

КВАЛИМЕТРИЧЕСКИЙ ПОДХОД В ОБРАЗОВАНИИ

УДК 378.14.015.62

М. Ю. Мамонтова

КАЧЕСТВО УЧЕБНЫХ ДОСТИЖЕНИЙ: ОЦЕНКА И ПРОГНОЗ НА ОСНОВЕ РЕЗУЛЬТАТОВ КРИТЕРИАЛЬНО-ОРИЕНТИРОВАННОГО ТЕСТИРОВАНИЯ

В статье анализируется проблема содержательной интерпретации наблюдаемых результатов обучения, сочетания их количественной и качественной оценок. Качество учебных достижений, его уровни и способы оценки рассмотрены с позиции квалиметрии. Обосновано использование биномиальной модели критериально-ориентированного тестирования и методов статистической группировки, позволяющее оптимизировать описание наблюдаемых результатов обучения как объекта оценивания и прогнозирования, дополнить количественную оценку содержательной интерпретацией.

Ключевые слова: результат обучения, учебные достижения, качество учебных достижений, оценка качества, уровни качества, критериально-ориентированный тест, квалиметрия.

The article covers the issue of profound interpretation of the training outcomes under study, and combination of their quantitative and qualitative assessment. Quality of academic progress, its level and assessment are considered in terms of quality control. Use of binomial test model of criterial oriented test and statistical grouping methods enables to optimize description of the training outcomes as an appraisal and forecast object, enrich their quantitative assessment with comprehensive analysis.

Key words: training outcome, academic progresses, quality of academic progresses, quality assessment, level of quality, criterial oriented test, quality control.

Проблема качества результатов обучения относится к ключевым проблемам образования. В последние годы существенно меняются подходы к измерению и оцениванию учебных достижений, совершенствуются существующие методы, разрабатываются и внедряются в образовательную практику новые методы и средства. С развитием компьютерных технологий стали возможны массовые обследования в сфере образования – сравнительные международные исследования, аттестация выпускников образовательных учреждений, мониторинг и т. п. Особое место в системе оценки качества результатов обучения заняли государственные образовательные стандарты, фиксирующие в требованиях к подготовке выпускников обязательный результат обучения. Использование математико-статистических методов для оценки результатов обучения способствует разработке основ статистики качества в образовании. Количественная оценка качества дает возможность изучать распределение учащихся по уровням

усвоения учебного материала. Вместе с тем остается нерешенной проблема содержательной интерпретации наблюдаемых результатов обучения, качественного описания различных уровней усвоения дисциплины.

Традиционный подход к оцениванию результатов обучения предполагает использование ранговой шкалы, разные позиции которой ассоциируются с различными уровнями усвоения дисциплины (отлично, хорошо, удовлетворительно, неудовлетворительно). Уровень усвоения учебного материала определяется с учетом характера и количества ошибок, допущенных учащимся при проверке. Качество подготовки различных групп (классов, параллелей и т. д.) характеризуется распределением учащихся по уровням усвоения. Такой подход к оценке качества, в рамках которого изучается статистика брака (характер допущенных ошибок и их количество), сортность (доли учащихся с разным уровнем усвоения дисциплины, выделение группы учащихся, не усвоивших дисциплину на должном уровне), называется в квалиметрии дефектным. Статистические данные о результатах проверок (средний балл, доли учащихся разных уровней усвоения) дополняются описанием ошибок и выделением среди них типичных, т. е. наиболее часто встречающихся.

Применение тестовых методов оценивания результатов обучения позволяет дополнить статистические характеристики данными о выполнении отдельных заданий теста (коэффициент решаемости), распределить задания по категориям трудности. Выделение трудных заданий, так же как и типичных ошибок, дает представление о недостатках в подготовке учащихся и, безусловно, позволяет принимать обоснованные решения о необходимых изменениях в учебном процессе. Для целей же аттестации, когда обучение завершено, представляет интерес разработка измерителей, дающих возможность не только получить количественную оценку наблюдаемого результата, отнести ее к определенному уровню, но и дать содержательную характеристику каждому уровню как совокупности усвоенных знаний и умений.

Поставленная задача, на наш взгляд, может быть решена в рамках междисциплинарного подхода. Результат обучения как объект оценивания и прогнозирования рассматривается в работе как дидактическая, квалиметрическая и статистическая категория.

Выбор методов оценивания учебных достижений определяется рамками концепции, принятой при проведении международных сравнительных педагогических исследований. Согласно этой концепции, результат обучения необходимо рассматривать на трех уровнях: на макроуровне (система образования) как планируемый – в виде стандартов и государственных программ обучения; на мезоуровне (образовательное учреждение, класс) как реализуемый – в виде образовательной программы учреждения, а также используемых для обучения методов и средств, и на микроуровне (учащийся) как достигнутый – в виде знаний, умений, навыков, усвоенных учащимися в ходе изучения дисциплины [3, с. 35]. Такой подход позволяет рассматривать планируемый результат как эталон, стандарт оценивания и сравнивать с ним достигнутые результаты на индивидуальном и массовом уровнях.

Как средство измерения результатов обучения предлагается использовать критериально-ориентированные тесты. Под качеством учебных достижений в определенной области содержания понимается совокупность знаний и умений, усвоенных учащимися в процессе изучения дисциплины. Оценка качества рассматривается в двух аспектах: 1) количественном – определение доли усвоенного учебного материала и соотнесение доли с определенным уровнем качества; 2) содержательном – описание совокупностей знаний и умений, характерных, типичных для разных уровней качества. Знания и умения, типичные для разных уровней качества, определяются с помощью методов классификации и теории вероятностей. Под классом понимается совокупность учащихся, отличающаяся определенным уровнем качества учебных достижений. Классификация основывается на априорном анализе содержательных представлений о характере типа объектов. Применение названных методов в сочетании с эвристическими приемами позволяет свести в один класс однотипные в содержательном плане, т. е. похожие в некотором отношении друг на друга, объекты и сократить число признаков для описания выделенных классов. Ниже приводятся характеристика методов и обоснование их использования для решения поставленной задачи.

В зарубежной образовательной практике критериально-ориентированные тесты используются с 60-х гг. прошлого века [1, с. 92–93]. В России такие тесты нашли применение в последние годы в государственной аттестации выпускников образовательных учреждений как процедуре установления соответствия уровня и качества подготовки требованиям образовательного стандарта (единый государственный экзамен, федеральный интернет-экзамен в сфере профессионального образования). Критериально-ориентированные тесты также могут быть использованы при мониторинге качества результатов обучения, для промежуточной аттестации учащихся (уровень образовательного учреждения, муниципалитета), проведения тематического контроля (уровень учителя).

Принципиальным отличием критериально-ориентированного от нормативно-ориентированного тестирования является то, что его результаты интерпретируются с точки зрения определенной заранее области содержания учебной дисциплины, а не с точки зрения положения учащегося в исследуемой выборке. Для толкования результата используется конкретная область содержания – тема, раздел, дисциплина в целом.

Планируемый результат обучения представлен в образовательном стандарте и программе учебной дисциплины в виде требований к знаниям и умениям обучающихся. То из запланированного, что фактически усвоено учащимися, рассматривают как достигнутый результат или учебные достижения. По классификации, разработанной в теории учебных задач, требования к знаниям и умениям относятся к так называемым критериальным задачам (что и на каком уровне должно быть усвоено) [2, с. 138]. Для каждой критериальной задачи подбирается набор контрольных задач-индикаторов, позволяющих выявить у учащихся наличие выделенных для проверки знаний и умений.

Основной целью критериально-ориентированного тестирования является определение степени усвоения обучающимся учебного матери-

ала. Показателем качества усвоения выступает относительный тестовый балл, равный отношению числа набранных испытуемым баллов к максимально возможному числу баллов. Оценка качества усвоения учебного материала формируется путем сравнения относительного тестового балла с так называемым стандартом оценивания (критериальным баллом) – если относительный балл выше критериального, то испытуемого относят к категории аттестованных, если ниже – к категории неаттестованных. Достаточная полнота и значимость отображенного в заданиях теста проверяемого содержания дисциплины, с одной стороны, и выбор стандарта оценивания адекватно поставленным целям обучения – с другой, обеспечивают критериальную и прогностическую валидность теста, создают предпосылки для объективного оценивания качества результатов обучения.

К важным свойствам критериально-ориентированных тестов также относят отсутствие необходимости в репрезентативной выборке обучающихся для определения тестовых норм и более «простой» по сравнению с нормативно-ориентированным тестированием математический аппарат построения модели тестирования и обработки его результатов.

Для решения задачи содержательной интерпретации полученных в ходе тестирования результатов и построения на их основе вероятностной прогнозной модели наиболее адекватной представляется биномиальная модель Милмана [5, с. 345]. Свойственные ей ограничения (событие имеет только два возможных исхода, события не зависят друг от друга и повторяются некоторое число раз, вероятность благоприятного исхода неизменна для каждого события) могут быть описаны в терминах теории тестирования. Приведем краткое описание модели в терминах теории вероятностей и теории тестирования.

Пусть число заданий в тесте равно n (число независимых испытаний). Каждое задание может быть выполнено учащимся или не выполнено. Пусть событие A – задание выполнено. Вероятность события A в каждом испытании принимается независимой от исходов других испытаний. Тестовые задания считаются локально независимыми – правильное выполнение одного задания не влияет на вероятность правильного выполнения других заданий. В общем случае вероятность правильного выполнения разных заданий (вероятность события A в разных испытаниях) может быть различной либо иметь одну и ту же величину. Вероятность правильного ответа рассматривается в данной модели как величина постоянная для всех тестовых заданий для данного испытуемого и соотносится с уровнем его знаний. Итак, вероятность наступления события A (правильное выполнение заданий) во всех испытаниях постоянна и равна p , вероятность ненаступления события A (задание не выполнено) также постоянна во всех испытаниях и вычисляется по формуле $q = 1 - p$. События «задание выполнено» и «задание не выполнено» будем считать несовместными – появление одного из них исключает появление другого. Событие A будем считать простым. Совмещение нескольких простых событий A (правильное выполнение нескольких заданий из предложенных в тесте) будем считать сложным (совместным) событием.

Число правильно выполненных учащимся заданий x (индивидуальный тестовый балл) рассматривается как дискретная случайная величина. Определим возможные значения величины x и их вероятности. Очевидно, что учащийся может не выполнить ни одного задания, может выполнить одно задание, два задания ... все n заданий. Вероятность сложного события (вероятность того, что при n испытаниях событие A наступит k раз и не наступит $(n - k)$ раз) можно рассчитать по формуле Бернулли:

$$P_n(k) = C_n^k p^k q^{n-k}, \quad (1)$$

где $C_n^k = \frac{n!}{k!(n-k)!}$,

$$k = 0, 1, 2, \dots, n. \quad (2)$$

Формула Бернулли является аналитическим выражением закона биномиального распределения дискретной величины x .

Результат выполнения отдельного задания измеряется в шкале наименований («выполнил» – 1 балл, «не выполнил» – 0 баллов). Результат выполнения теста в целом определяется путем сложения полученных баллов, после чего рассчитывается относительный тестовый балл. Относительный тестовый балл принимается за точечную оценку истинного индивидуального тестового балла.

Критериальный балл назначается экспертами. При оценивании результатов усвоения содержания общеобразовательных дисциплин критериальный балл устанавливается, как правило, равным 0,7. Усвоение такого объема проверяемого содержания считается в дидактике минимально достаточным для дальнейшего изучения дисциплины.

Поскольку по результатам тестирования принимается важное для учащегося решение об аттестации, ошибка неправильной аттестации должна быть минимальной. Если истинный индивидуальный балл значительно выше или ниже критериального, решение об аттестации или неаттестации принимается однозначно. Если же относительный тестовый балл близок к критериальному, вероятность неправильной аттестации возрастает. В модели Милмана такая вероятность определяется как условная вероятность истинного балла по отношению к критериальному баллу и зависит от количества заданий в тесте. Для минимизации ошибки может быть использован метод «нейтральной зоны» [7, с. 430], позволяющий определить интервал относительных баллов вблизи критериального балла, для которого вероятность неправильной аттестации минимальна.

Вероятность выполнения k заданий из теста, состоящего из n заданий, испытуемым, истинный балл которого равен z , определится следующим образом:

$$P_n(k) = C_n^k z^k (1 - z)^{n-k}, \quad (3)$$

Пусть z_0 – критериальный балл, z – истинный балл учащегося. Интервал от z_1 до z_m включает значение z_0 и называется «нейтральной зоной». Значения z_1 и z_m таковы, что при $z < z_0$ минимальная вероятность правильной аттестации α соответствует точке z_1 , а при $z > z_0$ минимальная

вероятность правильной аттестации β соответствует точке z_m . Успешный тестовый балл (число правильно выполненных заданий, необходимое для аттестации) равен n_0 . Значения α и β вычисляются по формулам:

$$\alpha = \sum_{k=0}^{n_0-1} C_n^k z_l^k (1-z_l)^{n-k}, \quad (4)$$

$$\beta = \sum_{k=n_0}^n C_n^k z_m^k (1-z_m)^{n-k}. \quad (5)$$

При заданном числе заданий в тесте (n) можно определить минимальные вероятности правильной аттестации для испытуемых с разным истинным баллом и ширину «нейтральной зоны».

Таким образом, каждый учащийся по результатам тестирования будет отнесен к одной из групп, для которых $z < z_l$ (однозначно не аттестован), $z_l \leq z \leq z_m$ (аттестован с минимально возможной ошибкой), $z \geq z_m$ (однозначно аттестован). Назовем «пограничной» группу аттестованных с минимально возможной ошибкой. Доля учащихся, входящих в пограничную группу, как правило, составляет 0,5–0,6.

Описанная выше процедура дает возможность повысить объективность оценивания результатов обучения. При таком подходе решение об аттестации принимается только на основе количества правильно выполненных учащимся заданий. Вместе с тем один и тот же тестовый балл учащиеся получают за выполнение разных заданий, отличающихся своими дидактическими характеристиками и выявляющих разные знания и умения. Результаты тестирования учащихся, попавших в одну группу по итогам аттестации, разнятся не только количеством выполненных заданий, но и их характеристиками. Насколько совпадают такие наборы заданий по своим дидактическим характеристикам? Можно ли выделить наборы заданий, типичные для каждой группы (неаттестованы, аттестованы, аттестованы с минимально возможно ошибкой)? Какой из наборов заданий наиболее вероятен в определенной группе учащихся и может ли он быть использован для качественной интерпретации полученных результатов тестирования и прогнозирования? Можно ли такие типичные наборы заданий рассматривать как свойства, входящие в состав качества результатов обучения?

Для качественной характеристики индивидуальных наблюдаемых результатов как совокупности выявленных знаний и умений достаточно ограничиться перечнем выполненных заданий (какие знания и умения учащийся продемонстрировал). Вместе с тем учащийся по итогам тестирования становится представителем определенной группы (аттестован, неаттестован и т. п.). В этом случае качество результатов обучения (совокупность выявленных свойств) следует рассматривать как статистическую категорию. Для качественной оценки групповых результатов количественного показателя недостаточно. Принципиально важным представляется не только число правильно выполненных учащимися заданий, но и их дидактические характеристики (какие знания и умения проверяются с помощью этих заданий), а также наиболее вероятное сочетание (или

несколько сочетаний), принимаемое за типичное, характерное для исследуемой группы учащихся. Для решения такой задачи могут быть использованы как статистические методы (типологический анализ, например), так и методы теории возможностей. Названные методы описаны в многочисленных монографиях и публикациях.

Как было отмечено выше, результат выполнения учащимся k заданий из n заданий теста рассматривается как совместное событие, вероятность которого можно определить по формуле (1). Каждому учащемуся ставится в соответствие определенное совместное событие. Совокупности учащихся, полученные в ходе тестирования, неоднородны. Задача заключается в том, чтобы выявить структуру совокупности, т. е. выделить внутри нее такие группы учащихся, для которых наборы заданий одинаковы (одинаковы совместные события), и определить совместное событие, имеющее наибольшую вероятность.

Использование метода комбинационной группировки с учетом всех заданий теста может привести к выделению большого числа групп малой численности. Для построения группировки целесообразно использовать наиболее информативные с точки зрения поставленной задачи признаки, снизив тем самым размерность пространства признаков. В этой связи для выявления типичного совместного события (набора заданий) наиболее информативными следует считать задания, имеющие высокие значения коэффициента решаемости, т. е. те, которые правильно выполнены подавляющим большинством представителей исследуемой совокупности учащихся и могут рассматриваться как характерные для данной группы, типичные. По результатам выполнения таких заданий выделенная группа учащихся должна быть однородна в статистическом смысле.

Поскольку результаты выполнения отдельных заданий рассматриваются как признаки, измеренные в шкале наименований, то для определения степени однородности группы можно использовать коэффициент вариации качественных признаков. Расчет коэффициента вариации качественных признаков Q основан на сопоставлении числа различных пар событий:

$$Q = \frac{[N^2 - \sum_{i=1}^b f_i^2]b}{(b-1)N^2}, \quad (6)$$

где b – число градаций признака (в нашем случае $b = 2$ – «выполнил задание» и «не выполнил задание»);

f_i – число объектов с i -й градацией признака (число учащихся, выполнивших и не выполнивших задание);

N – общее число объектов совокупности (число испытуемых).

Максимальное значение коэффициента вариации равно 1 и наблюдается в случае равенства частот классов (половина учащихся выполнили задание, другая половина – не выполнили). Нулевое значение коэффициента соответствует случаю, когда вся совокупность образует по данному признаку один класс (например, все выполнили задание или задание не выполнил ни один из учащихся). Малое значение коэффициента вариации

ации свидетельствует о том, что объекты по данному признаку распределены крайне неравномерно – большинство объектов обладают одной градацией признака и лишь небольшое число объектов имеют другую градацию. Группа считается однородной по определенному признаку, если коэффициент вариации равен или меньше 0,30.

Задания, с которыми не справился ни один учащийся или не справилось подавляющее большинство учащихся и которые имеют коэффициент вариации меньше 0,3, можно рассматривать как малоинформативные и в дальнейшем не использовать. Такой прием позволяет уменьшить количество заданий для построения группировки и, следовательно, число подгрупп учащихся с идентичными комбинациями заданий. На основе оставшихся заданий строится группировка, в которой учащиеся с идентичными комбинациями правильно выполненных заданий объединяются в подгруппы. Вероятности появления выделенных подгрупп рассчитываются с помощью правил умножения и сложения вероятностей и в сумме, в соответствии с теоремой полной вероятности, равны единице. Комбинационные группировки строятся для трех групп учащихся, имеющих разный уровень качества достигнутого результата обучения, – однозначно аттестованных, однозначно не аттестованных, аттестованных с минимально возможной ошибкой при данном числе заданий в тесте.

Построенные группировки могут быть использованы для содержательной интерпретации наблюдаемых результатов тестирования. На их основе выделяются комбинации заданий, которые встречаются чаще других и могут рассматриваться как характерные, типичные для определенного уровня качества результата обучения. Вероятность появления некоторых комбинаций очень низкая (0,01–0,03), поэтому их можно считать нехарактерными для данного уровня и не учитывать при описании.

Таким образом, количественная оценка качества учебных достижений, рассчитываемая как относительный тестовый балл, уровень качества достигнутого результата, определяемый путем сравнения с критериальным баллом (однозначно аттестован, однозначно не аттестован, аттестован с минимально возможной ошибкой при данном количестве заданий в тесте), дополняются содержательной характеристикой – совокупностями заданий, которые наиболее вероятны, характерны для каждого уровня.

Полученные в ходе анализа данные о распределении учащихся на группы с разным уровнем качества, а также наиболее вероятные комбинации заданий для каждой группы могут рассматриваться как переменные модели прогноза и использоваться для прогнозирования результатов критериально-ориентированного тестирования других контингентов учащихся. Речь идет о так называемом поисковом ретроспективном прогнозировании, когда период упреждения уже прошел и данные по другим контингентам рассматриваются как фактические данные прогнозируемых показателей. Цель поискового прогноза – выявить спектр возможных состояний объекта прогнозирования и установить вероятность этих состояний. Полученная ретроспективно ошибка прогноза характеризует точность применяемой модели прогнозирования.

Вместе с тем необходимо отметить, что вероятностное описание может быть использовано только в условиях стохастической устойчивости, когда частоты появлений тех или иных событий в разных выборках совпадают (эмпирические распределения согласуются). Вероятностная модель не прогнозирует индивидуальные результаты тестирования, но дает возможность предвидеть структуру обследуемой совокупности учащихся (деление на группы с разным уровнем качества учебных достижений), а также наиболее характерные, типичные для выделенных групп свойства, входящие в состав качества учебных достижений.

Предлагаемое в работе сочетание критериально-ориентированного тестирования на основе биномиальной модели, статистических методов группировки, эвристических приемов и вероятностных методов отбора информативных признаков и их комбинаций позволяет оптимизировать описание качества достигнутого результата обучения как объекта оценивания и прогнозирования, обеспечить заданную достоверность и точность прогноза при минимальных затратах на его разработку. Использование дихотомической шкалы дает возможность оптимизировать измерение единичных показателей качества (результатов выполнения отдельных заданий), наименее затратными способами извлечь информацию, достаточную для решения задачи содержательного анализа результатов обучения. Применение методов классификации и типологии в сочетании с эвристическими приемами позволяет сократить размерность пространства признаков при описании различных уровней качества благодаря отбору наиболее информативных признаков с точки зрения задачи оценивания и прогнозирования результатов обучения, оптимизировать степень формализации описания наблюдаемых результатов обучения.

Литература

1. Анастаси А. Психологическое тестирование. М.: Педагогика, 1982. Кн. 1. 318 с.
2. Балл Г. А. Теория учебных задач. М.: Педагогика, 1990. 184 с.
3. Пломп Т. Концептуализация сравнительных педагогических исследований // Перспективы. Вопр. образования. 1993. № 3. С. 29–41.
4. Руководство по применению стандарта ИСО 9000 : 2000 в области обучения и образования / пер. с англ. А. Л. Раскина. М.: РИА «Стандарты и качество», 2002. 128 с.
5. Millman J. Criterion-referenced measurement // Evaluation in Education: Current applications / ed. W. J. Popham. Barkeley CA: McCutchan, 1974. P. 311–397.
6. Willcox R. P. Determining the length of a criterion-referenced test. Applied Psychological Measurement. 1980. № 4. P. 425–446.