# Существующие подходы к коллегиальному оцениванию. Достоинства и недостатки.

## Основные термины

МООС-платформа – сайт (платформа), на которой размещаются массовые открытые онлайн-курсы.

Обучающийся – персона, проходящая обучение на курсе.

Грейдер – обучающийся, который оценивает работы, выполненные другими обучающимися.

Рубрики – это руководство, в котором перечислены конкретные критерии оценки данного задания и уровни соответствия данному критерию с указанием начисляемых баллов за соответствие этому критерию.

Инструктор – организатор курса, который ведет лекции на курсе и проверяет задания, выполненные обучающимися.

Относительная шкала оценивания – это шкала оценивания, на основании которой оценка за работу обучающегося вычисляется относительно лучшей работы в группе.

## 1.2 Количественное коллегиальное оценивание

Количественное коллегиальное оценивание – это метод оценивания заданий, суть которого заключается в том, что оценка за выполненную работу является численным количественным значением. В настоящее время данный метод является наиболее распространенным в МООС-платформах.

Как правило, количественный подход к оценке работ используется вместе с рубриками.

К таким методам относятся: PeerRank[ссылка], Vancouver Algorythm[ссылка]

Преимущества данного метода:

1. Возможность вводить рубрики, поясняющие тот или иной критерий оценивания. Такая возможность крайне необходима для упрощения процесса выставления баллов за задание, потому что большинство обучающихся не имеют нужных навыков выставления оценок. Так же многие студенты склонны ставить более высокие отметки своим сокурсникам. Исходя из вышесказанного, рубрики повышают объективность процесса рецензирования.
2. Инструктор получает больше информации о выполненных работах, так как использование рубрик помогает ему оценить то, какие знания и навыки студенты усвоили хуже, чтобы впоследствии на лекциях уделить этому вопросу большее внимание. В целом, данных подход увеличивает обратную связь между студентами и инструкторами. По большей части это заслуга рубрик, которые идеально подходят для их совместного использования с количественным подходом.
3. Использование абсолютной шкалы оценки заданий. Это означает, что оценки не зависят от качества работ студентов в этой группе. То есть если все студенты выполнили задание не удовлетворительно, то все они получат плохие оценки (баллы высчитываются не относительно лучшей работы в группе, как в порядковом подходе).

Недостатки данного метода:

1. Большая предвзятость у оценщиков. Для более-менее объективного оценивания нужен достаточный опыт в рецензировании работ. Каждый оценщик по-разному может оценить одну и ту же работу.
2. Абсолютная шкала оценки сама по себе довольно сложна для освоения. Людям по большей части гораздо проще сравнивать две работы и выявлять, какая из них лучше, а какая хуже (такой подход используется в порядковом методе коллегиального оценивания).

## Порядковое коллегиальное оценивание

Порядковое коллегиальное оценивание – это метод оценивания, основанный на попарном сравнении работ, который осуществляет ранжирование всех выполненных студентами заданий. Данный подход менее распространен и на практике он не применяется, однако, существует несколько работ, которые рассматривают данный метод в теории и даже приводят формализованную математическую модель этого порядкового подхода.

К таким методам относятся: Bayesian Peer Grading[ссылка], Mallows Model[], Bradley-Terry Model[].

Преимущества данного подхода:

1. Не требует от оценщика каких-либо специальных навыков оценивания. Для человека гораздо проще указать, что проект А хуже или лучше проекта Б, чем выставлять баллы в абсолютной шкале.
2. В некоторых математических моделях, которые использует в порядковом коллегиальном оценивании, есть специальные методы, позволяющие выявить недобросовестного оценщика.
3. Порядковый подход к оценке работ снижает нагрузку на оценщика, ему не приходится тратить дополнительное время на выставление баллов по абсолютной шкале. Следовательно, такой подход уменьшает вероятность того, что в какой-то момент ему надоест оценивать назначенные ему работы, и он выставит оценки наугад. Благодаря этому повышается качество обратной связи от оценщиков.
4. Так как люди лучше справляются со сравнением предметов, оценщик может быть в состоянии предложить более проницательные замечания по положительным и отрицательным аспектам решения студента.

Недостатки данного подхода:

1. Оценка, полученная при использовании порядкового метода, выражается в перцентилях (т.е. оценка вычисляется относительно самой лучшей работы в группе). Такая оценка несет в себе намного меньше информации о работе студента, так как оценивает не его знания, а его рейтинг в группе.
2. Необходимость переводить оценки в баллы по абсолютной шкале, потому что во всех МООС-платформах используются баллы в качестве вознаграждения за каждое выполненное задание.
3. Порядковая оценка содержит в себе меньше информации, как для инструктора, так и для оценщика. Чаще всего количественные рубрики более информативны и позволяют выяснить то, как обучающийся освоил тот или иной навык.
4. Не подходит для оценивания объемных работ таких как, эссе на более чем 150 слов и т.д.
5. Обучающимся гораздо проще оценивать, используя рубрики, т.к. они помогают прояснить некоторые детали процесса рецензирования.
6. Каждый оценщик охватывает лишь небольшую часть упорядочиваний работ.

## Сравнение количественного и порядкового подхода к коллегиальному оцениванию

В таблице 1.1 приведено сравнение количественного и порядкового подходов к коллегиальному оцениванию.

Таблица 1.1 Сравнение количественного и порядкового подходов

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Метод** | **Простота оценивания для студента** | **Методы определения недобросовестных оценщиков** | **Относительная шкала оценивания** | **Возможность использования рубрик** | **Подходит для оценивания объемных работ** | **Предполагает стимулы для грейдеров оценивать корректно** |
| **Количественное** | Требует навыков оценивания | Нет | Нет | Да | Да | Нет |
| **Порядковое** | Просто и интуитивно | Да | Да | Нет | Нет | Нет |

Исходя из этой таблицы, можно сделать вывод, что количественный метод больше подходит для большинства практических задач из-за того, что он предполагает использование рубрик, подходит для оценивания достаточно объемных работ (например, эссе на 150 слов).

Порядковый метод так же может использоваться для ряда заданий с использованием коллегиального оценивания, но из-за того, что с помощью него нельзя вычислить оценку по абсолютной шкале, и, учитывая, что в МООС-платформах обычно используется абсолютная шкала оценивания, применять этот подход будет не целесообразно.

Несмотря на то, что для коллегиального оценивания больше подходит количественный подход, в нем присутствуют несколько недостатков. Количественный подход никак не стимулирует грейдеров оценивать корректно, в нем отсутствуют методы наказания недобросовестных грейдеров, и так же в большинстве алгоритмов, которые высчитывают итоговую консолидированную оценку, полученную с помощью количественного подхода, используется среднее арифметическое всех оценок, которые грейдеры поставили за работу. Недостаток, связанный с вычислением итоговых баллов, как среднего от всех оценок, является самым существенным. Он сильно влияет на погрешность вычисления оценки обучающегося.

## PeerRank

PeerRank – алгоритм, который совершенствует количественный подход к коллегиальному оцениванию. Он был предложен Тоби Уолшем, профессором, изучающим искусственный интеллект, из австралийского Университета Нового Южного Уэльса[]. Название PeerRank было выбрано из-за того, что принцип работы алгоритма похож на принцип работы алгоритма ссылочного ранжирования PageRank[].

С помощью этого алгоритма оценка обучающегося вычисляется на основе оценок, которые ему поставили другие обучающиеся. При чем этот метод учитывает те оценки, которые поставил студент другим студентам, поскольку оценка – мера способности обучающегося правильно оценивать работы. Так же этот метод обеспечивает стимулы для студентов ставить оценки корректно, так как его собственная оценка зависит от оценок, которые он выставил другим обучающимся. Метод в том числе увеличивает оценку студентам, которые добросовестно подходят к процессу оценивания и уменьшает оценку тем студентам, которые недобросовестно оценивают назначенные им работы.

Преимущества данного алгоритма:

1. Более точное вычисление оценки, по сравнению с обычными методами вычисления консолидированной оценки, где просто вычисляется среднее между баллами, которые выставили оценщики за работу.
2. Обеспечение стимулов для студентов, для того чтобы они стремились оценивать лучше и подходили к этому процессу более ответственно. Если студент выставляет слишком заниженные баллы за работы других студентов, то его собственная оценка так же страдает от этого.
3. Использование абсолютной шкалы оценивания. Это означает, что оценка не зависит от лучшей работы в группе.
4. Данный алгоритм очень просто реализовать.

Недостатки данного алгоритма:

1. Несмотря на то, что метод PeerRank позволяет определить «плохих» грейдеров, которые выставляют чрезмерно заниженные баллы, чрезмерно завышенные оценки он определить не может[].
2. Относительно большое количество студентов, которые должны оценить каждую работу других студентов. Для нормальной работы алгоритма требуется не менее 5 оценок для каждой работы. Рекомендуемое количество рецензий на каждого студента равняется 10[]. В целом же метод не зависит от количества оценок, поставленных каждому студенту.