

В. В. Редненко, Г. Д. Коробов

КВАЛИМЕТРИЧЕСКИЙ ПОДХОД К МЕТОДОЛОГИИ ОБЪЕКТИВИЗАЦИИ ОЦЕНИВАНИЯ ПРИ ПРОВЕДЕНИИ ОБЪЕКТИВНОГО СТРУКТУРИРОВАННОГО КЛИНИЧЕСКОГО ЭКЗАМЕНА

УО «Витебский государственный ордена дружбы народов
медицинский университет»

В работе представлена универсальная методика создания инструмента оценки объективного структурированного клинического экзамена. В основу инструмента оценки положен чек-лист со штрафными баллами. Алгоритм предполагает поэтапное решение задачи создания инструмента оценки, условием для использования которой является обязательные расчеты согласованного мнения разработчиков и независимых экспертов из числа профильных специалистов, для чего используется расчет коэффициента конкордации Кенделла. Окончательное решение о пригодности инструмента оценки принимается после анализа надежности оценок с помощью критерия альфа Кронбаха после предварительной апробации.

Для облегчения регистрации действий экзаменуемых, чек-листы записываются в электронных таблицах с использованием «облачных» технологий. Автоматизация расчетов итоговых оценок осуществляется путем прописывания специальных формул с помощью мастера функций.

Рассматриваемая в статье методика, является результатом многолетнего поиска оптимального решения создания адекватной и удобной системы оценки ОСКЭ. Разработанный инструмент оценки внедрен в практику учебной деятельности симуляционно-аттестационного центра учреждения образования «Витебский государственный ордена Дружбы народов медицинский университет» и успешно апробирован при проведении объективных структурированных клинических экзаменов в ходе текущих и итоговых аттестаций.

Ключевые слова: Альфа Кронбаха, валидность, инструмент оценки, конкордация Кенделла, надежность оценок, ОСКЭ, чек-лист, экспертная оценка.

V. V. Rednenko, G. D. Korobov

QUALIMETRIC APPROACH TO THE METHODOLOGY OF OBJECTIVIZATION OF ASSESSMENT IN CONDUCTING AN OBJECTIVE STRUCTURED CLINICAL EXAM

The paper presents a universal methodology for creating an assessment tool for an objective structured clinical examination. The assessment tool is based on a checklist with penalty points. The algorithm assumes a step-by-step solution to the problem of creating an assessment tool, the condition for the use of which is the mandatory calculation of the agreed opinion of the developers and independent experts from among the specialized specialists, for which the calculation of the Kendall concordance coefficient is used. The final decision on the suitability of the assessment tool is made after analyzing the reliability of the assessments using the Cronbach's alpha criterion after preliminary testing.

To facilitate the registration of examinees' actions, checklists are recorded in spreadsheets using "cloud" technologies. Automation of calculations of final grades is carried out by writing special formulas using the function wizard.

The methodology considered in the article is the result of many years of searching for an optimal solution to create an adequate and convenient OSCE assessment system. The developed assessment tool has been implemented in the practice of educational activities of the simulation and certification center of the educational institution “Vitebsk State Order of Friendship of Peoples Medical University” and has been successfully tested during objective structured clinical examinations during current and final certifications.

Key words: *Cronbach's alpha, validity, assessment tool, Kendall's concordance, reliability of assessments, OSCE, checklist, expert assessment.*

Объективная оценка профессионального уровня специалистов здравоохранения – это оценка, которая не зависит от личности экзаменуемого или субъективного впечатления экзаменатора, основана на численных (объективных) параметрах оценки и/или выполнении надлежащих действий и манипуляций в ходе реалистичного моделирования клинической ситуации, диагностической или лечебной манипуляции с помощью механических, электронных и компьютерных (виртуальных) моделей или стандартизированных пациентов [1].

Объективная оценка позволяет быть уверенным в том, что медицинский персонал действительно готов к надлежащему выполнению профессиональных обязанностей.

Традиционный подход, основанный на субъективной оценке преподавателем знаний и навыков обучаемых, постепенно изживает себя в современном образовательном процессе. Применительно же к высшей школе, такой подход оказывается более вредным, чем полезным. Именно при обучении взрослых, какowymi в большинстве своём являются студенты, на первое место выходит объективизация оценочной деятельности преподавателей, которая всё больше наполняется корректирующей функцией, позволяющей обучаемым активно использовать оценки своего обучения для последующей самостоятельной работы над ошибками.

Объективный структурированный экзамен (ОСКЭ) в последние годы прочно вошел в систему медицинского образования Республики Беларусь. ОСКЭ, впервые описанный Р. М. Харденом в 1975 году, считается «золотым стандартом» для оценки успеваемости обучаемых в медицинских университетах и колледжах всего мира. Данный экзамен воспринимается как справедливый экзамен, как студентами, так и экзаменаторами, и является приемлемым инструментом для оценки клинической компетентности. В отличие от традиционного

экзамена ОСКЭ считается надежным и валидным инструментом оценки [2].

Отвечает ли критериям надежности и валидности ОСКЭ, проводимый в учреждениях образования Республики Беларусь. Анализ паспортов станций ОСКЭ, включающих в себя чек-листы, шкалы, таблицы сопряженности и др. позволили нам выявить определенные особенности, влияющие на надежность и валидность инструментов оценивания ОСКЭ.

Прежде всего, обращает внимание на себя то, что строки чек-листа имеют разный весовой коэффициент (коэффициент вклада), для каждого чек-листа разрабатывается свой инструмент оценивания с индивидуальной формулой и различными уровнями штрафных баллов, разные чек-листы имеют разный коэффициент сложности. Использование чек-листа не гарантирует межэкспертной надежности при отсутствии подробного описания критериев выставления оценок для каждого чек-листа, эмпирического назначения штрафов и оценки за станцию.

Одной из наиважнейших современных задач по совершенствованию ОСКЭ можно определить разработку объективных методов количественной оценки компетенций. Одним из путей преодоления субъективизма в качественной оценке учебной деятельности обучающихся является квалитетический подход, основанный на использовании шкал оценок полноты выполнения отдельных элементов учебного задания, из которых складывается итоговый результат.

Фиксация полноты и правильности выполнения обучающимся элементов задания в наилучшей степени может быть обеспечена при использовании чек-листов. Но, следует отметить, что, несмотря на широкое использование чек-листов, вопросы методологии их разработки в отечественной научной литературе продолжают оставаться недостаточно освещенными. Как результат на практике используются инструменты оценки с недостаточной доказательной базой.

Цель исследования

Разработать и апробировать универсальную, открытую квалитетическую методику создания инструмента оценки ОСКЭ.

Для достижения заявленной цели были сформулированы следующие задачи:

1. Создать алгоритм формирования инструмента оценки ОСКЭ на основе универсального шаблона с использованием согласованных экспертных мнений.

2. На основе электронных таблиц разработать универсальный инструмент оценки ОСКЭ.

3. Разработать оценочные шкалы для универсального инструмента оценки путем конкордации множества независимых коэффициентов и штрафных баллов к единому значению.

4. Подобрать адекватные статистические методы, обеспечивающие системный анализ валидности инструментов оценки результатов объективного структурированного клинического экзамена, в целях дальнейшего его совершенствования.

Методы

Для аналитического исследования были использованы архивные материалы электронных чек-листов, содержащих сведения об оценках, полученных студентами при проведении текущей и итоговой аттестации с использованием методологии ОСКЭ за 2018–2024 годы.

Статистический анализ проводился и использованием лицензионного пакета прикладных статистических программ STATISTICA 10-RUS (лицензия STA999K347156W, принадлежит ВГМУ). В частности, были использованы следующие модули пакета:

1. Описательные статистики – расчеты средних значений.

2. Непараметрическая статистика – расчет коэффициента конкордации Кендалла.

3. Разведочный анализ – анализ надежности (альфа Кронбаха) и позиционный анализ.

Результаты

С учетом зарубежного опыта в качестве регистрационно-оценочного инструмента был принят чек-лист, в котором экзаменатор может фиксировать действия экзаменуемого в процессе выполнения им медицинской манипуляции.

На первом этапе в качестве чек-листа был принят бумажный список с подробными эле-

ментами, полностью повторяющими протоколы лечебных или диагностических манипуляций, утвержденных Министерством здравоохранения. Каждому элементу присваивалось конкретное значение, что в сумме давало 100 баллов. В соответствии с требованиями учебно-методического совета университета положительными оценками были приняты 70 баллов и более. Большое количество фиксируемых элементов манипуляции и необходимость подсчета баллов вручную сильно затрудняло работу преподавателя и не давало возможности сразу же объявить результат испытуемому.

На втором этапе нами был разработан чек-лист, в котором отдельные элементы были объединены в логически последовательные блоки, имеющие условную завершенность, сократилось количество оцениваемых элементов (строк). Однако, в качестве оценки по-прежнему использовалась 100-балльная шкала и сам чек-лист продолжал использоваться на бумажном носителе.

Сложности в расчете оценок удалось избежать после перехода от бумажного носителя регистрируемой информации к использованию электронных таблиц. Кроме того, было принято решение отказаться от сложной 100-балльной шкалы, а для регистрации правильности выполнения действий испытуемого стали использовать двухбалльную шкалу. Через короткий промежуток времени мы отказались от двухбалльной шкалы из-за сложности интерпретации «правильно-неправильно» оцениваемого элемента и перешли на 3-балльную шкалу оценивания, включающую три значения 2 – «правильно» / «владеет», 1 – «частично правильно» / «владеет удовлетворительно», 0 – «неправильно» / «не выполнил» / «не владеет». К каждому чек-листу прилагалась «шкала снятия баллов» – подробное описание критериев выставления 2, 1 или 0 по каждой позиции. Перевод полученных значений к 10 баллам осуществляется путем умножения полученных оценок за все элементы на соответствующие, специально рассчитанные коэффициенты и выставления штрафных баллов.

В целях обеспечения высокой информативности и достоверности, используемые в настоящее время чек-листы, разрабатываются функциональными рабочими группами, включающими преподавателей (экспертов) клинических кафедр и сотрудников симуляционно-

аттестационного центра. Разработка чек-листов функциональной группой проводится по одобранному и утвержденному алгоритму:

1. На основе утвержденного в университете описания практических навыков (группы практических навыков, объединенных в клиническую ситуацию) и технических возможностей их аттестации в процессе ОСКЭ методом «мозгового штурма» составляется список позиций оценивания (строк) чек-листа.

2. Для исключения эмпирического подхода экзаменаторов к оцениванию позиций чек-листа, разрабатывается шкала снятия баллов для каждого чек-листа (пояснение за что возможно внесение в таблицы данных: 2, 1, 0).

3. Для расчета формулы преобразования нами использованы следующие правила. Распределение вклада: равный, с формированием оценки 10 баллов при правильном выполнении по всем позициям и использовании понижающих коэффициентов 0,5 и 0, соответственно при выставлении в связанную ячейку отметки 1 или 0 (ноль).

Дополнительно выставляются штрафные баллы при наличии отметки 0 (ноль) в связанной ячейке. Штрафные баллы вычитаются из общей оценки за станцию. Штрафные баллы двух уровней: критический и базовый.

«Критический уровень» – невыполнение или неправильное выполнение данного элемента (получение отметки 0 (ноль) по данной позиции) может существенно негативно повлиять на здоровье пациента. Позиции, являющиеся критическими, указаны в чек-листе путем внесения записи «(является грубой ошибкой и влечет штраф)».

«Базовый уровень» – невыполнение или неправильное выполнение любого элемента (получение отметки 0 (ноль) по данной позиции) за исключением критических. Позиции, являющиеся базовыми, определяются по отсутствию записи «(является грубой ошибкой и влечет штраф)».

1. Валидность общей суммы штрафных баллов определялась путем расчета коэффициента конкордации Кендалла экспертных оценок [3].

$$W = \frac{12S}{m^2(n^3 - n)},$$

где m – число экспертов в группе, n – число факторов оценки действий, S – сумма квадратов разностей рангов (отклонений от среднего).

Для определения согласованного мнения экспертов нами была разработана шкала, в основу которой, нами был положен принцип оценки степени влияния отдельного элемента на достижение конечного положительного результата манипуляции и влиянии ошибки на здоровье пациента. Таким образом, оценка проводится по следующим категориям – не оказывает влияния, влияние не значительное, влияние среднее, влияние сильное.

На примере станции инструмента оценки станции «Базовая сердечно-легочная реанимация: непрямой массаж сердца», после проведения нескольких коррекций, полученные данные о согласованности выглядят следующим образом: коэффициент конкордации М. Кендалла = 0,68 при $p = 0,04$. Это свидетельствует о хорошей независимой согласованности оценок экспертов ($>0,4$). Следует отметить, что различия по оценкам экспертов по всему комплексу показателей, согласно дисперсионному анализу не было статистически не значимым χ^2 Пирсона = 14,623 при $p < 0,07$.

Таким образом, по согласованным элементам, ошибки при выполнении которых признана критической, присваивается больший вес при снятии штрафных баллов по сравнению с ошибками базового уровня. Для данного чек-листа проведена коррекция штрафов базового и критического уровня. Общая сумма штрафных баллов критического уровня установлена в 5 баллов. Штраф за одну критическую позицию равен 5 баллам, деленным на количество критических ошибок в чек-листе. Общая сумма штрафных баллов базового уровня – 1 балл. Штраф за одну позицию базового уровня равен 1 баллу, деленному на количество базовых ошибок в чек-листе.

2. Оценка внутренней согласованности (надежности) оценок по каждому элементу чек-листа анализировалась с помощью критерия альфа Кронбаха [3].

$$\alpha_{st} = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}},$$

где N – число элементов чек-листа; \bar{r} – усредненный коэффициент корреляции Пирсона.

При достижении значения альфа Кронбаха 0,7 и более элементы можно считать согласованными между собой.

Расчеты надежности оценок, полученных на станции «Базовая сердечно-легочная реанимация»

Таблица. Надежность оценок по чек-листу «Базовая сердечно-легочная реанимация: непрямой массаж сердца»

Элементы действий	Итог для шкалы: среднее = 33,9009; стандартное отклонение = 4,82706; число наблюдений: 723; альфа Кронбаха: 0,727653			
	среднее при удалении элемента	стандартное отклонение при удалении элемента	ранговая позиция элемента	альфа Кронбаха при удалении элемента
Обеспечение безопасности	31,99631	4,788096	0,041855	0,733987
Проверка сознания	31,99493	4,729518	0,223202	0,723535
Обеспечение проходимости дыхательных путей	32,36284	4,469771	0,429815	0,702703
Проверка дыхания до начала СЛР	32,09036	4,574360	0,448038	0,706010
Проверка пульса до начала СЛР	32,18995	4,615114	0,291915	0,717412
Зов на помощь	32,04610	4,774199	0,096878	0,730156
Проведение компрессий грудной клетки до доставки дыхательного мешка	32,02121	4,695385	0,305357	0,718924
Количество компрессий грудной клетки в цикле СЛР	32,07238	4,706887	0,233003	0,722531
Руководство циклами реанимации	32,09728	4,706352	0,208345	0,724016
НМС: Положение рук при компрессии	32,13324	4,600584	0,337003	0,713659
НМС: Декомпрессия	32,04749	4,759148	0,099829	0,730919
НМС: Глубина компрессий	32,44583	4,590978	0,231388	0,724839
НМС: Частота компрессий	32,40987	4,483915	0,374708	0,708636
Проверка дыхания в процессе СЛР	32,29368	4,483185	0,449270	0,701421
Проверка пульса в процессе СЛР	32,22314	4,610712	0,301791	0,716550
Финальный балл	26,08852	3,186750	0,971386	0,584882

нимация: непрямой массаж сердца» представлены в таблице.

Как видно из таблицы общая альфа Кронбаха по всем элементам оценивания составляет 0,73, что свидетельствует о хорошей (достаточной) надежности оценивания. Последняя графа таблицы содержит информацию об изменении альфы Кронбаха при условном исключении одного из оцениваемых элементов. Эти данные свидетельствуют, что все элементы оценивания по чек-листу имеют примерно одинаковый вклад в надежность итоговой (финальной) оценки действий экзаменуемого.

Дальнейший расчет показал, что установленную надежность можно повысить до 0,9, если ввести дополнительно, добавить еще 3 элемента оценки аналогичного качества уже имеющимся.

Принятый согласованный чек-лист утверждается и используется в учебном процессе на всех факультетах университета. Утвержденные чек-листы для доступа студентам выкладываются в текстовом формате в соответствующих разделах сайта дистанционного обучения университета.

Для обеспечения фиксации действий экзаменуемых и автоматического расчета оценки каждый чек-лист записывается на отдельный

лист электронной GOOGLE-таблицы. Таким образом, каждая книга электронной таблицы содержит отдельные чек-листы по всем станциям ОСКЭ. Расчет итоговой оценки за экзамен проводится так же автоматически на отдельном листе. Программирование расчетов осуществляется с помощью встроенного мастера функций и синтаксиса Apps Script. Подготовленная таким образом таблица размещается на электронном облаке, которое обеспечивает экзаменаторам коллективный удаленный доступ по заданному паролю.

Обсуждение

Предложенные правила разработки инструмента оценки позволяют исключить излишнее влияние отдельных позиций (строк) чек-листа на оценку путем равномерного положительного вклада. В то же время, исключен механистический подход за счет выделения ошибок критического уровня. Предложенная система потенцирует обучаемых детально изучать весь алгоритм навыка за счет введения штрафа ошибок базового уровня. Имеется возможность провести «настройку» сложности путем изменения двух коэффициентов.

Рассматриваемая в статье методика позволяет создать универсальный инструмент

оценки ОСКЭ. Данный инструмент применим к любому чек листу, независимо от количества строк и ошибок различного уровня. Данный инструмент обеспечивает высокую межэкспертную надежность.

Валидность инструмента основывается на учете согласованного мнения разработчиков и независимых экспертов с использованием расчета коэффициента конкордации Кендалла. Окончательное решение о пригодности инструмента оценки принимается после анализа надежности оценок с помощью критерия альфа Кронбаха после предварительной апробации.

Литература/References

1. Alshahrani, Y., Cusack L., Rasmussen P. Undergraduate nursing students' strategies for coping with their first clinical placement: descriptive survey study // NurseEduc. Today. – 2018. – Vol. 69. – P. 104–108. – doi: <https://doi.org/10.1016/j.nedt.2018.07.005>.
2. Harden, R. M. The definitive guide to the OSCE: The Objective Structured Clinical Examination as a performance assessment / R. M. Harden, P. Lilley, M. Patricio. – Elsevier, 2016. – 224 p.
3. Buul, A. SPSS: искусство обработки информации. Анализ статистических данных и скрытых закономерностей [SPSS: information processing skill. Analysis of statistical data and hidden patterns] / A. Buul, P. Zefel. – SPb.: ООО "DiaSoftUP", 2025. – 608 p. [in Russian].

Поступила 05.05.2025 г.