в России.** Таким образом, размещение баз данных пользователей будет полностью соответствовать законам РФ.
2. **Доступ к базе данных из любого места и устройства.** При утере авторской копии на носителях либо отсутствии ее под рукой пользователя всегда смогут загрузить себе свои данные.
3. **Каждой базе данных присваивается DOI (Digital Object Identifier)** с библиографическим описанием на этот материал.
4. **Репозиторий интегрирован с РИНЦ (Российский индекс научного цитирования):**
  - Каждая база данных является публикацией и индексируется в РИНЦ, увеличивая количество публикаций, что особенно актуально для молодых ученых.
  - База данных может ссылаться на использованную литературу, повышая ее цитируемость.
  - На базу данных пользователя можно ссылаться в других публикациях, что также повышает цитируемость и индекс Хирша.
  - В описании базы данных можно указать статьи, в которых использовались данные из базы. Если внешнего читателя заинтересует размещенная база данных, он скорее всего посмотрит и статьи ее автора.
5. **Минимальное количество полей, требуемых при публикации базы данных.** Центр PsyDATA сократил поля до самых необходимых, сделав многие не обязательными к заполнению.
6. **Возможность опубликовать базу в режиме «эмбарго», т.е. с отложенным во времени доступом к самим данным.** Авторизованный пользователь загружает базу, но доступ к самим данным будет только начиная с указанного им момента, например, после окончания проекта или публикации статьи по результатам исследования.
7. **Минималистический пользовательский интерфейс.** Только самые необходимые функции для работы с базами данных и ничего лишнего.
8. **Простая процедура регистрации.** Разместить свои базы данных в репозитории могут только авторизованные пользователи. **Кнопка**



регистрации в репозитории RusPsyDATA находится в правом верхнем углу сайта <https://ruspsydata.mgppu.ru/>. Инструкция пользователя открывается нажатием на кнопку в верхней строке меню сайта.

Примеры размещенных в репозиториях баз данных и статей с результатами верификации можно найти ниже в списке литературы в разделе **3.3. Статьи по тематике эмпирических исследований верификации программ и технологий**, в том числе: [Марголис А.А., Сорокова М.Г., Шепелева Е.А., Расходчикова М.Н., Шведовская А.А., & Радчикова Н.П., Датасет, 2023]. [Андерсон С.А., Барковская Ю.В., Кормухина Е.В., Клопотова Е.Е., Леонова О.И., Оглоблин А.А., Севостьянов Ю.О., Федоров В.В., Датасет, 2024]. [Токарчук, Ю. А., Саломатова, О. В., & Рубцова, О. В., Датасет, 2024].

### **Этап 6. Включение верифицированной программы или технологии в «Реестр психолого-педагогических программ и технологий с доказанной эффективностью» и распространение опыта в профессиональном сообществе и СМИ**

Включение в федеральный «Реестр психолого-педагогических программ и технологий с доказанной эффективностью для образования и социальной сферы» программы или технологии, успешно прошедшей верификацию, подтверждает высокий уровень качества профессиональной деятельности и профессионального продукта в работе специалистов. К настоящему времени данная проблема приобрела комплексный характер, так как касается и подготовки (повышения квалификации, переподготовки) специалистов, и их аттестации, сертификации. В этой связи «Реестр психолого-педагогических программ и технологий с доказанной эффективностью для образования и социальной сферы», по существу, является базовой платформой демонстрации соблюдения современных нормативных требований к профессиональной деятельности и подтверждения качества профессиональной деятельности и ее продукта.

Отметим, что для верифицированных программ и технологий включение в «Реестр ...» означает новый этап их развития уже на более высоком профессиональном уровне. Они могут дорабатываться, совершенствоваться, транслироваться, становиться основой для масштабированных экспериментов и перехода в более высокую категорию соответствия стандартам программ и технологий с доказанной эффективностью для сферы образования и социальной сферы.

**Основными внешними результатами размещения программ и технологии в федеральном реестре** выступают: повышение доверия целевой аудитории, рекомендации экспертов в публичном и профессиональном

пространстве, развитие методического обеспечения подготовки специалистов на основе лучших практик, развитие новых форм совместной деятельности разных организаций, в том числе межведомственного взаимодействия государственных органов, университетов, научных, образовательных и социальных организаций всех уровней, а также общественных организаций, социально ориентированных некоммерческих организаций, автономных некоммерческих организаций (АНО), фондов, работающих в сфере науки, образования, здравоохранения и социальной сфере.

С другой стороны, компенсация труда авторов программы или технологии, прошедшей верификацию, будет выражаться в личностном и профессиональном саморазвитии, самоактуализации, самореализации, поддержке и внимании, получаемых от государственных органов управления образованием и социальной сферой, грантодателей, расширении круга профессиональных знакомств, обмену опытом, приобретении статуса эксперта в ведущих профессиональных объединениях.

## Литература

### 1.1. Документы, справочные издания и руководства-прототипы

- 1) Национальный стандарт доказательности практик в сфере детства [Электронный ресурс]. — М., 2023.
- 2) Руководство по стандартизации психодиагностического инструментария: требования и оценка качества : учебное пособие / М.Г. Сорокова, Е.Ю. Карданова, Н.П. Радчикова, В.В. Федоров; под ред. Сороковой М.Г. — М. : ФГБОУ ВО МГППУ, 2024. — 48 с.
- 3) Система функционирования психологических служб в общеобразовательных организациях: методические рекомендации / Авт. коллектив. М.: Издательство ФГБОУ ВО МГППУ, 2020. — 120 с.
- 4) Сорокова М.Г., Леонова О.И., Ботова Ю.Б., Пятаков Е.О. Организация независимой оценки психолого-педагогических программ и стандартизации психодиагностических инструментов в парадигме доказательного подхода // Современное дополнительное профессиональное педагогическое образование. 2023. Т. 6. № 4(23). С. 143–156
- 5) Institute of Education Sciences (IES). Available at: <https://ies.ed.gov/> (Accessed 18.05.2022)
- 6) IES What Works Clearinghouse. Handbooks and Reviewer resources. Available at: <https://ies.ed.gov/ncee/wwc/Handbooks> (Accessed 10.05.2022)
- 7) Publication Manual of the American Psychological Association, 6th Edition, Available at: <http://www.apastyle.org> (Accessed 10.05.2022)
- 8) What Works Clearinghouse Procedures and Standards Handbook, Version 3.0, Mar 2014 Available at: <https://ies.ed.gov/ncee/wwc/Handbooks> (Accessed 10.05.2022)
- 9) What Works Clearinghouse Procedures Handbook, Version 4.1 Available at: <https://ies.ed.gov/ncee/wwc/Handbooks> (Accessed 10.05.2022)
- 10) What Works Clearinghouse Procedures and Standards Handbook, Version 5.0 Available at: <https://ies.ed.gov/ncee/wwc/Handbooks> (Accessed 10.05.2022)
- 11) Публикационное руководство Американской ассоциации психологов (АПА) на Портале психологических изданий PsyJournals.ru — [https://psyjournals.ru/info/homestyle\\_guide/article\\_requirements.shtml#article\\_requirements\\_3](https://psyjournals.ru/info/homestyle_guide/article_requirements.shtml#article_requirements_3) [III. Требования к содержанию статей — PsyJournals.ru]

## 1.2. Литература по методологии и методам психолого-педагогических исследований

- 12) Бусыгина Н.П. Качественные и количественные методы исследований в психологии: учебник для бакалавриата и магистратуры / Н.П. Бусыгина. М.: Издательство Юрайт, 2019. 423 с.
- 13) Бусыгина Н.П. Методологические основания качественных исследований в психологии: дисс. ... канд. психол. наук. М., 2010. 149 с.
- 14) Бусыгина Н.П., Подушкина Т.Г., Зевина Д.А. Методологические проблемы доказательного описания психологических и социальных практик // Социальные науки и детство. 2021. Том 2. № 1. С. 24–36. DOI: <https://doi.org/10.17759/ssc.2021020102>
- 15) Бусыгина Н.П., Подушкина Т.Г., Станилевский В.В. Доказательный подход в социальной сфере: основные понятия и принципы, история, перспективы [Электронный ресурс] // Социальные науки и детство. 2020. Том 1. № 1. С. 8–26. DOI:10.17759/ssc.2020010101
- 16) Василюк Ф.Е., Зинченко В.П., Мещеряков Б.Г., Петровский В.А., Пружинин Б.И., Щедрина Т.Г. Методология психологии: проблемы и перспективы. Учебное пособие. М.; СПб.: Центр гуманитарных инициатив, 2012 509 с.
- 17) Войскунский А.Е., Скрипкин С.В. Качественный анализ данных // Вестник Московского университета. Серия. 2001. Т. 14. С. 93–109.
- 18) Выготский Л.С. Игра и ее роль в психическом развитии ребенка // Вопросы психологии. 1966. № 6. С. 62–72.
- 19) Выготский Л.С. Собрание сочинений: В 6-ти т. Т.3. Проблемы развития психики / Под ред. А.М. Матюшкина. М.: Педагогика, 1983. 368 с.
- 20) Выготский Л.С. Психология развития человека. Вступительная статья А.А. Леонтьева. М.: Изд-во Смысл; Изд-во Эксмо, 2005. 1136 с.
- 21) Выготский Л.С. Лекции по педологии. Ижевск, Издательский дом «Удмуртский университет», 2001. 304 с.
- 22) Гальперин П.Я., Кабыльницкая С.Л. Экспериментальное формирование внимания. М.: Издательство Московского Университета, 1974. 102 с.
- 23) Давыдов В.В. Теория развивающего обучения / Рос. Акад. образования, Психологический ин-т, Междунар. Ассоциация «Развивающее обучение». М.: ИНТОР, 1996. 544 с.
- 24) Дружинин В.Н. Экспериментальная психология. 2-е изд., доп. М.: Юрайт, 2019. 386 с.
- 25) Корнилова Т.В., Смирнов С.Д. Методологические основы психологии. М.: Юрайт, 2024. 474 с.
- 26) Марголис А.А. Требования к модернизации основных профессиональных образовательных программ (ОПОП) подготовки педагогических

- кадров в соответствии с профессиональным стандартом педагога: предложения к реализации деятельностного подхода в подготовке педагогических кадров // Психологическая наука и образования. 2014. С. 105–126.
- 27) Марголис А.А. Что смешивает смешанное обучение? // Психологическая наука и образование. 2018. Том 23. № 3. С. 5–19. DOI: 10.17759/pse.2018230301
- 28) Марголис А.А. Зона ближайшего развития, скаффолдинг и деятельность учителя // Культурно-историческая психология. 2020а. Том 16. № 3. С. 15–26. DOI: 10.17759/chp.2020160303
- 29) Марголис А.А. Зона ближайшего развития (ЗБР) и организация учебной деятельности учащихся // Психологическая наука и образование. 2020b. Том 25. № 4. С. 6–27. DOI: 10.17759/pse.2020250402
- 30) Марголис А.А. Деятельностный подход в педагогическом образовании // Психологическая наука и образование. 2021. Том 26. № 3. С. 5–39. DOI: 10.17759/pse.2021260301
- 31) Марголис А.А., Сафронова М.А. Итоги комплексного проекта по модернизации педагогического образования в Российской Федерации (2014–2017 гг.) // Психологическая наука и образование. 2018. Том 23. № 1. С. 5–24. doi:10.17759/pse.2018230101
- 32) Мельникова О.Т., Хорошилов Д.А. Методологические принципы качественных исследований в психологии // Вестник Московского университета. Серия 14. Психология. 2013. № 3. С. 4–17.
- 33) Мельникова О.Т., Хорошилов, Д.А. Стратегии валидизации качественных исследований в психологии // Психологические исследования. 2015. Т. 8. № 44. С. 3. <https://doi.org/10.54359/ps.v8i44.504>
- 34) Наследов А.Д. Математические методы психологического исследования. Анализ и интерпретация данных. СПб.: Речь, 2004. 392 с.
- 35) Наследов А.Д. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. СПб.: Питер, 2013. 416 с.
- 36) Обухова Л.Ф. Основы общей (генетической) психологии. Теория П.Я. Гальперина и формирующий эксперимент: монография: научное наследие / под ред. Г.В. Бурменская, И.В. Шаповаленко, А.А. Шведовская. М.: ФГБОУ ВО МГППУ, 2022.—292 с.
- 37) Пахальян В.Э. Качество профессиональной деятельности и ее продукта в работе практического психолога: проблема определения понятий, критериев и способов оценки [Электронный ресурс] // Вестник практической психологии образования. 2021. Том 18. № 4. С. 9–21. doi:10.17759/bppe.2021180401

- 38) Рубцов В.В., Романова Е.С. Психолого-педагогическая практика в системе образования (итоги Всероссийского конкурса лучших психолого-педагогических программ и технологий в образовательной среде — 2021) [Электронный ресурс] // Вестник практической психологии образования. 2022. Том 19. № 1. С. 8–17. doi:10.17759/bpre.2022190101
- 39) Совместная учебная деятельность и развитие детей / Коллективная монография. Под редакцией В.В. Рубцова, И.М. Улановской. М.: ФГБОУ ВО МГППУ, 2021. 352 с.
- 40) Сорокова М.Г. Математические методы в психолого-педагогических исследованиях: Учебное пособие / М.Г. Сорокова. М.: Неолит, 2020а. 216 с. DOI 10.17759/psychlib/978-5-6043562-0-3
- 41) Сорокова М.Г. Методы математической статистики в психологии [Электронный ресурс]: учебное пособие. — Саарбрюкен: Palmarium Academic Publishing, 2014. 405 с. URL: <http://psychlib.ru/mgppu/SMm-2014/MMa-405.htm#hid2> (дата обращения: 20.05.2022).
- 42) Сорокова М.Г. Воспитание и обучение дошкольников по системе М. Монтессори: Учебное пособие / М.Г. Сорокова. М.: Неолит, 2020b. 440 с.
- 43) Сорокова М.Г., Сороков Д.Г. Математика по методу Монтессори для дошкольников и школьников: Учебное пособие / М.Г. Сорокова, Д.Г. Сороков. М.: Редкая птица, 2018. 400 с. URL: Электронная библиотека МГППУ: Сорокова, Сороков. Математика по методу Монтессори: для дошкольников и школьников: Учебное пособие (psychlib.ru) (дата обращения: 10.03.2023).
- 44) Хорошилов Д.А., Балашова Е.Ю. Качественные методы в социальной и клинической психологии: пролегомены к междисциплинарному диалогу // Социальная психология и общество. 2018. Т. 9. № 3. С. 21–30. doi:10.17759/sps.2018090303
- 45) Чудинова Е.В. Инициативная опора школьников на схемы в самостоятельном решении задач // Психологическая наука и образование. 2021. Том 26. № 2. С. 66–77. DOI: 10.17759/pse.2021260206
- 46) Box G.E.P., Anderson S.L. Permutation theory in the derivation of robust criteria and the study of departures from assumptions // Journal of the Royal Statistical Society, 1955. Vol. 17. P. 1–34.
- 47) Box G.E.P. Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variances in the one-way classification // Annals of Mathematical Statistics. 1954a. Vol. 25. P. 290–302.
- 48) Box G.E.P. Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effect of inequality of variances and

- of correlation of errors in the two-way classification // *Annals of Mathematical Statistics*. 1954b. Vol. 25. P. 484–498.
- 49) Ebrahim A. *Measuring Social Change: Performance and Accountability in a Complex World*. Stanford University Press, 2019. 320 p.
- 50) Ebrahim A., Rangan V.K. What Impact: A Framework for Measuring the Scale and Scope of Social Performance // *California Management Review*. 2014. Vol. 56(3). P. 118–41. DOI:10.1525/cm.2014.56.3.118
- 51) Gambrell E. Evidence-based practice and the ethics of discretion // *Journal of Social Work*. 2011. Vol. 11(1). P. 26–48.
- 52) Ho D.E., Imai K., King G., Stuart, E.A. Matching as nonparametric pre-processing for reducing model dependence in parametric causal inference // *Political Analysis*, 2007. Vol. 15, P. 199–236. <http://dx.doi.org/10.1093/pan/15.1.199>
- 53) Lindman H.R. *Analysis of variance in complex experimental designs*. San Francisco: W.H. Freeman & Co.1974. 352 p.
- 54) Lyons E., Coyle A. (Eds.). *Analysing qualitative data in psychology*. Sage Publications Ltd. 2007. <https://doi.org/10.4135/9781446207536>
- 55) Manuel J.I., Mullen E.J., Fang L., Bellamy J.L., Bledsoe S.E. Preparing social work practitioners to use evidence-based practice: A comparison of experiences from an implementation project // *Research on Social Work Practice*. 2009. Vol. 19. P. 613–627.
- 56) McNeece C.A., Thyer B.A. Evidence-based practice and social work // *Journal of Evidence-Based Social Work*. 2004. Vol. 1 (1). P. 7–25.
- 57) Melnikova O.T., Khoroshilov D.A. Priority research directions in the area of qualitative methodology // *Psychology in Russia: State of the Art. Scientific Yearbook*. M.: Russian Psychology Society, 2010. P. 46–72.
- 58) Mullen E.J., Bledsoe S.E., & Bellamy, J.L. Implementing evidence-based social work practice. // *Research on Social Work Practice*. 2008. Vol. 18 (4). P. 325–338.
- 59) Mullen E.J., Dumpson J.R. (Eds.) *Evaluation of Social Intervention*. San Francisco: Jossey-Bass, 1972. 282 p.
- 60) Mullen E.J., Streiner D.L. The Evidence For and Against Evidence-Based Practice // *Brief Treatment and Crisis Intervention*. 2004. Vol. 4(2). P. 111–121.
- 61) Pawson R. Evidence-based Policy: In Search of a Method // *Evaluation*. 2002. Vol. 8 (2). P. 157–181.
- 62) Proctor E.K., Rosen A. From knowledge production to implementation: research challenges and imperatives // *Research on Social Work Practice*. 2008. Vol. 18 (4). P. 285–291.
- 63) Pustejovsky J.E., Hedges L.V., Shadish W.R. Design-Comparable Effect Sizes in Multiple Baseline Designs: A General Modeling Framework //

- Journal of Educational and Behavioral Statistics. 2014. Vol. 39 (5). P. 368–393. <https://doi.org/10.3102/1076998614547577>
- 64) Rubin A. Efforts to bridge the gap between research and practice in social work: Precedents and prospects: Keynote address at the Bridging the Gap Symposium // Research on Social Work Practice. 2015. Vol. 25 (4). P. 408–414.
- 65) Rubin A., Parrish D. Challenges to the future of evidence-based practice in social work education // Journal of Social Work Education. 2007. Vol. 43. P. 403–424.
- 66) Sackett D.L., Rosenberg W.M., Gray J.A., Haynes R.B., Richardson W.S. Evidence-based medicine: what it is and what it isn't // British Medical Journal. 1996. Vol. 312 (7023). P. 71–72.
- 67) Sackett D.L., Straus S.E., Richardson W.S., Rosenberg W., Haynes R.B. Evidence-based medicine: How to practise and teach EBM (2nd ed.). New York: Churchill Livingstone. 2000. 261 p.
- 68) Straus S.E., Richardson W.S., Glasziou P., Haynes R.B. (Eds.). Evidence-based medicine. How to practice and teach EBM. Edinburgh: Churchill Livingstone. 2005. 336 p.
- 69) Willig C. Introducing Qualitative research in Psychology (2nd ed.). New York: Open University Press. McGraw — Hill Education. 2008. 250 p.
- 70) Zwet R van der. Implementing Evidence-Based Practice in Social Work: A Shared Responsibility. Ipskamp Drukkers BV, Enschede, the Netherlands, 2018. 172 p.

### **1.3. Статьи по тематике эмпирических исследований верификации программ и технологий**

- 71) Андерсон С.А., Барковская Ю.В., Кормухина Е.В., Клопотова Е.Е., Леонова О.И., Оглоблин А.А., Севостьянов Ю.О., Федоров В.В. Измерение наглядно-образного мышления дошкольников 6–7 лет по методике «Схематизация» (выборка стандартизации) [Электронный ресурс]: [Набор данных] // RusPsyData: Репозиторий психологических исследований и инструментов. Москва, 2024. DOI: 10.48612/MSUPE/zgz2-an91-h9nf
- 72) Зак А.З. Условия формирования познавательных метапредметных результатов у младших школьников [Электронный ресурс] // Психолого-педагогические исследования. 2018. Том 10. № 2. С. 11–20 doi: 10.17759/psyedu.2018100202c
- 73) Зарецкий В.К. Становление и сущность рефлексивно-деятельностного подхода в оказании консультативной психологической помощи // Консультативная психология и психотерапия. 2013. Том 21. № 2. С. 8–37.



- 74) Зарецкий В.К. Один шаг в обучении — сто шагов в развитии: от идеи к практике // Культурно-историческая психология. 2016. Том 12. № 3. С. 149–188. DOI: 10.17759/chp.2016120309
- 75) Зарецкий Ю.В., Зарецкий В.К., Кулагина И.Ю. Методика исследования субъектной позиции обучающихся разных возрастов // Психологическая наука и образование. 2014. Том 19. № 1. С. 99–110.
- 76) Марголис А.А., Сорокова М.Г., Шепелева Е.А., Гаврилова Е.В., Расходчикова М.Н., Шведовская А.А., Дробязько А.А., Вихристюк О.В., Прокопьева Л.М., Лисицына А.А., Радчикова Н.П., Виноградов А.В. Просветительские мероприятия по вакцинопрофилактике COVID-19: эффективны ли они? // Вопросы образования / Educational Studies Moscow. 2023. № 4. С. 151–178. <https://doi.org/10.17323/vo-2023-17036>
- 77) Марголис А.А., Сорокова М.Г., Шепелева Е.А., Расходчикова М.Н., Шведовская А.А., Радчикова Н.П. Доказательный подход: верификация программы просветительских мероприятий в области вакцинопрофилактики COVID-19 [Датасет]. RusPsyData: Репозиторий психологических исследований и инструментов. <https://doi.org/10.48612/MSUPE/fbu1-pnhb-9dbn>
- 78) Радчикова Н.П., Одинцова М.А., Сорокова М.Г. Предпосылки для принятия цифровой образовательной среды в новых культурно-исторических условиях // Культурно-историческая психология. 2021. Том 17. № 3. С. 115–124. doi:10.17759/chp.2021170315
- 79) Рубцова О.В., Посакалова Т.А., Соловьева А.Г. Театр как деятельностная технология воспитания и формирования личностных образовательных результатов // Психологическая наука и образование. 2022. Том 27. № 1. С. 52–64. DOI: 10.17759/pse.2022270105
- 80) Сорокова М.Г. Электронный курс как цифровой образовательный ресурс смешанного обучения в условиях высшего образования // Психологическая наука и образование. 2020с. Том 25. № 1. С. 36–50. doi:10.17759/pse.2020250104
- 81) Сорокова М.Г. Предметные результаты студентов в цифровой среде университета на разных уровнях высшего образования: так кто же более успешен? // Психологическая наука и образование. 2021. Том 26. № 1. С. 76–91. doi:10.17759/pse.2021260105
- 82) Сорокова М.Г. Цифровая образовательная среда университета: кому более комфортно в ней учиться? // Психологическая наука и образование. 2020d. Том 25. № 2. С. 44–58. doi:10.17759/pse.2020250204
- 83) Сорокова М.Г., Одинцова М.А., Радчикова Н.П. Образовательные результаты студентов в электронных курсах при смешанном и онлайн-обучении // Моделирование и анализ данных. 2021a. Том 11. № 1. С. 61–77. doi:10.17759/mda.2021110105

- 84) Сорокова М.Г., Одинцова М.А., Радчикова Н.П. Шкала оценки цифровой образовательной среды (ЦОС) университета // Психологическая наука и образование. 2021b. Том 26. № 2. С. 52–65. doi:10.17759/pse.2021260205
- 85) Токарчук Ю.А., Саломатова О.В., Рубцова О.В. Опрос родителей по использованию настольных и цифровых игр дошкольниками [Датасет]. RusPsyData: Репозиторий психологических исследований и инструментов. <https://doi.org/10.48612/MSUPE/236d-93e5-nta6>
- 86) Чумаченко Д.В., Бойкина Е.Э., Чиркина Р.В., Романова Н.М., Киселёв К.А. ШНПО-ПМ [Электронный ресурс]: [Набор данных] // RusPsyData: Репозиторий психологических исследований и инструментов. Москва, 2023. DOI:10.48612/MSUPE/he4e-h7rv-rhkt
- 87) Sorokova M.G. Skepticism and learning difficulties in a digital environment at the Bachelor's and Master's levels: are preconceptions valid? // Heliyon. 2020e. Vol. 6, Issue 11, E05335. DOI: <https://doi.org/10.1016/j.heliyon.2020.e05335>

## Приложение 1

# Стандарты и общие требования к проведению эксперимента по верификации психолого-педагогических программ и технологий и оценке его результатов

## 1. Дизайн исследования: основные виды дизайна

В исследованиях Института наук об образовании (Institute of Education Sciences, IES) (США), представленных на портале What Works Clearinghouse (WWC), рассматриваются 2 основных вида дизайна групповых исследований [WWC Procedures and Standards Handbook, V. 3.0], [WWC Procedures and Standards Handbook, V. 5.0] — рандомизированные контролируемые испытания или исследования (Randomized controlled trials, RCT, РКИ) и квази-экспериментальный дизайн (Quasi-experimental design, QED, КЭД). Наряду с ними в качестве допустимых рассматриваются и другие виды дизайна — дизайн с регрессионным разрывом (Regression discontinuity design, RDD, DPP) и дизайн с одним случаем (single-case design, SCD, кейс-стади, ДОС). Оценка эффективности вмешательства, формирующего воздействия для разных видов дизайна имеет свою специфику. Кратко охарактеризуем их.

### 1.1. Рандомизированные контролируемые испытания (РКИ)

Рандомизированные контролируемые испытания (РКИ) [WWC Procedures and Standards Handbook, V. 3.0, С. 9] предполагают, что участники исследования (испытуемые) распределяются случайным образом в две или более групп, которые различаются по тому, получают ли они изучаемое вмешательство в процессе формирующего эксперимента или нет. Если группа подвергается формирующему воздействию, она считается экспериментальной (или группой вмешательства), а если не подвергается этому воздействию или же получает какое-то другое вмешательство, отличное от изучаемого, то она считается контрольной (или группой сравнения). На констатирующем этапе эксперимента, на момент определения выборки (и до вмешательства) группы должны быть в среднем схожи как по наблюдаемым, так и по ненаблюдаемым характеристикам. Этот дизайн позволяет интерпретировать любые последующие (т. е. после вмешательства) различия в исследуемых психолого-педагогических характеристиках

(переменных, признаках, конструктах и др.) между группой вмешательства и группой сравнения исключительно как результат вмешательства.

В исследованиях может использоваться случайное распределение на разных уровнях: **на индивидуальном уровне** и **на уровне кластера**. В рамках многоуровневой структуры тип данных и уровень анализа могут различаться. Индивидуум — это отдельный испытуемый, относящийся к изучаемой категории: в педагогических исследованиях это чаще всего обучающийся. *Анализ на индивидуальном уровне* — это анализ, проводимый с использованием данных для каждого человека. *Кластер* — это группа людей; в исследованиях в области образования это часто классная комната или школа. *Анализ на уровне кластера* — это анализ, проводимый с использованием данных для каждого кластера, которые часто представляют собой совокупность данных отдельных лиц в кластере в определенный момент времени. По терминологии WWC [WWC Procedures and Standards Handbook, V. 3.0, С. 10], среди индивидуумов *внутри кластера* **стайеры** (stayers) — это те, кто был в выборке как до, так и после вмешательства; **выбывшие** (leavers) — это те, кто был в выборке до вмешательства, но не после него; а **присоединившиеся** (joiners) — это те, кто попал в выборку (из другого кластера) после вмешательства, но не до него. Анализ на уровне кластера предполагает коррекцию результата при вычислении размера эффекта (см. **Приложение 2, п. 2.3**).

В кластерном РКИ, в котором кластеры представляют собой случайно распределенные единицы, нет необходимости в случайном распределении отдельных лиц по кластерам. Анализ стайеров и присоединившихся на уровне кластера, используемый для ответа на исследовательский вопрос на уровне кластера, может быть допустимым. Если же анализ проводится на индивидуальном уровне, любое неслучайное перемещение или помещение людей в группы вмешательства или сравнения после случайного распределения ставит под угрозу план исследования со случайным распределением. Индивидуальные исследования стайеров или стайеров и присоединяющихся могут быть допустимыми с оговорками, если исследование способно продемонстрировать базовую эквивалентность аналитической выборки.

**Аналитическая выборка** — это выборка, полученная на контрольном этапе эксперимента и оставшаяся после отсева, на основании которой делаются выводы о размере эффекта. **Базовая эквивалентность** (Baseline Equivalence, БЭ) в самом общем виде понимается как отсутствие существенных различий между экспериментальной группой (другими словами — группой вмешательства) и контрольной группой (т.е. группой сравнения) на констатирующем этапе эксперимента, т.е. до вмешательства, формирующего воздействия.

## 1.2. Квази-экспериментальный дизайн (КЭД)

Исследования квази-экспериментального дизайна (КЭД) должны так же продемонстрировать базовую эквивалентность, однако даже в этом случае они имеют некоторые ограничения. Квази-экспериментальный план сравнивает результаты учащихся, классов или школ, которые имели доступ к вмешательству, с теми, кто не имел доступа, но имел сходные наблюдаемые характеристики. Группы участников и не участвующих в формирующем эксперименте могут образовываться по многим причинам. Например, административный округ или город может выбрать пилотную экспериментальную программу по математике в одних школах, а не в других; учителя некоторых классов могут согласиться включить в свою учебную программу дополнительные занятия по чтению или другим предметам, тогда как другие могут не согласиться; или группа учащихся может иметь право на участие в программе продленного дня или дополнительных занятий, тренингов и др., но только некоторые из родителей могут выбрать участие для своих детей в такой программе. В каждом случае характеристики групп вмешательства и невмешательства (сравнения) различаются. Они могут различаться по характеристикам, которые мы можем наблюдать, например, по результатам тестов, или по тому, что мы не можем четко наблюдать, например, по мотивации. Даже при эквивалентности наблюдаемых характеристик могут существовать различия в ненаблюдаемых характеристиках, которые могут внести погрешность в оценку эффекта вмешательства.

## 1.3. Дизайн с регрессионным разрывом (ДРР)

Дизайн с регрессионным разрывом (ДРР) [WWC Procedures and Standards Handbook, V. 5.0, С. 62] — это групповой план, в котором может применяться распределение испытуемых по группам (условиям эксперимента, вмешательства) на индивидуальном уровне или на уровне кластера, подобно другим групповым планам. Исследователи в области образования используют ДРР, когда вмешательства предоставляются отдельным лицам или группам на основе того, как они соотносятся с **пороговым значением по какому-либо известному показателю**. Группы, таким образом, формируются не случайно, а целенаправленно, конструктивно, с использованием отсечки по непрерывной переменной — так называемой **переменной «вынуждения» или «назначения»**.

В исследованиях ДРР эффект вмешательства оценивается путем сравнения двух линий регрессии: одна представляет взаимосвязь между вынуждающей переменной и результатом в группе вмешательства и аналогичным образом для группы сравнения. Разница в этих двух линиях регрессии при пороговом значении вынуждающей переменной представляет предполагаемый эффект вмешательства.

Выбор вынуждающей переменной и конкретного предела, используемого для назначения вмешательства, т.е. включения испытуемых в экспериментальную группу, обычно определяется руководством, администрацией учреждения или авторами исследования. Например, администрация района может направить учащихся в программу летней школы, если их баллы по стандартизированному тесту ниже порогового значения, или школам может быть предоставлен грант на основе их баллов, набранных по заявке на грант.

Согласно стандартам WWC, процесс рассмотрения и оценки эффективности ДПП аналогичен процессу, используемому для рандомизированных контролируемых испытаний (РКИ) и квази-экспериментальных планов (КЭД), включая скрининг исследований на соответствие требованиям, оценку исследований в соответствии со стандартами WWC и отчет о результатах. Однако механизм проверки в некоторых случаях значительно отличается от рассмотрения WWC других групповых планов.

#### 1.4. Дизайн с одним случаем (кейс-стади, ДОС)

Дизайны с одним случаем (ДОС) [WWC Procedures and Standards Handbook, V. 5.0, С. 103] представляют собой экспериментальные планы с потенциалом демонстрации причинно-следственных эффектов, которые обычно включают **небольшое количество участников**. Исследования ДОС должны в целом придерживаться одних и тех же руководящих принципов с точки зрения их результатов и общих требований приемлемости. Однако, когда исследования ДОС сосредоточены на оценках причинно-следственных связей, они отличаются от групповых планов тем, как они генерируют оценки причинно-следственных связей. В результате исследования ДОС требуют другого процесса обзора с другими конкретными стандартами по сравнению с групповыми планами. Исследования ДОС также могут содержать более одного эксперимента, и каждый эксперимент должен получить собственную оценку.

Дополнительные квалификационные требования для ДОС:

1. Отдельный случай является единицей проведения вмешательства и анализа данных. Такой кейс чаще всего представляет собой **одного участника**. Это также может быть **группа участников**, например классная комната или школа.
2. В рамках дизайна исследования кейс (один участник или группа участников) может обеспечивать собственный контроль для целей сравнения. Например, серия повторных измерений результатов до вмешательства сравнивается с серией повторных измерений результатов во время и после вмешательства.
3. Переменная результата многократно измеряется в различных условиях. Эти различные условия часто структурированы как фазы, такие как

первая базовая фаза, первая фаза вмешательства, вторая базовая фаза и вторая фаза вмешательства.

Оценка эффективности психолого-педагогической, образовательной или социальной программы, технологии будет проводиться на основе количественного анализа данных, полученных в эксперименте, с учетом специфики дизайна исследования.

## 2. Отсев выборки

Хорошо известно, что психолого-педагогический эксперимент, в том числе, нацеленный на верификацию образовательных и социальных программ и технологий, всегда связан с отсевом первоначальной выборки. Отсевом называется потеря участников исследования, то есть невозможность для них по каким-либо причинам продолжение участия в исследовании. Различают **общий отсев** и **дифференциальный отсев**. Общий отсев — это процент всех участников исследования, которые отсеялись из общей выборки (из всех исследуемых групп). Вычисляется, как число отсеявшихся, деленное на общий объем выборки и умноженное на 100%. Дифференциальный отсев — это абсолютная величина разницы между отсевами в контрольной и экспериментальной группах. Вычисляется как разность между отсевами, вычисленными отдельно для контрольной группы и отдельно для экспериментальной группы. Оба вида отсева являются важными показателями и должны учитываться при проведении и анализе исследования.

### 2.1. Отсев в РКИ индивидуального уровня

Хорошим ориентиром для оценки показателей отсева для РКИ индивидуального уровня могут служить рекомендации WWC [WWC Procedures and Standards Handbook, V. 3.0, С. 11–15], которые определяют пороговые значения общего и дифференциального отсева и их соотношение. Для практических целей можно воспользоваться рисунком (Рис. 7) или таблицей (Табл. 3). Для конкретного исследования вычисляется общий и дифференциальный отсев, и по этим двум координатам точка помещается на Рис. 7. Если она попала в зеленую зону, то рисков для валидности исследования нет. Если точка находится в желтой зоне, то исследование не может считаться валидным при консервативных предположениях, но может считаться валидным при либеральных предположениях. Если точка попала в красную зону, то исследование не может считаться валидным: слишком высоки риски влияния отсева на результирующие показатели.

Уточним понятия «**либеральных**» и «**консервативных**» предположений. При разработке протокола рецензирования эксперты рассматривают типы выборок и **возможную взаимосвязь между отсевом и результатами**



**исследований в данной области.** Когда есть основания полагать, что большая часть отсева является экзогенной для рассматриваемых вмешательств, то есть не связана со статусом оказанного воздействия в формирующем эксперименте, могут быть уместны более либеральные предположения относительно взаимосвязи между отсевом и выходом. Например, руководство группы экспертов может выбрать либеральный стандарт, если считает, что отсев часто возникает из-за перемещения детей младшего возраста в из одних школ в другие из-за мобильности семьи или из-за случайного отсутствия в дни, когда проводится оценка, тестирование, опрос и т.д. И наоборот, если у руководства команды экспертов есть основания полагать, что отсев во многом является эндогенным для рассматриваемых вмешательств — например, когда обучающиеся старших классов выбирают, участвовать ли в программе дополнительных занятий для учеников с трудностями в обучении или в программе по коррекции девиантного поведения для подростков, склонных к такому поведению, — могут быть уместны более консервативные предположения.

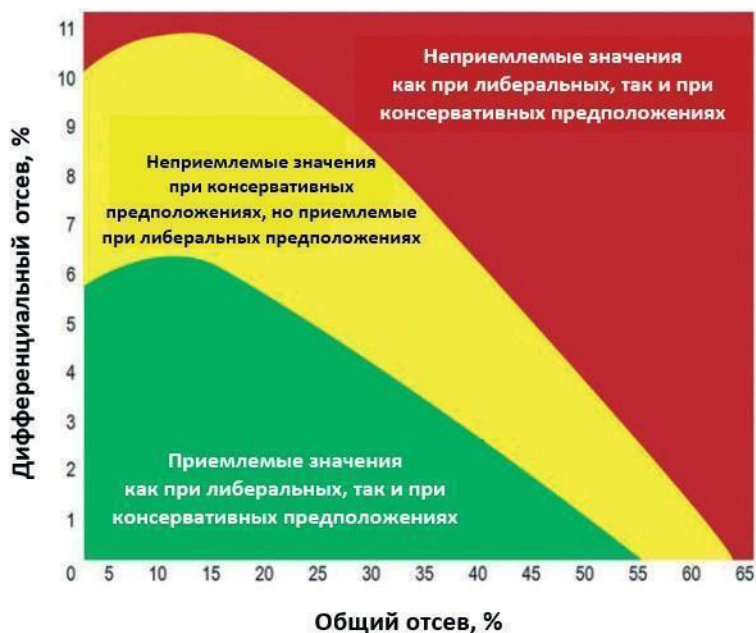


Рис. 7. Соотношение общего и дифференциального отсева при либеральных и консервативных предположениях

При использовании таблицы (Табл. 3) также определяется общий и дифференциальный отсев. Затем находят строку со значением общего



отсева и сравнивают вычисленный дифференциальный отсев с пороговыми значениями на этой строке. Если вычисленный дифференциальный отсев менее или равен пороговому значению дифференциального отсева при консервативном критерии, то исследование можно считать валидным. Если вычисленный дифференциальный отсев более порогового значения дифференциального отсева при консервативном критерии, но менее или равен пороговому значению дифференциального отсева при либеральном критерии, то исследование можно считать валидным при либеральных предположениях. Если вычисленный дифференциальный отсев более порогового значения дифференциального отсева при либеральном критерии, то исследование не может считаться валидным: слишком высоки риски влияния отсева на результирующие показатели.

Таблица 3

**Соотношение общего и дифференциального отсева  
при либеральных и консервативных предположениях**

Общий отсев	Дифференциальный отсев		Общий отсев	Дифференциальный отсев		Общий отсев	Дифференциальный отсев	
	Консервативный критерий	Либеральный критерий		Консервативный критерий	Либеральный критерий		Консервативный критерий	Либеральный критерий
0	5.7	10.0	22	5.2	9.7	44	2.0	5.1
1	5.8	10.1	23	5.1	9.5	45	1.8	4.9
2	5.9	10.2	24	4.9	9.4	46	1.6	4.6
3	5.9	10.3	25	4.8	9.2	47	1.5	4.4
4	6.0	10.4	26	4.7	9.0	48	1.3	4.2
5	6.1	10.5	27	4.5	8.8	49	1.2	3.9
6	6.2	10.7	28	4.4	8.6	50	1.0	3.7
7	6.3	10.8	29	4.3	8.4	51	0.9	3.5
8	6.3	10.9	30	4.1	8.2	52	0.7	3.2
9	6.3	10.9	31	4.0	8.0	53	0.6	3.0
10	6.3	10.9	32	3.8	7.8	54	0.4	2.8
11	6.2	10.9	33	3.6	7.6	55	0.3	2.6
12	6.2	10.9	34	3.5	7.4	56	0.2	2.3
13	6.1	10.8	35	3.3	7.2	57	0.0	2.1
14	6.0	10.8	36	3.2	7.0	58	-	1.9
15	5.9	10.7	37	3.1	6.7	59	-	1.6
16	5.9	10.6	38	2.9	6.5	60	-	1.4
17	5.8	10.5	39	2.8	6.3	61	-	1.1
18	5.7	10.3	40	2.6	6.0	62	-	0.9

Общий отсев	Дифференциальный отсев		Общий отсев	Дифференциальный отсев		Общий отсев	Дифференциальный отсев	
	Консервативный критерий	Либеральный критерий		Консервативный критерий	Либеральный критерий		Консервативный критерий	Либеральный критерий
19	5.5	10.2	41	2.5	5.8	63	-	0.7
20	5.4	10.0	42	2.3	5.6	64	-	0.5
21	5.3	9.9	43	2.1	5.3	65	-	0.3

## 2.2. Отсев выборки в кластерных РКИ

Многие исследования по верификации психолого-педагогических, образовательных и социальных программ и технологий основаны на проектах и дизайнах с несколькими уровнями, например, когда обучающиеся сгруппированы в классах или школах. Исследования, в которых кластеры, а не отдельные члены выборки, случайным образом распределяются по группам вмешательства и сравнения, называются **кластерными РКИ**. Систематическая ошибка в кластерных РКИ может быть вызвана не только потерей кластеров (например, школ), но и потерей членов выборки внутри кластеров (например, учащихся), если они уходят из-за своего статуса воздействия. Чтобы считаться РКИ с низким отсевом, кластерное РКИ, в котором проводится **анализ на индивидуальном уровне** (например, оценка влияния вмешательства на учащихся), **должно иметь низкий отсев на двух уровнях**. Во-первых, он должен иметь низкую убыль **на уровне кластера**, что определяется с помощью установленной выше границы убыли. Во-вторых, исследование должно иметь низкий уровень отсева **на уровне подкластера** (т. е. индивида в кластере), опять же с использованием установленной выше границы отсева, при этом отсев основан только на кластерах, оставшихся в выборке. То есть знаменатель **для расчета отсева подкластера** включает только членов выборки в кластерах (школах или классах), которые остаются в исследовании после отсева кластера.

Однако отсев для кластерного РКИ, в котором проводится **анализ на уровне кластера** (например, оценка влияния вмешательства на классы или школы), будет оцениваться **только на уровне кластера**. Оценки на уровне кластера отражают как воздействие на отдельных лиц (например, обучающихся) внутри кластера, так и изменения в составе отдельных лиц. Исследование будет считаться РКИ с низким отсевом, если оно имеет низкий отсев на уровне кластера с использованием границы отсева, определенной в протоколе (см. Рис. 7 и Табл. 3).

### **2.3. Потери выборки, которые не считаются отсевом**

Выборка, потерянная после первоначального случайного распределения из-за «стихийных явлений», таких как природные пожары или землетрясения, может быть исключена из первоначальной выборки для расчета убыли. Потеря выборки, вызванная стихийными бедствиями, скорее всего, не связана с результатами обучения и, следовательно, не создает возможности для систематической ошибки. Точно так же сбор данных о результатах только для подмножества исходной выборки не считается отсевом, если (1) выделение подвыборок последовательно применяется во всех группах вмешательства и сравнения и (2) подвыборка была отобрана либо случайным образом, либо на основе характеристик, которые были четко определены до случайного распределения (например, национальность, пол). В этих условиях потеря выборки не связана с вмешательством и не приводит к систематической ошибке.

Предполагается, что потеря выборки, возникающая из-за других причин, помимо стихийных бедствий или описанного выше процесса отбора подвыборок, может быть связана с результатами, и поэтому WWC учитывает потерю выборки при расчете отсева. Для данного исследования некоторые потери выборки, возможно, могут быть не связаны с результатами; например, решение об изменении школьной программы или замены учителей может привести к отсеву выборки, которое может быть связано или не быть связано с оцениваемым вмешательством. Такие соображения не учитываются в каждом отдельном исследовании; скорее, как обсуждалось выше, руководство группы экспертов по анализу принимает во внимание степень, в которой отсев в исследованиях, рассматриваемых в тематической области, может быть экзогенным, когда эксперты выбирают либеральный или консервативный стандарт отсева для этой области. Этот подход обеспечивает гибкость в разных областях, делая соответствующие предположения о взаимосвязи между отсевом и результатами, обеспечивая при этом единообразные, воспроизводимые оценки отсева во всех исследованиях в пределах области.

## **3. Базовая эквивалентность для групп в аналитической выборке**

### **3.1. Проверка базовой эквивалентности: рекомендации МГППУ**

Согласно протоколу исследования, в отчете об эксперименте по верификации психолого-педагогических, образовательных и социальных программ и технологий должно быть представлено описание выборки, в том

числе: регион(ы) проведения исследования, учреждения — участники исследования, категория испытуемых — целевые группы детей, в отношении которых в общеобразовательных организациях реализуются программы адресной психологической помощи, в соответствии с классификацией из документа **«Система функционирования психологических служб в общеобразовательных организациях. Методические рекомендации»** [Система функционирования психологических служб, 2020], общий объем выборки (N). Описание экспериментальной группы (ЭГ, группы вмешательства) и контрольной группы (КГ, группы сравнения) приводится с указанием следующих характеристик.

Экспериментальная группа (ЭГ): общая характеристика, объем выборки (N1), пол, возраст (среднее, стандартное отклонение, медиана, мода), другие контекстные характеристики, важные для исследования, например, социо-экономический статус, этнокультурная принадлежность, статус беженца, успеваемость в школе, наличие инвалидности или ОВЗ, нозологическая группа и др.

Контрольная группа (КГ): общая характеристика, объем выборки (N2), пол, возраст (среднее, стандартное отклонение, медиана, мода), другие контекстные характеристики, важные для исследования.

Особое внимание следует уделить **проверке эквивалентности КГ и ЭГ** в аналитической выборке на констатирующем этапе эксперимента, т.е. до оказания вмешательства, формирующего воздействия. По терминологии WWC [WWC Procedures and Standards Handbook, V. 3.0, С. 15], она называется базовой эквивалентностью (Baseline Equivalence, БЭ). **Аналитическая выборка** — это выборка, полученная на контрольном этапе эксперимента и оставшаяся после отсева, на основании которой делаются выводы о размере эффекта.

Есть **разные подходы** к проверке базовой эквивалентности. В российской психологии и образовании это означает, что на констатирующем этапе выборки не должны статистически значимо различаться по важным для исследования параметрам. Если БЭ подтверждается, а на контрольном этапе между ЭГ и КГ выявляются статистически значимые различия по данным параметрам, то эти различия с большой вероятностью можно отнести на счет оказанного вмешательства. При этом необходимо также проверить наличие достоверных сдвигов показателей исследуемых переменных «после минус до». Во всех случаях следует указать размер эффекта в абсолютных и стандартных единицах. В современных стандартных статистических пакетах есть опция вычисления стандартизованного размера эффекта, например, *d* Коэна, *g* Хеджеса,  $\Delta$  Гласса и др. Расчет размера эффекта в стандартных показателях (см. также в **Приложении 2**). Наконец, все результаты должны быть интерпретированы и подлежат обсуждению. Если на констатирующем этапе выявляются статистически значимые различия между группами (т.е., строго говоря, базовая

эквивалентность не установлена), что более вероятно для больших выборок, интерпретация и обсуждение результатов должны быть проведены особенно тщательно с учетом этого факта.

### 3.2. Проверка базовой эквивалентности: рекомендации WWC

Другой подход представлен в **рекомендациях WWC** [WWC Procedures and Standards Handbook, V. 3.0]. РКИ с низким отсевом имеет право на получение наивысшего рейтинга соответствия стандартам дизайна WWC Group без оговорок. Для **РКИ с высоким отсевом** и всех **квази-экспериментальных планов (КЭД)** исследований эквивалентность групп вмешательства и групп сравнения по наблюдаемым характеристикам на исходном уровне (т. е. до периода вмешательства) должна быть установлена для **аналитической выборки** (т.е. учащиеся, школы или классы, которые остаются в конце исследования, когда оцениваются результаты), а не для первоначальных групп в исследовании. Протоколы обзора для каждой тематической области определяют наблюдаемые характеристики, эквивалентность которых должна быть продемонстрирована.

Если зарегистрированная разность какой-либо исходной характеристики **превышает 0,25 стандартных отклонений** в абсолютном значении (на основе вариации этой характеристики в объединенной выборке), группы вмешательства и сравнения считаются **неэквивалентными**. Стандарт, ограничивающий различия между группами до вмешательства до 0,25 стандартных отклонений, основан на работе [Ho D.E., Imai K., King G., Stuart E.A., 2007]. Для различий в базовых характеристиках, которые составляют от 0,05 до 0,25 стандартных отклонений, анализ должен включать статистическую корректировку базовых характеристик, чтобы соответствовать требованию базовой эквивалентности. Различия менее или равные 0,05 не требуют статистической корректировки (Табл. 4).

Таблица 4

**Абсолютная разность размера эффекта  
(Absolute Effect Size Difference) — разность между средними значениями для групп ЭГ и КГ до вмешательства**

<b><math>0.00 \leq</math> Абсолютная разность средних в ЭГ и КГ до вмешательства <math>\leq 0.05</math></b>	<b><math>0.05 &lt;</math> Абсолютная разность средних в ЭГ и КГ до вмешательства <math>\leq 0.25</math></b>	<b><math>0.25 &lt;</math> Абсолютная разность средних в ЭГ и КГ до вмешательства</b>
Удовлетворяет базовой эквивалентности	Требуется статистическая корректировка для обеспечения базовой эквивалентности	Не соответствует базовой эквивалентности

РКИ с высоким отсевом или исследование КЭД может в лучшем случае получить рейтинг «Соответствует стандартам дизайна группы WWC с оговорками», если оно соответствует базовому требованию эквивалентности. **Если базовая эквивалентность не установлена, исследование не соответствует стандартам дизайна WWC Group.** Существует ряд дополнительных соображений, касающихся установления исходной эквивалентности в РКИ с высоким отсевом и исследованиях КЭД:

- Характеристики, по которым должна быть установлена эквивалентность, указываются в протоколе проверки. Эквивалентность исходного уровня часто устанавливается с помощью **теста перед вмешательством для академических показателей**. В обзорах без аналогичных предварительных мер (например, для не окончивших среднюю школу) часто требуется исходная эквивалентность **для демографических характеристик**, связанных с интересующим результатом.
- Если показано, что различия в исходных характеристиках находятся в пределах диапазона, требующего статистической корректировки (**от 0,05 до 0,25 стандартных отклонений**), можно использовать ряд различных методов, включая **регрессионную корректировку и ковариационный анализ (ANCOVA)**. Критическим фактором является то, что базовые характеристики, указанные в протоколе, должны быть включены в анализ на индивидуальном уровне.
- В исследованиях **кластерного дизайна** (например, исследованиях, в которых единицей вмешательства является классная комната или школа, а единицей анализа является учащийся) допускается установление эквивалентности между кластерами группы вмешательства и группы сравнения (например, классы или школы) с использованием либо (а) той же когорты из более раннего момента времени или (б) более ранней соседней когорты, измеренной в том же классе, что и когорта, использованная в анализе воздействия. Измерение на уровне когортного кластера не может использоваться для установления эквивалентности для анализа на индивидуальном уровне.
- Если есть доказательства того, что выборки группы вмешательства и сравнения были взяты из разных условий, руководство группы экспертов по анализу для предметной области может по своему усмотрению решить, что условия слишком различны, чтобы обеспечить адекватные условия сравнения.

## 4. Представление результатов и отчетность

### 4.1. Стандарты и структура «Отчета о результатах верификации»: требования МГППУ

В соответствии с «Протоколом верификации психолого-педагогических программ и технологий» (см. Глава 2, Этап 3), в «Отчете о результатах верификации» должны быть представлены:

#### Отчет о результатах верификации

1. Актуальность исследования и теоретико-методологическое обоснование программы
  - 1.1. Название и авторы образовательной или социальной программы/технологии. Если название экспериментальной программы/технологии переводится с иностранного языка, приведите название на языке оригинала. Опишите, была ли проведена модификация или адаптация зарубежной методики для условий РФ.
  - 1.2. Ссылка на веб-сайт с информацией о программе/технологии (если есть).
  - 1.3. Теоретико-методологическое обоснование программы/технологии. Кратко сформулируйте, какие теоретические положения и научные труды каких ученых, исследователей, авторов лежат в основе программы/технологии.
  - 1.4. Сформулировать, почему исследование эффективности этой программы актуально.
2. Краткий обзор предшествующих современных исследований и методических изданий (если есть). Приведите, пожалуйста, ссылки на современные публикации (статьи в научных и научно-методических журналах, диссертации, методические пособия и др.) с краткой характеристикой, какие именно положения, методы, понятия и др. важны для вашей Программы (до 5-ти публикаций).
3. Исследовательские вопросы. Например:  
ИВ1: Эффективна ли Программа «Дружный класс» для обучающихся 6–7-х классов в отношении улучшения климата в классе, снижения агрессивности подростков, профилактики буллинга?
4. Научное направление исследования в соответствии с государственным рубрикатором научно-технической информации (ГРНТИ <https://grnti.ru/>) (по третьему уровню иерархии). Например: 15.31.31 «Детская и юношеская психология» или 14.25.05 «Воспитание в общеобразовательной школе».
5. Цель и задачи исследования. Например:  
Цель: Проверить в эксперименте эффективность программы «Дружный класс».

Задачи:

- 1) Оценить климат в классе, уровень агрессивности подростков и риск буллинга с помощью психодиагностических методик и качественных методов на констатирующем этапе.
  - 2) Провести формирующий эксперимент по программе «Дружный класс».
  - 3) Оценить климат в классе, уровень агрессивности подростков и риск буллинга с помощью психодиагностических методик и качественных методов на контрольном этапе.
  - 4) Сравнить показатели до и после формирующего этапа и оценить эффект реализации программы «Дружный класс».
6. Дизайн исследования
- 6.1. Описание выборки: регион(ы) проведения исследования, учреждения — участники исследования, категория испытуемых, общий объем выборки (N).  
Экспериментальная группа (ЭГ): общая характеристика, объем выборки (N1), пол, возраст (среднее, стандартное отклонение, медиана, мода), другие контекстные характеристики, важные для исследования, например, социо-экономический статус, этнокультурная принадлежность, статус беженца, успеваемость в школе и др.  
Контрольная группа (КГ): общая характеристика, объем выборки (N2), пол, возраст (среднее, стандартное отклонение, медиана, мода), другие контекстные характеристики, важные для исследования.
  - 6.2. Психолого-педагогические характеристики (конструкты, параметры, признаки, переменные и др.), в отношении которых программа, предположительно, должна быть эффективна: перечислить, дать четкое определение, что под ними понимается.

Например:

К числу трудностей в учебной мотивации относится невнимательность при принятии и выполнении учебной задачи — *Общая типология трудностей в обучении у обучающихся, имеющих соответствующие риски неблагоприятных социальных условий — научно-исследовательский проект МГППУ «Разработка модели диагностики, профилактики и коррекции трудностей в обучении у обучающихся», 2019.*

Метапредметные результаты освоения программы начального общего образования должны отражать: овладение универсальными учебными познавательными действиями, универсальными учебными коммуникативными действиями, универсальными учебными регулятивными действиями — *Федеральный государственный образовательный стандарт начального общего образования.*

Овладение универсальными учебными познавательными действиями предполагает: 1) базовые логические действия 2) базовые



исследовательские действия 3) работу с информацией — *Федеральный государственный образовательный стандарт начального общего образования*.

Овладение универсальными учебными коммуникативными действиями предполагает: 1) общение 2) совместную деятельность — *Федеральный государственный образовательный стандарт начального общего образования*.

### 6.3. Психодиагностический инструментарий:

- стандартизованные инструменты (тесты, опросники, шкалы, диагностические методики и др.) для измерения исследуемых конструктов. Указать официальные названия инструментов, дать ссылки на статьи с результатами стандартизации в рецензируемых научных журналах. Ориентир — «Открытый реестр психодиагностических методик, вызывающих доверие профессионального сообщества» как составная часть документа «Система функционирования психологических служб в общеобразовательных организациях. Методические рекомендации» [Система функционирования психологических служб, 2020];
- другие инструменты (например, качественные методы) оценки исследуемых признаков (если есть).

Например:

1. **Опросник «Климат в классе»** направлен на выявление субъективного восприятия психологической атмосферы в классе и включает параметры, которые значимы для развития одаренности школьников, развития личности и продуктивности учащихся средних и старших классов школы. «Климат в классе» создан на основе русскоязычной версии опросника «Школьный климат» (Петрова, Щепланова, 2010) и представляет собой его стандартизованную и модифицированную версию (Шумакова, Щепланова, Сорокова, 2023).
2. **Опросник Р. Гудмана «Сильные стороны и трудности»**. Опросник оценки поведения, эмоциональной сферы и взаимоотношений со сверстниками детей и подростков. Предназначен для дошкольников и подростков до 16 лет. и др.

### 6.4 Этапы сбора эмпирических данных:

- 6.4.1. констатирующий этап: реальные сроки, длительность, содержание деятельности;
- 6.4.2. формирующий этап: реальные сроки, длительность, содержание формирующего воздействия, вмешательства (кратко);

- используемые цифровые устройства, приборы, дидактические материалы и др.;
- 6.4.3. контрольный этап: реальные сроки, длительность, содержание деятельности.
- 6.4.4. контроль отдаленных результатов (если предусмотрен): планируемые сроки, длительность, содержание деятельности.
- 6.5. Методы и процедуры сбора данных:
- 6.5.1. Общая характеристика процедур сбора данных
- Тип дизайна эксперимента (РКИ индивидуального уровня, кластерное РКИ, квази-экспериментальный дизайн (КЭД), дизайн с регрессионным разрывом (ДРР), дизайн с одним случаем (кейс-стади, ДОС)) (классификацию типов дизайна см. в **Приложении 1, п. 1**);
  - принцип группировки испытуемых при сборе данных (осуществлялось ли воздействие на испытуемых индивидуально, в составе малой или большой группы, например, школьного класса?). Опишите наименьшую единицу воздействия в формирующем эксперименте (индивид, рабочая группа, класс), результаты который будут анализироваться с точки зрения величины эффекта;
  - метод отбора, если был использован прием систематического отбора испытуемых;
  - общий и дифференциальный отсев выборки (указать, сколько участников эксперимента было на констатирующем этапе, сколько осталось на контрольном этапе, сколько отсеялось) (см. **Приложение 1, п. 2**);
  - условия и место, в которых происходил сбор данных;
  - все виды соглашений, которые заключались с участниками эксперимента, и выплаты, которые производились (если они имели место);
  - соглашения с Наблюдательными Советами учреждений (если они имели место); информация о соблюдении этических норм.
- 6.5.2. Условия проведения формирующего этапа эксперимента (вмешательства)
- количестве воздействий и их длительность (т.е. сколько было экспериментальных сессий, эпизодов и событий, связанных с воздействием на испытуемых и какова была их продолжительность),
  - частота, с которой оказывалось воздействие на испытуемых (например, завершался ли эксперимент после однократной экспериментальной сессии или участники неоднократно подвергались воздействию, и в этом случае,

каким был временной интервал между первой и последней сессией).

6.6. Информация о специалистах и персонале, осуществляющих экспериментальное воздействие (вмешательство)

- проводилось ли повышение квалификации и обучение специалистов для проведения формирующего эксперимента;
- сколько и какие специалисты осуществляли формирующее воздействие;
- конкретные действия и стимулы, с точки зрения их соответствия задачам исследования (т.е. какие функции, какую работу выполняли эти специалисты).

7. Методы анализа данных

- перечислить количественные методы анализа данных, указать использованные стандартные статистические пакеты для анализа данных научного исследования (SPSS, Statistica, R, KNIME Analytics Platform и др.)

Например:

- Методы описательной статистики (вычисление выборочных характеристик, построение графиков, диаграмм и т.д.);
- Критерий Манна — Уитни или Т-критерий Стьюдента для оценки статистической значимости различий между 2-мя независимыми выборками (между экспериментальной и контрольной группами) для количественных шкал. Оценка различий проведена на констатирующем («проверка базовой эквивалентности») и контрольном этапах эксперимента.
- Критерий знаковых ранговых сумм Уилкоксона или Парный критерий Стьюдента для оценки различий показателей «после минус до» отдельно для экспериментальной и контрольной групп и т.д.
- перечислить качественные методы (если применялись) Например: метод наблюдения, метод экспертных оценок и др.

8. Результаты анализа данных для аналитической выборки

*Примечание:* Рекомендованные учебные пособия по математическим методам в психологии и педагогике:

- 1) Сорокова, М.Г. Математические методы в психолого-педагогических исследованиях [Электронный ресурс]: учебное пособие / М.Г. Сорокова. — Москва: Неолит, 2020. — 216 с. — URL: <http://www.psychlib.ru/inc/absid.php?absid=390683> (дата обращения: 01.06.2023) [Сорокова М.Г., 2020a].
- 2) Сорокова, М.Г. Методы математической статистики в психологии [Электронный ресурс]: учебное пособие / М.Г. Сорокова. — Саарбрюкен: Palmarium Academic Publishing, 2014. — 405 с. — URL: [https://psychlib.ru/mgppu/SMm-2014/MMa-405.htm#\\$p1](https://psychlib.ru/mgppu/SMm-2014/MMa-405.htm#$p1) (дата обращения: 01.06.2023) [Сорокова М.Г., 2014].

- 3) Наследов А.Д. Математические методы психологического исследования: анализ и интерпретация данных: учебное пособие / А.Д. Наследов. — Издание 4-е, стереотипное. — Санкт-Петербург: Речь, 2012. — 392 с. См. также издания [Наследов А.Д., 2004], [Наследов А.Д., 2013].
- 4) Информация о вычислении стандартизованного размера эффекта и индекса улучшения представлена в данном учебном пособии в **Приложении 2**.

- 8.1. Описательные статистики для исследуемых переменных (конструктов, характеристик, параметров и т.д.): объемы выборок ЭГ и КГ, среднее ( $M$ ), стандартное отклонение ( $SD$ ), медиана ( $Me$ ), мода ( $Mo$ ), асимметрия ( $A$ ), стандартная ошибка асимметрии ( $\delta A$ ), эксцесс ( $E$ ), стандартная ошибка эксцесса ( $\delta E$ );  
Например:

Таблица 5

**Описательные статистики для экспериментальной ( $N1 =$ ) и контрольной ( $N2 =$ ) групп на констатирующем этапе**

№	Диагностическая методика	$M \pm SD$		$Me$		$Mo$		$A \pm \delta A$		$E \pm \delta E$	
		ЭГ	КГ	ЭГ	КГ	ЭГ	КГ	ЭГ	КГ	ЭГ	КГ
1.	Климат в классе										
	Шкала 1.1										
	Шкала 1.2										
	...										
2.	Тест Р. Гудмана ССТ										
	Шкала 2.1										
	Шкала 2.2										
	...										
3.	и др.										

- 8.2. Графики, гистограммы, диаграммы, иллюстрирующие результаты (если необходимо); все графики должны быть пронумерованы и озаглавлены, оси координат должны быть ясно обозначены;
- 8.3. Статистическая проверка распределений на нормальность (если нужно);
- 8.4. Проверка значимости различий между ЭГ и КГ по исследуемым параметрам на констатирующем этапе, т.е. до вмешательства (проверка базовой эквивалентности) и на контрольном этапе, т.е. после вмешательства; оценка размера эффекта вмешательства в абсолютных и стандартизованных значениях;

Информация о вычислении стандартизованного размера эффекта представлена в **Приложении 2** и в стандартных статистических пакетах.

- 8.5. Проверка значимости сдвигов показателей на контрольном этапе по сравнению с констатирующим отдельно для ЭГ и КГ: «после минус до»; оценка размера эффекта вмешательства в абсолютных и стандартизованных значениях.
- 8.6. Другие методы (если применялись) Выводы исследования:  
 — В отношении каких психолого-педагогических характеристик (переменных, признаков и т.д.) статистически доказана эффективность программы/технологии?

Таблица 6

**Проверка значимости различий между экспериментальной (N1 =) и контрольной (N2 =) группами на констатирующем этапе по критерию Манна—Уитни**

№	Диагностическая методика	M ± SD		Значение статистики U Манна—Уитни	Уровень значимости p
		ЭГ (N1 =)	КГ (N2 =)		
1.	Климат в классе				
	Шкала 1.1				
	Шкала 1.2				
	...				
2.	Тест Р. Гудмана ССТ				
	Шкала 2.1				
	Шкала 2.2				
	...				
3.	и др.				

Примечание: \*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$

Таблица 7

**Проверка значимости различий между экспериментальной (N1 =) и контрольной (N2 =) группами на контрольном этапе по критерию Манна—Уитни**

№	Диагностическая методика	M ± SD		Значение статистики U Манна—Уитни	Уровень значимости p
		ЭГ (N1 =)	КГ (N2 =)		
1.	Климат в классе				
	Шкала 1.1				
	Шкала 1.2				
	...				

№	Диагностическая методика	M ± SD		Значение статистики U Манна—Уитни	Уровень значимости p
		ЭГ (N1 = )	КГ (N2 = )		
2.	Тест Р. Гудмана ССТ				
	Шкала 2.1				
	Шкала 2.2				
	...				
3.	и др.				

Примечание: \*  $p < 0,05$  ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$

Таблица 8

**Проверка значимости различий показателей  
в экспериментальной (N1 = ) группе на контрольном этапе  
по сравнению с констатирующим по критерию Уилкоксона**

№	Диагностическая методика	M ± SD		Сдвиг средних «после минус до»	Значение статистики T Уилкоксона	Уровень значимости p
		До	После			
1.	Климат в классе					
	Шкала 1.1					
	Шкала 1.2					
	...					
2.	Тест Р. Гудмана ССТ					
	Шкала 2.1					
	Шкала 2.2					
	...					
3.	и др.					

Примечание: \*  $p < 0,05$  ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$

Таблица 9

**Проверка значимости различий показателей  
в контрольной (N2 = ) группе на контрольном этапе  
по сравнению с констатирующим по критерию Уилкоксона**

№	Диагностическая методика	M ± SD		Сдвиг средних «после минус до»	Значение статистики T Уилкоксона	Уровень значимости p
		До	После			
1.	Климат в классе					
	Шкала 1.1					
	Шкала 1.2					
	...					
2.	Тест Р. Гудмана ССТ					
	Шкала 2.1					
	Шкала 2.2					

№	Диагностическая методика	M ± SD		Сдвиг средних «после минус до»	Значение статистики Т Уилкоксона	Уровень значимости p
		До	После			
	...					
3.	и др.					

Примечание: \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$

- Каков размер эффекта?
  - В отношении каких исследованных психолого-педагогических характеристик эффект не выявлен?
  - Каковы ограничения результатов?
9. Ограничения применения программы и предложения по ее дальнейшему применению и развитию.

#### 4.2. Стандарты и требования WWC к представлению и оценке результатов верификации

Согласно стандартам и рекомендациям WWC [WWC Procedures and Standards Handbook, V. 3.0, С. 16–19], чтобы иметь право на рецензирование, результат должен (а) продемонстрировать валидность (face validity) и надежность (reliability), (б) не быть чрезмерно согласованным с вмешательством, и (в) быть собранным одинаковым образом для обеих групп вмешательства и сравнения. Предполагается, что **стандартизированные тесты**, в которых один и тот же тест проводится одинаковым способом для всех участников тестирования, **соответствуют этим критериям**, если они имеют отношение к теме.

Чтобы продемонстрировать доказательство **валидности** (face validity), WWC должно быть предоставлено достаточное описание способа измерения (шкалы измерения) результата, чтобы определить, что мера четко определена, имеет точную интерпретацию и измеряет именно тот конструкт, для измерения которого она предназначена. Например, количество произнесенных слов в течение определенного периода времени обладает валидностью для измерения беглости чтения, а процент учащихся, окончивших среднюю школу, мог бы быть результатом, имеющим валидность в качестве рейтинга окончания учебного заведения.

**Надежность** показателя результата может быть установлена путем соблюдения следующих минимальных стандартов: (а) внутренняя согласованность (такая как альфа Кронбаха) 0,50 или выше; (б) временная стабильность / ретестовая надежность 0,40 или выше; или (с) «межсудейская» надежность, то есть согласованность мнений экспертов (например, процентное согласие, корреляция или каппа Коэна) составляет 0,50 или выше. Протокол для рецензирования может указывать более высокие



стандарты для оценки надежности и/или может оговаривать, как обращаться с результатами, связанными с достижениями, которые вряд ли предоставят информацию о надежности. Примеры результатов, которые могут иметь только внешнюю валидность, включают школьные отметки, средние баллы, зачетные баллы или простые математические задачи для маленьких детей. Руководство группы рецензентов определяет, являются ли эти результаты приемлемыми, должны ли они быть подтверждены экспертом по контенту или не имеют права на проверку. Как правило, **WWC не рассматривает отметки или средние баллы как имеющие право на рецензирование**, поскольку критерии могут отличаться в зависимости от учителей, школ или районов.

Второе требование к шкалам измерения результата состоит в том, что они не должны быть *чрезмерно согласованы* с вмешательством. Когда показатели результата тесно связаны с вмешательством или адаптированы к нему (т.е. выполнены «на заказ», сделаны специально под него), результаты исследования могут не являться точным показателем эффекта вмешательства. Например, шкала измерения результата, базирующаяся на оценке, которая основывалась на материалах, использованных в условиях вмешательства, но не в условиях сравнения (например, конкретные отрывки чтения), скорее всего, будет оценена как чрезмерно согласованная. Решение о том, является ли показатель чрезмерным, принимается руководством группы рецензентов по тематической области.

Третье требование к показателям результатов состоит в том, что они должны быть *собраны одинаковым способом* для групп вмешательства и сравнения. WWC предполагает, что данные были собраны таким же образом, если информация не предоставлена. Тем не менее, рецензенты ищут комментарии в исследованиях о том, что (а) для групп использовались разные режимы, сроки или персонал, или (б) измерительные шкалы (инструменты для измерений) были построены по-разному для групп. Команды рецензентов могут отправлять вопросы авторам, чтобы уточнить, как были собраны данные. Когда данные результатов собираются по-разному для групп вмешательства и сравнения, оценки воздействия, представленные в исследовании, будут **смешивать различия из-за вмешательства с различиями из-за различных методов сбора данных**. Например, измерение показателей отсева на основе записей программы, реализуемой в образовательном учреждении, для группы вмешательства и административных записей школы для группы сравнения приведет к ненадежным оценкам воздействия, поскольку невозможно будет отделить истинное влияние вмешательства от различий в показателях отсева ввиду исключения из школы, которые получают из-за конкретного используемого критерия.

В исследованиях часто сообщается о полученных выводах для нескольких результатов, включая один и тот же результат, измеренный в разные моменты времени, альтернативных показателях одного и того же

конструкта или как о показателях на уровне элемента (пункта), так и о составных показателях. WWC установил следующие **руководящие принципы** для определения того, о каких результатах следует сообщать:

- ***Результаты, измеренные в разные моменты времени.***

Когда в исследовании сообщается как о немедленных, так и долгосрочных показателях результата, WWC выбирает один показатель (одну измерительную шкалу) в качестве основного вывода, который будет вносить вклад в оценку вмешательства; выводы для других результатов будут включены в дополнительные таблицы. Предпочтение определяется руководством группы рецензентов и описывается в протоколе проверки (рецензии).

- ***Выводы для всей группы и для подгруппы.***

Когда исследование представляет выводы отдельно для нескольких групп обучающихся без представления совокупного результата, WWC будет запрашивать авторов, чтобы выяснить, проводили ли они анализ всей выборки учащихся. Если WWC не может получить совокупные результаты от автора, он усредняет значения по подгруппам в рамках исследования, чтобы использовать их в качестве основного вывода, и представляет результаты подгруппы в качестве дополнительных таблиц.

- ***Меры на уровне отдельных пунктов и комплексные меры.***

Когда в исследовании сообщается о комплексных тестовых мерах (итоговых шкалах теста) и их компонентах, WWC считает, что комплексная мера является основным выводом, который вносит вклад в оценку вмешательства. Результаты субтестов компонентов или результаты на уровне отдельных пунктов включаются в дополнительные таблицы.

- ***Категориальные порядковые шкалы измерений.***

Для некоторых категориальных порядковых результатов WWC может «сворачивать» (объединять) категории, чтобы создать сопоставимые величины эффекта в разных исследованиях. Например, тест с пятью уровнями оценки с помощью объединения категорий может быть разбит на две категории «искусный (обученный, квалифицированный)» и «неквалифицированный», чтобы можно было сравнить его с другими показателями, которые сообщают только о двух возможных результатах.

- ***Фактические vs введенные (добавленные, imputed) измерения.***

Если определено, что рандомизированное контролируемое исследование имеет низкую степень отсева выборки, результаты анализа с приемлемыми методами учета недостающих (пропущенных) данных можно использовать при составлении отчетов о результатах исследования (они не влияют на рейтинг). Методы, перечисленные ниже, если они реализованы, как описано, являются приемлемыми для генерирования значений  $r$  или стандартных ошибок, о которых WWC может сообщить. Исследование может также использовать эти методы для расчета отсутствующих значений для ковариат или независимых переменных, но **добавления в базовые переменные** (переменные, для которых проверяется базовая

эквивалентность) **нельзя использовать для демонстрации базовой эквивалентности** (эквивалентности начальных условий).

- ***Полный анализ случая без корректировки методом регрессии.***

Самый простой подход к обработке пропущенных данных о результатах — **отбросить наблюдения с пропущенными результатами** анализа. Если очевидно, что в исследовании использовался этот подход, дополнительная информация не требуется для использования полученных в исследовании значений  $p$  и стандартных ошибок.

- ***Полный анализ случая с корректировкой методом регрессии для базовых ковариат (baseline covariates).***

Один из подходов к учету различий перед вмешательством между группами вмешательства и сравнения, которые могут возникнуть из-за отсева, заключается в проведении **статистической корректировки** для различий перед вмешательством (например, посредством регрессии или ANCOVA). Если очевидно, что в исследовании использовался этот подход, дополнительная информация не требуется для использования значений  $p$  исследования и стандартных ошибок.

- ***Метод максимального правдоподобия отдельно в зависимости от статуса вмешательства.***

Многие статистические пакеты используют методы максимального правдоподобия для учета отсутствующих данных. Это приемлемо, если ясно, что либо использовался стандартный статистический пакет (должно быть указано название пакета и процедуры или функции), либо цитируется рецензируемая методологическая статья в журнале или учебник. В противном случае, WWC запрашивает у автора информацию, чтобы определить, соответствует ли конкретный используемый метод максимального правдоподобия вышеуказанным условиям.

- ***Множественное добавление (внесение, введение).***

Множественное добавление включает в себя создание нескольких наборов данных, которые содержат добавленные значения для пропущенных данных о результатах (переменной, отражающей результат), которые **генерируются** посредством многократного применения алгоритма добавления (такого как добавление с помощью цепочки уравнений). Все подходы множественного добавления приемлемы, если (а) добавление проводится отдельно для групп вмешательства и сравнения и (б) либо использовался стандартный статистический пакет, либо приводилась цитата к рецензируемой методической журнальной статье или учебнике. Переменные, используемые в модели добавления, должны включать, по крайней мере, все ковариаты, которые использовались для статистической корректировки в оценке воздействия. Чтобы WWC использовал стандартные ошибки и  $p$ -значения, число добавлений должно быть больше единицы, а затем число добавлений должно учитываться при генерации общих стандартных ошибок и  $p$ -значений.

## 5. Смешивающие факторы

WWC вводит понятие «смешивающих факторов» (искажающих факторов, Confounding Factors) [WWC Procedures and Standards Handbook, V. 3.0, С. 19]. Учет смешивающих факторов может быть весьма полезен при оценке результатов верификации образовательных и социальных программ и технологий в Российской Федерации. В некоторых исследованиях компонент плана (дизайна) исследования или обстоятельства, при которых было реализовано вмешательство, полностью совпадают или смешиваются (перепутываются) либо с группой вмешательства, либо с группой сравнения. Другими словами, какой-то фактор присутствует у членов только одной группы и отсутствует у всех членов другой группы. В этих случаях невозможно сказать, **является ли само вмешательство или этот искажающий фактор причиной разницы в результатах**. Смешивающие факторы могут присутствовать в рандомизированных контролируемых исследованиях и квази-экспериментальных исследованиях.

Наиболее распространенный тип смешения возникает, когда **группа вмешательства или сравнения содержит одну учебную единицу**, например, когда всех учащихся, принимающих участие, обучает один учитель, все классы сравнения относятся к одной школе или когда все школы группы вмешательства из одного школьного округа. В этих ситуациях невозможно провести различие между эффектом вмешательства и этой единицей. Например, если все учащиеся, участвующие в программе по математике, квалифицируемой как вмешательство, обучаются у одного учителя, а не участвующие в этой программе — у других учителей, то любые последующие различия между результатами учащихся, использующих математическое вмешательство, и тех, кто этого не делает, **могут быть связаны с вмешательством, учителем или и тем, и другим**.

Другой пример смешения возникает, когда **характеристики единиц в каждой группе систематически различаются способами, которые связаны с результатами**. Например, небольшая группа преподавателей, имеющих степень магистра, осуществляет вмешательство, тогда как учеников в группе сравнения обучают преподаватели со степенью бакалавра. Если образование учителя не является компонентом вмешательства, то есть план эксперимента не указывает, что только учителя уровня магистра могут руководить вмешательством, то это потенциальный смешивающий фактор. В этом случае различия в результатах учащихся между группами вмешательства и группами сравнения **могут быть связаны с вмешательством, более высоким уровнем образования учителей группы вмешательства или их комбинацией**.

Смешение также происходит, **если вмешательство всегда предлагается в сочетании со вторым вмешательством**, потому что любые

последующие различия в результатах не могут быть отнесены исключительно к какому-либо вмешательству. Тем не менее, WWC может **рассматривать комбинацию как единое вмешательство** и сообщать о ее эффектах. Кроме того, если информация о экспериментальной группе поступает за один учебный год, тогда как информация о группе сравнения поступает за другой учебный год, то **время может рассматриваться как смешивающий фактор**.

В каждом приведенном выше примере смешивающий фактор может влиять на результат отдельно от вмешательства, которое не может быть устранено дизайном исследования. Поскольку невозможно разделить степень, в которой наблюдаемый эффект был обусловлен вмешательством, и степень, в которой он был обусловлен искажающим фактором, исследование со смешивающим фактором не может соответствовать стандартам WWC. В исследованиях с квази-экспериментальным дизайном смешение почти всегда является потенциальной проблемой из-за отбора выборки, поскольку на результат могли повлиять некоторые ненаблюдаемые факторы. WWC объясняет эту проблему тем, что **не позволяет квази-экспериментальным исследованиям получать наивысший рейтинг доказательности**.

Рецензенты WWC должны решить, достаточно ли информации, чтобы определить, что единственное различие между двумя группами, которое не контролируется планом или анализом, заключается в наличии вмешательства. В противном случае может иметь место искажающий фактор, и рецензент должен определить, может ли этот фактор повлиять на результат отдельно от вмешательства.

## **6. Выводы и оценка результатов верификации**

### **6.1. Выводы экспертизы и оценка результатов эксперимента по верификации: требования МГППУ**

Результаты эксперимента по верификации психолого-педагогических, образовательных и социальных программ и технологий оцениваются с помощью рейтинговой оценки. Эксперты заполняют **«Протокол экспертной оценки результатов верификации»** (см. Приложение 2 п. 1.3) на основе «Отчета о результатах верификации» или приравненного к нему документа (например, Автореферата диссертации на соискание ученой степени кандидата или доктора наук, Отчета по научно-исследовательскому проекту и др.). Критерии оценки соответствуют положениям доказательного подхода и общей структуре «Отчета о результатах верификации». В графе «Оценка» по каждому критерию ставится оценка: 0 — не соответствует, 1 — частично соответствует, 2 — полностью соответствует.

Итоговая рейтинговая оценка является суммой баллов по всем критериям (диапазон от 0 до 42 баллов), переведенной в процентную шкалу. Она выставляется в таблице в графе **Итого по протоколу**. В графе **Заключение экспертизы (нормативная оценка)** экспертами указывается одна из 4-х категорий соответствия стандартам программ и технологий с доказанной эффективностью: «Соответствует стандартам программ и технологий с доказанной эффективностью для сферы образования и социальной сферы», «Соответствует стандартам программ и технологий с доказанной эффективностью для сферы образования и социальной сферы **с учетом дополнительных рекомендаций**», «Соответствует стандартам программ и технологий с доказанной эффективностью для сферы образования и социальной сферы **на этапе пилотных проектов**» или «Не соответствует стандартам программ и технологий с доказанной эффективностью для сферы образования и социальной сферы». Кроме того, группа экспертов пишет краткий комментарий с замечаниями и, возможно, рекомендациями по развитию и совершенствованию программы/технологии, методики ее применения, по дополнительному обучению персонала, проведению масштабированного эксперимента по верификации и др.

Окончательные выводы экспертизы по «Протоколу экспертной оценки результатов верификации ...» эксперты согласовывают между собой и представляют на сессии секции «Доказательный подход в психологии и образовании» Научно-экспертного совета МГППУ (секция ДП НЭС) как результат консенсуса. После представления программы или технологии и обсуждения результатов ее экспертизы на сессии секции ДП НЭС авторам и разработчикам программы предоставляется **Экспертное заключение секции ДП НЭС**. В Экспертном заключении указываются рейтинговые оценки (проценты от максимальных баллов), набранные по «Протоколу качественного ассесмента ...» и «Протоколу экспертной оценки результатов верификации ...», Заключение экспертизы (нормативная оценка), а также комментарии и рекомендации экспертов. Экспертное заключение секции ДП НЭС **подписывают эксперты, руководитель и секретарь секции ДП НЭС**.

## 6.2. Завершение обзора: требования WWC

После того как исследование будет рассмотрено, чтобы определить, подходящий ли дизайн; существует ли по крайней мере одна релевантная, валидная и надежная мера результата; и есть ли какие-либо искажающие факторы, исследование получает одну из трех оценок: соответствует стандартам группового дизайна WWC без оговорок, соответствует стандартам группового дизайна WWC с оговорками или не соответствует стандартам группового дизайна WWC. Эти рейтинги относятся к степени уверенности, которую WWC придает способности исследования дать непредвзятую

оценку причинно-следственной связи между вмешательством и интересующими результатами. Исследования, которые не соответствуют стандартам, получают краткое описание как минимум одной причины, по которой исследование не соответствует стандартам WWC:

- **Качество дизайна.** Исследование представляет собой рандомизированное контролируемое испытание с высоким отсевом или квази-экспериментальное исследование с аналитическими группами, не являющимися эквивалентными.
- **Результаты и отчетность.** Недостаточно информации, чтобы определить, были ли показатели результата валидными или надежными, результаты чрезмерно совпадают с вмешательством или результаты измерялись по-разному для групп вмешательства и групп сравнения.
- **Смешивающий фактор.** По крайней мере одному из состояний была назначена только одна единица, или вмешательство всегда использовалось в сочетании с другим вмешательством.

Для каждого исследования, которое соответствует стандартам WWC с оговорками или без них, **WWC записывает информацию о вмешательстве и условиях сравнения в той мере, в какой они указаны в исследовании.** Например, группа сравнения также может получить вмешательство, такое как другая учебная программа; обычное предложение; или без воздействия. Воздействие вмешательства всегда связано с конкретным сравнением или описанием того, что не произошло, хотя и ожидалось (counterfactual), и выводы из результатов исследования должны учитывать контекст. WWC также документирует информацию об исследуемой выборке (включая учащихся, классы, учителей и школы); обстановку исследования; подходящие результаты, включенные в исследование, и как они были измерены; и детали, такие как обучение учителей или сотрудников, которые осуществили вмешательство. Это **важный контекст для интерпретации результатов исследования.**

Хотя WWC документирует, как было реализовано вмешательство и контекст, в котором оно было реализовано для выборки исследования, он не делает статистических корректировок или поправок на различия в реализации вмешательства (например, по сравнению с идеальным или средним внедрением). В исследованиях вмешательств следует ожидать изменений в реализации, потому что они происходят в реальных условиях, таких как классы и школы, и не обязательно в строго контролируемых условиях, за которыми наблюдают исследователи. Аналогично, WWC также **не вносит статистических поправок на неучастие** (т.е. членов группы вмешательства, получивших возможность участвовать в программе, но отказавшихся от этого) **или возможное загрязнение** (т.е. членов группы сравнения, которые получают вмешательство). Руководство группы экспертов по анализу имеет право по своему усмотрению определять, являются ли эти вопросы достаточно существенными, чтобы повлиять на рейтинг исследования, или считать их выходящими за рамки протокола обзора.



## Отчет о результатах верификации и размере эффекта

### 1. Требования и рекомендации МГППУ по оценке размера эффекта и статистической значимости результатов верификации программ и технологий

#### 1.1. Методы количественного анализа данных, рекомендуемые для оценки результатов верификации

Количественный анализ данных эксперимента по верификации психолого-педагогических, образовательных и социальных программ и технологий выполняется с использованием релевантных методов математической статистики, как параметрических, так и непараметрических. Применение параметрических методов требует проверки согласия распределений исследуемых переменных с нормальным распределением по критериям Колмогорова — Смирнова или Шапиро — Уилка. В случае измерения многих параметров для обобщения данных и оценки результата воздействия могут использоваться методы многомерной статистики. Если исследователей помимо проверки различий интересует оценка связи признаков, используются методы корреляционного анализа. Количественный анализ данных выполняется в стандартных статистических пакетах, таких как SPSS, Statistica, KNIME Analytics Platform, RStudio, Excel и др. Качественные методы исследования могут служить дополнением к количественным, а их анализ выполняется в соответствии с инструкциями к конкретным методам.

Выбор методов анализа определяется исследовательскими вопросами, типами измерительных шкал и количеством исследуемых выборок. **Измерительные шкалы** бывают номинативными, порядковыми и количественными [Сорокова М.Г., 2020а], [Сорокова М.Г., 2014]. Номинативные измерения представляют собой классификацию и распределяют испытуемых на категории. Порядковые шкалы позволяют выявить различные степени исследуемого свойства, а количественные — оценить, на сколько больше выражен признак у одного испытуемого по сравнению с другим в условных единицах — баллах, пунктах, физических единицах измерения и т.д. Самыми информативными являются количественные шкалы, однако все типы шкал позволяют проводить количественный анализ данных и выявлять статистически значимые различия. Для каждого типа шкал



разработан и активно применяется комплекс соответствующих статистических критериев и методов многомерной статистики.

**В классической схеме психолого-педагогического эксперимента**, т.е. при наличии данных замеров исследуемой переменной на констатирующем (до вмешательства) и контрольном (после вмешательства) этапах в экспериментальной и контрольной группах (ЭГ и КГ) при количественных измерениях чаще всего используются следующие методы. Сравнение ЭГ и КГ осуществляется дважды — до и после воздействия — по критериям Фишера (Ливена) проверки гипотезы о равенстве генеральных дисперсий и Стьюдента для 2-х независимых выборок для проверки гипотезы о равенстве генеральных средних. Можно использовать также непараметрический критерий Манна — Уитни. Ожидается, что на констатирующем этапе статистически значимых различий между ЭГ и КГ не будет, а на контрольном этапе они появятся. Кроме того, статистически оценивается достоверность сдвига значений исследуемой переменной до и после воздействия («после минус до») при помощи Парного критерия Стьюдента или непараметрического критерия Уилкоксона отдельно в ЭГ и в КГ. Ожидается, что в ЭГ различия будут статистически значимы, а КГ — в зависимости от наличия или отсутствия других воздействий. Однако в реальных исследованиях ожидания могут не подтвердиться как на одном, так и на обоих этапах, и тогда интерпретация результата и выводы об эффективности или неэффективности воздействия требуют более тщательного обдумывания и обсуждения. Пример анализа данных по классической схеме см. в [Марголис А. и др., 2023], [Сорокова М.Г., 2020с].

Если анализируется **несколько экспериментальных и/или контрольных групп**, для оценки различий на констатирующем и контрольном этапах применяется однофакторный дисперсионный анализ (1-way ANOVA), см. **Глава 2, п. 4.2**, или его непараметрического аналога — критерия Краскела — Уоллиса. Если помимо влияния на исследуемую переменную одного фактора — формирующего воздействия — необходимо оценить совокупное влияние 2-х факторов, — например, формирующего воздействия и пола, — применяется двухфакторный дисперсионный анализ (2-way ANOVA). В подобных задачах может быть также использован ковариационный анализ (ANCOVA). Например, в исследовании [Сорокова М.Г., Одинцова М.А., Радчикова Н.П., 2021а] проверка различий результатов итогового теста и итоговой оценки по ЭУК проводилась с помощью двухфакторного ковариационного анализа (ANCOVA) 2x2, где в качестве ковариаты рассматривались результаты входного теста. Ковариата использована для удаления посторонней вариации из зависимой переменной и выявления эффектов факторов. Факторами стали формат обучения (смешанное vs онлайн) и уровень высшего образования (магистратура и второе высшее vs первое высшее).

Среди методов **многомерной статистики**, наиболее часто используемых в подобных исследованиях, отметим факторный анализ (ФА), кластерный анализ (КА) и множественный регрессионный анализ (МРА). ФА и МРА предназначены для анализа количественных данных и являются параметрическими. КА имеет несколько разновидностей и может быть использован как для количественных шкал (кластерный анализ  $k$ -средними), так и для категориальных измерений (иерархический кластерный анализ). ФА основан на анализе структуры корреляционной матрицы, построенной из коэффициентов корреляции Пирсона, и применяется для снижения размерности исходного пространства и обобщения данных, чтобы обеспечить возможность сравнения результатов ЭГ и КГ по обобщенным параметрам — факторам. МРА используется в моделях прогнозирования с целью предсказания значений зависимой переменной (в том числе, отсутствующих) по совокупности значений независимых переменных.

## 1.2. Примерный перечень показателей результатов анализа данных

Результаты анализа данных для аналитической выборки:

- описательные статистики для исследуемых переменных (конструктов, характеристик, параметров и т.д.): объемы выборок ЭГ и КГ, среднее, стандартное отклонение, медиана, мода, асимметрия, эксцесс;
- графики, гистограммы, диаграммы, иллюстрирующие результаты (если необходимо);
- статистическая проверка распределений на нормальность;
- проверка значимости различий между ЭГ и КГ по исследуемым параметрам на констатирующем этапе, т.е. до вмешательства (проверка базовой эквивалентности) и на контрольном этапе, т.е. после вмешательства; оценка размера эффекта вмешательства в абсолютных и стандартизированных значениях (например,  $d$  Коэна,  $g$  Хеджеса,  $\Delta$  Гласса и др.);
- проверка значимости сдвигов показателей на контрольном этапе по сравнению с констатирующим отдельно для ЭГ и КГ: «после минус до»; оценка размера эффекта вмешательства в абсолютных и стандартизированных значениях.

Информация о вычислении стандартизованного размера эффекта и индекса улучшения представлена в **Приложении 2**. Пример расчета стандартизованного размера эффекта и индекса улучшения, в том числе, с поправкой на кластеризацию, можно найти в статье [Сорокова М.Г., 2021]. В стандартных статистических пакетах стандартизованный размер эффекта ( $d$  Коэна,  $g$  Хеджеса,  $\Delta$  Гласса и др.) вычисляется автоматически.

### 1.3. Протокол оценки результатов верификации: критерии экспертизы

Настоящий «Протокол оценки результатов верификации» (Табл. 10) заполняется экспертами на основе «Отчета о результатах верификации». Критерии оценки соответствуют положениям доказательного подхода (см. **Введение**) и общей структуре «Отчета о результатах верификации» (см. **Приложение 1, п. 4**).

#### **Комментарий для экспертов:**

Настоящий «Протокол экспертной оценки результатов верификации» заполняется экспертами на основе «Отчета о результатах верификации» или приравненного к нему документа (например, Автореферата диссертации на соискание ученой степени кандидата или доктора наук, Отчета по научно-исследовательскому проекту и др.). Критерии оценки соответствуют положениям доказательного подхода и общей структуре «Отчета о результатах верификации». В графе «Оценка» по каждому критерию ставится оценка: 0 — не соответствует, 1 — частично соответствует, 2 — полностью соответствует.

Итоговая рейтинговая оценка является суммой баллов по всем критериям (диапазон от 0 до 42 баллов), переведенной в процентную шкалу. Она выставляется в таблице в графе **Итого по протоколу**. В графе **Заключение экспертизы (нормативная оценка)** экспертами указывается одна из 4-х категорий соответствия стандартам программ и технологий с доказанной эффективностью. Кроме того, группа экспертов пишет краткий комментарий с замечаниями и, возможно, рекомендациями по развитию программы/технологии.

Окончательные выводы экспертизы по «Протоколу экспертной оценки результатов верификации ...» эксперты согласовывают между собой, оформляют в виде коллегиального Экспертного заключения и представляют на сессии секции «Доказательный подход в психологии и образовании» Научно-экспертного совета МГППУ (секция ДП НЭС) как результат консенсуса.

После представления программы или технологии и обсуждения результатов ее экспертизы на сессии секции ДП НЭС авторам и разработчикам программы предоставляется **Экспертное заключение секции ДП НЭС**. В Экспертном заключении указываются рейтинговые оценки (проценты от максимальных баллов), набранные по «Протоколу качественного ассесмента ...» и «Протоколу экспертной оценки результатов верификации ...», Заключение экспертизы (нормативная оценка), а также комментарии и рекомендации экспертов. Экспертное заключение секции ДП НЭС подписывают эксперты, руководитель и секретарь секции ДП НЭС.

Таблица 10

**Протокол экспертной оценки результатов верификации образовательной или социальной программы/технологии**

<b>Протокол экспертной оценки результатов верификации образовательной или социальной программы/технологии: название программы</b>		
<b>№</b>	<b>Критерий оценки</b>	<b>Оценка</b>
1.	Обоснована актуальность программы/технологии	
2.	Представлено методологическое обоснование программы/технологии	
3.	Перечислены психолого-педагогические характеристики (конструкты, параметры, признаки, переменные), в отношении которых программа, предположительно, должна быть эффективна	
4.	Использованы стандартизованные психодиагностические инструменты (методики), релевантные исследуемому конструктам	
5.	Эксперимент по верификации соответствует классической схеме: констатирующий — формирующий — контрольный этап	
6.	В эксперименте по верификации присутствуют экспериментальная(ые) и контрольная(ые) группы (группы вмешательства и группы сравнения)	
7.	Данные собраны не менее чем в 3-х регионах РФ или других стран	
8.	Использованы корректные процедуры сбора эмпирических данных	
9.	Общий и дифференциальный отсеv выборки не превышает порогового значения в соответствии с рекомендациями портала WWC (см. Приложение 1, п. 2)	
10.	Количественный анализ данных выполнен в стандартных статистических пакетах (SPSS, Statistica, R, KNIME Analytics Platform и др.)	
11.	Статистически подтверждена базовая эквивалентность ЭГ и КГ для аналитической выборки	
12.	Соответствие выбора и применения статистических методов схеме исследования и измеряемым переменным. Портал психологических изданий PsyJournals.ru — <a href="https://psyjournals.ru/info/homestyle_guide/article_requirements.shtml#article_requirements_4">https://psyjournals.ru/info/homestyle_guide/article_requirements.shtml#article_requirements_4</a> [III. Требования к содержанию статей — PsyJournals.ru]	
13.	Соответствие объема выборки поставленным задачам и применяемым статистическим методам	
14.	Корректность описания математического аппарата в разделе «Методы анализа данных»	
15.	Корректность описания результатов и величин эффекта в разделе «Результаты анализа данных для аналитической выборки»	
16.	Соответствие выводов полученным результатам количественного анализа данных	
17.	Проведен контроль отдаленных результатов, выявлен устойчивый эффект	

<b>Протокол экспертной оценки результатов верификации образовательной или социальной программы/технологии: название программы</b>		
№	Критерий оценки	Оценка
18.	Качественные методы исследования применены, а их результаты интерпретированы корректно (если они применялись)	
19.	Смешивающие (искажающие) факторы не оказывают существенного влияния на результат эксперимента отдельно от формирующего воздействия (вмешательства)	
20.	Корректно указаны ограничения результатов	
21.	Представлены предложения по дальнейшему развитию программы/технологии	
<b>Итого по протоколу</b>		<b>Балл, % от 42</b>
<p><b>Заключение группы экспертов:</b></p> <p><b>1. Заключение экспертизы (нормативная оценка):</b>                      «Соответствует стандартам программ и технологий с доказанной эффективностью для сферы образования и социальной сферы»,                      или                      «Соответствует стандартам программ и технологий с доказанной эффективностью для сферы образования и социальной сферы с учетом дополнительных рекомендаций»,                      или                      «Соответствует стандартам программ и технологий с доказанной эффективностью для сферы образования и социальной сферы на этапе пилотных проектов»,                      или                      «<b>Не соответствует</b> стандартам программ и технологий с доказанной эффективностью для сферы образования и социальной сферы»</p> <p><b>2. Комментарии и замечания экспертов, рекомендации</b> по развитию и продвижению программы/технологии, по размещению ее на цифровой платформе — ГИС Министерства Просвещения РФ</p>		

## 2. Рекомендации WWC по оценке размера эффекта вмешательства

По мере возможности, WWC [WWC Procedures Handbook, Version 4.1, С. 14 и далее] сообщает о величине и статистической значимости представленных в исследовании оценок эффективности вмешательств с использованием общих метрик и корректировок, таких как **поправки для кластеризации и множественных сравнений**, которые могут повлиять на результаты, представленные в исследовании. Затем применяется эвристика для характеристики результатов исследования таким образом, чтобы учитывать направление, величину и статистическую точность оценок воздействия. Наконец, в некоторых своих продуктах, включая систематические обзоры отчетов

о вмешательстве и практические руководства, WWC объединяет результаты отдельных исследований в сводные показатели эффективности, в том числе совокупные численные оценки величины воздействий, общие рейтинги эффективности и рейтинг для степени доказательности.

## 2.1. Результат индивидуального анализа

WWC определяет индивидуальный результат как измеренный эффект вмешательства относительно определенного условия сравнения на результат для выборки в определенный момент времени.

### 2.1.1 Величина результата

В целом, WWC сообщает о величине результатов исследования двумя способами: *размерами эффекта*, — то есть стандартизированными различиями в средних, — для непрерывных переменных (признаков) или наборов переменных, содержащих как непрерывные, так и дихотомические переменные, и через результаты воздействия, выраженные *с помощью процентных точек* для строго дихотомических переменных. WWC также иногда использует *индекс Кокса* (Cox index), чтобы преобразовать влияния на дихотомические признаки в величины эффектов и рассчитать *t*-статистики *для оценки статистической значимости* влияний на дихотомические признаки, чтобы можно было лучше сравнить их с влияниями на непрерывные признаки.

Кроме того, WWC может сообщать о величине воздействий, используя другие метрики, такие как  $U_3$  *Козна* — мера процента испытуемых группы вмешательства с результатами выше среднего по группе сравнения — или  $U_3$  *Козна минус 50 процентных пунктов*, который называется *«индекс улучшения»*. Более подробная информация для всех этих расчетов приводится ниже.

### 2.1.2. Размеры эффекта для исследований с групповым дизайном

Для всех исследований WWC фиксирует результаты исследований в единицах, о которых сообщили авторы исследования. Кроме того, для непрерывных переменных или дихотомических переменных, изучаемых в совокупности с непрерывными переменными, WWC вычисляет и записывает **размер (величину) эффекта**, связанного с результатами исследования и выраженного с помощью соответствующих показателей. В целом, чтобы улучшить сопоставимость оценок размера эффекта в разных исследованиях, WWC использует стандартные отклонения «на уровне учащегося» (т.е. без поправки на кластеризацию) при расчете величины эффекта независимо от единицы воздействия (т.е. индивидуума, группы, кластера и др.) или единицы вмешательства. Заметим, что в современных статистических пакетах SPSS, Jamovi и др. есть опции автоматического расчета стандартизованных размеров эффекта, таких как *d* Козна, *g* Хеджеса,  $\Delta$  Гласса и др.

Для непрерывных результатов WWC принял наиболее часто используемый **индекс размера эффекта**, стандартизированная разность в средних, известная как ***g* Хеджеса** (Hedges'), с поправкой на смещение для небольшой выборки. Для исследований по групповому дизайну этот размер эффекта определяется как разность между средним по группе вмешательства и средним по группе сравнения, деленная на объединенное внутригрупповое стандартное отклонение для показателя результата. Показатель ***g* Хеджеса** вычисляется по **формуле (1)**:

$$g = \frac{\omega(y_i - y_c)}{\sqrt{\frac{(n_i - 1)s_i^2 + (n_c - 1)s_c^2}{n_i + n_c - 2}}}. \quad (1)$$

где  $y_i$  и  $y_c$  — это выборочные средние для групп вмешательства и сравнения соответственно,  $n_i$  и  $n_c$  — объемы выборок (количество испытуемых в группах) вмешательства и сравнения,  $s_i$  и  $s_c$  — выборочные стандартные отклонения для групп вмешательства и сравнения.

WWC использует нескорректированные стандартные отклонения  $s_i$  и  $s_c$  «на уровне учащихся» в формуле (1), т.е. без учета объединения учащихся в классы, группы, кластеры и т.д. Когда нескорректированные

стандартные отклонения «на уровне учащихся» недоступны (например, когда сообщается о скорректированных стандартных отклонениях), WWC отправляет запрос автору предоставить нескорректированные стандартные отклонения. WWC отправляет запрос автору предоставить нескорректированные стандартные отклонения.

В формуле (1) используется поправка  $\omega$  на небольшую выборку, корректирующая размер эффекта, которая вычисляется по **формуле (2)**:

$$\omega = [1 - 3/(4N - 9)], \quad (2)$$

где  $N = n_i + n_c$ , где  $n_i$  и  $n_c$  — объемы выборок (количество испытуемых в группах) вмешательства и сравнения.

**Для дихотомических переменных** разность в групповых средних рассчитывается как разность в вероятности возникновения события. Показателем размера эффекта для дихотомических результатов является **индекс Кокса**, который дает значения размера эффекта, аналогичные значениям *g* Хеджеса, которые можно было бы получить, если бы были доступны средние значения для групп, стандартные отклонения и размеры выборки, предполагая, что мера дихотомической переменной основана на базовом логистическом распределении, подобном нормальному. Определяя  $p_i$  и  $p_c$  как вероятность события для учащихся в группах вмешательства и сравнения, **размер эффекта** определяется по **формуле (3)**

$$d_{Cox} = \left[ \ln\left(\frac{p_i}{1-p_i}\right) - \ln\left(\frac{p_c}{1-p_c}\right) \right] / 1.65. \quad (3)$$



Рекомендации WWC при расчете размера эффекта для непрерывных переменных:

- Если в исследовании сообщаются как нескорректированные, так и скорректированные средние значения после вмешательства, WWC сообщает скорректированные средние значения и нескорректированные стандартные отклонения и использует их при расчете величины эффекта.
- Для пре- и пост-тестов с использованием одного и того же показателя, когда сообщаются только нескорректированные групповые средние значения и информация о корреляции между тестами недоступна, WWC **вычисляет числитель размера эффекта** как разность между разностью средних значений пре- и пост-теста для группы вмешательства, а также разностью средних значений пре- и пост-теста для группы сравнения (т.н. «**разность между разностями**»). Однако эта **совокупная апостериорная (*post hoc*) корректировка** не является адекватной статистической корректировкой исходных различий в тех случаях, когда они попадают в диапазон стандартных отклонений от 0,05 до 0,25 для исследований с квази-экспериментальным дизайном и рандомизированных контролируемых исследований с высоким отсевом участников.
- Когда **в результатах пре- и пост-теста используются разные меры или показатели результатов являются дихотомическими** и авторы исследования сообщают только нескорректированные средние значения результатов для групп вмешательства и сравнения, WWC вычисляет размер эффекта разности между двумя группами по пре-тесту и размер эффекта разности между двумя группами по пост-тесту отдельно, при этом **окончательный размер эффекта определяется их разностью**.
- Если в исследовании представлены нескорректированные средние для пост-теста, WWC оценивает размер эффекта по формуле (1).

Когда WWC вносит корректировку «разность между разностями» в результаты, предоставленные автором исследования, WWC сообщает **уровни статистической значимости для скорректированных разностей**, которые отражают корректировку размера эффекта. Например, рассмотрим разницу в 0,2 до вмешательства в тесте достижений. Если бы разница после вмешательства составляла 0,3, скорректированный эффект разности между разностями был бы 0,1. Следовательно, статистическая значимость, сообщаемая WWC, будет основана на скорректированном результате 0,1, а не на нескорректированном результате 0,3.

Если сообщаемые автором и рассчитанные WWC размеры эффекта различаются, эксперты WWC пытаются определить источник различий и объяснить причины такого расхождения. В подобных случаях WWC сообщит о размере эффекта, рассчитанного WWC, потому что эти расчеты можно проверить, а использование показателей, рассчитанных WWC, обеспечивает сопоставимость результатов и исследований. Тем не менее,



WWC сообщит о представленном автором размере эффекта, который сопоставим с *g* Хеджеса, если он скорректирует исходные различия, а размер эффекта, рассчитанный WWC, не делает этого или основывается на апостериорной (*post hoc*) корректировке, описанной выше.

### **2.1.3. Размеры эффекта для исследований с дизайном с регрессионным разрывом (ДРР)**

Для исследований по дизайну с регрессионным разрывом (ДРР), которым присвоен рейтинг «Соответствует стандартам ДРР с оговорками» или «Соответствует стандартам ДРР без оговорок», WWC рассчитает размер эффекта таким же образом, как и для исследований с групповым дизайном. Как для непрерывных, так и для дихотомических переменных прогнозируемые средние значения или вероятности должны быть рассчитаны с использованием той же статистической модели, которая используется для оценки влияния на переменную в «точке отсечки» (т.е. для порогового значения).

*Для непрерывных переменных* числитель величины эффекта представляет собой **разность между прогнозируемыми средними для групп**, причем каждое среднее значение оценивается с использованием данных с соответствующей стороны порога («точки отсечки»). Стандартные отклонения и размеры выборок, используемые для стандартизации оценки воздействия, должны быть стандартными отклонениями групп вмешательства и сравнения от полной выборки (в отличие от только тех единиц внутри оптимальной ширины диапазона, который взвешивает наблюдения относительно их расстояния от порогового значения). Если можно было бы составить более одной группы вмешательства и сравнения (например, с нечетким ДРР), то группы вмешательства и сравнения должны быть сформированы на основе назначения вмешательства.

*Для дихотомических переменных* индекс Кокса должен рассчитываться с использованием прогнозируемых вероятностей при пороговом значении для групп вмешательства и сравнения, используя соответствующие данные выше и ниже порога.

### **2.1.4. Размеры эффекта для исследований с дизайном отдельного случая**

Для исследований с дизайном одного случая (кейс-стади, *single-case design*, ДОС), которые имеют рейтинг «Соответствует стандартам SCD WWC с оговорками» или «Соответствует стандартам SCD WWC без оговорок», WWC рассчитает размер эффекта, сравнимый с дизайном (*design-comparable effect size*, D-CES), там где это возможно и целесообразно по мнению руководства группы рецензентов. D-CES сопоставим с размером эффекта стандартизованной разности в средних, который предназначен для интерпретации аналогично показателю *g* Хеджеса, т.е. величине эффекта, используемой WWC для исследований в рамках группового дизайна [WWC Procedures Handbook, V. 4.1, С. 17].

Исследования ДОС включают **многочисленные наблюдения** в условиях вмешательства и сравнения для каждого человека. Несмотря на название, ДОС обычно включают данные **нескольких человек**. Для каждого человека есть несколько наблюдений в каждой фазе вмешательства (воздействия).

D-CES можно рассчитать для исследования, в котором участвуют **три или более участников** в вариантах дизайна, который имеет несколько базовых показателей (multiple baseline) для всех испытуемых, несколько проб (multiple probe) для всех испытуемых или дизайн обращения воздействия (AB)<sup>k</sup> (treatment reversal (AB)<sup>k</sup> design). В каждом случае числитель **размера эффекта** представляет собой среднее значение разности между наблюдениями в условиях вмешательства и сравнения, усредненных по отдельным лицам. Знаменатель размера эффекта — это оценка стандартного отклонения «между людьми в конкретных условиях» (between-person-within-condition standard deviation). Поскольку наблюдения для одного и того же человека коррелируют, вычисление степеней свободы знаменателя и дисперсии размера эффекта является более сложным, чем в обычных межсубъектных дизайнах. Кроме того, число степеней свободы в знаменателе, как правило, близко к числу субъектов, которое часто довольно мало, так что поправка смещения, аналогичная той, которая использовалась для вычисления **g Хеджеса**, является весьма важной.

Статистические подробности и формулы для расчета сравнимых по дизайну величин эффекта приведены в [WWC Procedures Handbook, V. 4.1, Приложение E].

Для вычисления D-CES требуется **доступ к сырым данным** о по каждому случаю, по ситуации наблюдения и по фазе вмешательства. Предпочтительный метод получения сырых данных, если он не представлен в подходящей форме в рецензируемой статье, **запрашиваются у авторов исследования**. Если авторы исследования не предоставляют сырые данные, но в статье представлены четкие графики, то рецензенты WWC могут также использовать программное обеспечение для оцифровки графиков для извлечения отдельных точек из графика (соответствующих отдельным испытуемым, условиям и т.д.).

При оценке **размера эффекта, сопоставимого с дизайном**, т.е.

1-C ES, WWC начнет со следующих **спецификаций по умолчанию**:

1. Используйте при оценке ограниченный метод максимального правдоподобия по умолчанию.
2. Укажите эффект вмешательства как фиксированный эффект.
3. Предположите, что «нет тренда» на базовом (начальном) или последующих этапах для оценки D-CES в дизайнах со множественными начальными условиями (multiple baseline designs).

Руководство группы рецензентов может определить на основе визуального анализа или соответствующего алгоритма, что основные данные не соответствуют вышеуказанным спецификациям. Группа рецензентов

может после консультации с экспертами по содержанию и методологии либо изменить вышеуказанные спецификации, либо не рассчитывать D-CES, если соответствующий метод недоступен. WWC будет документировать в отчете обоснование любых отклонений от стандартных спецификаций для вычисления D-CES.

## 2.2. Индекс улучшения

Чтобы помочь читателям оценить практическую важность размера эффекта вмешательства, WWC может **перевести размеры эффекта в значения индекса улучшения**. Индекс улучшения для отдельных результатов исследования представляет собой разность между рангом процентиля, соответствующим среднему значению результата (переменной) для группы вмешательства, и рангом процентиля, соответствующим среднему значению результата (переменной) для группы сравнения в распределении группы сравнения (последний, т.е. среднее по группе сравнения, составляет 50 процентов по определению). Подробная информация о вычислении индекса улучшения представлена в [WWC Procedures Handbook, V. 4.1, С. 18]. **Индекс улучшения может быть интерпретирован** как ожидаемое изменение в процентном ранге для среднего учащегося группы сравнения, если этот учащийся получил вмешательство.

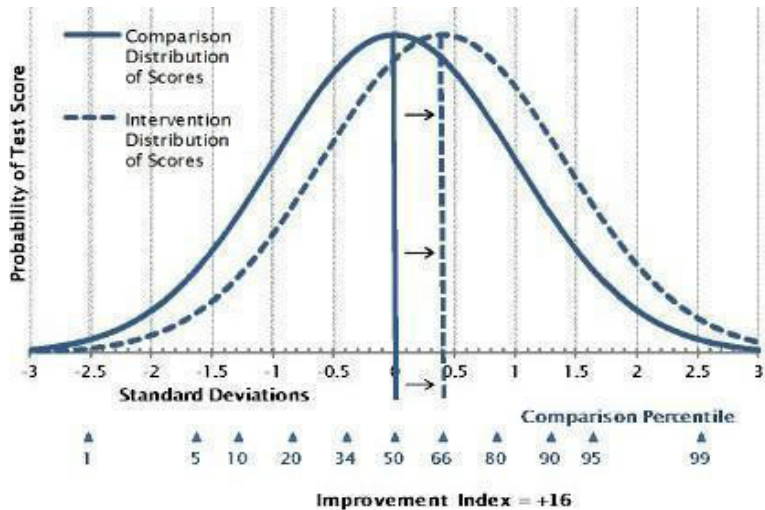


Рис. 4. Интерпретация индекса улучшения по WWC  
(по [WWC Procedures Handbook, V. 4.1, С. 18])

Рисунок 4 иллюстрирует **интерпретацию индекса улучшения**. В этом примере оцениваемое среднее влияние вмешательства представляет собой

улучшение на 0,4 стандартного отклонения результатов теста на чтение. Таким образом, в среднем ожидается, что учащийся в группе сравнения, результат которого близок к 50-му перцентилю для исследовательской выборки (сравнения), показал бы результат на 0,4 стандартного отклонения выше среднего (или примерно около 66-го перцентиля среди учащихся этой группы), если бы он или она получал вмешательство. Полученный индекс улучшения равен +16, что соответствует **изменению успеваемости среднего ученика** с 50-го до 66-го перцентиля распределения в группе сравнения. Для получения более подробной информации см. [WWC Procedures Handbook, V. 4.1, С. 18].

## 2.3. Статистическая значимость результата

### 2.3.1. Расчет статистической значимости при анализе на индивидуальном уровне

Чтобы адекватно оценить эффекты вмешательства, важно знать **статистическую значимость оценок эффектов** в дополнение к средней разности и размеру эффекта, как описано ранее. Для WWC статистически значимой оценкой эффекта является та, для которой нулевая гипотеза была проверена и отклонена с использованием ненаправленного теста и при уровне значимости (вероятности ошибки 1-го рода)  $\alpha = 0,05$  (в психологии принято обозначение  $p = 0,05$ ).

WWC обычно принимает уровни статистической значимости, **сообщаемые автором** (авторами) исследования. В некоторых случаях WWC потребуются вычислять статистическую значимость для результата, например, **если авторы исследования не сообщают о статистической значимости**. Чтобы вычислить статистическую значимость, WWC будет использовать размер эффекта и стандартные формулы ошибок, приведенные в [WWC Procedures Handbook, V. 4.1], [WWC Procedures and Standards Handbook, V. 5.0]. Например, t-статистика для оценки различий между групповыми средними при непрерывных измерениях рассчитывается с использованием **формулы (4)**:

где  $g$  — размер эффекта,  $n_i$  и  $n_c$  — объемы выборок (количество испытуемых в группах) вмешательства и сравнения, соответственно, для набора переменных.

$$t = \frac{g}{\sqrt{\frac{n_i+n_c}{n_i n_c} + \frac{g^2}{2(n_i+n_c)}}}, \quad (4)$$

Кроме того, t-статистика для результатов, основанных на дихотомических переменных, рассчитывается по **формуле (5)**:

$$t = 1.65 \frac{d_{\text{Cox}}}{\left( \frac{1}{P_i N_i} + \frac{1}{(1-P_i) N_i} + \frac{1}{P_c N_c} + \frac{1}{(1-P_c) N_c} \right)}, \quad (5)$$

где  $d_{\text{Сох}}$  — размер эффекта, основанный на индексе Кокса, а  $p_i$  и  $p_c$  — вероятности положительного результата для учащихся в группах вмешательства и сравнения соответственно. Эти рассчитанные или пересчитанные WWC оценки будут использоваться в продуктах (обзорах) WWC; техническая документация WWC будет включать обсуждение любых исправлений или модификаций значений вероятностей, сообщенных автором. Полный список всех рассчитанных WWC величин эффектов и их стандартных ошибок можно найти в [WWC Procedures Handbook, V. 4.1], [WWC Procedures and Standards Handbook, V. 5.0].

### 2.3.2. Кластерная коррекция для «несоответствующих» случаев анализа

В кластерных испытаниях (как случайных, так и неслучайных) группы участников, такие как классы или школы, в отличие от отдельных лиц, приписываются условиям эксперимента. **Участники ранее существовавших групп, как правило, больше похожи друг на друга, чем на участников других ранее существовавших групп.** Например, ученики одной школы больше похожи друг на друга, чем на учеников других школ. Это сходство людей в кластере означает, что **учащиеся в одном кластере не могут рассматриваться как независимые**, что является основным допущением, лежащим в основе большинства статистических тестов, которые проводятся в области образования, социальных наук и в медицине. Сравнительно часто анализ проводится на индивидуальном уровне, в то время как назначение (приписывание условиям) выполняется на уровне кластера. Неспособность учесть кластеризацию в анализе данных иногда называют **проблемой «несоответствия»**. Основная проблема заключается в том, что игнорирование корреляции между ответами среди людей в одних и тех же кластерах приводит к слишком маленьким стандартным ошибкам, и, следовательно, значения вероятности, возникающие из статистических тестов, также слишком малы. Другими словами, слишком высока вероятность, что нулевая гипотеза будет отвергнута, если анализ данных не учитывает кластеризацию.

Чтобы оценить статистическую значимость эффектов вмешательства в тех случаях, когда авторы исследования назначают условия на уровне кластера, но анализируют результаты на уровне учащегося, WWC рассчитывает **статистические оценки значимости с поправкой на кластеризацию** на основе руководства Хеджеса (2007). **Основной подход к коррекции кластеризации** состоит в том, чтобы сначала вычислить статистику  $t$ , соответствующую размеру эффекта, который игнорирует кластеризацию, а затем скорректировать как статистику  $t$ , так и соответствующие степени свободы для кластеризации на основе размеров выборки, количества кластеров и оценки внутриклассового коэффициента корреляции (intraclass correlation coefficient, ICC). В качестве значений по умолчанию, WWC

использует значения ICC, равные .20 для результатов достижений и .10 для всех других результатов, но будет использовать значения ICC, представленные в исследованиях, если они доступны. Если отклонение от этих значений по умолчанию оправдано, это будет указано в протоколе рецензии. Оценку статистической значимости, скорректированную для кластеризации, затем получают из t-распределения с использованием скорректированной t-статистики и степеней свободы. Каждый шаг процесса указан в [WWC Procedures Handbook, V. 4.1, Приложение F].

#### **2.4. Выводы из нескольких вариантов анализа (множественного анализа)**

В исследованиях часто представлено несколько результатов, полученных из разных видов анализа, в которых варьируют условие сравнения, показатель результата, выборка или момент времени. Например, анализ может включать всех участников исследования или подгруппы в популяции. Аналогичным образом, анализ может включать в себя несколько показателей результатов в одной и той же области, один результат, измеренный в несколько разных моментов времени, а также комплексный показатель и его компоненты [WWC Procedures Handbook, V. 4.1, С. 20].

Для исследования с множественным анализом WWC рецензирует все приемлемые основные результаты, как определено в соответствующем протоколе рецензирования исследования. **Рейтинг исследования** указан как **самый высокий рейтинг, полученный по всем приемлемым основным результатам**. В целом, главные выводы:

- Отвечают скорее на подтверждающие (конфирматорные), чем на исследовательские (эксплораторные) вопросы исследования.
- Соответствуют полной выборке, назначенной вмешательству, а не подмножествам этой выборки, за исключением тех случаев, когда полная выборка не подходит для рассмотрения в соответствии с протоколом обзора исследования.
- Полагаются на совокупные или составные (комплексные) показатели и основные результаты, как они определены в протоколе рецензии, а не на субшкалы или дополнительные результаты, которые также могут иметь право на рассмотрение.
- Сосредоточены скорее на анализе с учетом нормативных показателей, а не на анализе чувствительности (sensitivity analyses).
- Для вмешательств в 12-м классе или ниже фокусируют внимание на самом раннем моменте времени после получения вмешательства, если в протоколе рецензии исследования не указан другой момент времени.

Дополнительные выводы включают анализ, проведенный в дополнительных исследованиях, которые могут быть рассмотрены в соответствии с протоколом. WWC рассмотрит дополнительные выводы, только если

они указаны заранее для целей рецензии. Запросы к авторам проводятся по мере необходимости для оценки всех приемлемых представлений результатов анализа (основных и дополнительных), используемых для рецензирования исследования.

## Требования к заполнению «Паспорта верификации психолого- педагогических программ и технологий для образования и социальной сферы»

Отбор психолого-педагогических программ и технологий для организации и проведения их научной верификации осуществляется на основании заполнения экспертами-инициаторами «Паспорта верификации психолого-педагогических программ и технологий» (далее — Паспорт), структура которого основана на информации, необходимой для формирования открытого реестра психолого-педагогических программ и технологий федерального уровня. На основе этого Паспорта в случае положительных результатов эксперимента по верификации программа/технология вносится в «Реестр психолого-педагогических программ и технологий с доказанной эффективностью для образования и социальной сферы» Министерства просвещения РФ.

При этом экспертам рекомендуется представлять для верификации программы или технологии, которые прошли независимую профессиональную экспертизу (например, в рамках конкурсного отбора), основательно регламентированы в соответствии с требованиями к программам, вызывающим доверие профессионального сообщества (см. Раздел 4 «Реестр рекомендуемых программ психологической помощи» методических рекомендаций [Система функционирования психологических служб, 2020]), и заполнять Паспорт с учетом уточнений и предложений, внесенных ответственными за организацию формирующего этапа эксперимента экспертами и согласованных с авторами и разработчиками программ и технологий.

Основные структурные элементы Паспорта сформулированы в главе 2 «Рамочные условия организации научной экспертизы и исследований по верификации психолого-педагогических программ и технологий». В данном приложении приведем комментарии и требования, позволяющие авторам-разработчикам или экспертам-инициаторам представить необходимую информацию при заполнении Паспорта.

1. Авторы-разработчики или эксперты-инициаторы могут заявить для проведения научной верификации одну либо несколько разработок (программ, технологий), апробированных в условиях образовательных организаций либо учреждений социальной сферы.
2. В паспорте указывается **фамилия, имя, отчество ответственных экспертов** — инициаторов верификации, ученая степень, а также их должность и место работы.



3. Подтверждение об апробации программы/технологии должно присутствовать в структуре самой разработки. При этом в Паспорте указывается полное **наименование организации**, на базе которой была проведена апробация программы/технологии, получены и хранятся предварительные результаты, подтверждающие ее эффективность. Эти результаты могут быть основанием для проведения полномасштабного эксперимента по верификации. Если уже был проведен полномасштабный эксперимент под научным руководством представителей академического сообщества на большой выборке, его результаты могут быть представлены на экспертизу в МГППУ в соответствии с **Отчетом о результатах верификации** (см. **Приложение 1, п. 4.1.**) для получения рейтинговой и нормативной оценок в соответствии с **Протоколом оценки результатов верификации** (см. **Приложение 2, п. 1.3.**)
4. Авторы — разработчики программы/технологии.
5. Рекомендуется сохранять **наименование программы/технологии**, предложенное автором/авторским коллективом для фиксирования факта принятия разработки в федеральный реестр, возможно, ранее включенной в какие-либо реестры профессионального сообщества.
6. Для программы/технологии необходимо указать **уровень образования**, в рамках которого разработка применяется, руководствуясь классификацией уровней образования согласно № 273-ФЗ «Об образовании в РФ», ред. от 16.04.2022 (Общее образование: дошкольное образование, начальное общее образование, основное общее образование, среднее общее образование; Профессиональное образование: среднее профессиональное образование, высшее образование — бакалавриат, высшее образование — специалитет, магистратура, высшее образование — подготовка кадров высшей квалификации; Дополнительное образование: дополнительное образование детей и взрослых, дополнительное профессиональное образование) (или Социальная сфера).
7. Для выбора **вида программы/технологии** рекомендуется исходить из формулировок трудовых функций, соответствующих направленности программы, согласно профильному профессиональному стандарту.
8. Наименование **целевой группы** необходимо указать в соответствии с методическими рекомендациями «Система функционирования психологических служб в общеобразовательных организациях» раздел 5 «Реестр рекомендуемых психодиагностических методик для применения в общеобразовательных организациях», которые предусматривают следующий перечень целевых групп: 1. Норма (нормотипичные дети и подростки с нормативным кризисом взросления); 2. Дети, испытывающие трудности в обучении; 3. Уязвимые категории детей: 3.1. Дети, находящиеся в трудной жизненной ситуации: 3.1.1. Дети-сироты и дети, оставшиеся без попечения родителей, 3.1.2. Обучающиеся с ОВЗ, дети-инвалиды, 3.1.3 Дети с отклоняющимся поведением

- (девиантное поведение детей и подростков, суицидальное поведение детей и подростков) и 3.2. Одаренные дети; 4. Взрослые (с указанием основных характеристик, например, родители, педагоги и др.).
9. Для уточнения значимых характеристик целевой группы в Паспорте указывается **Целевая аудитория**, то есть конкретная более узкая группа, которой адресована программа/технология.
  10. Типология **сферы применения (сфера трудностей)** предложена в методических рекомендациях «Система функционирования психологических служб в общеобразовательных организациях» — Раздел 5 «Реестр рекомендуемых психодиагностических методик для применения в общеобразовательных организациях» — и включает следующие формулировки: Развитие основных психических функций, Метапредметные компетенции и универсальные учебные действия, Социальное развитие и морально-ценностная сфера, Эмоционально-волевая сфера, Коммуникативная сфера, Поведенческая сфера, Профессиональная направленность, мотивация, характерологические особенности, Профессиональные и личностные проблемы взрослых участников образовательного процесса, Профессиональное и/или дополнительное образование взрослых.
  11. В соответствии с указанной выше типологией также необходимо перечислить конкретные **типы трудностей**, на решение которых направлена программа.
  12. **Проблемную ситуацию**, на решение которой направлена программа, важно сформулировать наиболее подробно, с краткой характеристикой социальной ситуации развития ребенка.
  13. **Цели программы/технологии** могут быть уточнены экспертом-инициатором для точного соответствия специфической для участников программы/технологии предметной области.
  14. **Задачи программы/технологии**, то есть актуальные для организации конкретные задачи, используемые для достижения поставленных целей.
  15. **Краткая характеристика формирующего этапа эксперимента**, характеризующая четкое и последовательное содержание деятельности в соответствии с психологическими особенностями тех, на помощь кому направлена программа/технология, заявленными целями и задачами, демонстрирующая внутреннюю логику построения, содержательную взаимосвязь основных методов и алгоритмов взаимодействия.
  16. **Измеряемые конструкты и критерии оценки планируемых результатов** и характеристики (признаки, свойства, параметры и т.п.), на которые программа/технология должна оказывать эффект. Экспертам-инициаторам рекомендуется обратить внимание на содержательность и логичность критериев оценки достижения результатов (качественных и количественных) и способов/форм их использования.

17. **Ожидаемые результаты** реализации программы должны отвечать требованиям обоснованности определения и значимости результатов, подтверждающих эффективность реализации программы (технологии).
18. **Ограничения применения программы.** Авторам-разработчикам и экспертам-инициаторам рекомендуется указать, для какой аудитории и, возможно, при каких условиях и т.п., программу не следует применять или она требует для них определенной доработки, коррекции и дополнительных исследований эффективности.
19. **Психодиагностические методики**, используемые для оценки достижения планируемых результатов (с опорой на «Открытый реестр психодиагностических методик, вызывающих доверие профессионального сообщества»). Экспертам-инициаторам необходимо обратить внимание на научную обоснованность, доказательность и согласованность оснований выбора используемых методик, инструментария с планируемыми результатами, целями и задачами программы/технологии, на их соответствие измеряемым конструктам, а также на валидность и надежность выбранных автором(ами) психодиагностических методик. Использование стандартизованных инструментов — необходимое условие доказательности полученных результатов.
20. **Длительность реализации программы/технологии**, то есть период времени и/или количество акад. часов).
21. **Гиперссылка на программу/технологию**, если она размещена в реестре профессионального сообщества, рекомендующего программу к применению.
22. **Краткие сведения о проведении апробации** — приводится подтверждение полноты и конкретности данных о практической апробации программы (технологии), подтверждающей эффективность ее реализации в условиях образовательной организации.
23. Если авторский коллектив опубликовал результаты исследования эффективности программы/технологии, необходимо представить **ссылки на статьи, диссертации** и др.
24. **Ресурсные площадки** — наименование базы верификации программы/технологии (конкретные образовательные и социальные организации, учреждения, центры и т.д.), предоставляющие материально-технические ресурсы в целях создания открытого реестра рекомендованных на федеральном уровне программ/технологий.
25. **Планируемые сроки проекта по верификации программы/технологии** устанавливаются по согласованию с ответственными подразделениями — обеспечивающими анализ эмпирических данных верификации
26. **Результаты верификации** должны содержать краткие выводы и рекомендации по итогам проведения формирующего эксперимента и обработки и интерпретации его результатов. Этот пункт заполняется после проведения эксперимента по верификации и получения рейтинговой

и нормативной оценок в соответствии с **Протоколом оценки результатов верификации** (см. **Приложение 2, п. 1.3**).

**Статьи по результатам верификации** (библиографические описания, ссылки) публикуются в научных журналах и подтверждают проведение соответствующих процедур верификации программы/технологии. Этот пункт также заполняется после проведения эксперимента по верификации и получения рейтинговой и нормативной оценок в соответствии с **Протоколом оценки результатов верификации** (см. **Приложение 2, п. 1.3**).

Научное издание

**Доказательный подход: Руководство по верификации программ,  
технологий, практик в образовании и социальной сфере**

*Учебное пособие*

Под редакцией:

*Марголис А.А., Сорокова М.Г., Семья Г.В.*

Литературная редакция: *С.В. Гани*

Компьютерная верстка *М.В. Мазоха*

Подписано в печать: 20.06.24. Формат: 60х90/16.  
Гарнитура «Times». Усл. печ. п. 8,1. Усл.-изд. л. 7,3.  
Электронное издание. Печать по требованию.

Московский государственный психолого-педагогический университет 127051,  
г. Москва, ул. Сретенка, д. 29; тел.: (495) 632-90-77; факс: (495) 632-92-52

<http://mgppu.ru> <http://inclusive-edu.ru>