

общества. Для достижения указанной цели и необходимо преобразование существующих в образовательной системе институтов в соответствующие формы, а также введение новых институтов.

Таким образом, можно констатировать тот факт, что неуспешность и затяжной характер проводимых в области образования реформ обусловлены тем, что политическая власть в нашем государстве отражает интересы не самых широких слоев населения, но сравнительно малочисленной и наиболее богатой части нашего общества.

СПИСОК ЛИТЕРАТУРЫ

1. **Миронов В. В.** Размышления о реформе российского образования // Философия образования. – 2012. – № 1. – С. 3–43.

Принята редакцией: 15.08.2012

УДК 37.0 + 371 + 316.7

КОЛИЧЕСТВЕННЫЕ МЕТОДЫ ИССЛЕДОВАНИЯ ДАННЫХ ЭВАЛЮАЦИИ

М. В. Гуськова (Москва)

В статье подчеркнута актуальность проблем развития математизации педагогических процессов и исследований в России и за рубежом, описана специфика применения количественных методов оценки в системе эвалюации. Показаны возможности применения корреляционно-регрессионных подходов для целей анализа совокупности объектов эвалюации, которые являются однородными в отношении оцениваемых признаков или явлений. Кроме того, определены направления использования различных моделей для описания взаимосвязей между наблюдаемыми переменными (признаками) (регрессионные, многоуровневые, смешанные многомерные, латентные, путевого, многофакторного анализа и др.).

Ключевые слова: *система эвалюации, компетентостный подход, количественные методы, модели оценивания компетенций.*

QUANTITATIVE METHODS OF STUDYING THE EVALUATION DATA

M. V. Gus'kova (Москва)

The paper highlights the topicality of the problems of development of mathematization of pedagogical processes and studies in Russia and abroad. The features of using quantitative methods in the evaluation system are

© Гуськова М. В., 2012

Гуськова Марина Владимировна – кандидат технических наук, доцент, начальник Управления лицензирования, аккредитации, признания и подтверждения документов, Федеральная служба по надзору в сфере образования и науки.

E-mail: dep06@obnadzor.gov.ru

considered. There are demonstrated the possibilities of application of the correlation and regression approaches to the analysis of the set of evaluation objects that are homogeneous with respect to the evaluated characteristics or phenomena. Besides, there are determined the application directions for various models to describe the interrelation between the observed variables (attributes) (the regression models, the multilevel ones, the mixed multidimensional ones, the models of latent, path and multifactor analysis etc.).

Key words: *evaluation system, competence approach, quantitative methods, competence assessment models.*

Современный этап развития профессионального образования связан с внедрением Федеральных государственных образовательных стандартов (ФГОС). Известно, что в основу ФГОС заложен компетентостный подход как основа формирования компетенций у обучающихся и выпускников образовательных учреждений профессионального образования. В соответствии с действующим законодательством и нормативно-правовыми актами перед образовательными учреждениями стоит задача обеспечения функционирования систем мониторинга качества образования, формирования фондов оценочных средств, в том числе оценка формируемых компетенций. В связи с этим остро стоит необходимость формирования инструментария под данную задачу с использованием количественных методов оценки.

На пути развития количественных методов в исследованиях педагогических явлений в отечественном образовании имеется немало препятствий. Это и латентная природа исследуемых объектов, и отсутствие соответствующей подготовки у педагогов, а также специальных центров по обработке образовательной статистики. Но все же наиболее существенным препятствием являются традиции, сложившиеся в отечественном образовании: здесь на всех образовательных уровнях привыкли доверять субъективным оценкам педагогов, выставляемым в соответствии с их собственными критериями оценивания результатов своей работы. Педагоги, десятилетиями переносящие в профессиональную деятельность свой собственный опыт обучения, противятся неизвестному им количественному подходу, требующему обращения к зарубежной науке – теории педагогических измерений, приложения усилий по ее освоению и по разработке отсутствующих на сегодняшний день измерителей.

Возникновение прикладной статистики для работы с числовыми данными в социальных науках принято связывать с именем Ф. Гальтона, который в конце XIX в. впервые применил статистический анализ в психологии. В дальнейшем становление статистических методов происходило параллельно в естественных и в гуманитарных, социальных науках, во многом способствуя развитию последних с точки зрения требований строгости, доказательности, объективности, верифицируемости и пр. Первые попытки обращения к количественным методам в образовании, заслуживающие упоминания, появились в 20-х гг. XX в. – тогда в нашей стране и за рубежом были заложены основы научного аппарата классической теории тестов. Поскольку по результатам выполнения тестов выставлялся балл в соответствии с количеством правильно выполненных заданий, среди исследователей того времени бытовало мнение, что применение тестов по-

зволяет получить оценки испытуемых в метрической шкале. На протяжении периода 20–40-х гг., нередко называемого классическим в соответствии с классической теорией тестов, складывался фундамент теории педагогических измерений, определивший методы создания тестов и статистического анализа качества педагогических измерений. Для обоснования качества педагогических измерений – их надежности и валидности – стала использоваться теория корреляции, сформировались концепции параллельных и эквивалентных измерений, вводились основные аксиомы классической теории тестов.

В зарубежной практике в 20-х гг. XX в. стали появляться первые стандартизированные тесты для измерения результатов школьного обучения. Под влиянием работ известного американского психолога, профессора Колумбийского университета Э. Торндайка для их разработки использовались те же принципы, которые ранее легли в основу создания психологических тестов. В качестве примера первых стандартизированных измерений в образовании можно назвать шкалы оценок качества почерка и письменных сочинений, тесты на правописание, выполнение арифметических действий и др. В 1920 г. появились первые групповые тесты Томсона, Бурта и Спирмена. Вскоре были разработаны тесты интеллекта для лиц всех возрастов и категорий – от дошкольников до аспирантов.

На фоне прогресса в теории измерений развивались количественные и качественные исследования по выявлению возможностей измерителей и измерительных процедур, в том числе на эмпирической основе посредством соотнесения результатов тестирования с эмпирическими и экспертными данными. Поскольку результаты измерений не всегда подтверждались практикой, обнаруженные несоответствия использовались различными политическими течениями, в соответствии с потребностями которых при сравнении умственных способностей людей, принадлежавших к различным культурным, этническим и расовым группам, высказывались доводы «за» или «против» тестов.

С 40-х по 80-е гг. на развитие количественных методов в образовательных исследованиях значительное влияние оказали работы С. Стивенса, среди которых особо следует выделить его статью «О теории шкал измерений» (1946). Характерной чертой этого периода является значительное расширение сферы приложения теории измерений, когда центр тяжести прикладных исследований по измерениям переместился из точных наук в психологию, экономику, социологию и педагогику. Дальнейшее развитие получили специальные знания по вопросам применения факторного анализа в эмпирических исследованиях, все большее значение стала приобретать статистика, ее методы дисперсионного и корреляционного анализа для разработки тестов. Широкое развитие находят методы формирования репрезентативных выборок, необходимых для стандартизации тестов. Именно в эти годы к тестологам пришло осознание того, что использование тестирования в психологии и образовании не приводит автоматически к построению метрической шкалы. Задания теста имеют разную трудность, поэтому в шкале сырых тестовых баллов отсутствует единица измерения. В связи с этим в теории педагогических измерений активизировался поиск эффективных методов построения интервальных шкал

метрического характера, обладающих единицей измерения. Практически с середины 50-х вплоть до конца 70-х гг. было завершено формирование математико-статистического аппарата классической или традиционной (как ее называют иначе) теории тестов.

На смену классическим подходам С. Стивенса пришла современная математическая аксиоматика теории измерений. Определения и понятия уточнялись путем перевода их формулировок на язык математических терминов. В этот период появилась современная теория измерений, строящаяся на более строгой аксиоматической основе и значительно расширяющая многие традиционные представления о возможностях применения измерений в эмпирических науках. Сформировалась методология сочетания количественных и качественных оценок, обосновывающая для совокупности методов теоретические основы объединения количественных и качественных результатов в единой шкале.

Использование количественных методов в образовании имеет множество направлений: обработка результатов педагогических экспериментов и результатов контрольно-оценочной деятельности при построении шкал, создание измерителей, моделирование, диагностика и прогнозирование результатов образования. Количественные методы применяются также при анализе результатов качественных исследований, с теми или иными ограничениями, что обеспечивает корректность используемого аппарата.

Достаточно часто в наши дни в сфере количественных исследований применяется метод математического моделирования, классическим примером которого является теория ИРТ. В более общем контексте математическую модель можно трактовать как мысленно представленную и материально реализованную систему в форме функций, алгоритмов и закономерностей реализации определенных действий по их использованию; такая система адекватно отображает предмет исследования и способна замещать его для того, чтобы получить новую информацию об этом предмете.

Становление количественных методов в социальных науках за рубежом также сопровождалось определенными трудностями, поскольку оно всегда протекало на фоне непрекращающейся борьбы представителей различных философских парадигм. Противники количественных методов в образовании обычно аргументировали свою критику опасениями по поводу того, что переход к числам приведет к выхолащиванию сущности педагогических явлений. Сторонники их, наоборот, видели в количественных методах средство преодоления субъективизма в образовании. Компромисс между сторонниками и противниками был достигнут сравнительно недавно – в начале XXI в. в рамках парадигмы прагматизма, когда стали говорить не о противостоянии, а о взаимном дополнении двух подходов при проведении исследований.

Зарубежный опыт наших дней показывает, что математизация исследований в педагогике несет в себе огромный гносеологический потенциал. Она не только избавляет науку от одностороннего качественного описания, но и позволяет провести анализ эмпирических данных, значительно повышая надежность и валидность результатов исследования. Для корректной реализации количественных методов должны быть соблюдены важные условия, в число которых входит формулировка ясной непро-

тиворечивой гипотезы, построение аутентичной математической модели, включающей необходимое число переменных, наполнение компонентов модели экспериментальными фактами, собранными на представительной выборке объектов, применение математического аппарата для обработки данных, адекватного характеру их распределения и уровню измерения.

Работа по внедрению количественных методов в образование ведется и в России, но не с той интенсивностью, которая требуется для эффективного развития образовательной эвалюации. За последние годы было издано достаточно много работ, написанных отечественными авторами о применении количественных методов в социальных науках. В основном эта проблема рассматривалась с позиций потребности эвалюации в образовании, все эти работы обладают двумя крупными недостатками. Во-первых, в большинстве своем они посвящены количественным методам в социологии и психологии, а в меньшей степени – в образовании. Во-вторых, они охватывают простейшие и самые распространенные способы статистического анализа (описательную статистику, сопоставление данных двух выборок), а все выходящее за рамки этого обязательного минимума рассматривается только в ознакомительном формате [3].

В результате значительное число современных количественных методов, уже широко известных и активно используемых педагогами за рубежом, до настоящего времени известно лишь узкому кругу отечественных исследователей эвалюации и применяются крайне редко либо некорректно (например, дискриминантный анализ, анализ ковариаций, корреляционный анализ, мета-анализ, линейный иерархический анализ и т.д.).

Базовой программой, с помощью которой специалисты используют статистические методы в образовательной эвалюации, является SPSS. Помимо этого мощного статистического пакета для обработки данных, в эвалюации применяется программа STATISTICA. Эта более новая, но быстро развиваемая программа имеет множество дополнительных программных блоков, существенно расширяющих ее возможности. Для реализации конфирматорного факторного анализа и при использовании структурного моделирования обращаются к программе LISREL.

В целом количественные методы исследования включают описательную (дескриптивную) статистику, корреляционный, факторный, регрессионный, дискриминантный анализ, параметрическое и непараметрическое моделирование, мета-анализ, индуктивную статистику (генерализацию), предназначенную для распространения результатов выборочного исследования на генеральную совокупность объектов.

Роль количественных методов исследования в эвалюации трудно переоценить, поскольку в любой ситуации обращение к ним позволяет повысить надежность и валидность информации, получаемой при измерениях. Чаще всего количественные методы исследования в эвалюации применяются к данным, представленным в интервальной шкале, и используются в сочетании с математическими моделями. Однако к ним обращаются и при проведении качественных исследований, правда, в этом случае приходится уделять значительное внимание проблеме корректности выбора аппарата обработки данных, поскольку не все количественные методы применимы к результатам эвалюации, представленным в порядковой или номинальной шкале.

При выборе количественных методов в эвалюации прежде всего обращаются к описательной статистике. Анализ данных в ней обычно строится на графической интерпретации или на аналитических процедурах. Данные количественного анализа представляют в виде несгруппированного ряда произвольной формы, ранжированного ряда, частотного распределения или распределения сгруппированных частот. По распределениям проводится графическая интерпретация эмпирических данных, чаще всего, в виде гистограмм или сглаженных кривых. Нередко предпочтение отдают гистограмме, поскольку это представление является наиболее подходящим для визуального восприятия в том случае, когда изображается не более одного набора данных. Для сравнения двух или более распределений обычно используют полигоны частот, так как при наложении гистограмм получается довольно запутанная картина.

К дескриптивной статистике в первую очередь относят меры центральной тенденции совокупности данных, которые предназначаются для выявления «центрального положения», вокруг которого в основном группируется множество значений рассматриваемого распределения. Наиболее простой мерой центральной тенденции считается мода распределения. Достаточно просто осуществляется и подсчет среднего выборочного (среднего арифметического), на величину которого, в отличие от моды, влияют значения всех количественных данных. Поэтому среднее значение обобщает индивидуальные особенности составляющих распределения, в нем уравниваются отдельные значения рассматриваемой величины.

В состав показателей дескриптивной статистики также входят характеристики, служащие мерами изменчивости количественных данных. Они позволяют выявить существенные различия в структуре распределений. Наиболее важная характеристика указывает на особенности разброса эмпирических данных вокруг среднего значения баллов по тесту или среднего по другой совокупности данных. Отдельные значения из этой совокупности могут быть тесно сгруппированы вокруг своего среднего либо, наоборот, сильно удалены от него. Поэтому необходимы оценки характеристик распределения, отражающие вариацию данных, к которым относят размах, дисперсию и стандартное отклонение.

Вариационный размах легко вычисляется, но используется крайне редко, поскольку он является весьма приближенным показателем в силу того, что не зависит от степени изменчивости промежуточных значений, расположенных между крайними значениями в распределении данных, которые к тому же всегда довольно ненадежны. В связи с этим более удачной мерой считается дисперсия. Кроме нее, для характеристики меры изменчивости распределения удобно использовать стандартное отклонение, которое равно корню квадратному из дисперсии.

Стандартное отклонение является крайне полезной мерой вариации для случая нормального распределения данных, поскольку заранее приблизительно известно, какой процент результатов оценивания лежит внутри одного, двух и трех стандартных отклонений, откладываемых от центра распределения. Наиболее удобна единичная нормальная кривая, площадь под которой равна единице, среднее значение для нее равно нулю, а стандартное отклонение – единице.

В число показателей дескриптивной статистики также входят меры симметрии и островершинности кривых распределений, к которым относятся асимметрия (характеристика степени отклонения распределения наблюдаемых частот выборки от симметричного распределения, характерного для нормальной кривой) и эксцесс (характеристика островершинности полигона частот или гистограммы).

В заключение при анализе дескриптивной статистики осуществляют проверку значимости найденных статистик. Для этого добавляют информацию о принимаемом уровне риска допустить ошибку в статистическом выводе. Наиболее приемлемым для педагогических измерений является уровень в 5 %, который допускает ошибку в пяти случаях из ста.

После получения дескриптивной статистики в эвалюации нередко обращаются к аппарату корреляционного и регрессионного анализа. Корреляционный анализ предназначен для оценивания показателей связи между наборами данных. Корреляция в широком смысле слова означает связь между явлениями и процессами, сходный смысл при анализе связей имеет ковариация. Однако для изучения связи установить ее наличие недостаточно, необходимо также правильно выбрать ее вид и форму показателя для оценки меры связи между явлениями. Чаще всего используется классический коэффициент корреляции Пирсона, который имеет различные модификации в зависимости от характера распределений (дихотомические или полиномические данные), связь между которыми анализируется.

Достаточно близко с корреляционным анализом связан регрессионный анализ. Если первый оценивает силу связи между явлениями, то второй позволяет установить форму этой связи. Использование только одной связи в эвалюации приводит к излишней упрощенности, однако всестороннее изучение причинных отношений между объектами эвалюации требует проведения комплексного исследования. Поэтому нередко говорят о корреляционно-регрессионных связях и стараются анализировать такие совокупности объектов эвалюации, которые являются достаточно однородными в плане оцениваемых признаков или явлений.

Моделирование структурными уравнениями (SEM) использует различные виды математических моделей для описания взаимосвязей между наблюдаемыми переменными. Такие модели позволяют провести количественную проверку теоретической модели, которую разработал исследователь. Проверке могут подвергаться разнообразные теоретические модели, которые содержат гипотезы о том, как наборы переменных определяют логические структуры и как эти логические структуры связаны друг с другом. В SEM широко применяется конфирматорный факторный анализ, который позволяет проверять частные гипотезы о факторной структуре для множества переменных в одной или нескольких выборках.

Истоки теории SEM заложены достаточно давно и восходят к моделям регрессионного анализа, метода путевого анализа, основанного на изучении структуры причинных связей между переменными, и конфирматорного факторного анализа. Первый шаг на пути создания линейной регрессионной модели был сделан в 1896 г. Карлом Пирсоном [4], предложившим формулы для коэффициента корреляции. Немного позже Спирмен [5] использовал коэффициент корреляции, чтобы установить, какие ком-

поненты коррелируются для построения факторной модели и впервые использовал термин «факторный анализ» для определения двухфакторной модели в теории интеллекта.

В 40-х гг. XX в. Лоули развил идеи Спирмена о применении факторных моделей и предложил инструментарий для получения выводов, подтверждающих логические построения. Большинство из используемых в последующие годы анкет, вопросников и методик, определяющих способности и достижения обучаемых, были созданы с помощью факторного анализа [6]. В последующие годы появился термин «конфирматорный факторный анализ», широко используемый в настоящее время [7].

Модели метода путевого анализа, которые нередко в русскоязычной литературе называются диаграммами путей факторной модели, были созданы S. Wright (Сьюэлл Райт) в 40-х гг. XX в. Они построены на основе использования коэффициентов корреляции и регрессионного анализа с целью моделирования сложных взаимосвязей между переменными в биологии, но не нашли широкого применения вплоть до конца 1950-х гг., когда эконометристы переосмыслили их возможности для представления модели в виде системы одновременных уравнений. Социологи не открыли их заново в 1960-х гг.

Модели структурных уравнений, совмещающие модели метода путевого анализа и конфирматорного факторного анализа и включающие латентные переменные, получили развитие в 70-х гг. XX в. в трудах Йорескогу, Кислингу и Уилли. Первоначально модели структурных уравнений были известны как JKW-модели, но в дальнейшем приобрели известность под именем «модели линейных структурных взаимосвязей» (LISREL) с одноименным программным продуктом LISREL для их использования.

Постепенно к концу XX в. расширилась предметная область моделирования с использованием структурных уравнений. Помимо LISREL с середины 1980-х гг. стали интенсивно разрабатываться другие программные продукты, такие как Amos, EQS, позволяющие реализовать множество различных приложений SEM, включая регрессионные модели, модели путевого анализа, модели многофакторного анализа, многоуровневые модели, смешанные модели, многомерные модели, латентные криволинейные модели прироста и динамические модели.

СПИСОК ЛИТЕРАТУРЫ

1. Шмелев А. Г. Тесты для старшеклассников и абитуриентов // Первое сентября (приложение). – М., 2000.
2. Boekkooi F., Timminga, E. The construction of parallel tests from IRT – based banks // Journal of Educational Statistics. – 1991.
3. Найденова Н.Н. Формирование репрезентативной выборки: учеб. пособие. – М.: Логос, 2003.
4. Чельшкова М. Б., Шмелев А. Г. Шкалирование результатов Единого государственного экзамена : проблемы и перспективы // Вопросы образования. – 2004. – № 4.
5. Anderson D. A., & Aitkin M. Variance component models with binary response: Inter-viewer variability // Journal of the Royal Statistical Society. – 1985. – N 47(2).

Принята редакцией: 27.08.2012