

**Программа учебной дисциплины «Экспертиза качества тестов, опросников и других инструментов оценивания»**

Автор	Антипкина Инна Вениаминовна
Число кредитов	3
Контактная работа (час.)	44
Самостоятельная работа (час.)	70
Курс	1
Формат изучения дисциплины	Очный, full-time

**I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ**

Предлагаемый курс призван повысить психометрическую компетентность у тех студентов магистратуры (в частности, будущих психологов, педагогов, специалистов отделов кадров, социологов, экономистов и исследователей из разных областей науки), кто предполагает работать с тестами, опросниками, методиками оценивания. Под «работой с тестами» понимается использование их в профессиональной деятельности или в исследовательских целях (например, при написании магистерской диссертации), создание или адаптация инструментов, анализ их функционирования и результатов, полученных с их помощью. От качества измерений зависит справедливость и точность принимаемых по итогам оценивания решений. В России сегодня наблюдается дефицит специалистов-психометриков. В обществе не хватает поддержки работы психометриков, недостает понимания важности использования качественных инструментов измерения.

В результате освоения дисциплины студент должен:

**знать:**

- концепции психометрических теорий: надежность, валидность, принципы оценивания тестовых заданий
- правила разработки тестов, опросников и других инструментов оценивания
- типы профессиональных стандартов и где их искать
- современные представления о справедливости тестов.

**уметь:**

- анализировать качество тестовых заданий в рамках классической теории тестирования
- интерпретировать результаты анализа данных теста
- построить дизайн валидационного исследования
- соблюдать этапы разработки инструментов оценивания
- соблюдать правила, предъявляемые к адаптации инструментов оценивать
- выявлять недостающую информацию, необходимую для установления качества инструмента
- проводить экспертизу тестовых заданий

**владеть:**

- навыками использования психометрических теорий в важнейших практических приложениях;
- навыками обработки и интерпретации результатов тестирования;
- навыками коммуникации на тему разработки и использования тестов;

Изучение дисциплины «Экспертиза качества тестов и опросников» базируется на следующих дисциплинах:

- математика в объеме средней школы;

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- Уметь читать регламентирующие документы (стандарты качества) и анализировать их применимость.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин: научно-исследовательские семинары для проведения исследований для КР и ВКР

## **СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ**

### **Тема 1. История экспертизы качества тестов. Основные этапы психометрики в мире и в России. Обзор стандартов для измерений в социальных науках**

Движение за права потребителей. Ежегодник Оскара Бурса. Развитие стандартов в США. Стандарты в странах Европы: Голландия, Англия, Германия. Международные стандарты EFPA, ИТС, АЕА-Еуроге. Корпоративные стандарты (ETS). Стандарты в России (психология и оценка персонала). Наука психометрика в мире и в России. Классическая и современная теория тестирования.

Дискуссия «Психометрика в моей профессиональной области», Задание-кейс «Знакомимся со стандартами EFPA, классифицируем тест».

### **Тема 2. Требования к надежности и валидности инструмента**

Понятие надежности в классической теории тестирования. Предположения (аксиомы) КТТ. Показатель надежности и коэффициент надежности. Коэффициенты надежности в стандартах EFPA.

Как рассчитать показатель надежности Кронбаха альфа на данных реальных тестов. Связь коэффициента альфа с длиной теста и с характеристиками выборки. Работа в MS Excel. Дискуссия: критические интервалы надежности в разных стандартах, почему они различаются?

Как и чем оценивают надежность в современной теории тестирования (IRT). Демонстрация. Обсуждение кейсов: чем грозит низкая надежность теста.

Групповая работа: экспертиза надежности теста на основании психометрического отчета. Ищем недостающую информацию. Обсуждение результатов.

Валидность. Понятие. Современный подход к определению валидности.

Практика Кейс: строим дизайн валидационного исследования (работа в группах над заданием-кейсом и общая дискуссия по итогам).

Связь валидности и надежности. Валидность и надежность в разных профессиональных областях.

Кейсы: оцениваем характеристики надежности и валидности инструмента. Презентация результатов и групповое обсуждение.

### **Тема 3. Шкалы и шкалирование. НОРТ и КОРТ**

Многозначность термина «шкала» и «шкалирование». Уровни измерительных шкал. Преобразование шкал. Шкалы в классической теории тестирования и в современной теории тестирования.

Стандарты качества шкалирования. Сопоставимость шкал. Выравнивание.

НОРТ и КОРТ: разные типы тестов, разные типы решений.  
Стандарты качества для тестов НОРТ и КОРТ.

#### **Тема 4. Требования к разработке инструмента. Типы заданий. Разработка заданий. Конструирование инструмента.**

Что нужно предусмотреть и задокументировать при планировании инструмента. Роль теоретических рамок. Операционализация конструкторов. Идеальная команда разработчиков. Типы заданий.

Составление спецификаций для разработчиков заданий. Экспертиза соответствия заданий спецификациям.

Оценка деятельности, устных экзаменов и проектов. Составление критериев (рубрик). Анализ заданий.

Анализ заданий. Кейсы.

Какие возможности для анализа необходимо заложить на этапе «сборки» теста. Этапы реализации теста (экспертиза, когнитивные лаборатории, апробации)

Качество тестовых материалов. Реализация теста. Документация инструмента. Сообщение результатов оценивания.

#### **Тема 5. Искажения в измерениях**

Требования к исследованиям возможных искажений, заложенные в стандартах EFPA. Искажения на уровне функционирования заданий (DIF). Искажения вследствие нарушения процедур конструирования теста (нарушение параллельности, ошибки целевой аудитории, ошибки сопоставимости процедур проведения оценивания, шкалирования, интерпретации результатов). Возможности выявления искажений в рамках классической и современной теорий тестирования.

Эмпирические процедуры выявления искажений: исследование DIF методом Мантеля-Хэнцеля.

#### **Тема 6. Локализация и адаптация измерительных инструментов**

Требования к адаптации и локализации тестов

Построение дизайна адаптационных работ, отчетная документация (будем обсуждать статью Malda M. et al. Adapting a cognitive test for a different culture..

Нарушение требований адаптации и локализации в практике пользователей тестов. Кейсы.

#### **Тема 7. Опросники и их специфика**

Разнообразие типов и форм опросников. Ответные шкалы. Методы предварительного оценивания опросника (рецензирование экспертами, фокус-группы, когнитивные лаборатории, апробации). Анализ опросников в рамках классической и современной теорий тестирования: преимущества IRT.

Анализ функционирования опросника, построение шкалы методами IRT (демонстрация)

Оценивание качества опросника на основании предоставленных материалов и критериев EFPA.

#### **Тема 8. Роль экспертов в психометрике. Эксперты-разработчики. Эксперты-рейтеры. Роль экспертов при шкалировании результатов.**

Роль экспертов на разных этапах разработки инструмента: разработка инструмента, анализ функционирования заданий, интерпретация результатов. Эксперты-рейтеры (оценивании результатов), важность «межэкспертной надежности».

Тренинг рейтеров. Требования к процедурам работы рейтеров. Разбор кейсов оценивания работы рейтеров при оценивании языковых компетенций.

Игра: выявление «эффектов эксперта» у студентов.

### **Тема 9. Специфика оценивания в отдельных областях (психология, образование, оценка персонала, проведение научных исследований)**

Специфика оценивания в образовании. Национальные и межстрановые проекты. Связь между оценкой в образовании и образовательной политикой.

Специфика оценивания в психологии (гостевой лектор)

Специфика оценивания персонала. Стандарт оценивания персонала. Стандарт центра оценки. (Гостевой лектор)

Использование тестов и опросников при проведении научных исследований.

Оценивание как бизнес в России и в мире.

### **Тема 10. Стандарты проведения («администрирования») оценивания, этика оценивания, сообщение результатов, документация (ИТС)**

Стандартизация условий тестирования. Справедливость оценивания.

Разработка мануала для администраторов оценивания. Разработка формы информированного согласия.

### **Тема 11. Синтез требований к качеству измерений: специфика разных стандартов (APA, AERA & NSME, COTAN)**

Повторяем и систематизируем. Основные разделы критериев экспертизы теста. Собираем «архетип стандартов». Дискуссия: Зачем психометрикам много стандартов?

### **Тема 12. Включение психометрики в профессиональную траекторию студентов. Острые проблемы психометрики в России.**

Итоговое оценивание. Дискуссия: «Психометрика в моей профессиональной траектории». Дискуссия: стереотипы и мифы о тестах и опросниках. Проблема поддержки психометрики в обществе.

## **III. ОЦЕНИВАНИЕ**

Для каждой темы предусмотрены тесты с заданиями для самоконтроля. Они являются формирующим оцениванием и их результаты не учитываются в итоговой оценке.

В диплом выставляется результирующая оценка по учебной дисциплине.

$$O_{\text{итоговая}} = 0,7 * O_{\text{итоговый проект}} + 0,3 * O_{\text{экзаменационная}}$$

## **IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ**

Оценочные средства для текущего контроля студента

1. Расставьте этапы разработки теста в правильном порядке.

а) Разработка тестовой спецификации	1.
б) Когнитивные лаборатории + правки	2.

c) Определение целей	3.
d) Написание заданий	4.
e) Подготовка тестовых материалов (программирование или верстка)	5.
f) Определение конструкта (обзор литературы!+эксперты)	6.
g) Экспертная оценка заданий + правки	7.
h) Операционализация конструкта (обзор литературы+эксперты)	8.
i) Анализ результатов	9.
j) Администрирование теста (подготовка мануалов/тренинга+стандартизация процедур)	10.
к) Пилотирование (на маленькой выборке)	11.
l) Анализ результатов, подготовка отчетов и форм обратной связи, техотчета	12.
m) Правки в задания и/или дизайн теста	13.

Пример -- задание к дискуссии в группах *«Прочитайте два примера использования проективных методик оценивания [раздатки к заданию: 1. Екатерина Мурашова, глава 12 «Вперед и вверх» или «Стоять и не двигаться» из книги «Любить или лечить». М.: Самокат, 2017; 2. Приложение к монографии Захарова Е. И. Психология освоения родительства : монография. М.: ИИУ МГОУ, 2014. 262 с ]. Сопоставьте контексты оценивания. Можно ли применять критерии оценивания качества проективных методик EFPA к предложенным примерам? Обоснуйте свой ответ».*

Задание для групповой работы: *«Прочитайте отчет об анализе заданий теста. Какие задания относятся к «проблемным»? Почему? Можно ли на основании качественного анализа заданий понять причины их неадекватного функционирования и исправить задания?»*

В качестве итогового оценивания будет предложен тест из 15 вопросов по темам «Надежность», «Валидность», «Шкалы и шкалирование», «Адаптация и локализация», «Выявление искажений» и кейс типа *«Перед вами инструмент оценивания и результаты его анализа. В рамках какой теории тестирования (современной или классической) был проведен анализ? Оцените качество инструмента. Есть еще информация об этом инструменте, которую вы хотели бы узнать у разработчиков или отчет достаточно полный?»* На итоговом оценивании можно пользоваться стандартами EFPA или другими и всеми заметками к курсу.

## V. РЕСУРСЫ

### 5.1 Основная литература

- Стандарты EFPA (Европейской федерации психологических ассоциаций, 2013) на русском или английском языке

- Стандарты ETS (на английском языке)
- Стандарты ИТС (на английском языке)
- Стандарты СОТАН (на английском или русском языке)

Все стандарты находятся в открытом доступе или доступны по подписке библиотеки ВШЭ.

## 5.2 Дополнительная литература

Анастаси А., Урбина С., Алексеев А. А. Психологическое тестирование. Питер, 2009.

- Arne Evers, The Internationalization of Test Reviewing: Trends, Differences, and Results, *International Journal of Testing*, Volume 12, Issue 2, 2012, pages 136-156, DOI: 10.1080/15305058.2012.658932

Brown, A. (2000). An investigation of the rating process in the IELTS Speaking Module. In R. Tulloh (Ed.), *Research reports* (1999, Vol. 3, pp. 49-85). Canberra, Australia: IELTS Australia

- Dave Bartram Concluding Thoughts on the Internationalization of Test Reviews, *International Journal of Testing*, Volume 12, Issue 2, 2012, DOI: 10.1080/15305058.2011.649869
- DeVon H. A. et al. A psychometric toolbox for testing validity and reliability // *Journal of Nursing scholarship*. – 2007. – Т. 39. – №. 2. – С. 155-164.
- Dorans N. J., Holland P. W. DIF detection and description: Mantel-Haenszel and Standardization // *ETS Research Report Series*. – 1992. – Т. 1992. – №. 1.
- Ercikan K. Translation effects in international assessments, *International Journal of Educational Research* 29 (6), 543-553
- Evers A. et al. Assessing the quality of tests: Revision of the EFPA review model // *Psicothema*. – 2013. – Т. 25. – №. 3.
- Hambleton R. K., Bartram D., Oakland T. Guidelines for Improving Testing Practices // *IAAP Handbook of applied psychology*. – 2011. – С. 338.
- Kane M.T. Current Concerns in Validity Theory // *Journal of Educational Measurement*. 2001. Vol.38(4). P.319-342.
- Kolen M. J., Brennan R. L. Test equating, scaling, and linking. – 2004.
- Kurt F. Geisinger, Worldwide Test Reviewing at the Beginning of the Twenty-First Century *International Journal of Testing*, Volume 12, Issue 2, pages 103-107 2012 DOI: 10.1080/15305058.2011.651545
- Lumley T. Assessment criteria in a large-scale writing test: What do they really mean to the raters? // *Language Testing*. – 2002. – Т. 19. – №. 3. – С. 246-276.
- Lumley T., McNamara T. F. Rater characteristics and rater bias: Implications for training // *Language Testing*. – 1995. – Т. 12. – №. 1. – С. 54-71.
- Meiron, B. E. (1998). Rating oral proficiency tests: A triangulated study of rater thought processes. Unpublished masteris thesis, University of California at Los Angeles.
- Muniz, J., & Hambleton, R. K. (1997). Directions for the translation and adaptation of tests. *Papeles del Psicologo*, August, 63-70.
- Rattray J., Jones M. C. Essential elements of questionnaire design and development // *Journal of clinical nursing*. – 2007. – Т. 16. – №. 2. – С. 234-243.
- Stehr-Green P. A. et al. Developing a questionnaire // *Focus on field epidemiology*. – 2003. – Т. 2. – С. 2-6.
- Traub R. E. Classical test theory in historical perspective // *Educational Measurement: Issues and Practice*. 1997. No.16. P. 8-14.
- Velleman P. F., Wilkinson L. Nominal, ordinal, interval, and ratio typologies are misleading // *The American Statistician*. – 1993. – Т. 47. – №. 1. – С. 65-72.
- Walker, C. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, 29, 364-376

## 5.3 Программное обеспечение

№ п/п	Наименование	Условия доступа
1.	Microsoft Windows 7 Professional RUS Microsoft Windows 10	<i>Из внутренней сети университета (договор)</i>
2.	Microsoft Office Professional Plus 2010	<i>Из внутренней сети университета (договор)</i>

#### **5.4 Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)**

№ п/п	Наименование	Условия доступа
1.	Электронные ресурсы библиотеки НИУ ВШЭ	<i>Из внутренней сети университета (договор)</i>

#### **5.5 Материально-техническое обеспечение дисциплины**

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);
- мультимедийный проектор с дистанционным управлением.