

MEDICAL EDUCATION AND PROFESSIONAL DEVELOPMENT
International Journal for Medical Teachers

МЕДИЦИНСКОЕ ОБРАЗОВАНИЕ И ПРОФЕССИОНАЛЬНОЕ РАЗВИТИЕ

Журнал включен в Перечень российских рецензируемых научных журналов, которые рекомендованы Высшей аттестационной комиссией при Министерстве науки и высшего образования (ВАК) для публикации результатов диссертаций на соискание ученой степени кандидата и доктора наук.

Журнал индексируется в Ulrich's International Periodicals Directory, Российском индексе научного цитирования (РИНЦ, elibrary)

ЖУРНАЛ СООБЩЕСТВА МЕДИЦИНСКИХ ПРЕПОДАВАТЕЛЕЙ

Том 10, № 4 (36), 2019



ИЗДАТЕЛЬСКАЯ ГРУППА
«ГЭОТАР-Медиа»

СОДЕРЖАНИЕ

6 Редакционная статья

ТЕХНОЛОГИИ ОБУЧЕНИЯ

- 8 Пелл Г., Фуллер Р., Хомер М., Робертс Т.
Как измерить качество Объективного структурированного клинического экзамена: обзор показателей
- 30 Амлаев К.Р., Койчуева С.М., Ходжаян А.Б., Хрипунова А.А.
Технологии профессионального ориентирования и отбора в образовательные организации высшего медицинского образования в России и за рубежом
- 42 Булатов С.А., Лэйн Х.
Сравнительная оценка методологических подходов в оценке готовности выпускников медицинских вузов к самостоятельной деятельности
- 52 Запесоцкая И.В., Кузнецова А.А., Моргун Л.А., Данилова А.В.
Информационные, коммуникационные технологии в медицинском образовании
- 64 Кривых Е.А.
Обязательное медицинское страхование и приоритетные проекты системы здравоохранения как актуальные вопросы преподавания дисциплины «общественное здоровье и здравоохранение» в медицинских вузах Российской Федерации

КАДРОВОЕ РАЗВИТИЕ

- 74 Касимовская Н.А.
Реализация новых подходов в подготовке бакалавров сестринского дела: переход на уровень магистратуры «управление сестринской деятельностью»
- 84 Москвичева М.Г., Щепилина Е.С.
Анализ мнения руководителей медицинских организаций и медицинских работников по вопросам организации неотложной медицинской помощи
- 96 Трегубов В.Н., Якушина И.И., Романова И.Ю.
Оптимизация методики формирования жизненных целей при подготовке будущих врачей по специальности «медико-профилактическое дело»

ОПЫТ

- 106 Васева Е.М., Кныш О.И., Егорова А.О.
Биофармацевтическое исследование твердых лекарственных форм в учебном процессе с применением интерактивных методик обучения
- 116 Зорин К.В.
Этические аспекты концепции развития здравоохранения, подготовки и работы врача
- 126 Кузнецов В.В., Косилов К.В., Байрамов Р., Косилова Е.К., Ющенко Н.И., Смирнов Е.А.
Новый индекс здоровьесберегающего поведения студентов
- 144 Кныш О.И., Бышенко, В.В., Задираченко Л.Н.
О новом направлении подготовки провизоров для работы в ветеринарных аптечных организациях в Тюменском государственном медицинском университете
- 153 Избранные тезисы X международного форума РОСМЕДОБР-2019 «Инновационные обучающие технологии в медицине» (2–5 октября 2019 г., Санкт-Петербург)

162 **НОВОСТИ**

173 **АНОНСЫ НАУЧНЫХ СОБЫТИЙ**

CONTENT

Editorial	6
EDUCATIONAL TECHNOLOGIES	
Pell G., Fuller R., Homer M., Roberts T.	8
How to measure the quality of the OSCE: a review of metrics (translation into Russian edited by Alekseeva A.	
Amlaev K.R., Koychueva S.M., Khojayan A.B., Khripunova A.A.	30
Technologies of vocational guidance and occupational selection in graduate medical education organizations in Russia and abroad	
Bulatov S.A., Lane X.	42
Comparative evaluation of methodological approaches in assessing the readiness of medical graduates to independent activity	
Zapesotskaya I.V., Kuznetsova A.A., Morgun L.A., Danilova A.V.	52
Information and communication technologies in medical education	
Krivyh E.A.	64
Mandatory health insurance and priority projects of healthcare system as topical questions of teaching the discipline "public health and health care" at medical universities of the Russian Federation	
PROFESSIONAL DEVELOPEMENT	
Kasimovskaya N.A.	74
Implementation of new approaches in preparation of bachelors of nursing: transition to the level of masters "Management of nursing activities"	
Moskvicheva M.G., Shchepilina E.S.	84
Analysis of heads of medical organizations and medical stuff opinions on emergency medical care organization	
Tregubov V.N., Yakushina I.I., Romanova I.Yu.	96
Optimization of life goals setting methods during future doctors training in the specialty "medical and preventative care"	
EXPERIENCE	
Vaseva E.M., Knysh O.I., Egorova A.O.	106
Biopharmaceutical research of solid formulations in the educational process using interactive learning methods	
Zorin K.V.	116
Ethical implications of health development concept, physicians training and working	
Kuznetsov V.V., Kosilov K.V., Bayramov R., Kosilova E.K., Yuschenko N.I., Smirnov E.A.	126
The new index of health-preserving behavior of students	
Knysh O.I., Byshenko V.V., Zadirachenko L.N.	144
About a new direction of pharmacists` training at the Tyumen State Medical University for work in dispensaries of veterinary medicine	
Selected abstracts of the X international forum ROSMEDOBR-2019 "Innovative teaching technologies in medicine" (2–5th October 2019, St. Petersburg)	153
NEWS	162
ANNOUNCEMENTS	173

Уважаемые коллеги!

Подходит к концу 2019 г. – год, насыщенный изменениями в сфере медицинского образования, год, на который мы возлагали большие надежды в части формирования нормативной базы для непрерывного медицинского образования (НМО). К сожалению, не все проекты удалось реализовать в уходящем году: по-прежнему нет нового регулирования в сфере дополнительного образования в здравоохранении, но есть и определенный прогресс – запущен новый, очень удобный портал непрерывного медицинского и фармацевтического образования Минздрава России, в структуре НМО появились интерактивные обучающие инструменты, завершены работы по переносу личных кабинетов пользователей на единый портал НМО, что должно существенно облегчить участие в процессе медицинских работников.

Значительные изменения происходят и в сфере аккредитации специалистов: впервые в 2019 году была проведена первичная специализированная аккредитация выпускников ординатуры и программ переподготовки по 6 специальностям, разработаны контрольно-измерительные материалы по всем специальностям для проведения специализированной аккредитации в 2020 г.

Особенно важными в этих условиях становятся вопросы квалификации медицинских преподавателей. С 2010 г. АСМОК ведет активную деятельность в этом направлении: был основан журнал, ежегодно проводятся конференции, посвященные актуальным вопросам медицинского образования. В 2018 г. совместными усилиями АСМОК и Сеченовского университета был открыт Российский центр АМЭЕ, на базе которого регулярно проводятся курсы повышения квалификации для преподавателей.

В этом номере мы публикуем перевод очередного руководства АМЭЕ «Как измерить качество Объективного структурированного клинического экзамена: обзор показателей». Экзамен был внедрен во всех медицинских вузах с началом аккредитации специалистов, за прошедшие годы накопился материал для анализа, и мы надеемся, что данное руководство будет полезно для повышения качества экзамена.

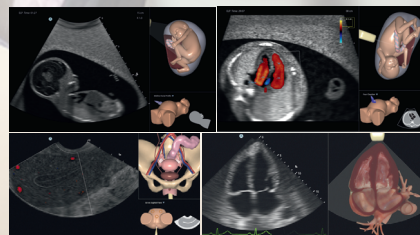
Мы желаем всем нашим читателям в наступающем 2020 г. совершенствоваться, получать удовольствие от работы и, конечно, помнить, что наше здоровье в руках тех, кого мы обучаем.

Залим Балкизов

СИМУЛЯТОР ДЛЯ ОБУЧЕНИЯ УЛЬТРАЗВУКОВЫМ ИССЛЕДОВАНИЯМ УЗИМЕНТОР



- Включает 12 валидированных модулей
- Опция виртуальной реальности позволяет ощутить полное погружение в клиническую среду, лучше понять ультразвуковую анатомию, освоить навыки манипулирования датчиком для получения оптимальных срезов
- Манекены мужского и женского тела с реалистичными пальпируемыми анатомическими ориентирами
- Имитация датчиков УЗИ
- УЗ-изображение со всеми деталями и артефактами



ГЭОТАР

решения для медицинского образования

115035, Москва, ул. Садовническая, д. 11, стр. 12.
Тел.: +7 (495) 921-39-07, +7 (916) 876-98-03.
E-mail: info@geotar-med.ru, www.geotar-med.ru

КАК ИЗМЕРИТЬ КАЧЕСТВО ОБЪЕКТИВНОГО СТРУКТУРИРОВАННОГО КЛИНИЧЕСКОГО ЭКЗАМЕНА: ОБЗОР ПОКАЗАТЕЛЕЙ*

Перевод с английского под редакцией эксперта Ассоциации профессиональных обществ по качеству медицинской помощи и медицинского образования Алексеевой А.Ю.

Пелл Г., Фуллер Р., Хомер М., Робертс Т.

Университет Лидса, Лидс, Великобритания

Все более частое использование методов оценки, основанных на применении специализированных критериев, в программах как додипломного, так и последипломного образования, ведет к необходимости обеспечить качество и надежность таких оценок. Очевидные вопросы встают перед теми, кто проводит оценку: как измерять это качество и какие механизмы позволили бы продемонстрировать улучшение качества оценки с течением времени? Есть несколько работ на эту тему, но лишь в отдельных статьях упоминают более 1-2 показателей качества при проведении объективного структурированного клинического экзамена (ОСКЭ).

В данном руководстве, предназначенном для специалистов по такой оценке, авторы ставят целью рассмотреть имеющиеся в настоящее время показатели, предназначенные для измерения качества оценки, и показывают, как можно получить обобщенную картину оценки качества ОСКЭ, используя несколько таких показателей, а также рассматривают, какие характеристики ОСКЭ и с помощью каких показателей лучше оценивать. Авторы обсудят вопросы качества как на уровне отдельных станций, так и в отношении процесса клинической оценки в целом, используя ряд рабочих примеров по проведению ОСКЭ в организации, в которой работают авторы.

HOW TO MEASURE THE QUALITY OF THE OSCE: A REVIEW OF METRICS

Pell G., Fuller R., Homer M., Roberts T.

University of Leeds, Leeds, West Yorkshire, England

With an increasing use of criterion based assessment techniques in both undergraduate and postgraduate healthcare programmes, there is a consequent need to ensure the quality and rigour of these assessments. The obvious question for those responsible for delivering assessment is how is this 'quality' measured, and what mechanisms might there be that allow improvements in assessment quality over time to be demonstrated? Whilst a small base of literature exists, few papers give more than one or two metrics as measures of quality in Objective Structured Clinical Examinations (OSCEs).

In this guide, aimed at assessment practitioners, the authors aim to review the metrics that are available for measuring quality and indicate how a rounded picture of OSCE assessment quality may be constructed by using a variety of such measures, and also to consider which characteristics of the OSCE are appropriately judged by which measure(s). The authors will discuss the quality issues both at the individual station level and across the complete clinical assessment as a whole, using a series of 'worked examples' drawn from OSCE data sets from the authors' institution.

* Данное руководство AMEE впервые было опубликовано в журнале «Medical Teacher»: Pell G., Fuller R., Homer M., Roberts T. How to measure the quality of the OSCE: A review of metrics. AMEE Guide #49. Medical Teacher. 2010; 32 (10): 802–11.

© AMEE 2011. ISBN: 978-1-903934-83-8

Как измерить качество Объективного структурированного клинического экзамена: обзор показателей

КЛЮЧЕВЫЕ МОМЕНТЫ

- Необходимо всегда оценивать качество экзаменов высокой важности, таких как объективный структурированный клинический экзамен (ОСКЭ), используя набор подходящих показателей.
- При оценке качества ОСКЭ очень важно использовать несколько показателей – это позволяет получить всестороннее представление о качестве оценки.
- Специалистам по оценке необходим набор инструментов, чтобы выявлять стандартные ошибки при оценке и избегать их в дальнейшем.
- Для значительного повышения качества оценки нужно сначала измерить его на уровне отдельных станций и внести необходимые улучшения, соотнося эти действия с общим контекстом всего процесса оценки в ходе ОСКЭ.
- Постоянное использование показателей оценки для повышения качества ОСКЭ позволяет четко оценить влияние вносимых изменений.
- Все более частое использование методов оценки, основанных на применении специализированных критериев, в программах до- и последипломного образования ведет к необходимости обеспечить качество и надежность таких оценок.

В обстановке все более пристально-го внимания к методам, которые используют при принятии решений высокого уровня в академических дисциплинах, критериальная оценка (КО) обеспечивает надежный структурированный методологический подход. В качестве методологии, основанной на компетенциях, КО позволяет выполнить итоговую оценку ответственных экзаменов (например, квалификационных экзаменов или экзаменов на получение ученой степени), обладает высокой надежностью и достоверностью. Это привлекательная методология оценки, имеющая ряд весомых преимуществ по сравнению с более традиционными, неструктурированными формами оценки (например, устные экзамены): она носит абсолютный характер, тщательно стандартизована для всех кандидатов, а порядок оценки четко спланирован и тесно увязан с целями профессиональной подготовки. Эти цели можно сравнить

с итогами обучения там, где они есть, со стандартами, которые разработаны регулирующими и лицензирующими органами как для студентов, так и для преподавателей. По этой причине методология КО широко применяется не только при итоговой оценке, но и в рамках целого ряда методов оценки на рабочем месте в самых разных академических дисциплинах (Norcini, 2007; Postgraduate Medical Education and Training Board, 2009). В настоящее время КО в Великобритании применяется также при наборе на работу врачей-ординаторов; при этом используют структурированное интервью, сходное с тем, что применяют при поступлении на программы высшего образования (Eva et al., 2004).

ОСКЭ использует принципы КО в рамках сложного процесса, который начинается с планирования содержания курса в соответствии с заранее определенными целями (Newble, 2004). Цель в данном случае – гарантировать, во-первых, что

оценивается верный стандарт, а во-вторых, что содержание ОСКЭ объективно совпадает с итогами обучения в соответствии с учебным планом. Подготовка экзаменуемого оценивается на уровне отдельных станций с помощью чек-листа, в который заносятся действия (или последовательность действий) экзаменуемого, а также в виде глобальной оценки, опирающейся на менее детерминированную общую оценку экзаменаторов (Cohen, 1997; Regehr, 1998).

Ключевое значение для успешной КО имеют обеспечение достаточного качества оценки и разработка жестких стандартов на основе ряда показателей, которые обеспечивают оценку в целом, а не просто сосредоточены на результатах, продемонстрированных кандидатами (Roberts, 2006). Рассматривать оценку крайне важно, так как проведение ОСКЭ – сложное и ресурсоемкое мероприятие, в котором обычно задействовано много экзаменаторов, кандидатов, тренажеров и пациентов и которое часто происходит параллельно в нескольких центрах. Такая сложность означает, что возможны трудности со стандартизацией КО и она очень сильно зависит от поведения экзаменатора, даже при таком механизме контроля, как чек-лист. Невозможно значимо оценить качество проведения экзамена по какому-то одному показателю, так же как невозможно оценить, к примеру, клиническую компетентность студента по результатам одного экзамена. Именно поэтому при проведении КО критическую роль играют понимание сути показателей и умение эффективно их использовать как при оценке качества, так и при распределении ресурсов для соответствующих дальнейших исследований и разработок экзаменов (Wass, 2001).

КОНЦЕПЦИЯ КАЧЕСТВА ПРИ ПРОВЕДЕНИИ ОБЪЕКТИВНОГО СТРУКТУРИРОВАННОГО КЛИНИЧЕСКОГО ЭКЗАМЕНА – ОБЩИЕ ПРИНЦИПЫ

В данном руководстве рассматриваются применяемые в последние годы показатели с использованием результатов ОСКЭ у выпускников в качестве примеров того, насколько точно с помощью этих показателей можно измерить качество оценки. Важно осознавать, что обзор показателей, используемых при ОСКЭ, – это лишь часть общего процесса оценки качества ОСКЭ, что требует учета всех взаимосвязей в ходе процесса оценки в целом (рис. 1).

При использовании ОСКЭ как составной части национального экзамена станции разрабатываются централизованно, исходя из общего стандарта, и, как правило, централизованно представляются центрам, проводящим экзамены. Однако на местном уровне, при экзаменах, разработанных в конкретных медицинских институтах, возможны некоторые вариации, например в максимальных показателях для каждой станции, и это зависит от важности и сложности данной станции для тех, кто готовит экзамен. Эти абсолютные различия между оценочными станциями отрицательно скажутся на надежности показателя, делая часто упоминаемое значение 0,9 недостижимым. Стандартизировать данные ОСКЭ и таким образом добиться большей надежности показателя можно, но это не будет полным отражением оценки с учетом целей тех, кто ее проводит. Данное руководство в первую очередь нацелено на тех, кто участвует в клинической оценке на местном уровне в отдельных медицин-

Как измерить качество Объективного структурированного клинического экзамена: обзор показателей



Рис. 1. Обеспечение и повышение качества объективного структурированного клинического экзамена – комплексный процесс.

ских институтах, где оценка может проходить сразу в нескольких местах, но под единым руководством. У тех, кто участвует в клинической оценке на национальном уровне, может быть иная точка зрения.

Какой метод разработки стандартов использовать

Метод разработки стандартов определяет, какие показатели можно использовать для оценки качества. Стандарты могут быть относительными (например, нормативно-ориентированными) или абсолютными, основанными либо на предмете оценки (методы Ebel и Angoff), либо на

уровне подготовки кандидата (пороговые методы). Поскольку к стандартам предъявляют такие требования, как оправданность, научная обоснованность и приемлемость (Norcini, 2003), в большинстве случаев используют абсолютные стандарты. При этом все методы разработки стандартов включают ряд ретроспективных показателей: например, доля успешно сдавших экзамен на данной станции, фиксированные эффекты (время оценки, сравнение между центрами проведения экзамена) или распределение оценок по частоте, но важно выбрать метод разработки стандарта, который позволяет получить дополнительные

Таблица 1. Сравнение пороговых методов разработки стандартов

Пограничные и контрастирующие группы	Пороговая регрессия
• Простота расчетов. • Нужны только 3 глобальные оценки («не сдано», пороговый результат, «сдано»). • Используют лишь данные по пороговым результатам и только соотношение взаимодействий экзаменатора с кандидатом. • Необходимо достаточное количество кандидатов в пороговой группе (≥ 20). • Дает ограниченное число показателей качества	• Расчеты требуют более глубоких навыков. • Обычно 5 глобальных оценок (например, «не сдано», пороговый результат, «сдано», «сдано с отличием», «сдано с блеском»). • При анализе используются все взаимодействия экзаменатора с кандидатом. • Не требует наличия студентов с пороговыми результатами. • Более широкий спектр показателей качества

показатели качества. В настоящее время во многих организациях предпочитают пороговые методы, но только метод регрессии позволяет получить данные о связи между глобальной оценкой и результатами чек-листа, а также по уровню различий между слабыми и сильными студентами. В табл. 1 перечислены основные различия между несколькими пограничными методами оценки и описан их вклад в показатели качества.

Авторы предпочитают метод пороговой регрессии, так как при этом используются все взаимодействия между экзаменаторами и кандидатами, и эти взаимодействия реальны. Метод объективно основан на заранее выбранных критериях, с использованием большого числа экзаменаторов и дает широкий диапазон показателей.

Один из критических аргументов против метода пороговой регрессии – его возможная чувствительность к выпадающим показателям. Такие выпадающие показатели можно разбить на 3 основные группы:

- студенты с очень низкими результатами, у которых сумма баллов в чек-листе близка к нулю;
- студенты с достаточно высокими результатами по чек-листу, которые, однако, произвели на экзаменатора неблагоприятное впечатление;
- экзаменатор, выставяющий неверную глобальную оценку.

Эти вопросы будут подробнее обсуждаться в соответствующих разделах данного руководства.

Как получать показатели качества на уровне станций

В табл. 2 приведен стандартный набор показателей для типичного ОСКЭ (20 станций за 2 дня, общее время тестирования – около 3 ч, 4 центра проведения экзамена). При этом обычно задействованы около 250 кандидатов, 500 экзаменаторов и 150 стандартизованных пациентов (СП), пациенты-волонтеры со стабильными клиническими признаками (используются для физического обследования). Кандидаты должны выполнить требования сдачи экзамена, которые включают общий проходной балл, минимальное число станций (предотвращает избыточную компенсацию оценки за счет других станций и добавляет соответствие требованию «компетентный, всесторонне подготовленный врач») и минимальное число приемлемых оценок со стороны пациентов. Экзаменаторы заполняют чек-лист, а затем выставляют глобальную оценку (глобальные оценки в наших ОСКЭ записываются в численном виде следующим образом: 0 – однозначно не сдано; 1 – пороговый результат; 2 – сдано; 3 – очень хорошо сдано; 4 – сдано на отлично).

Как измерить качество Объективного структурированного клинического экзамена: обзор показателей

Для разработки стандартов использовали метод пороговой регрессии (Pell, Roberts, 2006). Как правило, такой ОСКЭ дает приблизительно 60 000 элементов данных (т.е. отметок в отдельных чек-листах), что в совокупности представляет ценный источник для измерения качества оценки и ее последующего повышения. В результате использования таких данных мы видели, как наши ОСКЭ все более модернизируются, при этом сохраняя или повышая уровень надежности.

При всех пороговых методах разработки стандартов с выставлением глобальной оценки в дополнение к оценкам в баллах из чек-листов сопутствующие показатели позволяют измерить качество оценки. Для других типов разработки стандартов, при которых подобная глобальная оценка не составляет часть процедуры разработки стандартов (например, методы Ebel и Angoff), показатели дискриминации между классами и коэффициента детерминации (R^2) не применяются (Cusimano, 1996).

При всех пороговых методах разработки стандартов с выставлением глобальной оценки в дополнение к оценкам в баллах из чек-листов сопутствующие показатели позволяют измерить качество оценки (см. табл. 2).

ПОКАЗАТЕЛЬ 1. КОЭФФИЦИЕНТ НАДЕЖНОСТИ АЛЬФА КРОНБАХА

Выборка общих суммарных показателей в этом руководстве будет нужна для иллюстрации использования психометрических данных в действии и для того, чтобы обрисовать подходы к выявлению низкого качества оценки на уровне отдельных станций и к принятию соответствующих мер. Мы выбрали более старые данные по ОСКЭ, что позволяет лучше ос-

ветить вопросы качества и последующие действия по улучшению результатов.

Это мера внутреннего постоянства (что часто, хотя не вполне точно расценивается как надежность): при хорошей оценке лучше подготовленные студенты должны показывать относительно лучшие результаты на всех станциях (показатели в контрольных перечнях для каждой станции). Коэффициент альфа можно рассчитывать в двух вариантах: без стандартизации или стандартизованный; и в данном руководстве мы имеем в виду коэффициент без стандартизации (вариант, по умолчанию используемый в Пакете программ обработки статистических данных социальных наук, Statistical Package for the Social Sciences – SPSS). Это показатель средней взаимной корреляции, взвешенной по дисперсиям, который дает то же значение, что и коэффициент G для простой модели с сопоставлением показателей оценки и кандидатов. Значение (общее) коэффициента альфа, которое обычно считается приемлемым при важных с использованием стандартизованных моделей и реальных пациентов, когда показатели для отдельных станций не стандартизованы, составляет $\geq 0,7$.

Если показатели станции стандартизованы, ожидаемое значение коэффициента альфа выше. Коэффициент альфа для этого набора станций составлял 0,754, и можно видеть (см. 2-й столбец в табл. 2), что ни одна из станций не уменьшала общей надежности оценки, хотя станции 17 и 20 мало что добавляли к результату.

Поскольку коэффициент альфа возрастает с увеличением числа элементов оценки, итоговые показатели «альфа при исключении элемента оценки» должны быть ниже общих коэффициентов альфа,

если элемент оценки/станция пройдена хорошо. Это может быть не так по следующим причинам:

- данный элемент оценки измеряет нечто отличное от остальных элементов набора;
- элемент оценки выбран неудачно;
- есть проблемы, связанные с преподаванием: либо оцениваемую тему преподавали плохо, либо разным группам кандидатов преподавали по разным стандартам;
- экзаменаторы проводят оценку по общим стандартам.

В подобных обстоятельствах нужно повысить качество оценки, а для этого пересматривают станции, дизайны чек-листа и самой станции либо оценивают качество преподавания в рамках учебного плана.

Однако полагаться только на коэффициент альфа как на показатель качества оценки нельзя. Как мы продемонстрировали, при возрастании числа элементов оценки возрастает и коэффициент альфа, поэтому шкала оценки может выглядеть более однородной, чем на самом деле, если она содержит достаточно элементов. Это означает, что при объединении 2 шкал, измеряющих разные наборы показателей, в одну длинную шкалу, в результате может быть получен неоправданно высокий коэффициент альфа. Кроме того, набор элементов может дать высокий коэффициент альфа и при этом быть многофакторным. Это случается при наличии отдельных наборов элементов (измеряющих разные параметры) с высокой взаимной корреляцией, притом что сами наборы данных не слишком коррелируют друг с другом.

Коэффициент альфа может быть слишком высоким (например, $>0,9$), что, вероятно, указывает на избыточность

оценки; в то же время низкие значения коэффициента альфа иногда могут быть обусловлены значительными различиями в средних баллах для разных станций, а не плохой разработкой станций.

Нужно подчеркнуть, что в медицинском вузе у авторов руководства и во многих других подобных институтах в Великобритании во время проведения ОСКЭ требуется свыше 1000 экзаменаторов (обычно масштаб проводимого экзамена в 2–3 раза превышает вышеописанный). Таким образом, набор достаточного количества экзаменаторов приемлемого качества – это вечная проблема, поэтому невозможно внедрить двойную оценку, которая сделает целесообразным использование коэффициента обобщаемости (G) для более точного количественного определения различий между экзаменаторами. Анализы такого типа сложнее, чем те, что рассмотрены в данном руководстве, и часто требуют использования дополнительного, более сложного программного обеспечения. Решение на уровне организации: что использовать – коэффициент G или коэффициент альфа Кронбаха, – должно учитывать требования к характеру учебного процесса и недостатки, присущие методу.

Эффект «ястребов и голубей»*, и в рамках одной станции, и в целом, приводит к нежелательным эффектам, в том числе к завышенному значению коэффициента альфа. Однако, вероятно, что это приведет к неудовлетворительным значениям коэффициента детерминации, дисперсии между группами внутри одной станции и, возможно, к фиксированным различиям между центрами проведения экзамена, которые будут рассмотрены в данном руководстве ниже. Мы полагаем, что одного показателя, в том числе коэффициента альфа, всегда недостаточно

Как измерить качество Объективного структурированного клинического экзамена: обзор показателей

для оценки качества, и в случае ОСКЭ с высоким значением коэффициента альфа, но с низкими значениями других показателей о высоком качестве оценки говорить нельзя.

В качестве другого способа оценки, отличного от определения коэффициента альфа при исключении элемента, можно использовать корреляцию результата на конкретной станции с общим показателем за вычетом результата на данной станции. Это позволит получить более протяженную шкалу, но значение данных (т.е. корреляция) между вкладом в надежность или снижение надежности до некоторой степени зависят от дизайна оценки, поэтому труднее поддаются интерпретации.

ПОКАЗАТЕЛЬ 2. КОЭФФИЦИЕНТ ДЕТЕРМИНАЦИИ R2

Коэффициент R2 представляет пропорциональное изменение зависимой переменной (результат в чек-листе), обусловленное изменениями в независимой переменной (глобальная оценка). Это позволяет оценить степень (линейной) корреляции между результатом в чек-листе и глобальной оценкой для каждой станции с ожиданием, что более высокие глобальные оценки в целом будут соответствовать более высоким результатам по чек-листам. Квадратный корень из коэффициента детерминации представляет собой простой коэффициент корреляции по Пирсону. SPSS и другие статистические программные пакеты также позволяют рассчитать скорректированное значение R2, учитывающее размер выборки и число прогностических факторов в модели (в данном случае один); в идеале это

значение должно быть близко к нескорректированному значению.

Хорошая корреляция ($R^2 > 0,5$) указывает на достаточную связь между результатами в чек-листе и глобальной оценкой, но необходимо убедиться, что чрезмерно детализированные общие показатели не переводятся экзаменаторами автоматически в соответствующий результат в чек-листе, искусственно завышая R2. В табл. 2 у станции 14 (оценка практических и медико-юридических навыков) высокое значение коэффициента R2 (0,697) означает, что 69,7% различий в глобальных оценках студентов объясняется различиями в их результатах по чек-листам.

Каждое решение на уровне организации: что использовать – коэффициент G или альфа Кронбаха, – должно учитывать требования к характеру учебного процесса и существующие ограничения.

Станция 19 отличается худшими показателями – значение R2 равно 0,404. Это была новая станция, посвященная обеспечению безопасности пациентов и действиям при травме от укола иглой. Чтобы понять, почему значение R2 низкое, полезно представить связь между результатами по чек-листу и глобальными оценками графически (например, используя функцию SPSS «Оценка кривых»), чтобы исследовать точный характер этой связи (см. рис. 2). На рис. 2 глобальные оценки, выставленные экзаменаторами, отложены по оси абсцисс, а общие результаты по чек-листам – по оси ординат. Сгруппированные результаты чек-листов обозначены размером черных кружков (см. рис. 2). Программа SPSS может рассчитывать коэффициент R2 для многочленов разной степени и, таким образом, обеспе-

**Тенденция одних преподавателей завышать, а других занижать оценки всем экзаменуемым.*

Таблица 2. Показатели объективного структурированного клинического экзамена последнего года обучения

Стан-ция	Коэффициент альфа Кронбаха при исключении элемента оценки	R2	Дискриминация между классами	Количество результатов «не сдано»	Различия между группами, %
1	0,745	0,465	4,21	53	31,1
2	0,742	0,590	5,23	24	30,1
3	0,738	0,555	5,14	39	33,0
4	0,742	0,598	4,38	39	28,0
5	0,732	0,511	4,14	29	20,5
6	0,750	0,452	4,74	43	40,3
7	0,739	0,579	4,51	36	19,5
8	0,749	0,487	3,45	39	33,8
9	0,744	0,540	4,06	30	36,0
10	0,747	0,582	3,91	26	29,9
11	0,744	0,512	4,68	37	37,6
12	0,744	0,556	2,80	23	32,3
13	0,746	0,678	3,99	30	22,0
14	0,746	0,697	5,27	54	27,3
15	0,739	0,594	3,49	44	25,9
16	0,737	0,596	3,46	41	34,3
17	0,753	0,573	3,58	49	46,5
18	0,745	0,592	2,42	15	25,4
19	0,749	0,404	3,22	52	39,5
20	0,754	0,565	4,50	37	34,1

чивать дополнительную информацию по степени линейности взаимосвязи. Мы рекомендуем всегда строить диаграмму рассеяния для результатов по чек-листам относительно глобальных оценок как часть стандартной практики, независимо от показателей для данной станции.

На станции 19 мы видим две основные проблемы – широкий разброс результатов для каждого показателя глобальной оценки и очень широкий разброс результатов для глобальной оценки «не сдано» (0 на оси абсцисс). Это указывает на то, что некоторые студенты получили много положительных отметок по чек-листу,

однако общий результат экзамена не удовлетворил экзаменатора, что привело к глобальной оценке «не сдано».

Во введении мы уже говорили о влиянии выпадающих показателей на метод регрессии. Примеры низких оценок по чек-листам с достаточно высокими глобальными оценками показаны на рис. 3. На других станциях мы иногда видим, что кандидаты получили очень мало отметок в чек-листах. Это ведет к уменьшению отрезка, отсекаемого на вертикальной оси уравнением регрессии, и возрастанию угла наклона графика уравнения регрессии. Для данных в табл. 2 исключение выпадающих значений с перерас-

Как измерить качество Объективного структурированного клинического экзамена: обзор показателей

четом общего проходного балла и проходного балла для отдельных станций очень мало влияет на результаты, увеличивая проходной балл менее чем на 0,2%.

Мы рекомендуем всегда строить диаграмму рассеяния для результатов по чек-листам относительно глобальных оценок в рамках стандартной практики независимо от показателей для данной станции.

Эта неудовлетворительная связь между оценками по чек-листам и глобальными оценками приводит к некоторой нелинейности, что видно из табл. 3 (по результатам расчета в программе SPSS), где хорошо заметно, что лучше всего данную связь можно описать кубическим уравнением. Следует от-

метить, что с математической точки зрения кубическое уравнение всегда обеспечивает наилучшее описание, однако принципы целесообразности требуют, чтобы различие между двумя кривыми было статистически значимо для выбора модели более высокого порядка. В данном примере приближение для кубического многочлена статистически значимо лучше, чем для линейной модели. В таких случаях важно понять, чем является кубическое уравнение – результатом взаимосвязи результатов в чек-листе с глобальной оценкой или это связано с выпадающими показателями вследствие плохо продуманного дизайна чек-листа либо неприемлемого образа действий экзаменатора

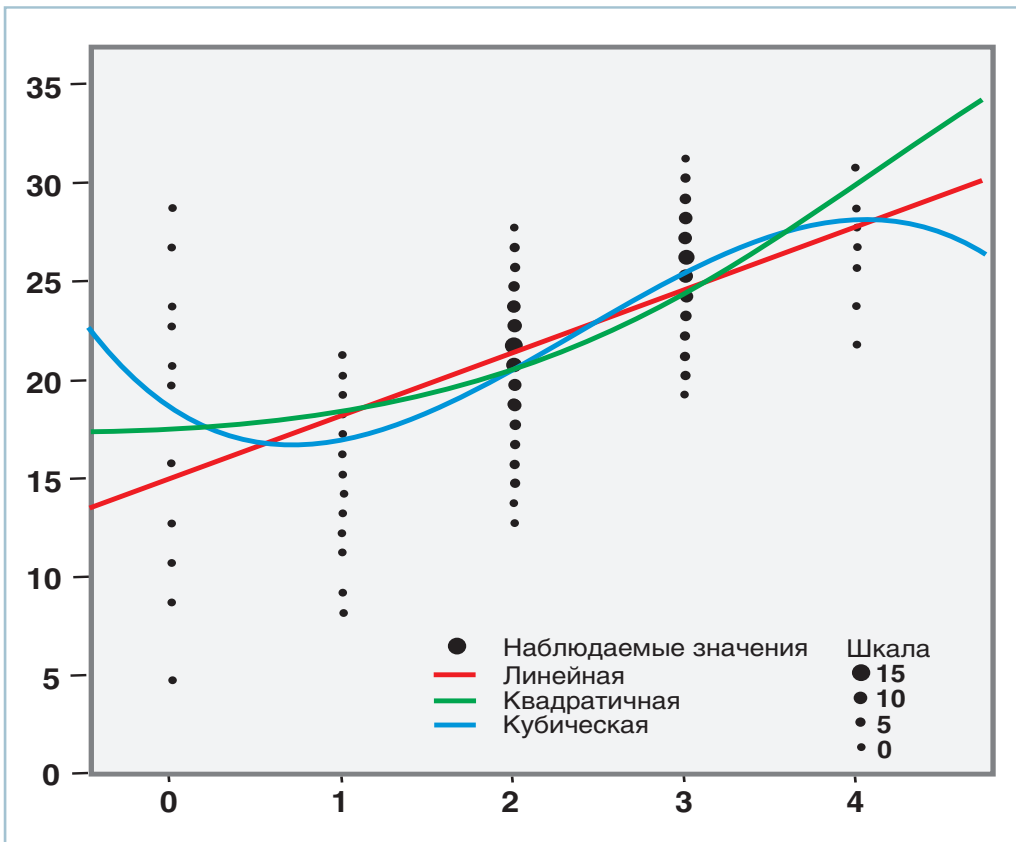


Рис. 2. Кривая оценок (станция 19) – результаты по чек-листам (y) и глобальная оценка (x)

при выставлении оценок в чек-листе. Оценивая это, читатели должны изучить распределение оценок, которое можно увидеть на диаграмме рассеяния. Наш опыт свидетельствует, что при хороших показателях на станциях отклонение от строгой линейности не должно вызывать беспокойство (см табл. 3).

Низкие значения R2 и/или широкий разброс баллов данной оценки для некоторых станций должны побудить к пересмотру чек-листа по данной теме либо дизайна станции. В данном случае, хотя и планировалось, что основной акцент при экзамене на данной станции будет сделан на безопасных, эффективных действиях, перепроверка чек-листа в свете полученных показателей продемонстрировала, что этот акцент был недостаточно четким. Более слабые кандидаты были в состоянии получить достаточно много отметок в чек-листе за выполнение, однако они не удовлетворяли более высоким ожиданиям (акцент на принятие решений). Ситуацию исправили, переписав станцию и чек-лист по ней, чтобы при следующем ОСКЭ заново использовать эту станцию с последующим анализом полученных на ней результатов.

ПОКАЗАТЕЛЬ 3. ДИСКРИМИНАЦИЯ МЕЖДУ КЛАССАМИ

Данный статистический показатель дает угол наклона кривой регрессии и показывает среднее возрастание оценки в чек-листе при возрастании на один уровень по шкале глобальной оценки. Хотя

четкого руководства или идеальных значений не существует, мы рекомендуем, чтобы показатель дискриминации составлял около 0,1 максимальной возможной оценки в чек-листе (по нашим данным, это обычно составляет 30–35).

Низкое значение дискриминации между классами часто сопровождается неудовлетворительными значениями других показателей для данной станции, например, низкими значениями R2 (указывает на недостаточную связь между глобальной оценкой и результатами в чек-листе) или высокой дисперсией ошибки у экзаменатора (см. ниже показатель 5), в тех случаях, когда экзаменаторы не использовали общий стандарт. Слишком высокие значения дискриминации между классами могут указывать либо на слишком низкий проходной балл, либо на отсутствие линейности вследствие небольшого числа не сдавших студентов (они делают подъем кривой регрессии более крутым).

При очень низких оценках студентов по баллам в чек-листе следует рассмотреть вопрос, нужно ли исключать эти очень низкие баллы из разработки стандарта, чтобы избежать избыточного влияния на общий проходной балл экзамена в сторону снижения.

Возвращаясь к табл. 2, мы видим, что показатели дискриминации между классами в целом приемлемы для разных станций (максимальные оценки для станций находятся в диапазоне 30–35 пунктов), хотя есть 3 станции, показатели дискриминации для

Таблица 3. Таблица кривой оценок (оценочная станция 19)

Подобранный многочлен	R-квадрат	F	df1	df2	Sig.
Линейный	.401	159.889	1	239	.000
Квадратичный	.435	91.779	2	238	.000
Кубический	.470	70.083	3	237	.000

Как измерить качество Объективного структурированного клинического экзамена: обзор показателей

которых выше 5 (например, станция 14 – станция оценки навыков, включающая заполнение формы для кремации умершего).

Если есть сомнения о прохождении станции, основанные на показателе дискриминации, зачастую информативными оказываются возвращение к показателю дисперсии R2 и повторная оценка кривой. В табл. 2 у станции 14 наиболее высокий показатель дискриминации между классами, а из рис. 3 видно, что для большинства показателей глобальной оценки вновь очень широк разброс баллов по чек-листам, особенно для оценки «сдано» – значение 2 на оси абсцисс в диапазоне от 4 до 27; однако самые низкие из этих значений явно выпадают. Поскольку остальные показатели для этой станции приемлемы, в эту станцию можно не вносить изменения, однако она требует пристального наблюдения при дальнейшей оценке.

ПОКАЗАТЕЛЬ 4. ЧИСЛО РЕЗУЛЬТАТОВ «НЕ СДАНО»

Было бы ошибочно автоматически считать, что необычно высокое число результатов «не сдано» указывает на слишком высокую сложность станции. «Проверка на соответствие действительности» – неотъемлемая часть пороговых методов, в значительной степени она компенсирует сложность станции. Она представляет собой экспертное суждение, сделанное квалифицированными экзаменаторами, по определению глобальной оценки при результатах, ожидаемых от минимально компетентного студента.

Как уже было сказано выше, при изучении дизайна оценочной станции и получаемых на ней результатов для выявления проблем могут использоваться другие пси-

хометрические данные. Частоту отметок «не сдано» можно использовать для оценки влияния изменений в преподавании определенной темы: если частота таких результатов возрастает, пересмотр содержания курса и методов обучения может улучшить его преподавание. Значимых выпадающих значений для этого показателя в табл. 2 нет, но проблемы со станцией 19 позволили нам их выявить и модифицировать преподавание, касающееся вопросов безопасности пациента, в программе последнего года обучения, а также повысить внимание к этим вопросам безопасности в чек-листах.

ПОКАЗАТЕЛЬ 5. РАЗЛИЧИЯ МЕЖДУ ГРУППАМИ (ВКЛЮЧАЯ ЭФФЕКТЫ СО СТОРОНЫ ЭКЗАМЕНАТОРА)

При анализе данных, полученных в результате комплексных оценок, таких как ОСКЭ, где студентов в практических целях неизбежно приходится делить на группы, крайне важно, чтобы дизайн исследования был полностью рандомизирован. Однако это не всегда возможно по чисто организационным причинам, например наличие студентов с особыми потребностями, которым необходимо больше времени и которых необходимо экзаменовать в рамках отдельного цикла. Любые случайно подобранные подгруппы нужно исключать из статистических анализов, которые в качестве ключевого допущения опираются на случайный характер данных.

В идеальном процессе оценки все различия в них будут обусловлены различиями в подготовленности студентов, а не в окружении (например, местные вариации в устройстве станции или оборудовании), расположении (например, центры проведения экзамена в больницах

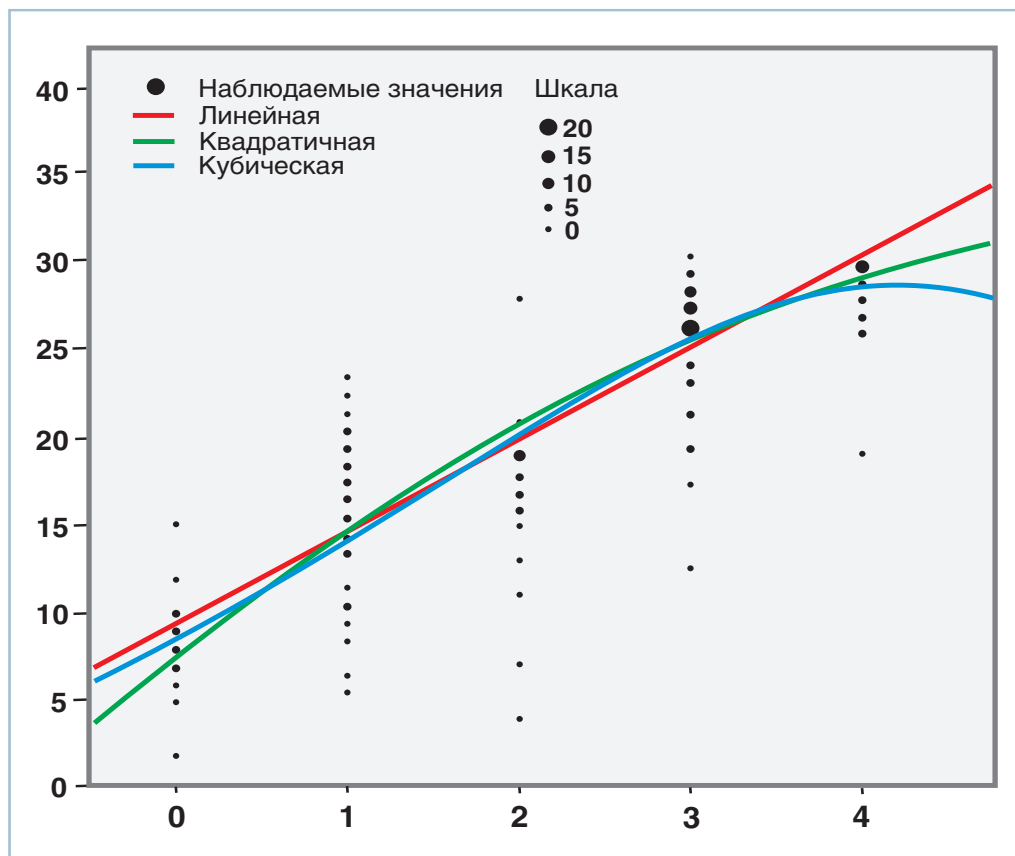


Рис. 3. Кривая оценок (оценочная станция 14) – результаты, выставленные экзаменатором в контрольном перечне (y), и общая оценка (x)

с разными местными подходами к ведению определенных клинических состояний) или в отношении экзаменаторов («ястребы и голуби»). Существуют два способа измерения таких эффектов: либо посредством однофакторного дисперсионного анализа данных по станции (например, с экзаменатором в качестве фиксированного эффекта), либо посредством расчета доли общей дисперсии, которая специфична для каждой группы. 2-й способ позволяет оценить долю различий баллов в чек-листах, которая обусловлена подготовкой студентов, в отличие от других возможных факторов, перечисленных выше, хотя обычно этот показатель

указывается как доля дисперсии, специфичная для данной цепочки.

Если компоненты дисперсии рассчитываются с использованием группы (т.е. цепочки) в качестве случайного эффекта, можно рассчитать процентную долю дисперсии, специфичную для группы. Это очень мощный критерий – он очень хорошо отражает единообразие процесса оценки для разных групп. Кроме того, он сравнительно просто рассчитывается. В идеале, различия между группами не должны превышать 30%, а значения, превышающие 40%, должны служить поводом для беспокойства, так как они указывают на возможные проблемы на

Как измерить качество Объективного структурированного клинического экзамена: обзор показателей

уровне станции, связанные с непоследовательным поведением экзаменатора и/или с другими специфическими характеристиками цепочки, а не с подготовленностью студентов.

Исходя из табл. 2 станции 6, 17 и 19 дают повод для беспокойства на основании этого показателя – различия между группами для них наиболее высоки. Кроме того, у станции 6 низкий коэффициент R^2 , а общее сочетание неудовлетворительных значений показателей для нее указывает на то, что низкий коэффициент R^2 , вероятно, обусловлен плохо разработанным чек-листом. Эти наблюдения привели к пересмотру дизайна станции 6: было обнаружено, что чек-лист содержал много критериев низкого уровня, благодаря которым плохо подготовленные кандидаты могут набрать высокие баллы, просто выполняя определенные действия. Иными словами, вероятно расхождение между содержанием чек-листа, целями и задачами данной станции в понимании экзаменаторов. Поэтому при переработке данной станции ряд критериев низкого уровня объединили, чтобы сформировать критерий более высокого уровня и облегчить оценку процессов более высокого уровня, как это было задумано изначально.

Со станцией 17 проблема иная – высокое значение R^2 в сочетании с высоким показателем различий между группами свидетельствует о том, что экзаменаторы проводят оценку внутри разных групп систематично, однако между группами заметен отчетливый эффект «ястребов и голубей». В подобном случае следует изучить этот результат посредством однофакторного дисперсионного анализа, чтобы определить, чем это обусловлено – конкретным экзаменатором или местом проведения экзамена. Дисперсия, обусловленная разными центрами про-

ведения экзамена, включается в простой расчет дисперсии между группами внутри одной станции, как описано выше. Однако ее значимость можно определить посредством однофакторного дисперсионного анализа с центрами проведения экзамена в качестве фиксированных эффектов.

Выносить суждения, основанные на одном показателе, надо с осторожностью, так как при достаточно крупных выборках дисперсионный анализ данных по отдельным станциям с большой вероятностью даст по крайней мере один значимый результат вследствие ошибки I типа, обусловленной множественными проверками значимости между большим числом групп (например, в результатах наших собственных ОСКЭ выборка составляла 250 студентов и приблизительно 15 цепочек в разных центрах проведения экзамена). Тщательный ретроспективный анализ выявил значимые эффекты «ястребов и голубей»; специфические группы необходимо проследить по другим станциям, чтобы оценить общий уровень подготовки. Если использовали модель полностью рандомизированной оценки студентов и экзаменаторов (следует помнить об оговорках о местных различиях в оборудовании и организации экзамена), многие из этих эффектов в значительной степени должны самоликвидоваться; специфичные для групп фиксированные эффекты важны в суммарной оценке, в этом случае они могут требовать корректирующих действий.

ПОКАЗАТЕЛЬ 6. РАЗЛИЧИЯ МЕЖДУ ГРУППАМИ (ДРУГИЕ ЭФФЕКТЫ)

Дисперсионный анализ может использоваться также при неслучайном распределении либо экзаменаторов, либо студентов, как бывает в некоторых медицинских вузах

с большим числом студентов и учебными клиниками, где оценка может происходить в нескольких экзаменационных центрах одновременно. При столь сложной организации возможно неслучайное распределение экзаменаторов по цепочкам, так как персоналу клиники зачастую трудно оставить свое место работы. Впоследствии это может привести к значительным различиям, обусловленным эффектами места проведения экзамена, которые можно выявить посредством соответствующих действий при анализе результатов.

Дисперсионный анализ поможет выявить и другие важные фиксированные эффекты. Например, эффекты в отношении подготовки экзаменатора, гендерные эффекты в отношении персонала/студентов и связанные взаимодействия – все они были описаны ранее (Pell, 2008) и подчеркивают необходимость всестороннего и углубленного обучения экзаменаторов, как уже было сказано выше (Holmboe, 2004).

ПОКАЗАТЕЛЬ 7. ОЦЕНКИ СТАНДАРТИЗОВАННЫХ ПАЦИЕНТОВ

В большинстве центров, где используются СП, от них требуют оценивать кандидатов, и этому обычно предшествует интенсивная программа обучения. В нашей организации СП задают вопрос, например: «Хотели бы вы снова консультироваться у этого врача?», с вариантами ответов («да, безусловно», «да», «ни да, ни нет», «нет», «нет, ни в коем случае»); 2 последних ответа рассматриваются как отрицательные. Как и в случае показателя 4 (число результатов «не сдано» для данной станции), доля кандидатов выше обычной (например, >10%) с отрицательными оценками от СП может указывать

на определенные проблемы. Литературных данных о том, что составляет приемлемый диапазон оценок СП на уровне станции, нет, поэтому мы выбрали произвольное предельное значение, равное 10%. Критическим вопросом в данном случае является необходимость оценить другие показатели для данной станции и то, как влияет на оценки СП их обучение или другие вмешательства.

В сочетании с более высокой частотой результатов «не сдано», чем обычно, это может быть результатом некачественного преподавания данной темы. Высокие значения этого показателя часто сопровождаются высокими значениями дисперсии между группами; экзаменаторы, сталкиваясь с кандидатами, демонстрирующими уровень компетенций ниже ожидаемого, часто не могут сохранять постоянство оценок.

Общую надежность оценки можно повысить, добавив оценку со стороны СП к общему баллу чек-листа; обычно оценка со стороны СП должна составлять 10–20% общего балла для станции (Homer, Pell, 2009). Другой подход, принятый в нашей организации при проведении выпускного ОСКЭ, – установить минимальные требования для комментариев СП как косвенный показатель удовлетворенности пациентов (при этом СП проходят интенсивное обучение).

КОМПЛЕКСНЫЙ ВЗГЛЯД НА КАЧЕСТВО ОБЪЕКТИВНОГО СТРУКТУРИРОВАННОГО КЛИНИЧЕСКОГО ЭКЗАМЕНА

Как уже было сказано, крайне важно оценить экзамен на каждой станции с учетом всех имеющихся показателей, перед тем как делать предположения о качестве и планировать улучшения.

Как измерить качество Объективного структурированного клинического экзамена: обзор показателей

Оценка показателей для станции 8 (посвящена консультированию, постановке диагноза и принятию решений) демонстрирует положительный вклад в надежность глобальной оценки (коэффициент альфа при исключении элемента оценки составляет 0,749). Как следует из кривой оценок (см. рис. 4), коэффициент R2 для этой станции низкий (0,4) с широким разбросом оценок в чек-листах в пределах одной и той же общей оценки и со значительным перекрытием среди более высоких оценок («сдано», «сдано с отличием», «сдано с блеском»).

В сочетании с высоким уровнем дисперсии между группами (33,8%) это указывает на несоответствие между ожиданиями экзаменатора и выставленными оценками и на необходимость изменений элементов чек-листа, чтобы дать возможность проводить оценку более высокого уровня. Это ведет к непостоянству результатов, полученных на одной станции и между станциями.

Для устранения этой проблемы обычно пересматривают содержание станции и, соответственно, чек-листа. Пересмотр описаний каждой оценки и вспомога-

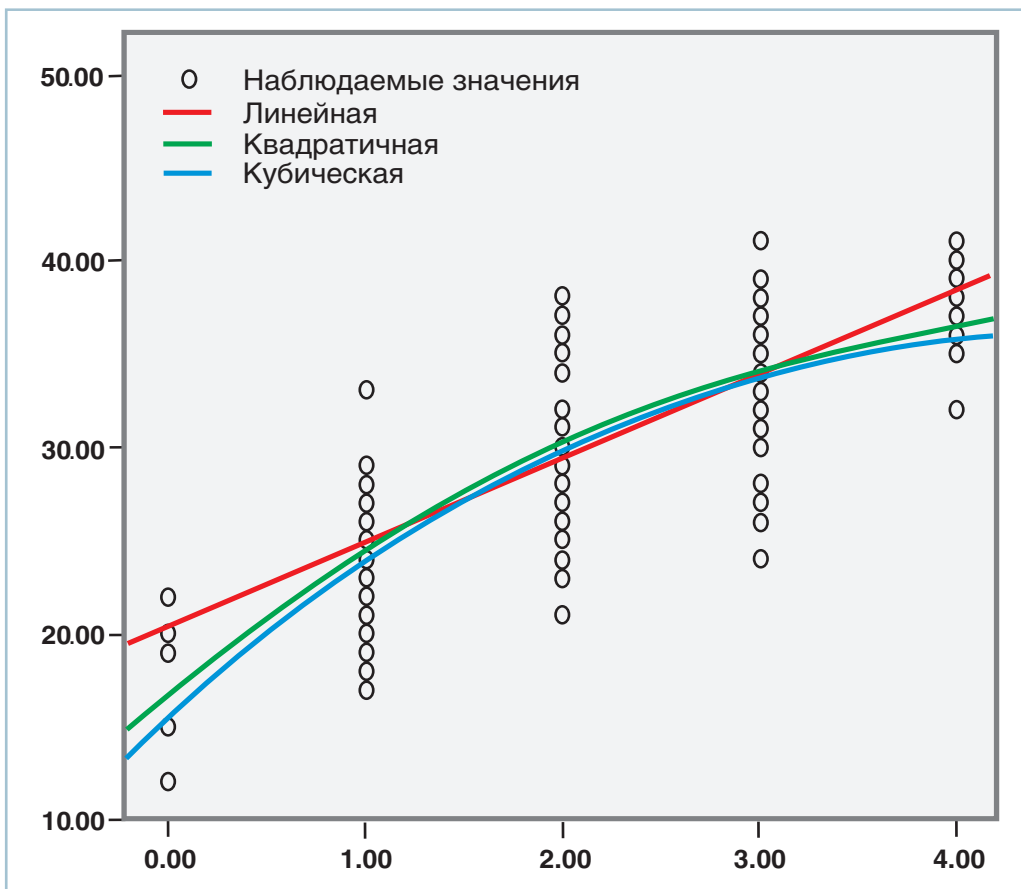


Рис. 4. Кривая оценок (станция 8) – оценки по чек-листу (y) и глобальная оценка (x)

ного материала для экзаменаторов на уровне станции должен помочь устранить несоответствие, на которое указывают низкое значение R2 и более высокая дисперсия ошибок.

Литературных данных о том, что составляет приемлемый диапазон оценок СП на уровне станции, нет.

Станция 9 представлена кривой оценок на рис. 5.

Здесь мы видим более высокий показатель надежности (коэффициент альфа при исключении элемента оценки составляет 0,74) и лучшие значения показателей станции. Величина коэффициента R2 приемлема (0,5), однако дисперсия между группами и здесь высока – 36%.

Кривая демонстрирует значительную дисперсию в подготовленности студентов для каждой глобальной оценки. Достаточно высокая величина R2 скорее свидетельствует о различиях в глобальных оценках, выставленных экзаменатором, чем в баллах по чек-листу, с наличием эффекта «ястребов и голубей».

Действия, направленные на изучение и улучшение этих результатов, стоит сосредоточить на вспомогательном материале для экзаменатора, связанном с глобальными оценками.

Контроль качества посредством наблюдения: выявление проблем при подготовке к объективному структурированному клиническому экзамену и в день экзамена

Для тех, кто стремится свести к минимуму дисперсию между группами, важно систематически наблюдать за проведением ОСКЭ. С учетом некоторых из причин ошибок между группами, все те, кто вовлечен в процесс проведения ОСКЭ

в целом (см. рис. 1), должны быть частью процесса контроля качества.

Многие факторы, вносящие вклад в дисперсию, можно предвидеть и скорректировать до начала ОСКЭ, применяя:

- проверку всех станций на соответствие друг другу по дизайну;
 - проверку того, что новые (а также более старые, уже устоявшиеся) станции соблюдают последние рекомендации в отношении дизайна чек-листов, весовых коэффициентов и опорных пунктов;
 - проверку организации параллельных мест цепочек ОСКЭ, например, при размещении дезинфицирующего геля до входа на станцию, экзаменатор не сумеет оценить навык студента по обработке рук;
 - проверку того, что станции одинаково обеспечены оборудованием (или допустить гибкость в этом вопросе, если студентов обучали разным подходам с использованием разного оборудования).
- При проведении ОСКЭ возможны другие источники дисперсии:
- опаздывающие экзаменаторы, которые пропускают инструктаж перед началом экзамена и, соответственно, не соблюдают должным образом предписанную методологию;
 - неуместные подсказки со стороны экзаменаторов (несмотря на обучение и инструктаж перед экзаменом);
 - неподобающее поведение экзаменаторов (например, изменение характера станции путем излишнего вмешательства);
 - слишком активные СП, задающие наводящие вопросы;
 - реальные пациенты с предубеждениями (например, в отношении пола или расовой принадлежности студентов). СП обучают тому, как взаимодействовать

Как измерить качество Объективного структурированного клинического экзамена: обзор показателей

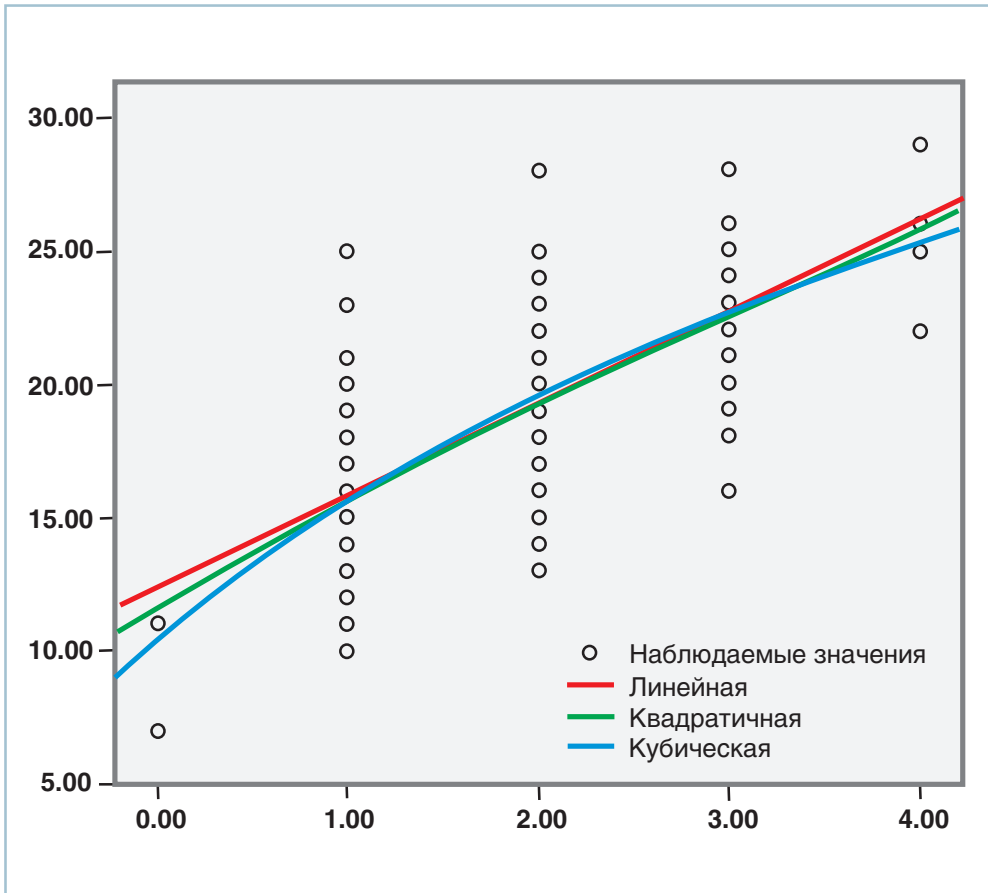


Рис. 5. Кривая оценок (оценочная станция 9) – оценки, выставленные экзаменатором в контрольном перечне (y), и общая оценка (x)

с кандидатами, но большинство реальных пациентов не достигают подобных результатов;

- экзаменаторы (или ассистенты), которые не возвращают оборудование в начальное или в нейтральное положение при смене экзаменуемых.

Исправления в соответствии с полученными результатами

При неудовлетворительных значениях показателей можно использовать ряд практических мер по исправлению ситуации.

1. Коррекция общих баллов с учетом эффектов места проведения экзамена: наиболее легкий способ – приведение к общему среднему показателю по всем центрам проведения экзамена. После такой коррекции профиль не сдавших студентов в данном центре проведения экзамена следует проверить, чтобы удостовериться, например, что все эти случаи не ограничены одним центром. При расчете величины поправки следует отбросить эффект, связанный с любыми группами студентов с особыми потребностями, выявленными в пределах одного центра (например, кан-

дидаты, которым требуется больше времени из-за проблем со здоровьем).

При неудовлетворительных значениях показателей можно использовать ряд практических мер по исправлению ситуации.

2. Коррекция на уровне станции необходима редко, так как противоположные эффекты обычно имеют тенденцию отменять друг друга. В редких случаях, когда это не так, можно провести процедуру на уровне станции, подобную описанной выше.

3. Исключение станции – тоже редкий случай; критериями для исключения обычно бывают низкие значения сразу нескольких показателей, что ставит студентов в такое невыгодное положение, что при апелляции невозможно отменить решения экзаменаторов по оценке.

ЗАКЛЮЧЕНИЕ

Используя ряд рабочих примеров и данные реальных наблюдений, это руководство освещает показатели, обычно используемые при оценке качества ОСКЭ, и то, как их можно использовать для выявления и устранения проблем, а также то, как данный подход помогает предвидеть проблемы на уровне вуза или организации. Таким образом, данная методология естественным образом встраивается в процессы оценки в более широком смысле, как показано на рисунке.

В организации, где работают авторы, существует тесная связь между теми, кто занимается анализом данных, и теми, кто разрабатывает и проводит клиническую оценку, а также занимается разра-

боткой учебных программ и преподаванием. Стандартный подробный разбор показателей на уровне станции выявил расхождения между чек-листами и глобальными оценками. Это привело к внесению изменений в некоторые станции ОСКЭ с последующим улучшением показателей. Некоторые такие изменения перечислены ниже:

- объединение ряда простых критериев в меньшее число критериев более высокого уровня;
- объединение, направленное на то, чтобы критерии более высокого уровня соответствовали уровню знаний студента, позволяя оценить навыки более высокого порядка, менее ориентированные на процессы;
- включение описаний промежуточной оценки в чек-листы для экзаменаторов;
- проверка содержания в критериях чек-листов 3 опорных пунктов вместо 2, там, где это необходимо, что позволяет экзаменаторам лучше дифференцировать результаты;
- более высокая степень единообразия в организации разных цепочек.

Высокая доля результатов «не сдано» для некоторых станций привела к пересмотру порядка преподавания некоторых предметов учебной программы, в результате улучшилась подготовка студентов, как следует из результатов последующих ОСКЭ.

Указания на недостаточное согласие между экзаменаторами в ряде случаев привели к изменениям, которые оказали положительное действие на качество оценки:

- совершенствование методов обучения экзаменаторов;

Как измерить качество Объективного структурированного клинического экзамена: обзор показателей

- актуализация (освежение) подготовки экзаменаторов, которые обучались некоторое время назад;
- обеспечение экзаменаторов более подробными вспомогательными материалами;
- улучшение инструктажа (брифинга) экзаменаторов перед началом экзамена;
- улучшение инструктажа СП перед началом экзамена;
- тренировочные экзамены перед настоящим экзаменом как для экзаменаторов, так и для СП (это действительно целесообразно только при сравнительно небольшом количестве студентов, на-

пример при пересдачах или при ОСКЭ у стоматологов, где группы студентов меньше).

Необходимость всех вышеперечисленных улучшений вряд ли будет очевидна при использовании только одного показателя надежности, например только коэффициента альфа Кронбаха или только коэффициента G. Получить реальную картину качества и выявить области, в которых качество неудовлетворительно, можно лишь при использовании группы показателей. Следуя этому подходу, мы будем вознаграждены стабильными улучшениями в порядке проведения и стандартах клинической оценки.

СВЕДЕНИЯ ОБ АВТОРАХ

Пелл Годфри (Pell Godfrey) – главный специалист по статистике с обширной подготовкой в области управления. Перед тем как начать работу в университете Лидса, он работал в Центре практики высшего образования при Открытом университете (Великобритания). Темы его текущих исследований включают разработку стандартов для практической оценки в высшем образовании и значимость краткосрочных программ в области образования.

Фуллер Ричард (Fuller Richard) – врач-консультант и руководитель программы по подготовке бакалавров медицины и бакалавров хирургии в Институте медицинского образования Лидса. Его исследовательские интересы включают клиническую оценку, особенно наблюдение за проведением объективного структурированного клинического экзамена и повышение его качества.

Хомер Мэттью (Homer Matthew) – стипендиат университета Лидса, работает в Институте медицины и в Институте образования. Ведет ряд исследовательских проектов и обеспечивает общую поддержку коллег в области статистики. Его исследовательские интересы включают статистическую сторону оценок, особенно в части объективного структурированного клинического экзамена.

Робертс Труд (Roberts Trudie) – врач-консультант, профессор в области медицинского образования и руководитель Института медицинского образования Лидса. Ее исследовательские интересы включают клиническую оценку.

ЛИТЕРАТУРА/REFERENCES

- Cohen D.S., Colliver J.A., Robbs R.S., Swartz M.H. A large-scale study of the reliabilities of checklist scores and ratings of interpersonal and communication skills evaluated on a standardized-patient examination. Adv Health Sci Educ. 1997; 1: 209–13.

- Cusimano M. Standard setting in medical education. *Acad Med.* 1996; 71 (10): S112–20.
- Eva K.W., Rosenfeld J., Reiter H., Norman G.R. An Admissions OSCE: the multiple mini-interview. *Med Educ.* 2004; 38: 314–26.
- Field A. *Discovering Statistics (using SPSS for windows)*. London: Sage Publications, 2000: 130 p.
- Holmboe E. Faculty and the observation of trainees' clinical skills: problems and opportunities. *Acad Med.* 2004; 79 (1): 16–22.
- Homer M., Pell G. The impact of the inclusion of simulated patient ratings on the reliability of OSCE assessments under the borderline regression method. *Med Teach.* 2009; 31 (5): 420–5.
- Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ.* 2004; 38: 199–203.
- Norcini J. Setting standards on educational tests. *Med Educ.* 2003; 37 (5): 464–9.
- Norcini J., Burch V. Workplace-based assessment as an educational tool: AMEE guide No. 31. *Med Teach.* 2007; 29 (9): 855–71.
- Pell G., Homer M., Roberts T.E. Assessor training: its effects on criterion based assessment in a medical context. *Int J Res Method Educ.* 2008; 31 (2): 143–54.
- Pell G., Roberts T.E. Setting standards for student assessment. *Int J Res Method Educ.* 2006; 29 (1): 91–103.
- Postgraduate Medical Education Training and Education Board. *Workplace Based Assessment. A Guide for Implementation*. London, 2009. URL: www.pmetb.org.uk/fileadmin/user/QA/assessment/PMETB_WPBA_Guide_20090501.pdf. [date of access May 11, 2009]
- Regehr G., Macrae H., Reznick R.K., Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med.* 1998; 73 (9): 993–7.
- Roberts C., Newble D., Jolly B., Reed M., et al. Assuring the quality of high-stakes undergraduate assessments of clinical competence. *Med Teach.* 2006; 28 (6): 535–43.
- Stevens J. *Applied multivariate statistics for the social sciences*. 2nd ed. Hillside, NJ: Erlbaum, 1992; Ch. 4: 151–82.
- Wass V., McGibbon D., Van Der Vleuten C. Composite undergraduate clinical examinations: how should the components be combined to maximise reliability? *Med Educ.* 2001; 35: 326–330.