**И.Б. Шиян, С.А. Зададаев,
Т.Н. Ле-ван, О.А. Шиян**

Апробация шкал оценки качества дошкольного образования ECERS-R в детских садах города Москвы

В данной статье приводятся результаты исследования по апробации международных шкал оценки качества дошкольного образования ECERS-R в российских детских садах. Целью исследования было изучение возможности использования шкал ECERS-R в качестве инструмента оценки качества в российской системе дошкольного образования.

Ключевые слова: качество дошкольного образования; шкалы ECERS-R; инструмент оценки; российская система дошкольного образования.

Шкалы ECERS-R (Early Childhood Environment Rating Scales — Шкалы для комплексной оценки качества образования в дошкольных образовательных организациях) [3] представляют собой инструмент оценки качества образования детей дошкольного возраста. На сегодняшний день они являются одним из наиболее признанных и распространенных международных инструментов оценки качества дошкольного образования. Только за период с 2003 по 2010 год в мире было проведено 76 исследований с использованием ECERS-R [2]. С помощью шкал было проведено множество международных сравнительных исследований [5–8]. Выбор шкал ECERS в качестве инструмента оценивания и исследований представителями профессионального сообщества из столь отличающихся друг от друга стран (от США до Греции, от Швеции до Южной Кореи) свидетельствует о существовании определенного консенсуса в отношении того, что в шкалах ECERS отображено современное представление о качестве дошкольного образования.

Процедура апробации шкал ECERS-R в детских садах города Москвы

Для того чтобы понять, насколько применимы шкалы ECERS-R для решения задач по оценке качества дошкольного образования в России, была осуществлена

апробация шкал в 23 детских садах г. Москвы (всего в апробации приняли участие 26 групп), которая проводилась лабораторией развития ребенка Института системных проектов Московского городского педагогического университета под руководством И.Б. Шияна и О.А. Шиян.

Поскольку шкалы ECERS-R являются, прежде всего, инструментом внешней оценки качества образования, особое внимание во время апробации было уделено подготовке экспертов. В их число вошли специалисты в области детской психологии и дошкольной педагогики, старшие воспитатели, воспитатели, психологи из московских детских садов. Всего в оценке участвовали 32 эксперта. К подготовке экспертов были привлечены как отечественные специалисты в области детского развития и дошкольного образования, так и наши партнеры из университета Гётеборга (Швеция) под руководством проф. Сони Шеридан, имеющие большой опыт использования шкал ECERS.

В процессе подготовки экспертов, а также в период осуществления апробационной экспертизы проводились аналитические семинары, в ходе которых вносились корректировки в примечания к шкалам во избежание неоднозначности толкований и оценок.

Каждая группа детского сада оценивалась в рамках апробации как минимум двумя экспертами независимо друг от друга. В среднем наблюдение с использованием шкал занимало от 4 до 6 часов.

Выборка была разделена на три части. В группу А были отобраны три группы детских садов г. Москвы, качество образования в которых, по экспертным оценкам, было значимо выше, чем в среднем по г. Москве. В группу В вошли пять случайно выбранных групп из детских садов Троицкого и Новомосковского административного округа г. Москвы, в группу С вошли 18 случайно выбранных групп из детских садов, расположенных в других административных округах города. Таким образом, группу А можно считать контрастной, а группы В и С — контрольными. Более высокое качество предполагает работу в зоне ближайшего развития ребенка-дошкольника в сочетании с дружественным и поощряющим характером взаимодействия с ребенком. Задача исследования состояла в том, чтобы оценить способность ECERS зафиксировать различия в показателях детских садов из контрастных групп.

В ходе апробации предполагалось ответить на вопросы: обладают ли шкалы различающей силой в условиях российской системы дошкольного образования; совпадают ли характеристики образовательной среды, обнаруживающиеся при использовании шкал ECERS-R, с оценками экспертного профессионального сообщества?

Результаты апробации

Представим результаты, полученные в ходе апробации адаптированных шкал ECERS-R в образовательных организациях г. Москвы.

В первую очередь обращает на себя внимание факт различий средних значений отдельных шкал в группе А по сравнению с группами В и С (между двумя последними группами значимых отличий выявлено не было). Так, например,

типично расхождение значений в среднем на два пункта (см. табл. 1). При этом отличие между контрастной и контрольными группами значимо на уровне $\rho \leq 0{,}05$ по шкалам «Присмотр и уход за детьми», «Речь и мышление», «Взаимодействие» и «Структурирование программы». На наш взгляд, это объясняется тем, что по данным шкалам качество среды зависит от инициативы и представлений о должном обустройстве образовательного процесса у административных и педагогических работников дошкольной образовательной организации, тогда как позиция «Предметно-пространственная среда» во многом регламентирована СанПиН (Санитарно-эпидемиологическими требованиями к устройству, содержанию и организации режима работы дошкольных образовательных организаций 2.4.1.3049-13, утв. постановлением Главного государственного санитарного врача Российской Федерации от 15 мая 2013 г. № 26, http://monm.rk.gov.ru/file/1.Санпин_ДОУ.pdf).

Отсутствие значимых различий в группах А, В и С при оценке шкалы «Виды детской активности», вероятно, связано с тем, что образовательная деятельность всех муниципальных дошкольных образовательных организаций регламентируется не только федеральным государственным образовательным стандартом дошкольного образования (ФГОС ДО), но также и сохраняющимися традициями применения предшествующих нормативно-инструктивных и нормативно-методических документов, подробно описывающих требования к условиям организации образовательного процесса в детских садах, то есть фактически — к условиям проведения организованной и свободной образовательной деятельности.

Таблица 1

Средние значения шкал ECERS-R для групп А, В и С и стандартные отклонения в значениях каждой группы

Шкалы	Группа А		Группа В		Группа С	
	Среднее значение	Стандартное отклонение	Среднее значение	Стандартное отклонение	Среднее значение	Стандартное отклонение
Предметно-пространственная среда	5,38	1,09	3,19	1,44	3,36	1,00
Присмотр и уход за детьми	5,86	0,21	3,83	1,37	3,73	1,15
Речь и мышление	5,88	0,99	3,20	1,64	3,05	1,39
Виды детской активности	4,97	1,23	2,70	1,34	2,57	0,84
Взаимодействие	6,40	0,66	4,38	1,51	3,80	1,42
Структурирование программы	5,25	0,75	3,30	0,99	2,47	0,96
Родители и персонал	5,06	1,17	3,78	1,39	3,37	1,20

Как видно из приведенных данных, четыре из семи шкал получили при экспертной оценке в группе А балл 5 (с точностью до единицы) или значение, приближенное к 5 (при сохранении точности до сотых долей), — то есть стандартное

отклонение среднего значения от значений каждого из экспертов колеблется в пределах 1, точнее — от минимального 0,21 (по шкале «Присмотр и уход за детьми» в группе А) до 1,64 (по шкале «Речь и мышление» в группе В). Самые высокие значения в группе А зафиксированы по позициям «Взаимодействие» (6,40) и «Речь и мышление» (5,88). Эти показатели свидетельствуют о высоком уровне качества педагогического общения в детских садах контрастной группы, тогда как в группах В и С показатели шкалы «Речь и мышление» оценены как «минимально допустимые» и имеют значения 3,20 (группа В) и 3,05 (группа С). При этом наивысшее среднее значение в группах В и С получено по шкале «Взаимодействие», и составляет 4,38.

Более наглядно различия между контрастной (А) и контрольными группами (В и С) в оценке среды с применением шкалы ECERS-R по шкалам можно продемонстрировать с помощью диаграммы (рис. 1).

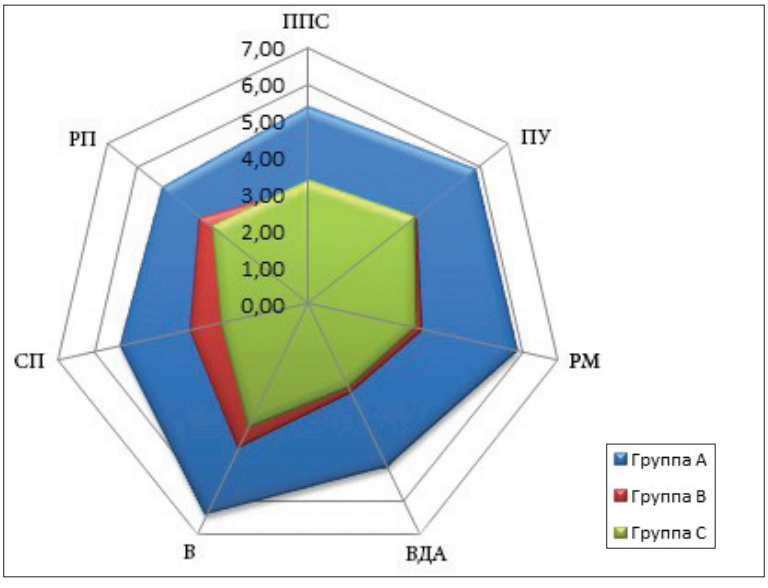

Рис. 1. Средние значения шкал ECERS-R для групп А, В и С:
ППС — предметно-пространственная среда, ПУ — присмотр и уход за детьми, РМ — речь и мышление, ВДА — виды детской активности, В — взаимодействие, СП — структурирование программы, РП — родители и персонал

Рассмотрим суть различий между дошкольными образовательными организациями контрастной и контрольных групп более детально. Для этого обратимся к анализу данных подшкал каждой из шкал, имеющим статистически достоверные различия (на уровне $\rho \leq 0{,}05$). Причем нас будут интересовать в первую очередь позиции, демонстрирующие отличия обеих контрольных групп (В и С) от контрастной (А), поскольку они, на наш взгляд, демонстрируют характерные тенденции, не зависящие от территориальных, административных, кадровых и других особенностей образовательной организации.

Отметим, что значимых различий по всем подшкалам, кроме «Групповые занятия» (шкала «Структурирование программы»), в детских садах двух контрольных групп (B и C) выявлено не было. Итак, различия между контрольными и контрастной группами наблюдаются по следующим подшкалам: «Места для уединения», «Встреча / прощание», «Сон / отдых», «Использование речи для развития мыслительных навыков», «Музыка / движение», «Математика / счет», «Распорядок дня», «Свободная игра», «Взаимодействие и сотрудничество персонала» (табл. 2).

Таблица 2

Значимые различия между контрольными (B и C) и контрастной (A) группами по подшкалам ECERS-R

Подшкалы	Группа A		Группа B		Группа C	
	Среднее значение	Стандартное отклонение	Среднее значение	Стандартное отклонение	Среднее значение	Стандартное отклонение
Места для уединения	6,00	0,87	2,10	2,19	1,87	1,62
Встреча / прощание	6,67	0,58	2,60	2,07	3,01	2,15
Сон / отдых	4,67	0,76	2,80	0,84	2,60	1,13
Использование речи для развития мыслительных навыков	5,83	1,04	2,30	0,97	2,94	1,57
Музыка / движение	5,83	1,26	2,60	1,95	2,17	1,25
Математика / счет	5,17	0,29	2,50	1,58	2,85	1,49
Распорядок дня	6,67	0,58	4,00	1,87	3,26	1,26
Свободная игра	6,83	0,29	4,80	1,30	3,99	1,38
Взаимодействие и сотрудничество персонала	6,67	0,58	3,70	2,05	3,70	1,46

Как видно из приведенных нами данных, подшкалы, по которым имеются значимые различия между контрольными и контрастной группами, были оценены в детских садах группы A как соответствующие хорошему и даже отличному уровню (балл 5 и 7 — с точностью до единицы, или приближенное к 5 и 7 — при сохранении точности до сотых долей). Наивысшее значение в группе A достигнуто по подшкалам «Свободная игра» (6,83), «Встреча / прощание», «Распорядок дня», «Взаимодействие и сотрудничество персонала» (6,67 каждый), тогда как в детских садах контрольных групп этим показателям свойственно значение на уровне «минимально» («Свободная игра», «Распорядок дня», «Взаимодействие и сотрудничество персонала» — в обеих контрольных группах, «Встреча / прощание» — в группе C) или недостаточно («Встреча / прощание» — в группе B). Обращает на себя внимание тот факт, что в контрольных группах на уровне «недостаточно» получили оценку такие важные для психологического комфорта детей показатели, как «Места для уединения», «Сон / отдых» (для групп B и C), «Встреча / прощание» (в группе B). Также не достигают минимально допустимого уровня значения подшкалы «Использование речи для развития мыслительных навыков», «Музыка /

движение», «Математика / счет». На практике это означает, что педагоги практически не обсуждают с детьми логические взаимосвязи (часто игнорируют вопросы детей и их любопытство относительно причин разных явлений, не обращают внимание детей на последовательность событий, произошедших в течение дня и т. д.), и, что самое типичное, воспитатель чаще предпочитает давать готовые ответы, а не помогает детям самостоятельно находить решения. Не удовлетворяются, как правило, потребности детей в музицировании или движении под музыку вне музыкальных занятий (именно в группе, во время свободной игры), и даже на соответствующих занятиях часто дети поют или слушают музыку сидя, редко импровизируют с движениями под музыку и пробуют извлекать звуки при помощи различных музыкальных инструментов. В группах недостаточно разнообразных материалов для обучения детей счету и числам, различным действиям с ними (счет, измерение, сравнение количества, узнавание формы, постижение письменных цифр и т. д.).

Если сравнивать результаты, полученные в группе С и в группе А, то достоверных различий наблюдается гораздо больше. Так, недостаточный уровень в группе С на фоне хорошего в группе А (оценка 6,00) был зафиксирован по показателям: «Предметы обстановки для отдыха и комфорта», «Прием пищи / перекусы», «Стимулирование общения», «Искусство». Следовательно, можно наблюдать как типичную ситуацию, когда необходимые для комфортного проживания детства мягкие предметы и игрушки не присутствуют в окружающем детей пространстве или недоступны им в любое время (например, используются исключительно воспитателем на занятиях в качестве дидактического материала). Еще одна ситуация, требующая анализа, — возможность более гибкого графика приема пищи детьми, учет их вкусов и потребностей, особенно возможность предусмотреть особое меню для детей с пищевой аллергией. Важным аспектом приема пищи является социальная атмосфера: в садах группы С типичной является ситуация, в которой воспитатели жестко требуют от детей соблюдения манер, заставляют доедать порции, поевших детей заставляют дожидаться остальных и т. п. Обращает на себя внимание и недостаточный уровень по показателю «Стимулирование общения» в группе С. Многие специалисты отмечают недоразвитие речи у значительного количества детей дошкольного возраста [1; 4], в связи с этим важно организовать особую речевую среду в детском саду: использовать определенные методы и приемы стимулирования детей к высказываниям, ролевой игре с разыгрыванием речевых ситуаций и т. п. Кроме того, современный социокультурный контекст российского дошкольного образования определяется национальным многообразием (особенно это касается мегаполисов), и дети, для которых русский язык не является родным (например, дети из семей мигрантов, семей беженцев), нуждаются в педагогической поддержке в плане освоения русского языка. Однако должных условий для этого в детских садах группы С обнаружено не было. И, наконец, прокомментируем недостаточный уровень обеспечения условий для занятий художественным творчеством (показатель «Искусство»). Шкалы ECERS-R указывают на необходимость присутствия в помещении группы и доступности во время

свободной деятельности детей материалов для рисования (бумаги, цветных карандашей, мелков, фломастеров, красок и др.), материалов для создания объемных произведений (пластилина, глины, дерева и др.), материалов для склеивания (клея, скотча и проч.). Этих материалов должно быть достаточно для индивидуального самовыражения каждого ребенка, причем не только в течение занятий по изобразительной деятельности, конструированию и лепке, но и для свободной деятельности детей.

Специфика использования отдельных показателей при оценке российских детских садов

Поскольку в ходе апробации нас, прежде всего, интересовала возможность применения шкал в российских условиях, важно проанализировать возникшие трудности в оценке тех или иных показателей. Так, например, при работе с показателем «Определены и удобно оборудованы… по крайней мере три центра интересов» (показатель «Обустройство пространства для игр») эксперты отмечали, что в ряде организаций, где они проводили оценку, центры по интересам номинально присутствуют: есть уголок домашнего хозяйства, «парикмахерская», «магазин», «ремонтная мастерская», центр театрализации, уголок формирования математических представлений, уголок для чтения, уголок для творчества и т. д., но создается впечатление, что детям не разрешено ими пользоваться в любое время, отведенное для свободных игр и самостоятельной образовательной деятельности. В итоге, в ходе апробации шкалы было принято решение исходить из конкретизированного описания показателя: засчитывать его только в случае, если наблюдается факт свободного использования данных центров детьми. Кроме того, возникла необходимость уточнить, что несколько разных центров для ролевых игр не являются различными центрами по интересам, поскольку склонность детей к ролевым играм может иметь проявление в различных сюжетах, однако, по сути, этот предпочтительный вид свободной игры основывается на одних и тех же механизмах развития личности и не способствует становлению многообразных форм самостоятельной деятельности детей. В противном случае, показатель оказывается неинформативным (то есть не позволяет зафиксировать реально существующие различия в детских садах). Другой пример неинформативности индикатора можно обнаружить в факте единодушия экспертов при оценке индикатора «Учитываются интересы детей, проснувшихся раньше других, и тех, которые не спят днем» (подшкала «Сон / отдых»): в подавляющем большинстве проэкспертированных детских садов (97,67 %) фиксировалось отсутствие возможности создать условия для рано встающих и незасыпающих детей. Эта ситуация, как правило, объективна, экспертная оценка отражает реальную действительность в детских садах, однако это не означает, что ситуацию не следует менять — создание таких условий связано с необходимостью учета индивидуальных особенностей детей, что зафиксировано в качестве одного из основополагающих принципов ФГОС ДО.

Аналогичной является ситуация с индикаторами соответствия расписания сна, туалета и мытья рук индивидуальным и возрастным особенностям детей, санитарным требованиям и нормам, связанным с гигиеной и чистотой.

Благополучие ситуации эксперты обоснованно отмечают по таким показателям, как наличие материалов и оборудования для переодевания для ролевых игр, отсутствие громкой фоновой музыки в течение дня, мешающей проведению ежедневных мероприятий, отсутствие крупных угроз безопасности внутри здания и на улице.

Аналогичная по единодушию, но в отрицательной оценке индикатора ситуация наблюдается в случае с использованием телевизора, видео и/или компьютера в образовательном процессе: показатель «Некоторые компьютерные программы развивают у детей творческий подход» 97,67 % экспертов оценили отрицательно, или как неприменимый (из-за отсутствия практики использования компьютерных программ в образовательном процессе). Здесь можно говорить о противоречии, с одной стороны, активной информатизации системы образования, в том числе дошкольного, через реализацию государственных программ, а с другой — неготовности дошкольных образовательных организаций эффективно использовать электронные ресурсы (неготовности и материально-технической, и профессиональной).

Обратимся к показателю «Условия для детей с ограниченными возможностями». Подавляющее большинство экспертов (88,37 %) сочли этот показатель в целом неприменимым в оцениваемых ими детских садах, отметив отрицательно все индикаторы. При этом в 97,67 % случаев у экспертов был обозначен «провал» во всех индикаторах, кроме первого: 11, 63 % оценок указало на то, что сотрудники дошкольных образовательных организаций «пытаются понять потребности детей, в том числе с помощью доступных способов диагностики» (напомним, мы говорим о детях с ограниченными возможностями здоровья, причем, согласно комментариям к показателю, применить и начать его оценивать можно лишь в случае, когда в группе есть ребенок с выявленной и диагностированной инвалидностью, но эксперт не должен требовать при этом от воспитателя указывать на такого ребенка, предоставлять его индивидуальный образовательный план и другую документацию, подтверждающую инвалидность).

Особенности образовательной среды московских детских садов, выявленные в ходе апробации

Наглядно результаты применения ECERS-R в детских садах групп A, B и C можно продемонстрировать с помощью диаграммы (рис. 2), где аббревиатуры указывают на конкретные показатели шкалы.

Как показано на диаграмме, общими тенденциями для всех трех групп (A, B и C) являются низкие значения по показателям «Условия для детей с ограниченными возможностями» (1,50; 0,20 и 0,24 баллов соответственно), «Условия для удовлетворения личных потребностей персонала» (3,33; 2,30 и 2,47 баллов соответственно), «Содействие принятию многообразия» (3,50; 0,90 и 1,16 баллов соответственно). Очевидно, данные позиции представляют собой проблемные поля современного дошкольного образования и требуют научно обоснованных и прак-

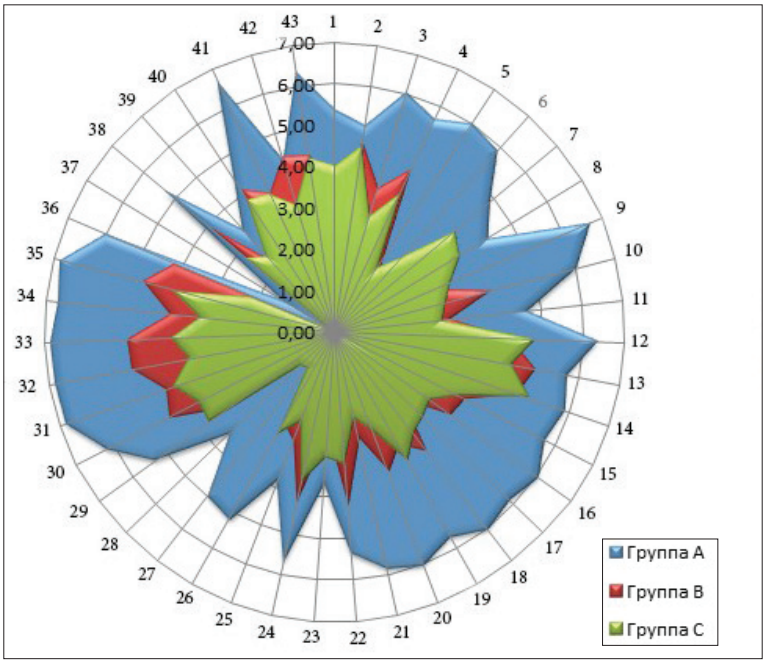

Рис. 2. Средние значения показателей ECERS-R для групп A, B и C (подшкалы обозначены их порядковыми номерами)

тически осмысленных решений. Об условиях для детей-инвалидов и детей с ограниченными возможностями здоровья, а также о поощрении принятия многообразия уже было сказано, представим наши соображения относительно обеспечения условий для удовлетворения личных потребностей персонала. Под таковыми подразумеваются, в частности, наличие мебели для взрослых за пределами мест для игр детей, мест для хранения личных вещей. Самым сложным для реализации на практике пунктом можно считать наличие минимум одного перерыва в течение рабочего дня, что связывается с необходимостью непрерывного наблюдения за безопасностью детей, проведения непрерывного образовательного процесса. Также представляет трудность организация совместного приема пищи педагога с детьми: это не оговаривается в обязанностях воспитателя (за исключением отдельных образовательных организаций), хотя совместная трапеза может служить эффективным способом приобщения детей к культуре питания, удачным временем для обсуждения с детьми прожитых впечатлений, планирования совместных дел и т. д.

Следует указать на практически полное совпадение значений контрастной группы A и контрольной группы B по показателю «Сопровождение работы и оценивание персонала» (4,33 и 4,40 соответственно, притом что значение группы C также имеет минимально допустимый уровень и равно 3,20). Здесь уместно прокомментировать, что в детских садах первых двух указанных групп производится ежегодная оценка деятельности персонала (общий

процент детских садов групп А, В и С, в которых данный индикатор отмечен положительно, равен 58,14 %). В процедуре оценки указываются как сильные, так и слабые стороны сотрудника, требующие улучшения (всего 41,86 % случаев), а также предпринимаются меры по реализации рекомендаций, сделанных на основе оценки деятельности сотрудников (всего 44,19 % случаев) — проводятся курсы повышения квалификации, при необходимости закупаются новые материалы и т. п.

Актуальные направления улучшения качества дошкольного образования

Еще раз подчеркнем, что целью экспертизы качества образования в детском саду с помощью адаптированных шкал ECERS-R является поиск актуальных направлений улучшения качества образовательной среды в детском саду. Результаты апробации приводят нас к выводу о том, что одним из важнейших аспектов в этом направлении является создание условий для повышения профессионализма воспитателей.

Одним из оснований для приведенного вывода послужило то, что при анализе двух групп в одной дошкольной образовательной организации контрольной группы С, работающих по одной образовательной программе и имеющих одни и те же ресурсы в организации предметно-пространственной среды, было обнаружено существенное различие в показателях по шкалам «Речь и мышление», «Взаимодействие», «Структурирование программы» (среднее значение — 2,00 и 6,00 баллов; 3,25 и 7,00 баллов; 2,67 и 7,00 соответственно). Содержание первых двух компонентов было достаточно детально описано нами выше. Обратимся к позиции «Структурирование программы»: в группе с более высокими баллами распорядок дня был более гибким, учитывались индивидуальные особенности детей, скорость процессов их жизнедеятельности (например, дети, которые уже переоделись после прогулки, слушают чтение сказки или обсуждают с воспитателем увиденное на улице), отводилось достаточное для потребностей детей время для свободных игр, педагоги наблюдали за детьми, чтобы помочь им в их играх, не навязывая свое мнение, групповая работа детей была организована по подгруппам с различными видами деятельности, причем дети имели возможность включиться в выбранную ими самими малую группу и т. д. Все эти позиции возможно организовать без специальных изменений в основной образовательной программе, без привлечения дополнительного кадрового или материального ресурса. В то же время их реализация становится возможной при высоком уровне профессионализма педагогов.

Таким образом, обобщая результаты, полученные в ходе апробации шкал ECERS-R, можно отметить следующие тенденции: с одной стороны, относительное благополучие в организации предметно-пространственной среды (наличие сообразной возрасту детей мебели, оборудования, игр и игрушек, пособий и дидактических материалов, обеспечения должной безопасности), а с другой — недостаточно оптимальное использование этих ресурсов, а также недостаточная готовность педагогов к организации педагогического взаимодействия с детьми, соответствующего ценностям, заявленным в ФГОС ДО.

Результаты апробации свидетельствуют о потенциальной эффективности использования шкал ECERS-R для обнаружения актуальных направлений развития как отдельного детского сада, так и всей системы дошкольного образования.

Математические показатели надежности и валидности результатов апробации шкал ECERS

В заключение укажем используемые в данном исследовании ECERS-R математические конструкции и критерии. В ходе апробации нам было важно оценить согласованность экспертных оценок, информативность параметров, а также наличие статистически значимых различий между контрастными группами.

Согласованность оценок экспертов

1. Обозначим для каждой пары экспертов, оценивающих одно и то же учреждение, $X = x_i$ — ответ (оценка) первого эксперта на i-й вопрос анкеты, а $Y = y_i$ — ответ (оценка) второго эксперта на i-й вопрос анкеты, где индекс i пробегает целые значения от 1 до N (общее число вопросов в анкете). Тогда в этих обозначениях используемый нами коэффициент парных корреляций Пирсона экспертных оценок (X, Y) принимает вид:

$$\rho(X,Y) = \frac{\sum_{i=1}^{N}(x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x}_i)^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}}.$$

Данный показатель не следует считать основным в оценке согласованности работы экспертов, так как близость коэффициента корреляции $\rho(X,Y)$ по модулю к единице свидетельствует всего лишь о наличии линейной связи между оценками экспертов:

$$Y = aX + b,$$

что еще не говорит о точном совпадении оценок:

$$Y = X.$$

Здесь возможны регулярные завышения или занижения баллов между экспертами. Данный показатель мы использовали как первичный индикатор разбалансированности оценок экспертов и уточняющий показатель вместе с количественными величинами, приводимыми ниже.

2. Доля точных совпадений K_0 и доля оценок с отличием в один балл K_1:

$$K_0 = \frac{N_0}{N}\ 100\ \%\ \text{и}\ K_1 = \frac{N_1}{N}\ 100\ \%,$$

где N, N_0 и N_1 — соответственно количество всех ответов, совпавших ответов и ответов с разницей в один балл. Средние доли совпадений в рассматриваемой работе оказались на уровне:

$$K_0 = 36{,}89\ \%,\ K_1 = 29{,}92\ \%\ \text{и}\ K_0 + K_1 = 66{,}81\ \%,$$

что хорошо согласуется с подобными исследованиями позиций ECERS-R[1].

3. Средний по всей анкете модуль расхождения в оценках двух экспертов, то есть модуль разности оценок экспертов на каждый вопрос анкеты в среднем:

$$R_A(X,Y) = \frac{\sum_{i=1}^{N} |x_i - y_i|}{N}.$$

Введенный выше показатель $R_A(X,Y)$ является индикатором разбалансированности в оценке конкретного учреждения парой экспертов.

4. Средний по всем экспертам модуль расхождения в оценке *i*-й позиции (вопроса) анкеты:

$$R_{E,i}(X,Y) = \frac{\sum_{j=1}^{M} |x_i(j) - y_i(j)|}{M},$$

где x_i (*j*) — оценка первого эксперта в *j*-й паре на *i*-й вопрос анкеты, y_i (*j*) — соответственно оценка второго эксперта в *j*-й паре на *i*-й вопрос; *M* — общее количество работавших пар экспертов.

$R_{E,i}(X,Y)$ — показатель, отвечающий за среднюю ошибку всех экспертов в данной позиции анкеты *i*. Последующее усреднение данного показателя по всем вопросам анкеты можно считать средней ошибкой экспертов по всему проекту. В рассматриваемом исследовании такое расхождение в оценках всех экспертов составило в среднем на каждый вопрос 1,23 балла, что является вполне приемлемым.

Информативность параметров

5. Общая информативность индикатора (общая различительная способность пункта анкеты):

$$Inform(A) = \frac{\min(M_0, M_1)}{\max(M_0, M_1)} 100\,\%,$$

где M_0 и M_1 — соответственно общее количество отрицательных (0) и общее количество положительных (1) ответов экспертов. Если все эксперты оказывались единодушны в оценке какого-либо конкретного индикатора, например, *A*, то его общая информативность была нулевой *Inform* (*A*) = 0 %, если мнения экспертов разделялись строго пополам — индикатор показывал максимальную дифференциацию *Inform* (*A*) = 100 %. Средняя общая информативность индикаторов в исследовании составила 33,35 %.

6. Доли положительных ответов (пиков) $K_+(A)$ и отрицательных ответов (провалов) $K_-(A)$ по индексам:

---

[1] Доли совпадений K_0 и K_1 по булевым (принимающим значения 0 или 1) индексам, каждой позиции ECERS-R получились более чем приемлемыми на уровне: K_0 = 76,95 % и K_1 = 23,05 % . В случае булевых значений индикаторов естественно сумма долей оказывается равна «единице»: $K_0 + K_1$ = 100 %.

$$K_+(A) = \frac{\sum_{i=1}^{N} x_i}{N} \ 100\ \% \text{ и } K_-(A) = 100\ \% - K_+,$$

где $x_i = \begin{cases} 0, \text{ если «нет»} \\ 1, \text{ если «да»} \end{cases}$ — ответ i-го эксперта по конкретному индикатору A.

Проверка статистических гипотез
о равенстве средних показателей ECERS-R

При проверке значимости различий в показателях ECERS-R был применен асимптотический t-критерий Стьюдента, используемый при квазинормальном распределении признака (в нашем случае средних значений оценок экспертов) при неизвестных и, вообще говоря, не равных между собой, генеральных дисперсиях. Обозначим: x_i — наблюдаемые оценки в первой группе (I); а y_i — во второй (II), $\bar{x}$ и $\bar{y}$ — их средние значения:

$$\bar{x} = \frac{\sum_{i=1}^{N_X} x_i}{N_X} \text{ и } \bar{y} = \frac{\sum_{i=1}^{N_Y} y_i}{N_Y},$$

σ (X) и σ (Y) — несмещенные выборочные оценки их среднеквадратичных отклонений:

$$\sigma(X) = \sqrt{\frac{\sum_{i=1}^{N_X} (x_i - \bar{x})^2}{N_X - 1}} \text{ и } \sigma(Y) = \sqrt{\frac{\sum_{i=1}^{N_Y} (y_i - \bar{y})^2}{N_Y - 1}},$$

а N_X и N_Y — соответственно количество экспертов, работавших в первой и второй группе. Тогда рассмотрим нулевую гипотезу о равенстве средних значений (математических ожиданий):

$$H_0\colon E(X) = E(Y)$$

против альтернативной гипотезы:

$$H_1\colon E(X) \neq E(Y)$$

на уровне значимости α = 0,05. Здесь использована альтернатива «≠», так как изначально нам неизвестно, какая из групп объективно лучше[2].

В этом случае наблюдаемая статистика:

$$t_{\text{набл.}} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma^2(X)}{N_X} + \frac{\sigma^2(Y)}{N_Y}}}$$

[2] В случае если, к примеру, первая группа объективно лучше, результаты сравнения можно усилить, выбрав одностороннюю альтернативу «>».

оказывается распределена по закону Стьюдента с m степенями свободы, где $m = \left[\frac{1}{r}\right]$ — целая часть от отношения $\frac{1}{r}$, а

$$r = \frac{\left(\frac{\frac{\sigma^2(X)}{N_X}}{\frac{\sigma^2(X)}{N_X} + \frac{\sigma^2(Y)}{N_Y}}\right)^2}{N_X - 1} + \frac{\left(\frac{\frac{\sigma^2(Y)}{N_Y}}{\frac{\sigma^2(X)}{N_X} + \frac{\sigma^2(Y)}{N_Y}}\right)^2}{N_Y - 1}.$$

Отличия считались значимыми в случае попадания наблюдаемого значения $t_{\text{набл.}}$ в критическую область

$$|\, t_{\text{набл.}} \,| \geq t_{\text{крит.}},$$

где $t_{\text{крит.}}$ — квантиль уровня $1 - \frac{\alpha}{2}$ для двусторонней альтернативной гипотезы; и соответственно в критическую область

$$t_{\text{набл.}} \geq t_{\text{крит.}},$$

где $t_{\text{крит.}}$ — квантиль уровня $1 - \alpha$ для односторонней альтернативы распределения Стьюдента с m степенями свободы[3].

Проверка статистических гипотез
о равенстве средних интегралов ECERS-R

При проверке значимости различий в интегралах ECERS-R, помимо указанного выше в общем случае асимптотического t-критерия Стьюдента, где это было возможно, был применен прямой t-критерий Стьюдента в предположении о равных неизвестных генеральных дисперсиях. Возможность его применения на одинаковость дисперсий проверялась критерием Фишера по частному наблюдаемых дисперсий $\frac{\sigma^2(X)}{\sigma^2(Y)}$. В этом случае вместо оценок $\sigma(X)$ и $\sigma(Y)$ в t-критерии использовалась объединенная оценка среднеквадратичного отклонения:

$$S^2 = \frac{(N_X - 1)\sigma^2(X) + (N_Y - 1)\sigma^2(Y)}{N_X + N_Y - 2},$$

а количество степеней свободы соответствовало

$$m = N_X + N_Y - 2.$$

[3] В данной статье мы не обсуждаем одностороннюю альтернативу и приводим дополнительную критическую область для полноты изложения, хотя группа А объективно и может быть признана лучшей.

ECERS-R продемонстрировал значительные различия между группами А и В/С в следующих шкалах: «Присмотр и уход за детьми», «Речь и мышление», «Взаимодействие и структурирование программы».

Таким образом, в результате проведенного исследования можно сделать вывод о возможности использования ECERS-R в качестве стандартизированного инструмента оценки и развития качества в российской системе дошкольного образования.

Литература

1. *Волосовец Т.В.* Преодоление общего недоразвития речи у дошкольников. М.: Сфера, 2008. 224 с.

2. *Шиян О.А., Воробьева Е.В.* Новые возможности оценки качества образования: шкалы ECERS-R апробированы в России // Современное дошкольное образование. Теория и практика. 2015. № 7. С. 38–49.

3. Шкалы для комплексной оценки качества образования в дошкольных организациях: переработанное издание / Тельма Хармс, Ричард М. Клиффорд, Дебби Крайер. М.: Национальное образование, 2016. 130 с.

4. *Фельдштейн Д.И.* Проблемы формирования личности растущего человека на новом историческом этапе развития общества // Образование и наука. 2013. № 9 (108). С. 3–23.

5. *Zachopoulou E., Grammatikopoulos V., Gregoriadis A., Gamelas A., Leal T., Pessanha M., Barros S., Liukkonen J., Loizou E., Henriksen C., Sanders Olesen L., Ciolan L.* Comparing aspects of the process quality in six europen early childhoog educational settings // 6th International Conference of Education, Research and Innovation. Conference abstracts. Seville. Spain, 2013.

6. *Sheridan S., Giotaa J., You-Me Hanb, Jeong-Yoon Kwonc.* A cross-cultural study of preschool quality in South Korea and Sweden: ECERS evaluations // Early Childhood Research Quarterly 24 (2009). P. 142–156.

7. *Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart. B.* (2004). The Effective Provision of Pre-School Education (EPPE Project). Technical Paper 12. London: Department for Education / Institute of Education, University of London.

8. *Tietze W., Cryer D., Bairrão J., Palacios J., Wetzel G.* Comparisons of observed process quality in early child care and education programs in five countries // Early Childhood Research Quarterly, Volume 11, Issue 4, December 1996. P. 447–475.

Literatura

1. *Volosovecz T.V.* Preodolenie obshhego nedorazvitiya rechi u doshkol'nikov. M.: Sfera, 2008. 224 s.

2. *Shiyan O.A., Vorob'yeva E.V.* Novy'e vozmozhnosti ocenki kachestva obrazovaniya: shkaly' ECERS-R aprobirovany' v Rossii // Sovremennoe doshkol'noe obrazovanie. Teoriya i praktika. 2015. № 7. S. 38–49.

3. Shkaly' dlya kompleksnoj ocenki kachestva obrazovaniya v doshkol'ny'x organizaciyax: pererabotannoe izdanie / Tel'ma Xarms, Richard M. Klifford, Debbi Krajer. M.: Izdatel'stvo «Nacional'noe obrazovanie», 2016. 130 s.

4. *Fel'dshtejn D.I.* Problemy' formirovaniya lichnosti rastushhego cheloveka na novom istoricheskom e'tape razvitiya obshhestva // Obrazovanie i nauka. 2013. № 9 (108). S. 3–23.

5. *Zachopoulou E., Grammatikopoulos V., Gregoriadis A., Gamelas A., Leal T., Pessanha M., Barros S., Liukkonen J., Loizou E., Henriksen C., Sanders Olesen L., Ciolan L.* Comparing aspects of the process quality in six europen early childhoog educational settings // 6th International Conference of Education, Research and Innovation. Conference abstracts. Seville. Spain, 2013.

6. *Sheridan S., Giotaa J., You-Me Hanb, Jeong-Yoon Kwonc.* A cross-cultural study of preschool quality in South Korea and Sweden: ECERS evaluations // Early Childhood Research Quarterly 24 (2009). P. 142–156.

7. *Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart. B.* (2004). The Effective Provision of Pre-School Education (EPPE Project). Technical Paper 12. London: Department for Education / Institute of Education, University of London.

8. *Tietze W., Cryer D., Bairrão J., Palacios J., Wetzel G.* Comparisons of observed process quality in early child care and education programs in five countries // Early Childhood Research Quarterly, Volume 11, Issue 4, December 1996. P. 447–475.

I.B. Shiyan, S.A. Zadadayev,
T.N. Le-van, O.A. Shiyan

Approbation of Scales of Quality Assessment of Pre-school Education (Early Childhood Environment Rating Scales — ECERS-R) in Moscow Kindergartens

This article presents the results of studies on the approbation of international Scales of Quality Assessment of Pre-school Education ECERS-R in Russian kindergartens. The aim of the study was to investigate the possibility of using ECERS-R scales as a quality assessment tool in the Russian system of preschool education.

Keywords: quality of pre-school education; ECERS-R scales; Assessment instrument; Russian system of pre-school education.