

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

Институт Образования
Центр мониторинга качества образования

**Рабочая программа дисциплины
Теория и практика компьютерного тестирования**

для образовательной программы Измерения в психологии и образовании
37.04.01 Психология
уровень магистр

Разработчик(и) программы
Авдеева С.М., к.т.н., savdeeva@hse.ru
Федерякин Д.А., dafederiakin@hse.ru

Утверждена Академическим советом образовательной программы
«28»августа 2017 г., № протокола 01

Москва, 2017

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

1 Область применения и нормативные ссылки

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 Теория и практика компьютерного тестирования, учебных ассистентов и студентов направления подготовки/специальности 37.04.01 Психология, обучающихся по образовательной программе [Измерения в психологии и образовании].

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом высшего образования федерального государственного автономного образовательного учреждения высшего профессионального образования «Национального исследовательского университета «Высшей школы экономики», по направлению 37.04.01 «Психология»
- Образовательной программой 37.04.01 Психология.
- Объединенным учебным планом университета по образовательной программе Измерения в психологии и образовании, утвержденным в 2013г.

2 Цели освоения дисциплины

Целями освоения дисциплины Теория и практика компьютерного тестирования являются Освоение базовых навыков работы в программной среде R;
Овладение различными моделями и технологиями компьютерного тестирования, умение выбирать наиболее оптимальную из них для конкретной практической задачи
Знакомство с технологически-усовершенствованными и инновационными типами заданий и подходами к их разработке.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

[Компетенции для программы учебной дисциплины берутся из стандартов: ФГОС/ ОС НИУ ВШЭ для соответствующего уровня и направления подготовки и из числа закрепленных за дисциплиной в матрице компетенций образовательной программы]

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

В результате освоения дисциплины студент осваивает компетенции:

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен принимать управленческие решения, оценивать их возможные последствия и нести за них ответственность	СК-М5	СД/МЦ	Понимает достоинства и недостатки различных моделей, технологий и программных решений; Может найти оптимальный компромисс между затра-	Лекции, чтение литературы, обсуждения, сравнения различных подходов к разработке и реализации компьютерных тестов	Проверочная работа (тест), домашнее задание

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
			тами на разработку компьютерного теста и предполагаемым результатом; Способен обосновать разным группам коллег и пользователей результатов выбранную стратегию разработки компьютерного теста в связи с поставленной задачей и имеющимися ресурсами.		
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	РБ	Способен обосновать разным группам коллег и пользователей результатов выбранную стратегию разработки компьютерного теста в связи с поставленной задачей и имеющимися ресурсами; Способен презентовать результаты своей работы и поддерживать научную коммуникацию с различными группами пользователей	Обсуждения, презентации, подробная обратная связь на домашние работы с последующим обсуждением комментариев	Экзамен, домашние задания
Способен использовать современные средства ИКТ для поиска и обработки информации, работы с базами данных профессиональной информации и сетевой коммуникации.	ПК-4	СД	Демонстрирует навыки владения основами программирования на языке R; Способен использовать различные программные пакеты для решения задач психометрической обработки данных тестирования; Может использовать различные программные решения для разработки и администрирования раз-	Освоение дистанционной части курса, лекции, чтение релевантной литературы, практические домашние задания, подробная обратная связь на домашние работы с последующим обсуждением комментариев	Предоставление сертификата об освоении онлайн-курса, домашние задания

				ры	ские заня- тия	виды работы	работа
1	Введение в компьютерное тестирование (КТ). Модели КТ		2				
2	Основы Компьютерного Адаптивного Тестирования (КАТ)		2		2		
3	КАТ: структура, механизмы, практические проблемы.		2		4		
4	КАТ и Многоступенчатое тестирование (МСТ)		2		2		
5	Формирование банка заданий		2		2		
6	Технологически-усовершенствованные и инновационные типы заданий. Подходы к их разработке.			2			
7	Специфические проблемы: многомерное и политомическое КАТ и МСТ			2			

6 Формы контроля знаний студентов

Тип контроля	Форма контроля	2 год				Кафедра/подразделение	Параметры **
		1	2	3	4		
Текущий	Контрольная работа	*	*				Бланковый тест на 10 минут
	Домашнее задание	*	*				Технический отчет не более 3 (5) страниц
	Другие формы (указать)	*					Сертификат об освоении онлайн-курса
Итоговый	Экзамен		*				Экзамен в виде защиты презентации по проделанной работе

7 Критерии оценки знаний, навыков

Сертификат: процент освоения материала онлайн-курса является прямым компонентом итоговой оценки. Это демонстрирует степень освоения материала онлайн-курса.

Проверочная работа №1 по моделям компьютерного тестирования: тест из 10 заданий. Балл за тест – итоговый балл за данную работу. Тест проверяет понимание студентами достоинств и недостатков различных моделей компьютерного тестирования, а также их умение выбирать наиболее оптимальную из них в описанных условиях.

Домашняя работа N1 по результатам работы с симулятором компьютерного адаптивного теста. Форма отчета: технический отчет размером не более чем на 3 страницы. Отчет должен содержать описательную информацию о результатах трех прогонов симулятора компьютерного адаптивного теста: с полностью ожидаемыми ответами, с небольшим количеством неожиданных ответов, с большим количеством неожиданных ответов. Данная домашняя работа проверяет, насколько точно терминологически и содержательно студент способен описать работу компьютерного адаптивного теста.

Система оценки:

Балл	Критерий
10	Работа точна терминологически и демонстрирует глубокое понимание студентом особенностей функционирования адаптивного теста при применении различных психометрических моделей
9	То же самое, что в 10, но содержатся единичные случайные содержательные неточности
8	То же самое, что в 9, но неточности носят систематический характер
7	Имеются нарушения в использовании терминов, или студент продемонстрировал непонимание содержания некоторых терминов
6	То же самое, что в 7, но респондент продемонстрировал систематическое непонимание терминов
5	Имеются ошибки в аргументации выводов
4	Имеются систематические нарушения в логике повествования
3	Текст нечитабелен, или отчет содержит только таблицы с незначительной или отсутствующей интерпретацией
2	
1	
0	Работа не сдана в срок

Проверочная работа N2 по техническим аспектам компьютерного адаптивного и многоступенчатого тестирования: тест из 10 заданий. Балл за тест – итоговый балл за данную работу. Тест проверяет, насколько студент понимает технические аспекты применения и разницу между компьютерными адаптивными и многоступенчатыми тестами.

Домашняя работа N2 калибровка предоставленной базы данных. Данная работа проверяет умение студентов справляться с различными трудностями в работе с данными, похожими на реальные (например, несогласующиеся с моделью профили, большое количество угадывания, неясная структура базы данных и т.д.). По результатам студенты предоставляют технический отчет не больше, чем на 3 страницы с кратким описанием диагностированных проблем, процедур, примененных для их преодоления, и результатами этих процедур, а также с описанием, результатов калибровки модели.

Система оценки:

Балл	Критерий
10	Работа точна терминологически и демонстрирует глубокое понимание студентом особенностей функционирования адаптивного теста при применении различных психометрических моделей
9	То же самое, что в 10, но содержатся единичные случайные содержательные неточности
8	То же самое, что в 9, но неточности носят систематический характер
7	Имеются нарушения в использовании терминов, или студент продемонстрировал непонимание содержания некоторых терминов
6	То же самое, что в 7, но респондент продемонстрировал систематическое непонимание терминов
5	Имеются ошибки в аргументации выводов
4	Имеются систематические нарушения в логике повествования

3	Текст нечитабелен, или отчет содержит только таблицы с незначительной или отсутствующей интерпретацией
2	
1	
0	Работа не сдана в срок

Домашняя работа N3: разработка ядра компьютерного адаптивного теста. Данная работа проверяет умение студентов разрабатывать наборы правил, необходимых для функционирования ядра КАТ, проверять их функционирование в совокупности на полностью симулированных самостоятельно и предоставленных массивов данных, а также аргументировать свое финальное решение. По результатам студенты предоставляют технический отчет не больше, чем на 5 страниц, в котором описывают аргументацию и последовательность действий и различных наборов правил, которые они протестировали в работе.

Система оценки:

Балл	Критерий
10	Работа точна терминологически и демонстрирует глубокое понимание студентом особенностей функционирования адаптивного теста при применении различных психометрических моделей
9	То же самое, что в 10, но содержатся единичные случайные содержательные неточности
8	То же самое, что в 9, но неточности носят систематический характер
7	Имеются нарушения в использовании терминов, или студент продемонстрировал непонимание содержания некоторых терминов
6	То же самое, что в 7, но респондент продемонстрировал систематическое непонимание терминов
5	Имеются ошибки в аргументации выводов
4	Имеются систематические нарушения в логике повествования
3	Текст нечитабелен, или отчет содержит только таблицы с незначительной или отсутствующей интерпретацией
2	
1	
0	Работа не сдана в срок

Экзамен: представление полного цикла разработки ядра компьютерного теста от получения массива данных и пожеланий заказчика до представления результатов различным группам пользователей. Экзамен проходит в виде защиты презентации на 10 минут с последующим 10-минутным обсуждением представленных результатов всей учебной группой.

Система оценки:

Балл	Критерий
10	Работа точна терминологически и демонстрирует глубокое понимание студентом особенностей функционирования адаптивного теста при применении различных психометрических моделей
9	То же самое, что в 10, но содержатся единичные случайные содержательные неточности
8	То же самое, что в 9, но неточности носят систематический характер
7	Имеются нарушения в использовании терминов, или студент продемонстрировал непонимание содержания некоторых терминов
6	То же самое, что в 7, но респондент продемонстрировал систематическое непонимание терминов
5	Имеются ошибки в аргументации выводов
4	Имеются систематические нарушения в логике повествования
3	Текст нечитабелен, или отчет содержит только таблицы с незначительной или отсутствующей интерпретацией
2	
1	

8 Содержание дисциплины

Содержание дисциплины приведено в соответствии с тематическим планом учебной дисциплины, представленным в разделе 5. Количество часов аудиторной работы и самостоятельной работы по каждой теме приведено в разделе 5 «Тематический план учебной дисциплины». Выполнение домашних заданий предполагается по разделам 2 (2 штуки) и 5. На выполнение домашнего задания 1 (работа с симулятором компьютерного адаптивного теста, раздел 2) отводится 5 часов. На выполнение домашнего задания 2 (калибровка банка заданий, раздел 5) отводится 10 часов. На выполнение домашнего задания 3 (разработка ядра КАТ; раздел 2) отводится 15 часов. Остальное время самостоятельной подготовки студентов предназначено для чтения литературы по курсу.

Раздел 1. Введение в компьютерное тестирование (КТ). Модели КТ

Описание использования технологий в ассессменте, их потенциал и эволюция. Взаимосвязь развития технологий в тестировании и эволюции целей и понимания тестирования. Модели компьютерного тестирования: от линейного к адаптивному тестированию с применением байесовских сетей.

Обязательная литература:

1. de Ayala, R.J. (2008). The Theory and Practice of Item Response Theory. NY: The Guilford Press.
2. Bond, T. & Fox, C. (2007). Applying the Rasch model: Fundamental measurement in the human sciences (2nd edition). Mahwah, NJ: LEA.
3. Robitzsch, A., Kiefer, T., & Wu, M. (2018). TAM: Test analysis modules. R package, version 2.13-15.
4. Chalmers, P.R., Pritikin, J., Robitzsch, A., Zoltak, M., Kim, K., Falk, C.F., Meade, A., Schneider, L., King, D., Liu, C.-W., Oguzhan, O. (2018). Package ‘mirt’. R package, version 1.29.
5. Chalmers, R.P. (2018). mirt: A multidimensional item response theory package for the R environment. Journal of Statistical Software, 48(6), 1-29.

Рекомендованная литература:

1. Baker, F.B., Kim, S.-H. (2017). The Basics of Item Response Theory Using R. Switzerland, Cham: Springer.
2. Mair, P. (2018). Modern Psychometrics with R. Switzerland, Cham: Springer.
3. Wu, M., Tam, H.P., Jen, T.-H. (2018). Educational Measurement for Applied Researchers: Theory into Practice. Singapore: Springer.
4. Desjardins, C.D., Bulut, O. (2018). Handbook of Educational Measurement and Psychometrics Using R. Florida, Boca Raton: CRC Press.
5. van der Linden, W. J., Breithaupt, K., Chuah, S. C., & Zhang, Y. (2007). Detecting differential speededness in a multistage testing. Journal of Educational Measurement, 44, 117–130.

Раздел 2. Основы Компьютерного Адаптивного Тестирования (КАТ)

Основы КАТ. Походы к КАТ: традиционное линейное понимание конструктора и нелинейное моделирование конструктора с помощью байесовских сетей. Идеология КАТ, статистические и психометрические основания КАТ. Методы оценки параметров модели. Программное обеспечение для разработки ядра и администрирования КАТ.

Обязательная литература:

1. Becker, K.A., Bergstrom, B.A. (2013). Test Administration Models. Practical Assessment, Research and Evaluation, 18(14), 1-7.
2. de Ayala, R.J. (2008). The Theory and Practice of Item Response Theory. NY: The Guilford Press.

- Bond, T. & Fox, C. (2007). Applying the Rasch model: Fundamental measurement in the human sciences (2nd edition). Mahwah, NJ: LEA.
- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). Practical considerations in computer-based testing. New York: Springer.

Рекомендованная литература:

- Wim J. van der Linden, Cees A.W. Glas (Eds.). Elements of Adaptive Testing, Springer, 2010.
- Deng, H., Ansley, T., & Chang, H. H. (2010). Stratified and maximum information item selection procedures in computer adaptive testing. Journal of Educational Measurement, 47, 202–226.
- Magis, D., Yan, D., von Davier, A.A. (2017). Computerized Adaptive and Multistage Testing with R. Cham: Springer.

Раздел 3. КАТ: структура, механизмы, практические проблемы

Ядро КАТ: правила КАТ. Правило начала тестирования, правило выбора следующего задания, правило оценки параметров (начисления баллов), правило остановки тестирования. Комбинирование правил. Способы проверки функционирования наборов правил. Симуляции: “истинные” и реальные. Программное обеспечение для разработки ядра КАТ.

Обязательная литература:

- Becker, K.A., Bergstrom, B.A. (2013). Test Administration Models. Practical Assessment, Research and Evaluation, 18(14), 1-7.
- de Ayala, R.J. (2008). The Theory and Practice of Item Response Theory. NY: The Guilford Press.
- Bond, T. & Fox, C. (2007). Applying the Rasch model: Fundamental measurement in the human sciences (2nd edition). Mahwah, NJ: LEA.
- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). Practical considerations in computer-based testing. New York: Springer.

Рекомендованная литература:

- Wim J. van der Linden, Cees A.W. Glas (Eds.). Elements of Adaptive Testing, Springer, 2010.
- Deng, H., Ansley, T., & Chang, H. H. (2010). Stratified and maximum information item selection procedures in computer adaptive testing. Journal of Educational Measurement, 47, 202–226.
- Magis, D., Yan, D., von Davier, A.A. (2017). Computerized Adaptive and Multistage Testing with R. Cham: Springer.

Раздел 4. КАТ и Многоступенчатое тестирование (МСТ)

Многоступенчатое тестирование. Психометрические детали МСТ: локальная зависимость заданий. Способы повышения точности измерений с использованием МСТ. Относительные достоинства и недостатки различных моделей тестирования. Выбор модели тестирования под задачу.

Обязательная литература:

- Thompson, N.A., Weiss, D.J. (2011). A Framework for the Development of Computerized Adaptive Tests. Practical Assessment, Research and Evaluation, 16(1), 1-9.
- Thompson, N.A. (2007). A Practitioner’s Guide for Variable-length Computerized Classification Testing. Practical Assessment, Research and Evaluation, 12(1), 1-13.
- Almond, R.G., Mislevy, R.J. (1999). Graphical Models and Computerized Adaptive Testing. Applied Psychological Measurement, 23(3), 223-237.
- Liu, C., Li, J. (2012). Exploration and Reflection of the Design Models of Computerized Testing. AASRI Procedia, 101-106.
- Luecht, R.M., Sireci, S.G. (2011). A Review of Models for Computer-Based Testing. College Board Research Report 2011-12.
- Chalmers, R.P. (2016). Generating Adaptive and Non-Adaptive Test Interfaces for Multidimensional Item Response Theory Applications. Journal of Statistical Software, 71(5), 1-38.

Рекомендованная литература:

1. Adema, J.J., van der Linden, W.J. (1989). Algorithms for Computerized Test Construction Using Classical Item Parameters. *Journal of Educational and Behavioral Statistics*, 14(3), 279-290.
2. Bunderson, C.V., Inouye, D.K., Olsen, J.B. (1988). The Four Generations of Computerized Educational Measurement. ETS Research Report Series.
3. van der Linden, W.J., Glas, G.A.W. (2000). *Computerized Adaptive Testing: Theory and Practice*. Dordrecht: Springer.
4. Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J. (2000). *Computerized Adaptive Testing: a Primer* (2nd edition). NY: Routledge.
5. Parshall, C.G., Spray, J.A., Kalohn, J., Davey, T. (2002). *Practical Considerations in Computer-Based Testing*. NY: Springer.
6. Wainer, H., Kiely, G.L. (1987). Item Clusters and Computerized Adaptive Testing: A Case for Testlets. *Journal of Educational Measurement*, 24(3), 185-201.
7. Cheng, Y. (2009). When Cognitive Diagnosis Meets Computerized Adaptive Testing: CD-CAT. *Psychometrika*, 74(4), 619-632.
8. Adema, J. J., & Van der Linden, W. J. (1989). Algorithms for computerized test construction using classical item parameters. *Journal of Educational Statistics*, 14(3), 279-290.
9. Eggen, T. (2007). Choices in CAT Models in the Context of Educational Testing. Paper presented on the Conference on Computerized Adaptive Testing.

Раздел 5. Формирование банка заданий

Банк заданий: определение, характеристики. Различные метрики банка заданий. Способы защиты банка заданий. Сверхпредъявление заданий: диагностика и способы предотвращения. Способы контроля содержания при предъявлении заданий из банка, и их воздействие на результаты измерения.

Обязательная литература:

1. Way, W.D. (2005). Protecting the Integrity of Computerized Testing Item Pools. *Educational Measurement: Issues and Practice*, 17(4), 17-27.
2. Stocking M.L., Swanson, L. (1998). Optimal Design of Item Banks for Computerized Adaptive Tests. *Applied Psychological Measurement*, 22(3), 271-279.
3. Belov D.I., Armstrong, R.D. (2009). Direct and Inverse Problems of Item Pool Design for Computerized Adaptive Testing. *Educational and Psychological Measurement*, 69(4), 533-547.

Рекомендованная литература:

1. Wise, S. L., Kingsbury, G. G., & Webb, N. L. (2015). Evaluating content alignment in computerized adaptive testing. *Educational Measurement: Issues and Practice*, 34(4), 41-48.
2. Magis, D., Raiche, G., Barrada, J.R. (2018). Package 'catR'. R package, version 3.16.
3. Magis, D., Barrada, J.R. (2017). Computerized Adaptive Testing with R: Recent Updates of the Package catR. *Journal of Statistical Software*, 71, 1-19.
4. Adema, J.J., van der Linden, W.J. (1989). Algorithms for Computerized Test Construction Using Classical Item Parameters. *Journal of Educational and Behavioral Statistics*, 14(3), 279-290.
5. Magis, D., Yan, D., von Davier, A.A. (2017). *Computerized Adaptive and Multistage Testing with R*. Cham: Springer.
6. Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(4), 311-327.
7. Bjorner, J.B., Kosinski, M., Ware J.E. Jr. (2005). Computerized Adaptive Testing and Item Banking. In book *Assessing Quality of Life in Clinical Trials: Methods and Practice*, eds. Fayers P.M., Hays, R. NY: Oxford University Press.

Раздел 6. Технологически-усовершенствованные и инновационные типы заданий. Подходы к их разработке

Возможности для использования технологий при разработке и модификации тестовых заданий. Технологически-усовершенствованные задания: определения и классификации. Сложности в разработке и валидации технологически-усовершенствованных заданий. Evidence-Centred Design как способ преодоления трудностей при разработке и валидации технологически-усовершенствованных заданий.

Обязательная литература:

1. Magis, D., Raiche, G., Barrada, J.R. (2018). Package 'catR'. R package, version 3.16.
2. Chalmers, P.R., Pritikin, J., Robitzsch, A., Zoltak, M., Kim, K., Falk, C.F., Meade, A., Schneider, L., King, D., Liu, C.-W., Oguzhan, O. (2018). Package 'mirt'. R package, version 1.29.
3. Robitzsch, A., Kiefer, T., & Wu, M. (2018). TAM: Test analysis modules. R package, version 2.13-15.
4. Cowley, B., Fantato, M., Jennett, C., Ruskov, M., & Ravaja, N. (2014). Learning When Serious: Psychophysiological Evaluation of a Technology-Enhanced Learning Game. *Educational Technology & Society*, 17(1), 3-16.
5. Bryant, W. (2017). Developing a Strategy for Using Technology-Enhanced Items in Large-Scale Standardized Tests. *Practical Assessment, Research & Evaluation*, 22(1).
6. Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical education*, 46(8), 757-765.

Рекомендованная литература:

1. van der Linden, W.J., Glas, G.A.W. (2000). *Computerized Adaptive Testing: Theory and Practice*. Dordrecht: Springer.
2. Drasgow, F. (2015). *Technology and Testing: Improving Educational and Psychological Measurement*. NY: Routledge.
3. Ally Thomas (2016). *Evaluating the Validity of Technology-Enabled Educational Assessment Items and Tasks: An Empirical Approach to Studying Item Features and Scoring Rubrics*. The Graduate Center, City University of New York.
4. Alghazo, I. M. (2006). Student Attitudes toward Web-Enhanced Instruction in an Educational Technology Course. *College Student Journal*, 40(3).

Раздел 7. Специфические проблемы: многомерное и политомическое КАТ и МСТ

Трудности использования многомерного моделирования в КАТ. Изменение параметров заданий и их влияние на результаты КАТ. Политомическое КАТ в психологическом и образовательном тестировании. Использование технологий при проверке открытых заданий. Использование технологий КАТ в работе экспертов.

Обязательная литература:

1. Bryant, W. (2017). Developing a Strategy for Using Technology-Enhanced Items in Large-Scale Standardized Tests. *Practical Assessment, Research & Evaluation*, 22(1).
2. Dodd, B. G., De Ayala, R. J., & Koch, W. R. (1995). Computerized adaptive testing with polytomous items. *Applied psychological measurement*, 19(1), 5-22.
3. Paap, M. C., Kroeze, K. A., Terwee, C. B., van der Palen, J., & Veldkamp, B. P. (2017). Item usage in a multidimensional computerized adaptive test (MCAT) measuring health-related quality of life. *Quality of life research*, 26(11), 2909-2918.
4. Smits, N., Paap, M. C., & Böhnke, J. R. (2018). Some recommendations for developing multidimensional computerized adaptive tests for patient-reported outcomes. *Quality of Life Research*, 27(4), 1055-1063.

Рекомендованная литература:

1. Choi, S. W. (2009). Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement*, 33, 644-645.
2. Choi, S. W., Swartz, R. J. (2009). Comparison of CAT item selection criteria for polytomous items. *Applied Psychological Measurement*, 32, 419-440.

3. van Groen, M. M., Eggen, T. J., & Veldkamp, B. P. (2016). Multidimensional Computerized Adaptive Testing for Classifying Examinees with Within-Dimensionality. *Applied psychological measurement*, 40(6), 387-404.
4. Seo, D. G., & Weiss, D. J. (2015). Best design for multidimensional computerized adaptive testing with the bifactor model. *Educational and psychological measurement*, 75(6), 954-978.
5. Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61(2), 331-354.

9 Образовательные технологии

Лекции, семинары, самостоятельная работа, проверочные работы, практические занятия, работа по группам.

10 Оценочные средства для текущего контроля и аттестации студента

10.1 Оценочные средства для оценки качества освоения дисциплины в ходе текущего контроля

Пример вопроса из проверочной работы 1:

В отличие от линейной модели тестирования, компьютерное адаптивное тестирование с использованием байесовских сетей допускает:

- А) Нелинейное понимание природы измеряемого конструкта;
- Б) Невзаимозаменяемость заданий в тесте;
- В) Неоднородность инструмента;
- Г) Использование мультимедиа-элементов в стимульном материале.

Пример вопроса из проверочной работы 2:

Банк вопросов для компьютерного адаптивного теста имеет тестлет-дизайн: несколько заданий относятся к одному стимульному материалу. Среди следующих альтернатив выберите ту, которая НЕ подходит для дизайна данного теста:

- А) Многоступенчатое тестирование;
- Б) Использование модели, учитывающей локальную зависимость;
- В) Адаптивный тест с последовательным предъявлением заданий;
- Г) Многомерное моделирование конструкта.

Пример ситуации для экзаменационного задания и домашних заданий:

К вам обратился заказчик – компания, занимающаяся разработкой и продажей тестов и опросников. Компания разрабатывает дихотомический тест, и желает сделать его компьютерным адаптивным. Штатный ведущий психометрик этого проекта уволился, оставив проект на середине разработки. К сожалению, все проектные файлы с компьютера предыдущего психометрика недоступны по техническим причинам, вам могут предоставить только базу апробационного тестирования.

Разработайте ядро компьютерного адаптивного теста под предоставленный банк заданий.

Детали заказа:

1. Тест состоит из дихотомических заданий и предназначен для сбора информации об отношении пользователей популярного онлайн-сервиса к новым функциям.
2. Всем респондентам апробационной выборки предъявлялись 10 случайных заданий из общего неоткалиброванного банка.
3. Онлайн-сервис хочет использовать этот банк заданий в дальнейшем для мониторинга удовлетворенности пользователей этими функциями и их изменениями. Требуются ваши рекомендации по улучшению банка заданий.

11 Порядок формирования оценок по дисциплине

Дисциплина преподается 2 модуля на 2 курсе (1 и 2 модули), но не имеет оценки в конце первого модуля. Итоговая оценка складывается из домашних и проверочных работ и экзаменационной презентации.

$$O_{\text{накопленная}} = 0,33 * O_{\text{сертификат}} + 0,0825 * O_{\text{проверочная 1}} + 0,0825 * O_{\text{ДЗ 1}} + 0,0825 * O_{\text{проверочная 2}} + 0,0825 * O_{\text{ДЗ 2}} + 0,33 * O_{\text{ДЗ 3}}$$

Способ округления накопленной оценки текущего контроля: в пользу студента.

Результирующая оценка за дисциплину рассчитывается следующим образом:

$$O_{\text{результ}} = 0,6 * O_{\text{накопл}} + 0,4 * O_{\text{экз}}$$

Способ округления итоговой оценки: в пользу студента.

12 Учебно-методическое и информационное обеспечение дисциплины

12.1 Базовый учебник

[Укажите базовый учебник. Норматив обеспеченности студентов базовым учебником - не менее 50 %. Ридеры указываются при отсутствии базового учебника или неполного покрытия базовым учебником основных тем программы один из предложенных вариантов формирования оценки, которая идет в диплом.

Укажите, если доступна электронная версия базового учебника].

12.2 Основная литература

1. Almond, R.G., Mislevy, R.J. (1999). Graphical Models and Computerized Adaptive Testing. *Applied Psychological Measurement*, 23(3), 223-237.
2. Becker, K.A., Bergstrom, B.A. (2013). Test Administration Models. *Practical Assessment, Research and Evaluation*, 18(14), 1-7.
3. Belov D.I., Armstrong, R.D. (2009). Direct and Inverse Problems of Item Pool Design for Computerized Adaptive Testing. *Educational and Psychological Measurement*, 69(4), 533-547.
4. Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd edition). Mahwah, NJ: LEA.
5. Bryant, W. (2017). Developing a Strategy for Using Technology-Enhanced Items in Large-Scale Standardized Tests. *Practical Assessment, Research & Evaluation*, 22(1).
6. Chalmers, P.R., Pritikin, J., Robitzsch, A., Zoltak, M., Kim, K., Falk, C.F., Meade, A., Schneider, L., King, D., Liu, C.-W., Oguzhan, O. (2018). Package 'mirt'. R package, version 1.29.
7. Chalmers, R.P. (2016). Generating Adaptive and Non-Adaptive Test Interfaces for Multidimensional Item Response Theory Applications. *Journal of Statistical Software*, 71(5), 1-38.
8. Chalmers, R.P. (2018). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29.
9. Cowley, B., Fantato, M., Jennett, C., Ruskov, M., & Ravaja, N. (2014). Learning When Serious: Psychophysiological Evaluation of a Technology-Enhanced Learning Game. *Educational Technology & Society*, 17(1), 3-16.
10. de Ayala, R.J. (2008). *The Theory and Practice of Item Response Theory*. NY: The Guilford Press.
11. Dodd, B. G., De Ayala, R. J., & Koch, W. R. (1995). Computerized adaptive testing with polytomous items. *Applied psychological measurement*, 19(1), 5-22.
12. Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical education*, 46(8), 757-765.

13. Liu, C., Li, J. (2012). Exploration and Reflection of the Design Models of Computerized Testing. *AASRI Procedia*, 101-106.
14. Luecht, R.M., Sireci, S.G. (2011). A Review of Models for Computer-Based Testing. College Board Research Report 2011-12.
15. Paap, M. C., Kroeze, K. A., Terwee, C. B., van der Palen, J., & Veldkamp, B. P. (2017). Item usage in a multidimensional computerized adaptive test (MCAT) measuring health-related quality of life. *Quality of life research*, 26(11), 2909-2918.
16. Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). Practical considerations in computer-based testing. New York: Springer.
17. Robitzsch, A., Kiefer, T., & Wu, M. (2018). TAM: Test analysis modules. R package, version 2.13-15.
18. Smits, N., Paap, M. C., & Böhnke, J. R. (2018). Some recommendations for developing multidimensional computerized adaptive tests for patient-reported outcomes. *Quality of Life Research*, 27(4), 1055-1063.
19. Stocking M.L., Swanson, L. (1998). Optimal Design of Item Banks for Computerized Adaptive Tests. *Applied Psychological Measurement*, 22(3), 271-279.
20. Thompson, N.A. (2007). A Practitioner's Guide for Variable-length Computerized Classification Testing. *Practical Assessment, Research and Evaluation*, 12(1), 1-13.
21. Thompson, N.A., Weiss, D.J. (2011). A Framework for the Development of Computerized Adaptive Tests. *Practical Assessment, Research and Evaluation*, 16(1), 1-9.
22. Way, W.D. (2005). Protecting the Integrity of Computerized Testing Item Pools. *Educational Measurement: Issues and Practice*, 17(4), 17-27.

12.3 Дополнительная литература

1. Adema, J.J., van der Linden, W.J. (1989). Algorithms for Computerized Test Construction Using Classical Item Parameters. *Journal of Educational and Behavioral Statistics*, 14(3), 279-290.
2. Alghazo, I. M. (2006). Student Attitudes toward Web-Enhanced Instruction in an Educational Technology Course. *College Student Journal*, 40(3).
3. Ally Thomas (2016). Evaluating the Validity of Technology-Enabled Educational Assessment Items and Tasks: An Empirical Approach to Studying Item Features and Scoring Rubrics. The Graduate Center, City University of New York.
4. Baker, F.B., Kim, S.-H. (2017). *The Basics of Item Response Theory Using R*. Switzerland, Cham: Springer.
5. Bjorner, J.B., Kosinski, M., Ware J.E. Jr. (2005). Computerized Adaptive Testing and Item Banking. In book *Assessing Quality of Life in Clinical Trials: Methods and Practice*, eds. Fayers P.M., Hays, R. NY: Oxford University Press.
6. Bunderson, C.V., Inouye, D.K., Olsen, J.B. (1988). *The Four Generations of Computerized Educational Measurement*. ETS Research Report Series.
7. Cheng, Y (2009). When Cognitive Diagnosis Meets Computerized Adaptive Testing: CD-CAT. *Psychometrika*, 74(4), 619-632.
8. Choi, S. W. (2009). Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement*, 33, 644-645.
9. Choi, S. W., Swartz, R. J. (2009). Comparison of CAT item selection criteria for polytomous items. *Applied Psychological Measurement*, 32, 419-440.
10. Deng, H., Ansley, T., & Chang, H. H. (2010). Stratified and maximum information item selection procedures in computer adaptive testing. *Journal of Educational Measurement*, 47, 202-226.
11. Desjardins, C.D., Bulut, O. (2018). *Handbook of Educational Measurement and Psychometrics Using R*. Florida, Boca Raton: CRC Press.
12. Drasgow, F. (2015). *Technology and Testing: Improving Educational and Psychological Measurement*. NY: Routledge.

13. Eggen, T. (2007). Choices in CAT Models in the Context of Educational Testing. Paper presented on the Conference on Computerized Adaptive Testing.
14. Magis, D., Barrada, J.R. (2017). Computerized Adaptive Testing with R: Recent Updates of the Package catR. *Journal of Statistical Software*, 71, 1-19.
15. Magis, D., Raiche, G., Barrada, J.R. (2018). Package 'catR'. R package, version 3.16.
16. Magis, D., Yan, D., von Davier, A.A. (2017). *Computerized Adaptive and Multistage Testing with R*. Cham: Springer.
17. Mair, P. (2018). *Modern Psychometrics with R*. Switzerland, Cham: Springer.
18. Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(4), 311-327.
19. Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61(2), 331-354.
20. Seo, D. G., & Weiss, D. J. (2015). Best design for multidimensional computerized adaptive testing with the bifactor model. *Educational and psychological measurement*, 75(6), 954-978.
21. van der Linden, W. J., Breithaupt, K., Chuah, S. C., & Zhang, Y. (2007). Detecting differential speededness in a multistage testing. *Journal of Educational Measurement*, 44, 117-130.
22. van der Linden, W.J., Glas, G.A.W. (2000). *Computerized Adaptive Testing: Theory and Practice*. Dordrecht: Springer.
23. van Groen, M. M., Eggen, T. J., & Veldkamp, B. P. (2016). Multidimensional Computerized Adaptive Testing for Classifying Examinees with Within-Dimensionality. *Applied psychological measurement*, 40(6), 387-404.
24. Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J. (2000). *Computerized Adaptive Testing: a Primer* (2nd edition). NY: Routledge.
25. Wainer, H., Kiely, G.L. (1987). Item Clusters and Computerized Adaptive Testing: A Case for Testlets. *Journal of Educational Measurement*, 24(3), 185-201.
26. Wim J. van der Linden, Cees A.W. Glas (Eds.). *Elements of Adaptive Testing*, Springer, 2010.
27. Wise, S. L., Kingsbury, G. G., & Webb, N. L. (2015). Evaluating content alignment in computerized adaptive testing. *Educational Measurement: Issues and Practice*, 34(4), 41-48.
28. Wu, M., Tam, H.P., Jen, T.-H. (2018). *Educational Measurement for Applied Researchers: Theory into Practice*. Singapore: Springer.

12.4 Программные средства

п/п	Наименование	Условия доступа
1	Ядро статистического пакета R	<i>Свободное лицензионное соглашение</i>
2	Оболочка статистического пакета RStudio	<i>Свободное лицензионное соглашение</i>
3	Платформа Concerto	<i>Свободное лицензионное соглашение</i>

12.5 Дистанционная поддержка дисциплины

Онлайн-курс расположен на платформе Stepik: <https://stepik.org/497>.

Язык преподавания: Русский.

Длина курса: 19 часов.

Для прохождения курса не нужно обладать никакими предварительными навыками программирования. Тем не менее, для успешного завершения курса нужно иметь установленный язык R (версии 3.2.0 или новее). В качестве среды разработки в видеороках используется RStudio.

Организация-разработчик: Институт биоинформатики (<http://bioinformaticsinstitute.ru/>).