

**СЕРГЕЙ АНАТОЛЬЕВИЧ БЕЛЯКОВ, ВЛАДИМИР ЖАНОВИЧ КУКЛИН /  
SERGEY BELYAKOV, VLADIMIR KUKLIN**

## **СРАВНИТЕЛЬНАЯ ОЦЕНКА В ОБРАЗОВАНИИ: МЕТОДИКИ И ПРОЦЕДУРЫ**

## **COMPARATIVE EVALUATION IN EDUCATION: METHODS AND PROCEDURES**

### **Аннотация / Abstract**

В работе рассматриваются система сравнительной оценки, основанная на алгоритме ранжирования (формирования рейтинга), использующем порядковые шкалы, и модельные результаты применения процедуры для сравнительной оценки подсистем дошкольного и общего образования региональных образовательных систем. Представленные результаты получены в рамках выполнения научно-исследовательской работы «Проведение оценки состояния региональных систем образования и их сопоставление» в Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации. Приводится и обсуждается обобщенная модель системы сравнительной оценки, порядок ее функционирования и последовательность этапов формирования рейтинга с привлечением лиц, принимающих решения, экспертов по направлениям деятельности образовательных систем и специалистов по обработке и анализу информации. Приводятся результаты апробации предложенного подхода для сравнительной оценки региональных систем дошкольного и общего образования.

In this paper we consider a system of comparative evaluation based on the ranking algorithm (rating system) that uses ordinal scales and simulation results of the procedure for comparative evaluation of subsystems namely preschool and general education of regional educational systems. The presented results were obtained as an outcome of the research project “Comparative evaluation of regional educational systems and their comparison” at the Russian Presidential Academy of National Economy and Public Administration. We here propose and discuss a general model of comparative evaluation system, its functioning and a sequence of steps to form the rating involving decision makers, experts in the areas of educational systems and specialists in data processing and analysis. Finally, we give the testing results of the proposed approach for the comparative evaluation of regional systems of pre-school and general education.

**Ключевые слова:** системы сравнительной оценки, ранжирование,  
региональные системы образования, ранговые процедуры  
сравнительной оценки

**Key words:** system of comparative evaluation, ranking, regional educational  
systems, ranking procedures of comparative evaluation

В последние годы процедуры сравнительной оценки – «мониторинги» и рейтинги – из методов косвенного информационного воздействия на участников образовательной деятельности превратились в механизмы принятия управленческих решений. При этом из всего спектра существующих процедур сравнительной оценки наибольшее распространение получила процедура формирования рейтинга на основе «взвешенной суммы нормированных показателей (индикаторов)».

Известно, что система образования в целом и ее подсистемы (региональные системы образования, образовательные организации, другие учреждения системы образования) являются, в терминологии системного анализа, сложными адаптивными саморегулирующимися системами с целеполаганием [1; 2]. При этом цели системы в целом и цели подсистем различного уровня в большинстве случаев различаются. Поэтому достижение своих целей подсистемами более высокого уровня иерархии не может быть достигнуто за счет прямого (административного, директивного) управления и требует использования более тонких механизмов целенаправленного воздействия на управляемые объекты. В сложных системах средством согласованного достижения целей выступает целенаправленное формирование среды функционирования подсистем нижних уровней иерархии, в которой цели подсистем оказываются согласованными с целями системы в целом.

В настоящее время в число механизмов «надзора, контроля» и управленческого воздействия на образовательные системы входят разнообразные системы сравнительной оценки – рейтинговые системы, системы мониторинга с рейтинговой составляющей и процедуры сопоставительного анализа. При этом во многих случаях результаты сравнительной оценки не только становятся известными широкому кругу заинтересованных потребителей указанной информации, но и начинают использоваться в качестве оснований для принятия управленческих решений на различных уровнях управления, вплоть до федерального уровня.

В этих условиях задача анализа существующих методик и процедур, определения условий их адекватного использования, формирования требований к обоснованности и корректности их применения, а также условий и форм использования результатов выходит на первый план.

В традиционной постановке задача формирования интегральных показателей деятельности обычно рассматривается в контексте процедуры метрического измерения – процедуры, основанной на обработке входной количественной (числовой) информации в выходной результат – число, используемое далее как количественная мера качества деятельности. При этом для интеграции первичных характеристик в интегральные показатели используется алгоритм аддитивного взвешивания, в котором учет относительной значимости частных показателей для интегральной оценки осуществляется через выбор весовых коэффициентов. Такой подход, показавший свою эффективность в задачах оценивания технических систем, последние годы начинает активно использоваться и в системах интегральной сравнительной оценки сложных социальных и социально-экономических систем. При этом неявно предполагается, что выполняются условия корректности использования процедуры метрического измерения. Однако

такой подход ограничен множеством метрических показателей и не может использоваться в ситуациях, когда частные рейтинги получены в порядковых шкалах. Обычно только часть первичных показателей имеет метрический характер, то есть позволяет количественно сравнивать значения показателей в шкале интервалов. Однако одновременно с метрическими показателями часто используются показатели, измеренные в порядковой шкале (в первую очередь это относится к результатам экспертного оценивания). В связи с этим представляет интерес использование методов ранжирования, позволяющих строить интегральное ранжирование в случаях, когда часть показателей имеют порядковый характер [3; 4].

Задача формирования рейтинга как интегральной характеристики деятельности объекта является типичной задачей определения отношения порядка на множестве объектов при заданном множестве частных показателей деятельности. Сложность формирования интегральной характеристики определяется:

- разнородностью показателей, измеряемых в различных шкалах;
- сложностью оценки и учета относительной значимости отдельных показателей (большим разбросом «весов» показателей).

В то же время существенным для принятия решений фактором во многих случаях является именно порядок на множестве объектов, при том что конкретное значение интегральной характеристики не может быть содержательно интерпретировано в терминах показателей деятельности сравниваемых объектов.

В силу этого целесообразно в качестве основы для определения рейтинга объекта рассматривать его место (порядковый номер) в упорядоченной последовательности объектов. Основной сложностью при формировании рейтинга является необходимость учета относительной значимости разнородных частных показателей. Наиболее распространенный в настоящее время подход, исторически связанный с «технократическим» восприятием показателей как элементов метрического пространства и основанный на вычислении взвешенной суммы или взвешенного произведения, обеспечивает учет относительной значимости показателей за счет нормирования значений показателей с последующим определением соответствующих весовых коэффициентов.

- Такой подход имеет ряд недостатков, среди которых выделим следующие:
- невозможность отражения немонотонной (уни- или полимодальной) зависимости значения рейтинга от значения частного показателя;
  - сложность интерпретации весовых коэффициентов в терминах рассматриваемой проблемной области и обусловленная этим сложность обоснования достоверности полученных результатов;
  - сложность априорного определения весовых коэффициентов и связанная с этим необходимость моделирования результатов ранжирования с использованием реальных данных;
  - необходимость выбора либо вычислительной модели, неустойчивой относительно вариации значений частных показателей и весовых коэффициентов, либо нарушения принципов согласованного выбора Эрроу [3].

На основании анализа обобщенной модели системы сравнительной оценки, включающего исследование порядка ее функционирования и последовательности этапов формирования рейтинга, с учетом участия в этом процессе лиц, принимающих решения (ЛПР), экспертов по направлениям деятельности образовательных систем и специалистов по обработке и анализу информации, для исключения части из приведенных выше недостатков, предлагается использовать ранговую процедуру формирования рейтинга на основе следующего алгоритма:

- для каждого показателя определяются число групп для расслоения (слоев) и критериальные (пороговые) значения показателя для разделения слоев;
- множество объектов упорядочивается по каждому частному показателю, формируются частные рейтинг-листы, при этом при одинаковом значении показателя объекты получают одинаковое место (т. е. строится отношение нестрогого порядка);
- для каждого объекта исходного множества строится вектор групповых номеров;
- исходное множество упорядочивается на основе векторов групповых номеров (содержательно алгоритм можно описать как «выбор лучших из оставшихся»: на каждом шаге из исходного множества выбираются объекты с самым высоким рейтингом, которые формируют очередную группу в итоговом рейтинге);
- при необходимости полученное упорядочение (в большинстве практических случаев частичное) еще раз ранжируется внутри каждого из полученных классов.

Иначе говоря, формирование итогового рейтинг-листа на основе предложенного алгоритма распадается на два этапа:

- на первом производится расслоение множества на несколько подмножеств;
- на втором этапе внутри полученных подмножеств определяется строгий порядок элементов, если это необходимо.

В общем случае алгоритм не накладывает ограничения на число слоев, однако из теоретических соображений и практики использования методов ранжирования целесообразно использовать от четырех до семи слоев для каждого частного показателя.

Заметим, что в ряде случаев необходимость во втором этапе отпадает. Применительно к задачам, возникающим в образовательных системах, можно отметить, что необходимость во втором этапе возникает, как правило, только для объектов, попавших в «группу лучших» либо в «последнюю группу».

Приведем последовательность основных этапов формирования сравнительной оценки на примере региональных подсистем дошкольного образования и подсистем общего образования субъектов Российской Федерации. Более детальное обсуждение полученных результатов будет представлено в докладе на конференции, где будут рассмотрены результаты моделирования ранжирования регионов по состоянию региональных подсистем дошкольного образования – по четырем вариантам выбора

количества слоев, с использованием экспертной коррекции критериальных значений показателей, в том числе:

- с разбиением показателей на 4 слоя с определением критериальных (граничных) значений для слоев на основе порядковых статистик (базовый вариант);
- с использованием экспертной коррекции;
- количества слоев разбиения (от 4 до 7);
- критериальных (граничных) значений для слоев на основании содержательного анализа информационных массивов.

На подготовительном (нулевом) этапе определяются источники первичной информации – для рассматриваемого примера это статистические данные, представленные в информационной системе Росстата:

*d1* – численность детей в дошкольных образовательных организациях (ДОО);

*d2* – число мест в ДОО;

*d3* – численность педагогических работников ДОО;

*d4* – численность детей, стоящих на учете для определения в ДОО.

На первом этапе с участием ЛПР и экспертов в сфере образования определяются показатели, характеризующие состояние рассматриваемых подсистем (с учетом доступности, достоверности и полноты соответствующей первичной информации):

*P1* – доля детей, стоящих на учете для определения в ДОО, из общей численности детей, заинтересованных в услугах ДОО (по показателю *P1* значение 0 является лучшим: соответствует отсутствию очереди в ДОО);

*P2* – численность педагогических работников ДОО, отнесенная к численности детей;

*P3* – число мест в ДОО, отнесенное к общей численности детей, заинтересованных в услугах ДОО;

*P4* – численность детей в ДОО, отнесенная к общей численности детей, заинтересованных в услугах ДОО (этот показатель можно рассматривать как интегральный показатель удовлетворения потребности населения в услугах ДОО).

На втором этапе при участии экспертов в сфере образования и специалистов в сфере обработки и анализа информации определяется степень статистической независимости показателей состояния (для рассматриваемого случая она оказывается достаточной для продолжения процесса – корреляция между показателями *P1-P4* лежит в допустимых пределах, в противном случае могла бы возникнуть необходимость в возвращении к первому этапу и пересмотру или коррекции его результатов).

На третьем этапе эксперты в сфере образования определяют по каждому показателю число слоев (групп, на которые разбиваются объекты по данному показателю) и критериальные (пороговые) значения, разделяющие группы. Определение порогов проводится в несколько итераций. В случае наличия содержательных требований пороги определяются из этих требований, и на основе первичного моделирования, при необходимости, корректируются. В случае отсутствия подобных требований на первой итерации используются порядковые статистики – для варианта «4 слоя» – квартили соответствующего

распределения, которые далее корректируются на основе первичного моделирования. Далее строятся частные рейтинг-листы по каждому из параметров.

На четвертом этапе с участием ЛПР и экспертов в сфере образования для формирования интегрального рейтинга определяется число слоев и критериальные значения для разбиения объектов в интегральном рейтинге. С использованием информационно-программных средств производится моделирование результатов для нескольких вариантов сравнительной оценки, анализируется устойчивость результатов ранжирования и их содержательная валидность.

На пятом, завершающем, этапе проводится анализ результатов сравнительной оценки и формируются рекомендации для объектов, расположенных в «нижней» части интегрального рейтинга, по возможным направлениям улучшения их позиции в рейтинге. Разработанный инструментальный визуальный представление частных рейтинг-листов и интегрального рейтинга позволяет представить необходимую для принятия решений информацию в компактном виде.

При необходимости этапы 3 и 4 проводятся повторно, например, для выбора другого числа слоев по одному или нескольким показателям, либо для иного набора критериальных (пороговых) значений. Следует отметить, что использование подобной вариативности для анализа позволяет дать более обоснованные рекомендации для объектов ранжирования.

Практика использования предложенного подхода на реальных и специально сконструированных тестовых данных показывает, что при незначительных ограничениях на характер элементов множества параметров и корректном выборе процедуры формирования итогового ранжирования имеет место устойчивость – «малое изменение точек расслоения по любому частному показателю ведет к малому изменению итогового рейтинга». При более сильных ограничениях на характер исходных данных можно получить даже «поглощение» малых ошибок, т. е. неизменность итогового рейтинга при малых изменениях значений порогов частного показателя либо малой вариации значений показателей объектов исходного множества. Анализ процедуры и результатов апробации позволяет сделать вывод, что использование предложенного подхода для сравнительной оценки образовательных систем обеспечивает условия оперативного и объективного сравнения, а также позволяет определить направления работ по улучшению состояния конкретного объекта сравнения. Следует отметить также, что еще одним существенным отличием предложенного варианта является принципиальная возможность учитывать немонотонные изменения показателей (случай, когда начиная с некоторого значения большее значение показателя соответствует меньшему значению качества).

В рамках продолжения исследований предполагается рассмотреть задачу разработки системы поддержки принятия решений по сравнительной оценке образовательных систем в форме информационно-аналитического комплекса, и в первую очередь – «Автоматизированного рабочего места эксперта» как основного элемента этой системы.

## Литература / References

1. Бурков, В. Н., Новиков, Д. А. Модели и механизмы теории активных систем в управлении качеством подготовки специалистов. – М. : ИЦ, 1998.
2. Дружинин, В. В., Конторов, Д. С. Проблемы системологии. – М. : Сов. радио, 1976.
3. Литвак, Б. Г. Экспертная информация: Методы получения и анализа. – М. : Радио и связь, 1982.
4. Пфанцаль, И. Теория измерений / пер. с нем. – М. : Мир, 1976.

Сергей Анатольевич Беляков  
доктор экономических наук, доцент  
Российская академия народного хозяйства и государственной службы при  
Президенте Российской Федерации  
Москва, Россия

Sergey Belyakov  
PhD in Economics (Doctor of Science), Associate Professor  
The Russian Presidential Academy of National Economy and Public Administration  
Moscow, Russia

Владимир Жанович Куклин  
доктор технических наук, доцент  
Российская академия народного хозяйства и государственной службы при  
Президенте Российской Федерации  
vzh.kuklin@gmail.com

Vladimir Kuklin  
PhD in Technical Sciences (Doctor of Science), Associate Professor  
The Russian Presidential Academy of National Economy and Public Administration  
Moscow, Russia