

Современные подходы к оцениванию качества результатов высшего образования

Звонников Виктор Иванович

доктор педагогических наук, директор Департамента качества управленческого образования, Москва, Государственный университет управления

Чельшкова Марина Борисовна

доктор педагогических наук, заместитель директора Департамента качества управленческого образования, Москва, Государственный университет управления, inf@guu.ru

Ключевые слова: оценивание качества образования, компетенция, надёжность, валидность, адаптивное тестирование, модель измерителя, моделирование структурными уравнениями.

На совершенствование подходов к оцениванию качества результатов образования влияет целая совокупность факторов, как замедляющих, так и ускоряющих этот процесс. С одной стороны, позитивное стимулирующее воздействие оказывает практика проведения массовых оценочных процедур, примером которых в России могут быть Единый государственный экзамен, сертификация специалистов и выпускников и их аккредитация в сфере здравоохранения. С другой стороны, совершенствованию оценочных процедур в направлении повышения надёжности, валидности и сопоставимости их результатов препятствуют те эволюционные процессы, которые происходят в трактовке качества результатов образования.¹ Рост приоритетов творческих компонентов подготовки обучающихся, внедрение компетентностного подхода в систему высшего профессионального образования, потребность в выявлении талантливых обучающихся приводят к тому, что на смену количественным оценкам учебных достижений приходят качественные оценки, надёжность и сопоставимость которых всегда ниже, чем у количественных оценок.

Таким образом, новые требования общества и государства к выпускникам образовательных организаций системы высшего образования, заложенные во ФГОС, инициируют создание современных подходов к оцениванию качества результатов образования, базирующихся на бипарадигмальной методологии измерений в образовании. Базовые идеи бипара-

¹ Звонников В.И. Качество образования, или Кое-что новое о вечной проблеме. Высшее образование сегодня. — № 8. — 2009.

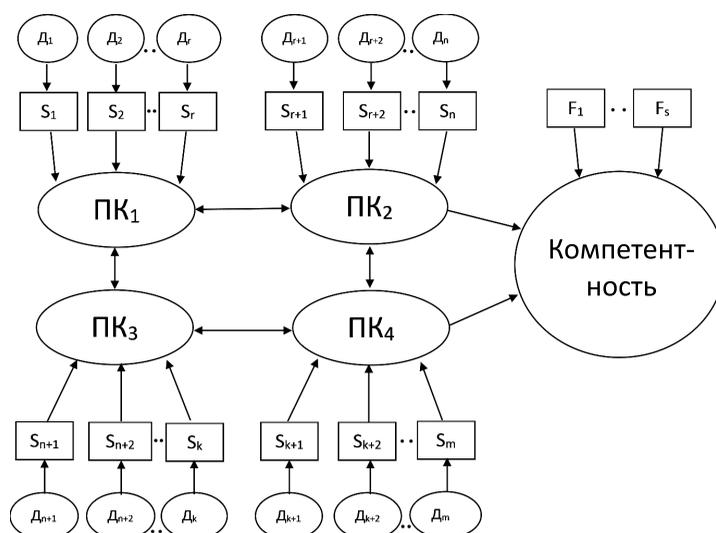


Рис. 1. Структурная модель связи дисциплин и компетенций

дигмальной методологии, утверждающие возможность построения единых шкал для интеграции количественных и качественных данных, вносят свою специфику в трактовку всех компонентов педагогических измерений, к которым обычно относят: переменные измерения и их эмпирические референты, измерительные процедуры, измерительный инструментарий, шкалы и методы анализа полученных результатов измерения.²

Для выбора планируемых переменных измерения следует построить целевую модель, определяющую структуру взаимосвязи количественных и качественных переменных. На рис. 1 приведена структурная модель для оценивания четырёх компетенций, носящая упрощённый гипотетический характер и называемая диаграммой пути в том случае, если используется метод моделирования структурными линейными уравнениями (МСУ), позволяющий проверить причинные гипотезы о связи между латентными переменными, факторами влияния и резуль-

татами образования на основе корреляционных данных.³

Символом «Д» с соответствующим индексом обозначены дисциплины, а стрелки указывают на то, что они вносят свой вклад в формирование компетенций. На рисунке изображён упрощённый вариант, когда базой для формирования каждой профессиональной компетенции являются разные дисциплины. Однако в реальной практике обучения часто могут встречаться случаи, когда одна и та же дисциплина участвует в формировании многих компетенций, тогда стрелки, соединяющие квадратики и овалы с номерами компетенций, будут многократно пересекаться.

Символ «S» предназначается для обозначения того вклада, который вносит соответствующая дисциплина с тем же индексом в формирование профессиональной компетенции. Символ «ПК» обозначает профессиональную компетенцию, а \$F_1\$ и \$F_s\$ выбраны для факторов, влияющих на формирование компетентности у выпускника вуза.

² Звонников В. И. Актуальные проблемы высшего профессионального образования: сборник научных статей/под ред. проф., д. э. н. Е. М. Козакова, Екатеринбург: УрГУПС, 2010. Вып. 75 (158) с. 47–58.; Звонников В. И., Нарбут В. В. Управленческое образование в университетах России: перемены на горизонте/ журнал «Высшее образование сегодня». — № 1. — 2016 г. — С. 18–24.

³ Остапенко Р. И., Остапенко А. И. Использование методов моделирования структурными уравнениями в области управления образованием. Журнал «Государственный советник» Выпуск № 4 / 2013. — С. 24–32.; Bender P. M. EQS structural equations program manual. Los Angeles: BMDP Statistical Software. — 1989. — 284 с.

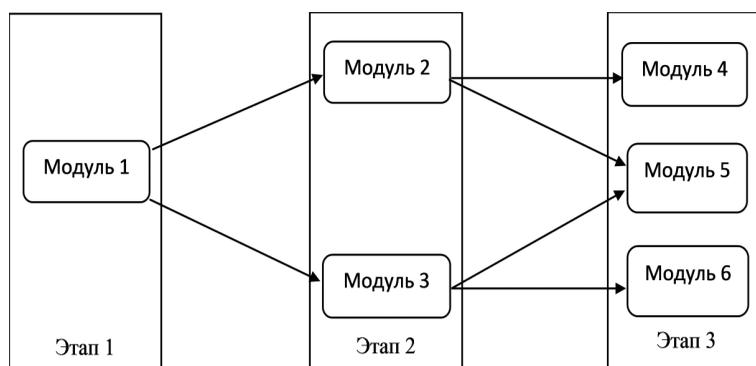


Рис. 2. Трёхэтапные измерения

В методе МСУ набор переменных и факторов связывается определёнными логическими моделями, которые называются диаграммами пути и носят постепенно усложняющийся характер. Цель метода МСУ состоит в выявлении степени подтверждения теоретической модели выборочными данными. Сначала строится простая логическая модель, включающая только несколько латентных переменных и факторов влияния. При наличии достаточной степени подтверждения модель постепенно усложняется за счёт введения дополнительных переменных или связей между ними, которые, как правило, носят причинно-следственный характер. В противном случае, если подтверждение не наблюдается, меняется исходная логическая модель за счёт введения других переменных или связей между ними. В результате применения метода МСУ создаётся оптимальная модель, отображающая связи между переменными измерения и факторами влияния.⁴

Несмотря на то что статистические методы анализа связей совокупности переменных МСУ были описаны и доступны достаточно давно, в течение долгого времени эти методы имели ограниченное применение для проверки причинно-следственных предположений. В течение последних 10–15 лет ситуация изменилась, и метод МСУ стал рассматриваться как наиболее адекватный для анализа причинно-следственных предположений

в экономике и социальных науках. В наши дни он широко используется в ряде зарубежных стран для анализа конструктивной валидности компетентностно-ориентированных измерителей и анализа факторов, влияющих на качество результатов образования.

Процесс применения метода МСУ включает несколько этапов:

- на первом этапе рисуется упрощённая «диаграмма путей», включающая минимальное число всех необходимых переменных, входящих в состав моделируемой причинно-следственной системы. В модели должны входить как зависимые, так и независимые переменные, связи между которыми изображаются направленными стрелками. Наблюдаемые переменные обычно изображаются квадратами, а латентные — овалами или кругами. Направление стрелок соответствует направлению влияний;
- на втором этапе диаграммы переводятся на язык уравнений множественных регрессий. При этом записывается столько уравнений, сколько модель содержит в себе переменных, требующих объяснения, т.е. количество уравнений соответствует количеству зависимых переменных;
- на третьем этапе собираются выборочные данные, описывающие анализируемые связи;
- на четвёртом этапе системы уравнений подвергаются статистическому анализу для проверки соответствия модели, выраженной посредством системы линейных уравнений, и эмпирических данных;
- на пятом этапе принимается решение об усложнении модели путём ввода допол-

⁴ Joreskog K.C., Sorbom D. LISREL 17, a guide to the program and applications. Chicago: SPSS, 2007. — 500 p.



Рис. 3. Пример трёхуровневой шкалы

нительных переменных или её полного изменения.⁵

Для практического применения метода MCSU обычно используются стандартные статистические пакеты типа LISREL или EQS. Чаще всего предпочтение отдаётся EQS в связи с «дружелюбным интерфейсом» этого пакета.⁶

Метод MCSU и базовые положения бипарадигмальной методологии адекватны ситуации измерения для совокупности различных переменных (знаний, умений, компетенций и т. д.), которые могут носить как количественный, так и качественный характер и объединяться в единой уровневой шкале.⁷ Приложение идей бипарадигмальности к выбору моделей измерителей приводит к необходимости проведения многостадийных измерений, включающих несколько этапов. В частности, на рисунке 2 показаны три этапа измерений, каждый из которых требует создания своего инструментария для оценивания знаний

и умений или компетенций. Как правило, в многостадийных измерениях этапы оценивания соотносят с отдельными диапазонами уровневой шкалы, показывающими нарастающие уровни освоения знаний и умений или компетенций. В частности, для этапов, изображённых на рис. 2, диапазон минимальной компетентности можно соотнести с первым этапом, диапазон базовой компетентности — со вторым этапом и диапазон высокой компетентности — с третьим этапом. В этом случае нарастание трудности измерителей также будет происходить слева направо, хотя обычно принято располагать уровневую шкалу вертикально, как показано далее на рис. 3.

При построении модели измерителя приходится принимать во внимание характер измеряемых переменных, их особенности, отражающиеся в процессе операционализации переменных в их эмпирических референтах — наблюдаемых признаках проявления. В частности, в рамках компетентностного подхода следует учитывать, что компетенции имеют:

- отсроченный характер проявления в профессиональной деятельности выпускников образовательных учреждений после окончания обучения;
- причинный характер связей с эффективностью профессиональной деятельности;
- мета-латентную природу.

Эти характеристики позволяют рассматривать компетенции, сформированные по результатам обучения в вузе, как глубоко лежащие устойчивые поведенческие

⁵ Improving decision making in organizations. The opportunity to reinvent finance business partners, Chartered Institute of Management Accountants, 2009. — 167 с.

⁶ Joreskog K. C., Sorbom D. LISREL 17, a guide to the program and applications. Chicago: SPSS, 2007. — 500 p.

⁷ Звонников В. И., Чельшкова М. Б. Оценка качества результатов обучения при аттестации. М.: Логос, 2012, 279 с.; Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов. Под ред. В. И. Звонникова, М. Б. Чельшковой. М.: Логос, 2010, 667 с.; Computerized Multistage Testing. Theory and Applications/Ed. by D. Yan, A. von Davier, C. Lewis. CRC Press. — 2014. — 504 p.

свойства человеческой личности, прогнозирующие эффективность профессиональной деятельности на основе освоенных знаний, умений и навыков. Перечисленные ключевые особенности компетенций приводят к нескольким выводам.

Во-первых, в качестве переменных на начальном этапе измерения можно выбрать знания и умения, освоение которых обеспечивает необходимые, но не достаточные условия для формирования компетенций. Во-вторых, для оценивания уровня освоения компетенций нужны ситуационные задания, имитирующие профессиональную деятельность. В-третьих, измерители для оценивания уровня освоения компетенций должны обладать высокой прогностической валидностью.⁸ В четвёртых, необходимо проводить тщательный анализ соответствия измеряемых переменных планируемыми конструктам, обеспечивая высокую конструктивную валидность результатов измерений.⁹

Необходимость обращения к ситуационным заданиям, нередко требующим при выполнении от обучающихся развёрнутых ответов, проведения анализа и привлечения дополнительной информации, приводит к появлению такого недостатка, как маленькая длина измерителя. В отличие от заданий с выбором ответов выполнение каждого ситуационного компетентностного задания требует значительного времени, поэтому обычно невозможно включить в тест не менее 25 заданий, которые необходимы для обеспечения достаточной надёжности измерений. Преодолению этой проблемы способствует введение элементов адаптивного тестирования в процесс многостадийного измерения. Благодаря адаптивности появляется возможность оптимизировать подбор трудности заданий

измерителя, минимизировав тем самым число заданий в нём.

Адаптивность отражена на рис. 2 с помощью модулей, число которых нарастает по мере продвижения по этапам измерения. На первом этапе весь тест состоит из одного модуля, по результатам выполнения которого группа обучающихся делится на две части: лучше и хуже подготовленную группу. Порогом для такого деления обычно выступает 50% уровень выполнения оценочного средства, выдаваемого на первом этапе и имеющего задания с множественным выбором, содержание которых предназначается для оценивания знаний и умений, лежащих в основе формирования компетенций.

Благодаря использованию формы заданий с выбором между первым и вторым этапами не нужен организационный перерыв, поскольку автоматизированная проверка результатов выполнения заданий первого модуля позволяет разделить группу обучающихся на две подгруппы за считанные минуты. Возможен усложнённый вариант деления группы обучающихся, когда для первого этапа устанавливается пороговый (проходной) балл и на две группы делятся только те обучающиеся, чьи результаты превысили его, а остальные считаются не соответствующими требованиям минимальной компетентности и выбывают из дальнейших испытаний.

На втором этапе оценочное средство включает два модуля: второй и третий. Второй модуль содержит более трудные компетентностные задания с конструируемыми ответами, а третий — более лёгкие задания той же формы. По результатам выполнения двух модулей две группы испытуемых делятся на три подгруппы. Поскольку для проверки ответов на задания с конструируемыми ответами необходимы эксперты, то между вторым и третьим этапами необходим организационный перерыв, после которого принимаются решения о выделении подгрупп обучающихся.

В первую подгруппу попадают обучающиеся, успешно выполнившие второй модуль (превысившие 50% порог или иной пороговый балл). На третьем

⁸ Taylor, Amy M. «The validity of personality trait interactions for the prediction of managerial job performance». Theses and Dissertations. University of South Florida Scholar Commons. 2008, 316 p.

⁹ Звонников В.И., Чельшкова М.Б. Оценка качества результатов обучения при аттестации. М.: Логос, 2012, 279 с.; Остапенко Р.И., Остапенко А.И. Использование методов моделирования структурными уравнениями в области управления образованием. Журнал «Государственный советник» Выпуск № 4 / 2013, с 24–32.

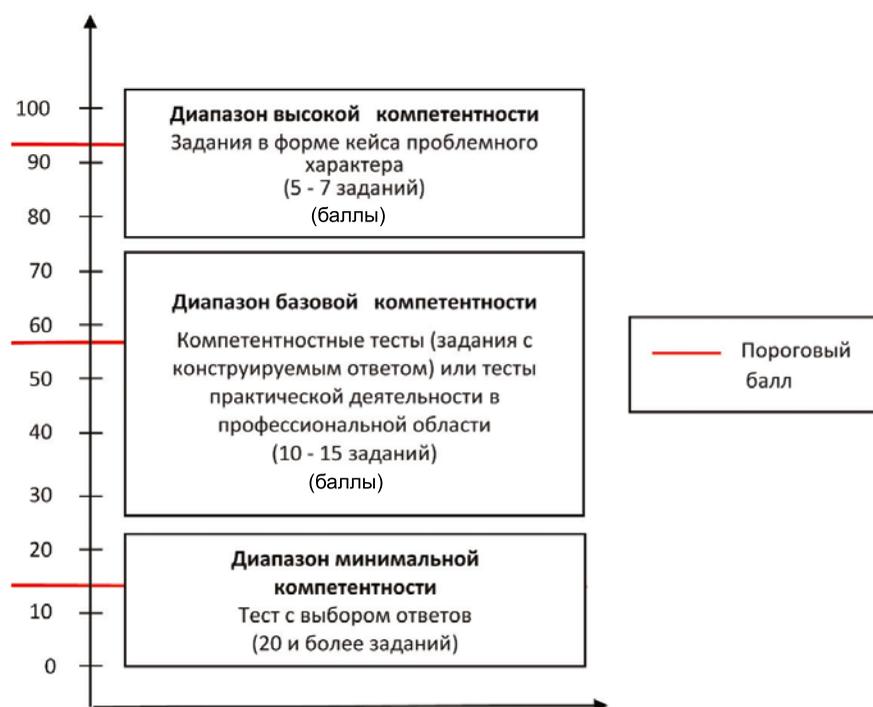


Рис. 4. Модель трёхстадийного измерителя

этапе они получают наиболее трудные задания в форме мини-кейсов четвертого модуля.

Вторая подгруппа объединяет худших испытуемых, не прошедших за пороговый балл по результатам выполнения второго модуля, и лучших испытуемых, прошедших за пороговый балл при выполнении заданий третьего модуля. Эти обучающиеся получают мини-кейсы средней трудности, включённые в пятый модуль.

Наконец, третья подгруппа предназначена для самых слабых обучающихся, не справившихся с заданиями третьего модуля. На третьем этапе им выдаются самые лёгкие мини-кейсы.

Благодаря адаптивности каждый испытуемый в подгруппах обучающихся не выполняет слишком лёгкие задания, с которыми он наверняка справится, или слишком трудные задания, с которыми он наверняка не справится. Вклад таких слишком лёгких или слишком трудных заданий в общую надёжность измерений невелик, поэтому оптимизация подбора заданий по трудности и минимизация их чис-

ла для каждого испытуемого не снизит общую надёжность теста.¹⁰

Таким образом, благодаря выделению подгрупп обучающихся и оптимизации трудности заданий, выдаваемых им при оценивании, обеспечивается общая высокая надёжность измерений, несмотря на незначительное число заданий измерителя, как, например, показано на рис. 4.

С учётом накопленного опыта зарубежных исследований и в контексте современных трактовок качества результатов образования можно выделить ряд наиболее актуальных направлений исследований, способствующих развитию многостадийных измерений в образовании. К ним относятся:

- развитие методологии совмещения качественных и количественных шкал при измерениях в образовании;
- выбор адекватных методов анализа данных при интеграции результатов об-

¹⁰ Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов. Под ред. В.И. Звонникова, М.Б. Челышковой. — М.: Логос, 2010. — 667 с.

учающихся при выполнении инновационных форм заданий измерителей, позволяющих оценить компетенции, творческие и практико-ориентированные аспекты подготовленности;

- развитие новых моделей IRT для параметрических и непараметрических методов оценивания в педагогических измерениях;

- разработка методик, алгоритмов и математико-статистического аппарата для обработки результатов тестирования на основе многомерных и инновационных моделей IRT в условиях многостадийных измерений;

- поиск путей повышения надёжности, содержательной и конструктивной валидности измерителей для обеспечения высо-

кой обоснованности оценок обучающихся по результатам педагогических измерений;

- разработка теоретико-методологического обеспечения адаптивных измерений в образовании;

- развитие методов анализа размерности пространства измерений в образовании.

В целом, можно сделать вывод о том, что проведение масштабных исследований по приведённым направлениям и применение аппарата метода моделирования структурными уравнениями открывают новые возможности в оценивании качества результатов образования в соответствие с существующими федеральными государственными образовательными стандартами.