Федеральное государственное автономное образовательное учреждение высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"

Институт образования

Рабочая программа дисциплины по выбору общеуниверситетского пула (МАГОЛЕГО) [Экспертиза качества тестов, опросников и других инструментов оценивания]

для уровня подготовки - магистратура

Разработчик(и) программы:
Антипкина Инна Вениаминовна, младший научный сотрудник, iantipkina@hse.ru

Одобрена на заседании комиссии
«___»_____ 201_ г.
Председатель комиссии [Введите И.О. Фамилия]_____ [подпись]

Утверждена «___»_____ 201_ г.
Руководитель Методического центра ДООП
«___»_____ 201_ г.
[Введите И.О. Фамилия] _____ [подпись]

Москва, 2018

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения кафедры-разработчика программы.

1. Область применения

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Разработка и экспертиза качества тестов», и студентов магистерских программ, выбравших данную дисциплину из общеуниверситетского пула.

Программа разработана в соответствии с Образовательным стандартом высшего образования федерального государственного автономного образовательного учреждения высшего образования «Национальный Исследовательский Университет «Высшая Школа Экономики» Уровень высшего образования: Магистратура Направлению подготовки 37.04.01 Психология.

2. Цели освоения дисциплины

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.

Обучение на курсе даст студентам навыки по оцениванию качества тестов и опросников.

В результате освоения учебной дисциплины студенты должны овладеть следующими образовательными результатами:

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции[1]	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен предлагать модели, изобретать и апробировать способы и инструменты профессиональной деятельно-	СК-2 СК-М2	СД	Применяет критерии для оценивания качества инструмента, выстраивает порядок работ по разработке или апробации ин-	Практические занятия-симуляции, групповые дискуссии, решение кейсов	Решение итогового кейса

[1] Уровни формирования компетенций:
РБ — ресурсная база, в основном теоретические и предметные основы (знания, умения);
СД – способы деятельности, составляющие практическое ядро данной компетенции;
МЦ – мотивационно-ценностная составляющая, отражает степень осознания ценности компетенции человеком и готовность ее использовать

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции[1]	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
сти			струмента		
Способен организовать многостороннюю (в том числе межкультурную) коммуникацию	СК-7 СК-М7	СД	Коммуницирует на темы психометрики в профессиональном понятийном поле, учитывает интересы разных сторон (разработчиков тестов, пользователей, испытуемых) при общении на темы психометрики и разработки тестов	Групповая работа по созданию заданий; групповая работа по анализу результатов	Анализ презентации групповой работы
Способен вести профессиональную, в том числе научно-исследовательскую деятельность в международной среде.	СК-8 СК-М8	СД	Применяет международные стандарты качества для оценивания измерительных инструментов;	Решение российских и международных кейсов о применении инструментов оценивания	Итоговое оценивание
Способен организовать самостоятельную проектную, научную, исследовательскую, консультационную и прикладную деятельность на основе правовых и профессиональных норм и обязанностей.	ПК-1 ИК-М1.2п/н/и/к/пр_6.1	СД	Понимает специфику использования измерительных инструментов в своей профессиональной деятельности (в том числе научной); знает и соблюдает правовые, этические требования	Выполнение заданий занятия «Инструменты оценивания в исследовательской деятельности»	Итоговое оценивание, решение кейсов
Способен использовать конкретные концепции, модели, методы, способы и инструменты работы для решения комплексных задач в государственных, общественных и бизнес- организациях, административ-	ПК-6 ИК-М7.1_7.7	СД	Оценивает качество инструмента измерения/оценивания, с которым предполагает работать, понимает ограничения использования инструментов оценивания в своей профессиональной деятельности	Групповая дискуссия	Итоговое оценивание

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции[1]	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
ных органах, научно-исследовательских и консалтинговых организациях, а также организациях, оказывающих консультационные и психотерапевтические услуги населению.					
Способен анализировать, верифицировать, оценивать полноту информации в ходе профессиональной деятельности и работать в условиях неопределённости.	СК-6 СК-М6	СД	Оценивает полноту информации о качестве инструмента, используемого в работе	Анализ качества тестов Анализ адекватности применения инструментов в предлагаемом контексте	Решение кейсов, итоговое оценивание
Способен определять, транслировать общие цели в профессиональной и социальной деятельности	9 ПК-8 СЛК-М3	СД	Понимает важность использования качественны тестов и опросников, транслирует эти ценности в своей профессиональной деятельности, критически анализирует существующие в обществе стереотипы о психометрике и тестах	Дискуссии, работа со стандартами, решение кейсов	Эссе
Способен строить профессиональную деятельность, бизнес и делать выбор, руководствуясь принципами социальной ответственности	ПК-11 СЛК-М7	СД	Соблюдает этические требования при проведении измерений и требования к справедливости оценивания	Дискуссии, работа со стандартами, решение кейсов	Итоговое оценивание
Способен описывать проблемы и ситуации профессиональной деятельности,	ПК-5 ИК-М5.3_5.4_5.6_2.	СД	Встраивает психометрические знания в свою профессиональную деятель-	Групповые обсуждения	Эссе

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции[1]	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
используя язык и аппарат гуманитарных, экономических и социальных наук для решения проблем на стыке наук, в смежных профессиональных областях.	4.1		ность		
Способен оформить и представить результаты своей деятельности в виде отчета по научной и практической работе и презентации на русском (государственном) и иностранном языке с использованием современных средств ИКТ.	ПК-3 ИК-М3.1_3.2_2.4.1_2.4.2_2.5.1_2.5.2_4.2	СД	Презентует результаты поведенного анализа качества и инструментов на русском и английском языках	Презентация результатов групповых обсуждений; презентация результатов групповых проектов	Презентация результатов с использованием ИКТ

3. Тематический план учебной дисциплины

[Тематический план отражает содержание дисциплины (перечень разделов), структурированное по видам учебных занятий с указанием их объемов]

№	Название раздела	Всего часов	Аудиторные часы			Самостоятельная работа
			Лекции	Семинары	Практические занятия	
1	История экспертизы качества тестов. Основные этапы психометрики в мире и в России. Обзор стандартов для измерений в социальных науках.	5	1		1	3

2	Требования к надежности и валидности инструмента	14	4	2	2	6
3	Шкалы и шкалирование. НОРТ и КОРТ	11	3		2	6
4	Требования к разработке инструмента. Типы заданий. Разработка заданий. Конструирование инструмента.	15	4	1	2	9
5	Искажения в измерениях	7	1		1	5
6	Локализация и адаптация измерительных инструментов	5	2	1		2
7	Опросники и их специфика	7	1	1	1	4
8	Роль экспертов в психометрике. Эксперты-разработчики. Эксперты-рейтеры. Роль экспертов при шкалировании результатов.	7	1	1	1	5
9	Специфика оценивания в отдельных областях (психология, образование, оценка персонала, проведение научных исследований)	20		5		15
10	Стандарты проведения («администрирования») оценивания, этика оценивания, сообщение результатов, документация (ITC)	8	1	1	1	4
11	Синтез требований к качеству измерений: специфика разных стандартов (APA, AERA & NSME, COTAN, ETS)	4		1	1	2
12	Включение психометрики в профессиональную траекторию студентов. Острые проблемы психометрики в России.	11		1	1	8
		114	18	14	13	69

4. Формы контроля знаний студентов

Тип контроля	Форма контроля	модуль		Параметры **
		3	4	
Текущий	Контрольная работа	*		Тесты для самопроверки
	Эссе			600-1000 слов по теме «Психометрика в моей профессио-

				нальной траектории» или «Проблема поддержки психометрики в обществе».
	Домашнее задание			Чтение материалов
	Лабораторная работа			Расчет коэффициента альфа Кронбаха; разработка заданий; анализ DIF методом Mantel-Haenzel
	Проект			Презентация анализа заданий из 15 слайдов, отчет
Промежуточный	Экзамен			Решение кейсов
Завершающий	Экзамен			15 вопросов с множественным выбором и кейс

[Удалите соответствующие строки, если какой-либо контроль не предусмотрен. * - Для промежуточного и завершающего контроля - отметка, в каком модуле/ семестре проводится.

** В графе Параметры указывается регламент (свод постоянных или временных правил, регулирующих внутреннюю организацию и формы деятельности) проведения контроля (заполняется для каждого контроля соответственно): формат работы (письменная, устная, тест, тест в компьютерной программе и другое), время, отведенное на аудиторные работы, количество дней проведения контроля, количество дней оценки результатов контроля (только для завершающего контроля), объем письменных работ для домашних работ, сроки сдачи письменных работ (число), время на самостоятельную подготовку письменных работ и другая информация, носящая регламентирующий характер.]

5. Критерии оценки знаний, навыков

Выполняя задания текущего контроля студент должен продемонстрировать 1. знание требований к качеству инструментов оценивания/проведения оценивания и понимание этих требований, 2. знание этапов разработки заданий/тестов и понимание важности этих этапов

Выполняя задания текущего контроля студент должен продемонстрировать 1. умение интерпретировать показатели качества инструментов оценивания, 2. умение работать со стандартами и технической документацией тестов,

Выполняя задания итогового контроля студенты должны продемонстрировать 1. навык оценивания качества теста (любого инструмента измерения/оценивания) 2. Навык подготовки аргументированных отчетов по результатам экспертизы инструмента

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

По результатам проверки выполнения заданий студенты получают обратную связь от преподавателя.

6. Содержание дисциплины

Раздел представляется в удобной форме (список, таблица). Изложение строится по разделам и темам. Содержание темы может распределяться по лекционным и практическим занятиям.

<u>Раздел 1 История экспертизы качества тестов. Основные этапы психометрики в мире и в России. Обзор стандартов для измерений в социальных науках.</u>

Лекция (1 час): Движение за права потребителей. Ежегодник Оскара Буроса. Развитие стандартов в США. Стандарты в странах Европы: Голландия, Англия, Германия. Международные стандарты EFPA, ITC, AEA-Europe. Корпоративные стандарты (ETS). Стандарты в России (психология и оценка персонала). Наука психометрика в мире и в России. Классическая и современная теория тестирования.

Практика (1 час): Знакомство. Дискуссия «Психометрика в моей профессиональной области», Задание-кейс «Знакомимся со стандартами EFPA, классифицируем тест».

Самостоятельная работа (3 часа).

Литература по разделу:

- История развития стандартов оценивания (раздаточный материал в электронном виде)
- Стандарты Европейской федерации психологических ассоциаций, 2013
- Бурлачук Л. и др. Психодиагностика: Учебник для вузов. 2-е изд. – " Издательский дом"" Питер""", 2014. Глава 1.
- Анастази А., Урбина С. Психологическое тестирование.–7-е изд.–СПб.: Питер, 2007.–688 с.: ил.–(Серия «Мастера психологии»). – ISBN 978-5-272-00106-1. Сс. 48-60.
- Hambleton R. K., Bartram D., Oakland T. Guidelines for Improving Testing Practices //IAAP Handbook of applied psychology. – 2011. – С. 338.
- Evers A. et al. Assessing the quality of tests: Revision of the EFPA review model //Psicothema. – 2013. – Т. 25. – №. 3.
- Traub R. E. Classical test theory in historical perspective // Educational Measurement: Issues and Practice. 1997. No.16. P. 8-14.
- 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
- 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
- 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

Раздел 2. Требования к надежности и валидности инструмента

Лекция (1 час) Понятие надежности в классической теории тестирования. Предположения (аксиомы) КТТ. Показатель надежности и коэффициент надежности. Коэффициенты надежности в стандартах EFPA.

Практика (1 час): 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?

Лекция (1 час). Как и чем оценивают надежность в современной теории тестирования (IRT). Демонстрация. Обсуждение кейсов: чем грозит низкая надежность теста.

Семинар (1 час). Групповая работа: экспертиза надежности теста на основании психометрического отчета. Ищем недостающую информация. Обсуждение результатов.

Лекция (1 час). Валидность. Понятие. Современный подход к определению валидности.

Практика (1 час) Кейс: строим дизайн валидационного исследования (работа в группах над заданием-кейсом и общая дискуссия по итогам).

Лекция. (1 час) Связь валидности и надежности. Валидность и надежность в разных профессиональных областях.

Семинар (1 час) Кейсы: оцениваем характеристики надежности и валидности инструмента (кейсы). Презентация результатов и групповое обсуждение.

Самостоятельная работа (6 часов).

Литература по разделу:

- Стандарты ETS

- Стандарты EFPA
- Крокер Л. Алгина Джеймс. Введение в классическую и современную теорию тестов: учебник //М.: Логос. – 2010. Раздел II и III.
- Анастази А., Урбина С. Психологическое тестирование.–7-е изд.–СПб.: Питер, 2007.–688 с.: ил.–(Серия «Мастера психологии»). – ISBN 978-5-272-00106-1. Сс. 48-60. Глава 4 и 5.
- Kane M.T. Current Concerns in Validity Theory // Journal of Educational Measurement. 2001. Vol.38(4). P.319-342.
- DeVon H. A. et al. A psychometric toolbox for testing validity and reliability //Journal of Nursing scholarship. – 2007. – Т. 39. – №. 2. – С. 155-164.

Раздел 3. Шкалы и шкалирование. НОРТ и КОРТ

Лекция (1 час) Многозначность термина «шкала» и «шкалирование». Уровни измерительных шкал. Преобразование шкал. Шкалы в классической теории тестирования и в современной теории тестирования.

Практика (1 час) Стандарты качества шкалирования. Сопоставимость шкал. Выравнивание.

Лекция (1 час) НОРТ и КОРТ: разные типы тестов, разные типы решений.

Практика (1 час) Стандарты качества для тестов НОРТ и КОРТ.

Самостоятельная работа (6 часов).

Литература по разделу:

- Стандарты EFPA
- Стандарты ETS
- Крокер Л. Алгина Д. Введение в классическую и современную теорию тестов: учебник //М.: Логос. – 2010. Глава 3.
- Velleman P. F., Wilkinson L. Nominal, ordinal, interval, and ratio typologies are misleading //The American Statistician. – 1993. – Т. 47. – №. 1. – С. 65-72.
- Kolen M. J., Brennan R. L. Test equating, scaling, and linking. – 2004.

Раздел 4. Требования к разработке инструмента. Типы заданий. Разработка заданий. Конструирование инструмента.

Лекция (1 час): Что нужно предусмотреть и задокументировать при планировании инструмента. Роль теоретических рамок. Операционализация конструктов. Идеальная команда разработчиков. Типы заданий.

Семинар (1 час) Составление спецификаций для разработчиков заданий. Экспертиза соответствия заданий спецификациям.

Лекция (1 час) Оценка деятельности, устных экзаменов и проектов. Составление критериев (рубрик). Анализ заданий.

Практика (1 час): Анализ заданий. Кейсы.

Лекция (1 час): Какие возможности для анализа необходимо заложить на этапе «сборки» теста. Этапы реализации теста (экспертиза, когнитивные лаборатории, апробации)

Лекция (1 час): Качество тестовых материалов. Реализация теста. Документация инструмента. Сообщение результатов оценивания.

Самостоятельная работа (9 часов).

Литература по разделу:

- Стандарты EFPA

- Стандарты ETS
- Крокер Л. Алгина Д. Введение в классическую и современную теорию тестов: учебник //М.: Логос. – 2010. Глава 4.
- Шмелев А. Г. Практическая тестология. Тестирование в образовании, прикладной психологии и управлении персоналом //М.: ООО «ИПЦ «Маска. 2013. Глава 2.
- Irwing P., Hughes D. Test Development (глава из книги)
- Cohen A, Wollack J, Helpful Tips for Creating Reliable and Valid Classroom Tests

Раздел 5. Искажения в измерениях

Лекция (1 час): 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.

Практика (1 час): Эмпирические процедуры выявления искажений: исследование DIF методом Мантеля-Хэнцеля.

Самостоятельная работа (5 часов).

Литература по разделу:

- Стандарты ETS
- Стандарты EFPA
- Zumbo, B.D. (2007). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. Language Assessment Quarterly, 4, 223–233.
- 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
- Dorans N. J., Holland P. W. DIF detection and description: Mantel-Haenszel and Standardization //ETS Research Report Series. – 1992. – Т. 1992. – №. 1.

Раздел 6. Локализация и адаптация измерительных инструментов

Лекция (1 час): Требования к адаптации и локализации тестов

Семинар (1 час): построение дизайна адаптационных работ, отчетная документация (будем обсуждать статью Malda M. et al. Adapting a cognitive test for a different culture..

Лекция (1 час): Нарушение требований адаптации и локализации в практике пользователей тестов. Кейсы.

Самостоятельная работа (2 часа).

Литература по разделу:

- International Test Commission (2005). International Guidelines on Test Adaptation
- Стандарты ETS
- Стандарты EFPA
- Muniz, J., & Hambleton, R. K. (1997). Directions for the translation and adaptation of tests. Papeles del Psicologo, August, 63-70.
- Ercikan K. Translation effects in international assessments, International Journal of Educational Research 29 (6), 543-553
- (кейс) Malda M. et al. Adapting a cognitive test for a different culture.

Раздел 7. Опросники и их специфика

Лекция (1 час): 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.

Практика (1 час): анализ функционирования опросника, построение шкалы методами IRT (демонстрация)

Семинар (1 час): оценивание качества опросника на основании предоставленных материалов и критериев EFPA.

Самостоятельная работа (4 часа).

Литература по разделу:

- Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System
- Stehr-Green P. A. et al. Developing a questionnaire //Focus on field epidemiology. – 2003. – Т. 2. – С. 2-6.
- Rattray J., Jones M. C. Essential elements of questionnaire design and development //Journal of clinical nursing. – 2007. – Т. 16. – №. 2. – С. 234-243.
- Embretson S.E., Reise, S.P. Item Response Theory for Psychologists. New Jersey: Lawrence Erlbaum Associates, Publishers, 2000. Гл. 1-3. 2.

Раздел 8. Роль экспертов в психометрике. Эксперты-разработчики. Эксперты-рейтеры. Роль экспертов при шкалировании результатов.

Лекция (1 час): Роль экспертов на разных этапах разработки инструмента: разработка инструмента, анализ функционирования заданий, интерпретация результатов. Эксперты-рейтеры (оценивании результатов), важность «межэкспертной надежности».

Семинар (1 час) Тренинг рейтеров. Требования к процедурам работы рейтеров. Разбор кейсов оценивания работы рейтеров при оценивании языковых компетенций.

Практика (1 час) Игра: выявление «эффектов эксперта» у студентов.

Самостоятельная работа (5 часов).

Литература по разделу:

- 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.
- 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
- Meiron, B. E. (1998). Rating oral proficiency tests: A triangulated study of rater thought processes. Unpublished masterís thesis, University of California at Los Angeles.

Раздел 9. Специфика оценивания в отдельных областях (психология, образование, оценка персоала, проведение научных исследований)

Семинар (1 час) Специфика оценивания в образовании. Национальные и межстрановые проекты. Связь между оценкой в образовании и образовательной политикой.

Семинар (1 час) Специфика оценивания в психологии (гостевой лектор)

Семинар (1 час) Специфика оценивания персонала. Стандарт оценивания персонала. Стандарт центра оценки. (Гостевой лектор)

Семинар (1 час) Использование тестов и опросников при проведении научных исследований.

Семинар (1 час) Оценивание как бизнес в России и в мире.

Самостоятельная работа (15 часов).

Литература по разделу:

- Анастази А., Урбина С., Алексеев А. А. Психологическое тестирование. Питер, 2009. Глава 5.
- Вучетич Е. и др. Российский стандарт центра оценки //Организационная психология. – 2013. – Т. 3. – №. 2.
- Батурин Н. А. и др. Российский стандарт тестирования персонала (временная версия, созданная для широкого обсуждения в 2015 году) //Организационная психология. – 2015. – Т. 5. – №. 2.
- Стандарты AEA-Europe

Раздел 10. Стандарты проведения («администрирования») оценивания, этика оценивания, сообщение результатов, документация (ITC)

Лекция (1 час): Стандартизация условий тестирования. Справедливость оценивания.

Практика (1 час): Разработка мануала для администраторов оценивания. Разработка формы информированного согласия.

Семинар (1 час) Кейсы

Самостоятельная работа (4 часа).

Литература по разделу:

- Стандарты ITC
- Стандарты APA, AERA & NSME

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

Практика (1 час): Повторяем и систематизируем. Основные разделы критериев экспертизы теста. Собираем «архетип стандартов»

Семинар (1 час): Дискуссия: Зачем психометрикам много стандартов?

Самостоятельная работа (2 часа).

Литература по разделу:

- Стандарты: EFPA
- Стандарты APA, AERA & NSME
- Стандарты COTAN
- Шмелев А. Г. Практическая тестология. Тестирование в образовании, прикладной психологии и управлении персоналом //М.: ООО «ИПЦ «Маска. 2013.

Раздел 12. Включение психометрики в профессиональную траекторию студентов. Острые проблемы психометрики в России.

Практика (1 час): Итоговое оценивание.

Семинар (1 час): Дискуссия: «Психометрика в моей профессиональной траектории». Дискуссия: стереотипы и мифы о тестах и опросниках. Проблема поддержки психометрики в обществе.

Самостоятельная работа (6 часов).

Литература по разделу:

- Стандарты EFPA

7. Образовательные технологии

Групповые дискуссии, разбор практических задач и кейсов, анализ данных. Для темы по специфике оценивания персонала предусмотрен гостевой преподаватель.

8. Методические указания студентам

Основная масса литературы по психометрике создана на английском языке, поэтому хорошие навыки чтения на английском помогут в овладении материалами курса. Тем не менее, ключевые тексты нашего курса есть на русском языке, что облегчит участникам курса овладение профессиональной терминологией.

9. Оценочные средства для текущего контроля и аттестации студента

Примеры заданий текущей и промежуточной аттестации

Задания для самоконтроля студентов.

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

Пример задания из занятия «Стандарты проведения («администрирования») оценивания, этика оценивания, сообщение результатов (ITC)»: *«Продолжите фразу. Отметьте все правильные варианты. Компетентный пользователь тестов будет:*

1. *Использовать тесты обоснованно*
2. *Внимательно прочитывать содержание теста и принимать решение об его использовании преимущественно на основании очевидной валидности*
3. *Следить, чтобы изначально иноязычные тесты были адаптированы в соответствии с тщательно-разработанной методологией в соответствии с лучшими практиками адаптации и локализации тестов*
4. *Предоставлять респондентам примеры, тренировочные задания или практические рекомендации*
5. *Проводить тренинг для администраторов тестирования (при массовом тестировании)*
6. *Для повышения мотивации дополнительно предупреждать респондентов о важности теста и возможных негативных последствиях*
7. *Читать респондентам инструкции отчетливо и спокойно*
8. *Предоставить респондентам максимально полные отчеты по итогам тестирования, подготовленные по стандартам профессионального языка оцениваемой области.*

   (правильные ответы 1, 3, 4, 5, 7)

Демонстрационные материалы для проведения текущего контроля (тестовые задания, кейсы, сценарии игр и пр.).

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».*

Тренировочные материалы по дисциплине, демонстрационные материалы для проведения промежуточного контроля (тестовые задания, кейсы, сценарии игр и пр.)

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?»*

Примерный перечень вопросов для самопроверки студентов к экзамену по всему курсу или к каждому промежуточному и завершающему контролю и другое.

В качестве итогового оценивания будет предложен тест из 15 вопросов с множественным выбором 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.

10. Порядок формирования оценок по дисциплине

Оценки за работу на семинарских и практических занятиях преподаватель выставляет в рабочую ведомость. Оценка по 10-ти балльной шкале за работу на семинарских и практических занятиях определяется перед промежуточным или итоговым контролем - $O_{аудиторная}$.

Преподаватель оценивает самостоятельную работу студентов: правильность интерпретации и использования критериев качества измерительных инструментов, полнота анализа, правильность оформления анализа (ссылки на соответствующие критерии и стандарты). Для эссе: логичность рассуждений, убедительность аргументов.

Накопленная оценка по дисциплине рассчитывается по формуле:

$$O_{накопленная} = 0{,}2 * O_{эссе} + 0{,}2 * O_{презентация\ групповой\ работы} + 0{,}3 * O_{кейсы} + 0{,}3\ O_{лабораторная\ работа}$$

В диплом выставляется результирующая оценка по учебной дисциплине.

$$O_{итоговая} = 0{,}6 * O_{накопленная} + 0{,}4 * O_{экзаменационная}$$

10. Учебно-методическое и информационное обеспечение дисциплины

10.1 Базовая литература

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

10.2 Основная литература

- Анастази А., Урбина С., Алексеев А. А. Психологическое тестирование. Питер, 2009.
- Крокер Л. Алгина Джеймс. Введение в классическую и современную теорию тестов: учебник //М.: Логос. – 2010.
- Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System
- International Test Commission (2005). International Guidelines on Test Adaptation

- Irwing P., Hughes D. Test Development (глава из книги)

10.3 Дополнительная литература

- 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

- Cohen A, Wollack J, Helpful Tips for Creating Reliable and Valid Classroom Tests
- 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.
- Embretson S.E., Reise, S.P. Item Response Theory for Psychologists. New Jersey: Lawrence Erlbaum Associates, Publishers, 2000. Гл. 1-3. 2.
- 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 masterís 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

- Zumbo, B.D. (2007). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. Language Assessment Quarterly, 4, 223–233.
- Бурлачук Л. и др. Психодиагностика: Учебник для вузов. 2-е изд. – " Издательский дом"" Питер"""", 2014.
- Вучетич Е. и др. Российский стандарт центра оценки //Организационная психология. – 2013. – Т. 3. – №. 2.
- Шмелев А. Г. Практическая тестология. Тестирование в образовании, прикладной психологии и управлении персоналом //М.: ООО «ИПЦ «Маска. 2013. Глава 2.

10.4 Справочники, словари, энциклопедии

- Fernández-Ballesteros R. (ed.). Encyclopedia of psychological assessment. – Sage, 2002.
- Salkind N. J. Encyclopedia of measurement and statistics. – Sage, 2007. – Т. 1.

10.5 Программные средства

- MS Excel

10.6 Дистанционная поддержка дисциплины

Дистанционная поддержка дисциплины осуществляется в дистанционной образовательной среде обучения LMS либо через обмен электронными письмами с преподавателем.

11. Материально-техническое обеспечение дисциплины

- Ноутбук и проектор

Приложение 1

Примеры расчета результирующей оценки за дисциплину

Результирующая оценка за дисциплину рассчитывается следующим образом:
Накопленная оценка за текущий контроль учитывает результаты студента по текущему контролю следующим образом:

$$O_{накопленная} = k_1 * O_{текущий} + k_2 * O_{ауд} + k_3 * O_{сам.работа}$$

где $O_{текущий}$ рассчитывается как взвешенная сумма всех форм текущего контроля, предусмотренных в РПУД; $O_{сам.работа}$ – накопленная оценка за самостоятельную работу студента

$$O_{текущий} = n_1 \cdot O_{эссе} + n_2 \cdot O_{к/р} + n_3 \cdot O_{реф} + n_4 \cdot O_{кол} + n_5 \cdot O_{дз};$$

[Оставьте те формы текущего контроля, которые предусмотрены вРПУД. сумма удельных весов должна быть равна единице: $\sum n_i = 1$] Способ округления накопленной оценки текущего контроля: [указывается способ – арифметический, в пользу студента, другое].

1. Если дисциплина преподается один модуль:

$$O_{результ} = k_1 * O_{накопл} + k * \cdot O_{экз}$$

Способ округления накопленной оценки промежуточного (завершающего) контроля в форме зачета: [указывается способ – арифметический, в пользу студента, другое].

2. Если дисциплина преподается несколько модулей (например, 3):

$$O_{промежуточная\ i} = m_1 \cdot O_{текущая\ i\ этапа} + m_2 \cdot O_{промежуточный\ экзамен}$$

Где $O_{текущая\ i\ этапа}$ рассчитывается по приведенной выше формуле

$$O_{накопленная\ завершающая} = (O_{промежуточная\ 1} + O_{промежуточная\ 2} + O_{накопленная\ 3}) : на\ число\ модулей$$

Где $O_{промежуточная\ 1} + O_{промежуточная\ 2}$ – промежуточные оценки этапов 1 и 2,
а $O_{накопленная\ 3}$ – накопленная оценка последнего этапа перед завершающим зачетом/экзаменом

Способ округления накопленной оценки промежуточного (завершающего) контроля в форме экзамена: [указывается способ – арифметический, в пользу студента, другое].

[Сумма удельных весов должна быть равна единице: $\sum m_i = 1$, при этом, **$0{,}2 \le m_1 \le 0{,}8$ (согласно Положению об организации промежуточной аттестации и текущего контроля успеваемости студентов НИУ ВШЭ, утвержденному приказом НИУ ВШЭ от 19.08.2014 №6.18.1-01/1908-02)**

На экзамене студент может получить дополнительный вопрос (дополнительную практическую задачу, решить к пересдаче домашнее задание), ответ на который оценивается в 1 балл.

[Оставьте те оценки, которые учитываются при выставлении результирующей оценки за промежуточный или завершающий контроль. Сумма удельных весов должна быть равна единице: $\sum k_i = 1$, при этом, **$0{,}2 \le k_1 \le 0{,}8$ После всех формул в обязательном порядке приводится способ округления полученного результата.**]

[Только для многомодульных дисциплин, по которым предусмотрен промежуточный контроль, укажите один из предложенных вариантов формирования оценки, которая идет в диплом]