

УДК 378+33

## КЛАССИФИКАЦИЯ МЕТОДОВ ОЦЕНКИ КАЧЕСТВА ОБРАЗОВАТЕЛЬНЫХ УСЛУГ

Н.С. БУГРО, В.Н. ЯРМОЛИК

*Белорусский государственный университет информатики и радиоэлектроники  
П. Бровка, 6, Минск, 220013, Беларусь*

*Поступила в редакцию 18 марта 2013*

Приведено обоснование актуальности задачи оценки качества образовательных услуг высших учебных заведений, выделена ее подзадача – оценка входных метрик. Для данной подзадачи приведен обзор классификации методов оценки (шкал), на основе которого сделано заключение о наиболее подходящей шкале для выделенной подзадачи.

*Ключевые слова:* оценка образовательных услуг, шкалы, методы оценки.

### Введение

В настоящее время в области образовательных услуг высших учебных заведений большое влияние приобретает стремительно развивающаяся конкуренция. Современное высшее образование – это глобальный продукт, предлагаемый различными образовательными учреждениями по всему миру. Их многообразие и широкий спектр предлагаемых ими образовательных услуг способствуют развитию конкуренции как за уровень подготовки абитуриентов, так и за их количественный состав. Для этого предлагаются новаторские, творческие подходы для удовлетворения потребностей и предпочтений студентов, как в части качества образовательных услуг, так и сопутствующих факторов.

В то же время студенты перестают ориентироваться исключительно на местные учреждения образования, в большей степени обращая внимание на мировой уровень. Экономические, социальные и технологические достижения способствуют постоянному улучшению учебного процесса в организациях, университетах и школах, и студенты ожидают получить соответствующий уровень образования.

Возможность определения уровня предоставляемого образования и степени удовлетворенности им студентов являются одними из ключевых факторов для обеспечения конкурентоспособных услуг в области высшего образования. Данные возможности позволяют оценить существующие учреждения образования, выявить их слабые стороны и, имея данную информацию, разработать стратегию эффективного повышения качества образовательных услуг.

Оценка входных данных – это одна из подзадач при определении уровня образовательного сервиса. При оценке данных возможно использование различных подходов, которые могут быть разделены на две категории: относительные методы оценки (comparative scale techniques) и абсолютные способы оценки (non-comparative scale techniques) [1, 2].

При относительных подходах к оценке респондента просят сравнить один объект (продукт, сервис, характеристику) с другим [1, 2]. Абсолютные методы базируются на непрерывных или детализированных шкалах оценок. Классификация методов оценки представлена ниже на рис. 1.

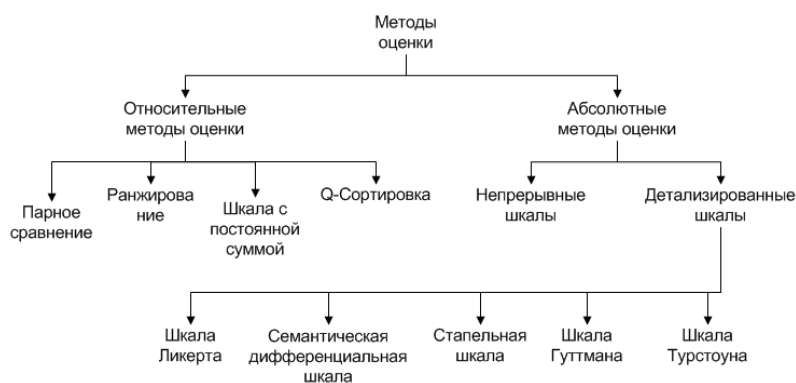


Рис. 1. Классификация методов оценки

### Относительные методы оценки

Как видно из рис. 1, относительные методы оценки могут быть разделены на следующие четыре типа оценок: оценка парным сравнением, оценка ранжированием, оценка с постоянной суммой и оценка Q-сортировкой [3].

*Парное сравнение.* Оценка парным сравнением (Paired Comparison Scale) – это относительный способ оценки, в котором респонденту представляют два объекта и его просят выбрать согласно заданного критерия один из них [4]. Например, А, В, С, D и E – 5 продуктов одного типа. Для каждого продукта респондент имеет разный уровень привлекательности, поэтому в каждой паре продуктов он может предпочесть один. Он, например, может предпочесть А другим продуктам. В общем случае, при наличии N продуктов получается  $N(N-1)/2$  сравнений. Полученные в результате сравнений данные могут быть представлены в виде таблицы размерностью  $N*N$ . В случае нашего примера результаты сравнения могут быть представлены в таблице.

Таблица 1. Пример таблицы, полученной в результате парных сравнений

Продукт	A	B	C	D	E
A	-	-	-	+	-
B	+	-	-	-	+
C	+	+	-	-	+
D	-	+	+	-	-
E	+	-	-	+	-
Количество предпочтений	3	2	1	2	2

Символ «+» означает, что продукт в колонке имеет предпочтение продукту представленному в строке. Количество раз, которое заданный продукт предпочтен, равно количеству знаков «+» в заданной колонке. Другой способ получения данных с помощью парного сравнения – получение пропорции предпочтения одного продукта другому.

Таблица 2. Пример таблицы пропорций предпочтения, полученной в результате парных сравнений

Продукт	A	B	C	D	E
A	-	10	45	69	27
B	90	-	20	16	85
C	55	80	-	30	90
D	31	84	70	-	47
E	73	15	10	53	-

Значения в ячейках отражают количества респондентов в процентах, которые предпочли тип продукта в колонке типу продукта в строке. К примеру, 90 % респондентов предпочли тип продукта А типу продукта В, и 10 % предпочли тип продукта В типу продукта А. Парное сравнение удобно, когда количество типов продуктов ограничено и пока необходимо прямое сравнение.

*Ранжирование.* Оценка ранжированием (Rank Order Scaling) – это относительный способ оценки, в котором респонденту представляют несколько объектов (продуктов,

сервисов) одновременно и просят его отсортировать или ранжировать их согласно заданного критерии [1]. Сортировка или ранжирование могут производиться на основе предпочтений, привлекательности, важности, эффективности и др. На основе предыдущего примера, респондента можно попросить отсортировать продукты А, В, С, D и Е, согласно его предпочтений. Начиная с выбора самого привлекательного типа продукта, респондент присваивает ему номер 1. Далее он находит второй тип продукта по степени предпочтительности для него и т.д. В итоге получаем следующие данные.

Таблица 3. Результаты ранжирования

Продукт	A	B	C	D	E
Место	1	4	5	3	2

Важно, чтобы два типа продукта не были на одном уровне. При оценке ранжированием должно быть принято только  $(N-1)$  решение

*Шкала с постоянной суммой.* В шкале с постоянной суммой (Constant Sum Scaling) [1] респондентам дается некоторая сумма единиц, таких как баллы, деньги, кредиты и далее респондентов просят распределить их между различными объектами. Например, вы можете попросить респондента отразить уровень предпочтения, важности или другой характеристики продукта и дать ему 100 баллов для распределения между различными продуктами. Если характеристика какого-либо продукта не важна для респондента – он может ассоциировать с ним даже 0 баллов. Если один продукт в два раза значимее для респондента, чем другой – он может ассоциировать с ним в два раза больше баллов, чем с другим. Когда распределение баллов закончено – сумма баллов, ассоциированных со всеми продуктами, должна равняться 100. Для нашего предыдущего примера могут быть получены следующие результаты.

Таблица 4. Результаты при использовании шкалы с постоянной суммой

Продукт	A	B	C	D	E
Уровень	40	6	15	19	20

Соотношение данных, полученное в ходе оценки с постоянной суммой, является самым мощным, поскольку позволяет охарактеризовать оценку ассоциированным с ней количеством баллов, что позволяет оценить различия между альтернативами.

*Q-сортировка.* Оценка Q-сортировкой (Q-Sort Scaling) – это относительный тип оценки, при котором респондента просят отсортировать значительное количество объектов или утверждений и классифицировать их на predetermined небольшое количество множеств согласно некоторого критерия, такого как предпочтительность, положение или поведенческие особенности. В целях повышения статистической достоверности, должно использоваться от 60 до 140 объектов для классификации по множествам (обычно не более 11 множеств). Этот подход хорош для быстрой обработки больших групп позиций.

### Абсолютные методы оценки

Абсолютные методы оценки (non-comparative scales) часто называют моно-дическими оценками (monadic scales), т.к. одновременно оценивается только один объект [1, 5, 6]. При таком подходе отсутствуют сравнения с другими объектами. Абсолютные методы оценки делятся на два класса: оценки с непрерывной шкалой (Continuous Rating Scales) и оценки с детализированной шкалой (Itemized Rating Scales) [3].

При оценке с непрерывной шкалой респондентов просят дать оценку путем выставления пометки на непрерывной линии. Он помечает, например знаком «\*», ощущаемый уровень характеристики конкретного продукта от негативного «Очень плохо» до позитивного «Очень хорошо». Шкала может быть напечатана на карте и показана респонденту в ходе интервью. На рис. 2 изображены два варианта непрерывных шкал.

При использовании первой шкалы, метка респондента делит линию на 2 части, определяя оценку соотношением пунктов в ее частях, либо длинами этих отрезков. В любом случае результат определяется выбранным интервалом.



дифференциальная шкала одна из наиболее широко используемых техник, используемых при оценке уровней величин.

Биполярные пары прилагательных могут использоваться для широкого круга объектов, и эта техника удобна для определения оценки человека. Недостатком семантических дифференциальных шкал является сложность определения соответствующих пар биполярных прилагательных, требуемых для создания шкалы.

*Стапельная шкала.* Стапельная шкала (Stapel Scale) предлагает респонденту оценить продукт или сервис согласно заданной характеристики в интервале от  $-5$  до  $+5$ , отражая насколько хорошо характеристика отражает продукт или сервис [3]. Обычно шкала изображается вертикально, с прилагательным, описывающим характеристику, по середине с вариантами от  $-5$  до  $+5$ . Пример данной шкалы представлен ниже.

- |   |   |
|---|---|
| <input type="radio"/> (+5) <i>Отлично описывает</i> | <input type="radio"/> (+5) <i>Отлично описывает</i> |
| <input type="radio"/> (+4)                          | <input type="radio"/> (+4)                          |
| <input type="radio"/> (+3)                          | <input type="radio"/> (+3)                          |
| <input type="radio"/> (+2)                          | <input type="radio"/> (+2)                          |
| <input type="radio"/> (+1)                          | <input type="radio"/> (+1)                          |
| <input type="radio"/> <i>Инновационный</i>          | <input type="radio"/> <i>Передовой</i>              |
| <input type="radio"/> (-1)                          | <input type="radio"/> (-1)                          |
| <input type="radio"/> (-2)                          | <input type="radio"/> (-2)                          |
| <input type="radio"/> (-3)                          | <input type="radio"/> (-3)                          |
| <input type="radio"/> (-4)                          | <input type="radio"/> (-4)                          |
| <input type="radio"/> (-5) <i>Ужасно описывает</i>  | <input type="radio"/> (-5) <i>Ужасно описывает</i>  |

Рис. 4. Стапельная шкала

Согласно данной шкале, респондент не может выбрать нейтральный ответ. Это создает некоторые затруднения при практической реализации данного подхода.

*Шкала Гутмана (Guttman scale)* [9]. Состоит из подмножества вопросов, имеющих возможные ответы «Да» и «Нет», отсортированных таким образом, что при положительном ответе на один из них все последующие также имеют положительный ответ. Это значит, что при согласии с одним утверждением, респондент автоматически соглашается со всеми нижеследующими. Этот подход значительно отличается от шкалы Ликерта.

*Шкала Турстоуна.* Одна из первых и самых продуктивных теорий оценки была предложена Луисом Турстоуном (Thurstone) [10]. На самом деле он изобрел 3 различных метода разработки одномерных шкал Турстоуна: метод равно-фигурирующих интервалов, метод успешных интервалов и метод парных сравнений [10, 11]. Три метода отличаются по способу конструирования значений шкал, но во всех трех случаях полученная шкала оценивается респондентами схожим образом.

## Заключение

Одной из подзадач при разработке методов оценки образовательных услуг является выбор входных метрик. Для их оценки необходимо определить шкалу, на которую отражаются их возможные значения.

Поскольку относительные шкалы позволяют получить оценку только в рамках заданного при сравнении множества вариантов, их использование накладывает ограничения на использование всей методики оценки образовательных услуг. Данное ограничение заключается в невозможности объединения результатов, полученных по итогам независимых оценок.

Вышеописанного недостатка не имеют абсолютные методы оценки. Наиболее предпочтительной выглядит шкала Ликерта, т.к. она обладает следующими преимуществами: подходит для оценки различных метрик, т.к. не требует подбора каких-либо пар терминов, описывающих крайние положения шкалы; содержит оптимальное количество вариантов ответа (человеческая кратковременная память, как правило, не может одновременно манипулировать более чем  $7 \pm 2$  элементами [12]); не требует большого количества времени для ответа; получила широкое распространение на данный момент, привычна респондентам и проста для их понимания.

# EDUCATIONAL SERVICES QUALITY ASSESSMENT METHODS CLASSIFICATION

N.S. BUGRO, V.N. YARMOLIK

## Abstract

Comprehensive analyses of different scaling techniques to be used for high education service quality estimation are provided. The most well-known scaling techniques have been described with simple illustrating examples. The Likert scaling approach has been pointed out as the more suitable approach for high education service quality estimation.

## Список литературы

1. DeVellis Robert F. Scale development: theory and applications. London, 2003.
2. Lodge Milton. Magnitude scaling: quantitative measurement of opinions. London, 1981.
3. Sridhar M.S. // Measurement and Scaling Techniques. 2007.
4. Krus D.J., Kennedy P.H. // Educational and Psychological Measurement. 1977. P. 189–193.
5. Stevens Stanley Smith // Science. 1946. P. 677–680.
6. Narens L. // Journal of Mathematical Psychology. 1981. P. 249–275.
7. Likert Rensis // Archives of Psychology. 1932. P. 1–55.
8. Osgood C.E., Suci G., Tannenbaum P. The measurement of meaning. Urbana, 1957.
9. Guttman L. // The American Soldier. New York, 1950.
10. Thurstone, L.L. // Journal of Abnormal and Social Psychology. 1928. № 21. P. 384–400.
11. Thurstone L.L. // American Journal of Sociology. 1928. № 33. P. 529–554.
12. George A. Miller // The Psychological Review. 1956. Vol. 63. P. 81–97.