

МОДЕЛИРОВАНИЕ И АНАЛИЗ ДАННЫХ ДЛЯ ЦИФРОВОГО ОБРАЗОВАНИЯ

Применение современной теории тестов (IRT) для анализа методики «Шкала оценки цифровой образовательной среды (ЦОС)»

Радчикова Н.П.

Московский государственный психолого-педагогический университет
(ФГБОУ ВО МГППУ), г. Москва, Российская Федерация
ORCID: <https://orcid.org/0000-0002-5139-8288>
e-mail: nataly.radchikova@gmail.com

Сорокова М.Г.

Московский государственный психолого-педагогический университет
(ФГБОУ ВО МГППУ), г. Москва, Российская Федерация
ORCID: <http://orcid.org/0000-0002-1000-6487>
e-mail: sorokovamg@mgppu.ru

Одинцова М.А.

Московский государственный психолого-педагогический университет
(ФГБОУ ВО МГППУ), г. Москва, Российская Федерация
ORCID: <https://orcid.org/0000-0003-3106-4616>
e-mail: mari505@mail.ru

Гусарова Е.С.

Московский государственный психолого-педагогический университет
(ФГБОУ ВО МГППУ), г. Москва, Российская Федерация
ORCID: <https://orcid.org/0000-0001-6968-5253>
e-mail: gusarovaes@mgppu.ru

С позиции современной теории тестов (Item Response Theory – IRT) рассмотрена методика «Шкала оценки ЦОС», отражающая шесть различных аспектов цифровой образовательной среды (ЦОС) университета: удовлетворенность и практическая польза; удовлетворенность коммуникативным взаимодействием; стресснапряженность; необходимость поддержки; нечестные стратегии; доступность. Ответы на 38 вопросов методике даются по 5-балльной шкале Ликерта. В исследовании приняли участие 406 студентов МГППУ, закончившие электронные курсы в онлайн-формате (90,1 % женского пола). Возраст участников в среднем составил $28,7 \pm 9,6$ лет (медиана 24 года, минимум 19 лет, максимум 72 года). Применение модели градуированных ответов (GRM) отдельно к каждой из шести субшкал методики показало, что все вопросы методики имеют дискриминативность пункта не ниже умеренной, а большинство вопросов – высокую. Пороговые значения для каждого вопроса методики, соответствующие относительному

индикатору трудности, монотонно возрастают. Графики информационной функции для каждой шкалы методики говорят о том, что практически все измерения компонентов Шкалы оценки ЦОС являются довольно точными в диапазоне средних значений, и только для шкалы 6 «Доступность ЦОС» точность оценки высока для самых низких значений и резко падает для средних и высоких значений.

Ключевые слова: Цифровая образовательная среда (ЦОС), Шкала оценки ЦОС, современная теория тестов (IRT), трудность пункта, дискриминативность пункта, информационная функция теста.

Финансирование. Работа выполнена при финансовой поддержке ФГБОУ ВО «Московский государственный психолого-педагогический университет» в рамках научно-исследовательского проекта «Цифровые технологии в высшем образовании: разработка технологии индивидуализации обучения средствами электронных учебных курсов».

Для цитаты:

Радчикова Н.П., Сорокова М.Г., Одицова М.А., Гусарова Е.С. Применение современной теории тестов (IRT) для анализа методики «Шкала оценки ЦОС» // Цифровая гуманитаристика и технологии в образовании (ДНТЕ 2021): сб. статей II Всероссийской научно-практической конференции с международным участием. 11–12 ноября 2021 г. / Под ред. В.В. Рубцова, М.Г. Сороковой, Н.П. Радчиковой. М.: Издательство ФГБОУ ВО МГППУ, 2021. 557–570 с.

Введение

Цифровая образовательная среда (ЦОС) как совокупность условий для реализации образовательных программ с применением электронного обучения и дистанционных образовательных технологий является сегодня неотъемлемой частью учебного процесса любого вуза. Быстрое развитие ЦОС университетов требует ее постоянной оценки, пересмотра педагогических технологий для обеспечения их соответствия ожиданиям всех участников образования, а это, в свою очередь, требует хорошего диагностического инструментария, позволяющего дать такой среде комплексную оценку. Одним из таких диагностических инструментов является методика «Шкала оценки ЦОС» [4], которая отвечает требованиям надежности, внутренней и внешней валидности и позволяет дать комплексную оценку на основании ряда индикаторов: удовлетворенность и практическая польза; удовлетворенность коммуникативным взаимодействием; стресснапряженность; необходимость поддержки; нечестные стратегии; доступность, и может использоваться наряду с другими объективными характеристиками ЦОС

(наполняемость учебных дисциплин, процент видео-лекций, обеспеченность контрольно-измерительными материалами, образовательные достижения и т.д.).

Проверка Шкалы оценки ЦОС [4] была проведена с позиции классической теории тестов, которая часто подвергается критике из-за невозможности оценить каждый из вопросов (пунктов) методики отдельно и определить трудность пунктов теста; за зависимость результатов тестирования от исследуемой выборки, а также за зависимость характеристики тестируемого или выборки от того, какой тест используется [1, 2]. Из-за этих недостатков в последнее время все чаще стала использоваться теория IRT (Item-Response Theory), которая называется теорией латентных черт, современной теорией тестов, теорией моделирования и параметризации педагогических тестов [3], математической теорией педагогических измерений [1]. Мы будем придерживаться названия «современная теория тестов», так как оно употребляется чаще всего и не привязано только к педагогическим измерениям. Целью данной работы было применение современной теории тестов к методике «Шкала оценки ЦОС».

Метод

Участники исследования. Всего в исследовании приняли участие 406 человек. В момент исследования все они были студентами Московского государственного психолого-педагогического университета (МГППУ), закончившими электронные курсы в онлайн-формате (90,1 % (N=366) женского пола). Возраст участников в среднем составил $28,7 \pm 9,6$ лет (медиана 24 года, мода 20 лет, минимум 19 лет, максимум 72 года). База данных собрана в сентябре – декабре 2020 года, когда университет работал в дистанционном режиме, и доступна в репозитории RusPsyDATA [11].

Методики исследования. Шкала ЦОС университета [4], позволяющая дать комплексную характеристику ЦОС на основании выделения шести индикаторов: удовлетворенность учебным процессом; удовлетворенность коммуникативным взаимодействием; стресснапряженность; необходимость поддержки; нечестные стратегии при контроле знаний; доступность.

Статистический анализ. Согласно современной теории тестов, исследователей интересует некая характеристика, которая не может быть измерена прямо (латентная характеристика), однако предполагается, что индивид обладает этой характеристикой в некоторой мере. Эта мера, определяющая положение индивида на шкале латентной характеристики и выраженная в количестве баллов, обозначается буквой Θ . Для каждого уровня такой характеристики индивид с определенной вероятностью $P(\Theta)$ даст правильный (нужный)

ответ на вопрос теста. Согласно современной теории тестов, каждый пункт теста может иметь две характеристики: трудность пункта и дискриминативность (различительная способность) пункта. Трудность пункта (β) характеризуется индексом, который соответствует доле лиц, правильно решивших задание (верно ответивших на вопрос). Дискриминативность (α) показывает, насколько хорошо можно разделять отдельных индивидов по уровню выполнения.

Еще одной характеристикой теста является информационная функция (test information curve $I(\Theta)$), которая вычисляется на основании суммирования информации, полученной от каждого пункта опросника. График данной функции показывает, как методика оценивает величину измеряемой характеристики в полном диапазоне значений этой характеристики. Ярко выраженный пик на этой кривой свидетельствует о неравномерности точности измерения: такая методика будет подходящей для точного измерения конструкта только для тех тестируемых, чьи значения находятся вблизи пика кривой. Плоские участки кривой свидетельствуют о том, что в данном диапазоне измерения можно считать достаточно точными.

Проверка моделей и вычисление показателей современной теории тестов (IRT) проводилась с помощью пакета MIRT в R.

Результаты и обсуждение

При применении IRT предполагается, что с помощью данного теста измеряется только один конструкт, то есть только одна латентная переменная, поэтому анализ осуществлялся для каждой субшкалы Шкалы оценки ЦОС отдельно.

Чтобы оценить диагностические свойства методики Шкала оценки ЦОС с психометрической точки зрения, сначала было проведено сравнение двух моделей: модели градуированного ответа (graded response model [10]) и обобщенной модели частичного доверия (generalized partial credit model [8]), так как именно эти две модели используются для методик, в которых ответы даются по шкале Ликерта. Модель градуированного ответа является простейшей моделью, в которой каждый возможный ответ сравнивается со всеми возможными ответами выше данного уровня. Например, β_1 является пороговым значением предпочтения ответа “1” над всеми другими ответами (например, “2”, “3”, “4” и “5”), β_2 является пороговым значением между ответами “2” и “3”, “4”, “5” и т.д. Обобщенная модель частичного доверия не предполагает, что ответы упорядочены, и вычисляет значения β , сравнивая только близлежащие ответы. Например, β_1 будет показывать порог выбора между ответами “1” и “2”, β_2 будет показывать порог выбора между ответами “2” и “3” и т.д. Для определения наилучшей модели в нашем случае были рассчитаны информационный

критерий Акаике (Akaike Information Criterion, AIC) и Байесовский информационный критерий (Bayesian information criterion, BIC). Известно, что предпочтительнее выбирать модель с более низкими AIC и BIC, так как они указывает на более близкое соответствие данных истинной модели [6]. Результаты вычислений показали, что модель градуированного ответа лучше соответствует данным (табл. 1).

Таблица 1

Сравнение модели градуированного ответа (GRM) и обобщенной модели частичного доверия (GPCM) для каждой субшкалы методики

Шкала методики	BIC		AIC	
	GRM	GPCM	GRM	GPCM
Шкала 1	10718,23	10820,61	10477,85	10580,22
Шкала 2	7691,61	7738,45	7551,39	7598,23
Шкала 3	8479,56	8616,37	8319,30	8456,12
Шкала 4	6455,08	6497,94	6334,89	6377,75
Шкала 5	6296,85	6317,62	6176,66	6197,43
Шкала 6	4128,54	4200,32	4032,39	4104,16

Примечания: BIC – Байесовский информационный критерий; AIC – информационный критерий Акаике; GRM – модель градуированного ответа; GPCM – обобщенная модель частичного доверия.

Пороговые значения (β) для каждого вопроса (пункта) методики, дискриминативность пунктов (α) и статистика критерия $S-\chi^2$ (только уровень статистической значимости) отражены в табл. 2–3. Результаты показывают (табл. 2–3), что дискриминативность пунктов опросника довольно высока [5, с. 34]: только семь пунктов имеют умеренную дискриминативность ($< 1,34$), а остальные – высокую ($1,35-1,69$) и очень высокую ($> 1,70$). Наименьшую дискриминативность имеет вопрос 24 «Все равно будут студенты, кто использует нечестные стратегии при контроле знаний» ($\alpha=0,93$) из шкалы «Нечестные стратегии в ЦОС», и шкала «Нечестные стратегии в ЦОС» содержит вопросы с самой низкой дискриминативностью. Наибольшую дискриминативность имеет вопрос 30 «Освоение электронного курса дает мне ощущение удовлетворенности учебным процессом» ($\alpha=3,19$), и шкала «Удовлетворенность учебным процессом в ЦОС» содержит вопросы с самой высокой дискриминативностью.

Результаты также показывают (табл. 2–3), что пороговые значения β для каждого вопроса методики, соответствующие относительному индикатору трудности, монотонно возрастают. Самый большой диапазон измеряемого латентного показателя перекрывают вопросы шкалы 5 «Нечестные стратегии в ЦОС» (от -5,37 до 4,10), а самый небольшой – вопросы шкалы 6 «Доступность ЦОС» (от -3,34 до 0,70).

Таблица 2

**Статистика по каждому вопросу (пункту)
субшкал 1–3 методики Шкала оценки ЦОС**

Вопрос	α	β_1	β_2	β_3	β_4	Уровень статистической значимости р критерия S- χ^2
Шкала 1. Удовлетворенность учебным процессом в ЦОС						
Вопрос 3	1,52	-2,92	-1,86	-0,77	0,92	н.з.
Вопрос 5	1,47	-2,96	-1,99	-0,84	0,84	н.з.
Вопрос 6	1,86	-2,46	-1,54	-0,81	0,20	н.з.
Вопрос 16	1,41	-4,52	-3,33	-2,11	-0,11	н.з.
Вопрос 20	1,93	-3,30	-2,65	-1,64	0,07	н.з.
Вопрос 21	2,48	-1,92	-1,09	0,08	1,33	н.з.
Вопрос 23	1,38	-3,84	-2,49	-1,39	0,62	н.з.
Вопрос 25	2,85	-1,89	-1,04	0,05	1,11	н.з.
Вопрос 28	2,71	-2,58	-1,92	-0,83	0,67	н.з.
Вопрос 29	3,12	-1,28	-0,63	0,25	1,06	н.з.
Вопрос 30	3,19	-1,74	-0,82	-0,08	1,00	н.з.
Вопрос 35	1,72	-2,92	-1,92	-0,73	0,96	н.з.
Шкала 2. Удовлетворенность коммуникативным взаимодействием и мотивация к учению в ЦОС						
Вопрос 2	2,14	-1,54	-0,53	0,85	1,94	н.з.
Вопрос 7	2,61	-1,61	-0,64	0,38	1,29	н.з.
Вопрос 11	1,56	-1,44	-0,35	0,81	1,83	0,011
Вопрос 18	2,56	-0,70	0,25	0,95	1,70	н.з.
Вопрос 22	1,56	-0,46	0,45	1,29	2,10	н.з.
Вопрос 29	2,40	-1,43	-0,68	0,31	1,15	н.з.
Вопрос 38	1,21	-1,62	-0,15	1,13	2,75	н.з.
Шкала 3. Стресснапряженность при обучении в ЦОС						
Вопрос 9	1,35	-2,25	-0,55	0,65	2,11	н.з.
Вопрос 17	2,20	-0,48	0,84	1,52	2,24	0,016
Вопрос 26	2,53	-1,02	0,06	0,95	1,81	н.з.
Вопрос 31	1,85	-1,18	0,23	1,30	2,24	н.з.
Вопрос 32	2,43	-0,77	0,15	1,01	1,86	н.з.
Вопрос 33	1,83	-1,58	-0,52	0,27	1,41	н.з.
Вопрос 34	1,94	-1,07	-0,01	0,83	1,77	0,019
Вопрос 36	1,68	-0,67	0,92	1,76	2,43	н.з.

Примечания: статистика критерия S- χ^2 является обобщенной статистикой хи-квадрат [7, 9], эмпирическое значение, р – уровень статистической значимости; α – параметр, отражающий дискриминативность утверждения; β – параметр, отражающий трудность утверждения

Соответствие пунктов реальным данным было оценено с помощью критерия $S-\chi^2$ [7, 9], который вычисляет сумму различий между наблюдаемыми и ожидаемыми частотами ответов для каждого утверждения (пункта) опросника. Статистически значимые различия указывают на потенциальное несоответствие. Тем не менее, известно, что хи-квадрат чувствителен к объему выборки, поэтому данный критерий не является абсолютно показательным для больших выборок. В нашем случае (табл. 2–3) для 32 утверждений из 38 получились приемлемые значения. Наибольшие отклонения обнаружены для вопроса 27 «Большинство (более половины) моих сокурсников справляются с контрольными заданиями самостоятельно» ($p=0,004$) и вопроса 11 «В электронном курсе на онлайн-занятиях с преподавателем я работаю гораздо интенсивнее, чем при традиционно-очном обучении» ($p=0,011$).

Таблица 3

**Статистика по каждому вопросу (пункту)
 субшкал 4–6 методики Шкала оценки ЦОС**

Вопрос	α	β_1	β_2	β_3	β_4	Уровень статистической значимости p критерия $S-\chi^2$
Шкала 4. Необходимость поддержки в учебной деятельности в ЦОС						
Вопрос 1	1,65	-0,55	1,04	2,50	3,13	н.з.
Вопрос 4	1,25	-1,83	-0,14	1,03	2,62	н.з.
Вопрос 8	1,29	-1,18	0,30	1,07	2,38	н.з.
Вопрос 12	2,61	-0,99	0,14	0,84	1,92	0,023
Вопрос 14	1,66	-0,30	0,99	1,63	2,83	н.з.
Вопрос 31	1,56	-1,29	0,24	1,41	2,45	н.з.
Шкала 5. Нечестные стратегии в ЦОС						
Вопрос 9	1,07	-2,60	-0,65	0,70	2,45	н.з.
Вопрос 13	2,77	-1,79	-0,23	0,83	1,99	н.з.
Вопрос 15	2,58	-1,35	0,01	1,22	2,10	н.з.
Вопрос 19	1,01	-1,69	0,34	1,82	4,10	н.з.
Вопрос 24	0,93	-5,37	-3,08	-0,96	1,27	н.з.
Вопрос 27	0,99	-1,22	1,26	3,34	5,17	0,004
Шкала 6. Доступность ЦОС						
Вопрос 10	1,94	-2,66	-2,11	-0,71	-	н.з.
Вопрос 14	1,40	-3,14	-1,84	-1,16	0,29	н.з.
Вопрос 20	2,00	-3,16	-2,57	-1,65	0,06	н.з.
Вопрос 36	2,54	-2,58	-1,90	-1,00	0,70	0,025
Вопрос 37	2,17	-3,34	-2,58	-1,75	-0,32	н.з.

Примечания: статистика критерия $S\text{-}\chi^2$ является обобщенной статистикой хи-квадрат, эмпирическое значение, p – уровень статистической значимости; α – параметр, отражающий дискриминативность утверждения; β – параметр, отражающий трудность утверждения.

Графики информационных функций для каждой шкалы представлены на рис. Для шкалы «Удовлетворенность учебным процессом в ЦОС» (шкала 1) точность оценки невысока для самых высоких значений, что говорит о том, что высокая удовлетворенность учебным процессом измеряется с помощью данного измерительного инструмента недостаточно точно. Средние значения удовлетворенности измеряются точно, а низкие – лучше, чем высокие. Следовательно, наиболее точные измерения и наибольшую дифференциацию респондентов можно получить среди тех, кто не очень удовлетворен учебным процессом в ЦОС.

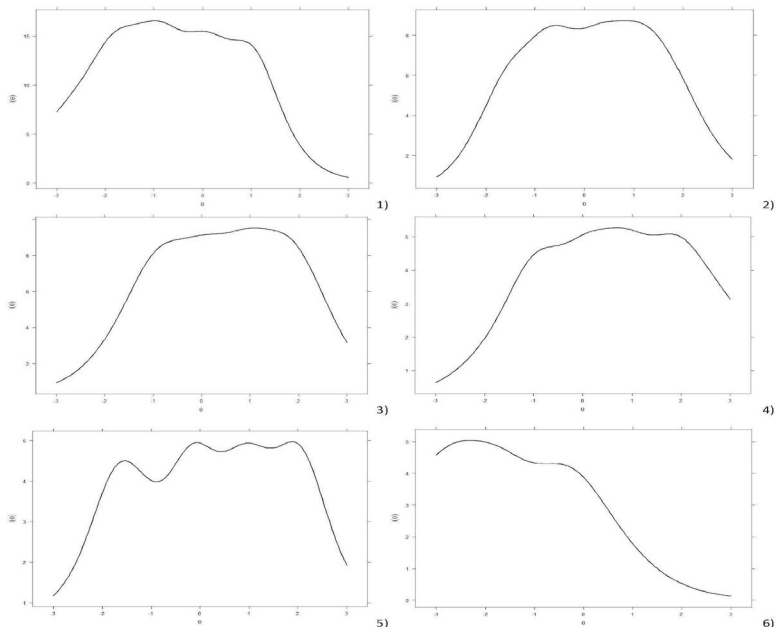


Рис. Графики информационных функций для шкал методики «Шкала оценки ЦОС»

Примечания: 1 – шкала 1 «Удовлетворенность учебным процессом в ЦОС»; 2 – шкала 2 «Удовлетворенность коммуникативным взаимодействием и мотивация к учению в ЦОС»; 3 – шкала 3 «Стресснапряженность при обучении в ЦОС»; 4 – шкала 4 «Необходимость поддержки в учебной деятельности в ЦОС»; 5 – шкала 5 «Нечестные стратегии в ЦОС»; 6 – шкала 6 «Доступность ЦОС».

Для шкалы «Удовлетворенность коммуникативным взаимодействием и мотивация к учению в ЦОС» (шкала 2) точность оценки высока для широкого диапазона средних значений: от -2 до 2. Только очень низкие и очень высокие значения удовлетворенность коммуникативным взаимодействием и мотивации к учению измеряются недостаточно точно. Для шкалы «Стресснапряженность при обучении в ЦОС» (шкала 3) ситуация практически такая же, как и для шкалы 2: точность оценки высока для достаточно широкого диапазона средних значений, хотя этот диапазон не так симметричен и охватывает значения примерно от -1,5 до 2,5. Для шкалы «Необходимость поддержки в учебной деятельности в ЦОС» (шкала 4) точность оценки невысока для самых низких значений, но очень высока для средних и высоких значений, что говорит о том, что высокая необходимость поддержки в учебной деятельности измеряется с помощью данного измерительного инструмента достаточно точно. Таким образом, средние значения необходимости поддержки измеряются точно, а низкие – гораздо хуже, чем высокие. Для шкалы «Нечестные стратегии в ЦОС» (шкала 5) точность оценки высока для широкого диапазона средних значений: от -2 до 2,5. Для шкалы «Доступность ЦОС» (шкала 6) точность оценки невысока для высоких значений, что говорит о том, что высокая доступность ЦОС измеряется с помощью данного измерительного инструмента недостаточно точно. Фактически, точно измеряются только значения ниже среднего, поэтому данный инструмент может быть использован для диагностики тех, кто низко оценивает доступность ЦОС, но не тех, кто удовлетворен этой доступностью. В общем, практически все шкалы методики дают неплохие оценки в диапазоне от -2 до 2. Шкалы 1 и 6 более точно оценивают низкие значения измеряемых показателей, а шкалы 3, 4 и 5 – более высокие.

Выводы

Рассмотрена методика «Шкала оценки ЦОС», предназначенная для комплексной оценки цифровой образовательной среды университета на основании ряда индикаторов: удовлетворенность и практическая польза; удовлетворенность коммуникативным взаимодействием; стресснапряженность; необходимость поддержки; нечестные стратегии; доступность. Для оценки методики в рамках современной теории тестов (IRT) применялась модель градуированных ответов. Результаты показали, что все вопросы методики имеют дискриминативность пункта не ниже умеренной, а большинство пунктов – высокую. Пороговые значения для каждого вопроса

методики, соответствующие относительному индикатору трудности, монотонно возрастают. Графики информационной функции для каждой шкалы методики показали, что практически все измерения компонентов Шкалы оценки ЦОС являются довольно точными в диапазоне средних значений, и только для шкалы 6 «Доступность ЦОС» точность оценки высока для самых низких значений и резко падает для средних и высоких значений. Таким образом, можно определить, в каком именно диапазоне измерения субшкалы методики «Шкала оценки ЦОС» отражают реальность точнее.

Литература

1. *Аванесов В.С.* Истоки и основные понятия математической теории педагогических измерений (Item Response Theory) // Педагогические измерения. 2007. № 3. С. 3–36.
2. *Крокер Л., Алгина Дж.* Введение в классическую и современную теорию тестов. М.: Логос, 2010. 668 с.
3. *Нейман Ю.М., Хлебников В.А.* Введение в теорию моделирования и параметризации педагогических тестов. М.: Прометей, 2000. 168 с.
4. *Сорокова М.Г., Одинцова М.А., Радчикова Н.П.* Шкала оценки цифровой образовательной среды (ЦОС) университета // Психологическая наука и образование. 2021. Том 26. № 2. С. 52–65. DOI: <https://doi.org/10.17759/pse.2021260205>
5. *Baker F.B.* The Basis of Item Response Theory, 2nd ed. ERIC Clearinghouse on Assessment and Evaluation, College Park, 2001. 186 p.
6. *Burnham K.P., Anderson D.R.* Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.). New York: Springer-Verlag, 2002. 488 p.
7. *Kang T., Chen T.T.* Performance of the generalized S-X2 item fit index for polytomous IRT models // Journal of Educational Measurement. 2008. № 45(4). P. 391–406.
8. *Muraki E.* A generalized partial credit model: Application of an EM algorithm. ETS Research Report Series, 1992. P. 159–176.
9. *Orlando M., Thissen D.* Likelihood-based item-fit indices for dichotomous item response theory models // Applied Psychological Measurement. 2000. № 24(1). P. 50–64.
10. *Samejima F.* Graded Response Model. In W.J. van der Linden, & R.K. Hambleton (Eds.). Handbook of Modern Item Response Theory. New York, NY: Springer, 1997. P. 85–100.
11. *Sorokova, Marina; Odintsova, Maria; Radchikova, Nataly (2021):* Scale for Assessing University Digital Educational Environment (AUDEE Scale): validation database. Psychological Research Data & Tools Repository. Dataset. <https://doi.org/10.25449/ruspsydata.14742816>

Информация об авторах

Радчикова Наталья Павловна, кандидат психологических наук, ведущий научный сотрудник Научно-практического центра по комплексному сопровождению психологических исследований PsyDATA, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация, ORCID: <http://orcid.org/0000-0002-5139-8288>, e-mail: nataly.radchikova@gmail.com

Сорокова Марина Геннадьевна, доктор педагогических наук, кандидат физико-математических наук, профессор кафедры Прикладной математики факультета Информационных технологий, руководитель Научно-практического центра по комплексному сопровождению психологических исследований PsyDATA, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация, ORCID: <http://orcid.org/0000-0002-1000-6487>, e-mail: sorokovamg@mgppu.ru

Одинцова Мария Антоновна, кандидат психологических наук, заведующий кафедрой психологии и педагогики дистанционного обучения, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0003-3106-4616>, e-mail: mari505@mail.ru

Гусарова Елена Сергеевна, магистр психологии, модератор Научно-практического центра по комплексному сопровождению психологических исследований PsyDATA, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0001-6968-5253>, e-mail: gusarovaes@mgppu.ru

MODELING AND DATA ANALYSIS FOR DIGITAL EDUCATION

The Application of Item Response Theory (IRT) for AUDEE Scale Psychometric Assessment

Nataly P. Radchikova

Moscow State University of Psychology & Education, Moscow, Russia
ORCID: 0000-0002-5139-8288
e-mail: nataly.radchikova@gmail.com

Marina G. Sorokova

Moscow State University of Psychology & Education, Moscow, Russia
ORCID: <http://orcid.org/0000-0002-1000-6487>
e-mail: sorokovamg@mgppu.ru

Maria A. Odintsova

Moscow State University of Psychology & Education, Moscow, Russia
ORCID: <https://orcid.org/0000-0003-3106-4616>
e-mail: mari505@mail.ru

Elena S. Gusarova

Moscow State University of Psychology & Education, Moscow, Russia
ORCID: <https://orcid.org/0000-0001-6968-5253>
e-mail: gusarovaes@mgppu.ru

Using item response theory (IRT), we examine AUDEE Scale that measures six different aspects of university digital educational environment (DEE): DEE Learning Process Satisfaction, DEE Communication satisfaction and Learning Motivation, DEE Stress Tension, Need for Support in DEE Learning Activity, DEE Dishonest Strategy Prevalence, and DEE Accessibility. Five-point Likert scale is used for all 38 questions of the AUDEE Scale. The study involved 406 MSUPE students who completed e-courses in the online format (90.1 % female). The average age of the participants was $28,7 \pm 9,6$ years (median = 24 years, minimum = 19 years, maximum = 72 years). The application of the graded response model (GRM) separately to each of the six subscales of the AUDEE Scale showed that all the items have at least moderate item discrimination, and most of the items have high item discrimination. The threshold values for each question of the AUDEE Scale, corresponding to the relative indicator of difficulty, increase monotonically. Test information functions for each subscale of the AUDEE Scale indicate that almost all measurements of the components of the AUDEE Scale are quite accurate in the range of mean values, and only for the scale 6 “DEE Accessibility” the estimation accuracy is high for the lowest values and drops sharply for medium and high values.

Keywords: Digital educational environment (DEE), AUDEE scale, Item Response Theory (IRT), item difficulty, item discrimination, test information function.

Funding. The reported study was funded by the Moscow State University of Psychology and Education (MSUPE) in the framework of the research project “Digital Technologies in Higher Education: Development of Technology for Individualizing Education Using E-Courses

For citation:

Radchikova N.P., Sorokova M.G., Odintsova M.A., Gusarova E.S. The Application of Item Response Theory (IRT) for AUDEE Scale Psychometric Assessment // Digital Humanities and Technology in Education (DHTE 2021): Collection of Articles of the II All-Russian Scientific and Practical Conference with International Participation. November 11–12, 2021 / V.V. Rubtsov, M.G. Sorokova, N.P. Radchikova (Eds). Moscow: Publishing house MSUPE, 2021. 557–570 p.

References

1. Avanesov V.S. Istoki i osnovnye ponyatiya matematicheskoi teorii pedagogicheskikh izmerenii (Item Response Theory) [The Origins and Basic Concepts of Item Response Theory] // *Pedagogicheskie Izmereniya = Pedagogical measurements*, 2007, no. 3, pp. 3–36. (In Russ.)
2. Kroker L., Algina D. Vvedenie v klassicheskuyu i sovremennuyu teoriyu testov [Introduction to classical and modern test theory]. Moscow: Logos, 2010. 668 p. (In Russ.)
3. Neiman Yu.M., Khlebnikov V.A. Vvedenie v teoriyu modelirovaniya i parametrizatsii pedagogicheskikh testov [Introduction to the Theory of Modeling and Parametrization of Educational Tests]. Moscow: Prometei, 2000. 168 p. (In Russ.)
4. Sorokova M.G., Odintsova M.A., Radchikova N.P. Shkala otsenki tsifrovoi obrazovatel'noi sredy (TsOS) universiteta [Scale for Assessing University Digital Educational Environment (AUDEE Scale)] // *Psikhologicheskaya nauka i obrazovanie = Psychological Science and Education*, 2021, Vol. 26, no. 2, pp. 52–65. DOI: <https://doi.org/10.17759/pse.2021260205> (In Russ., Abstr. in Engl.).
5. Baker F.B. The Basis of Item Response Theory, 2nd ed. ERIC Clearinghouse on Assessment and Evaluation, College Park, 2001. 186 p.
6. Burnham K.P., Anderson D.R. Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.). New York: Springer-Verlag, 2002. 488 p.
7. Kang T., Chen T.T. Performance of the generalized S-X2 item fit index for polytomous IRT models // *Journal of Educational Measurement*, 2008, no. 45 (4), pp. 391–406.
8. Muraki E. A generalized partial credit model: Application of an EM algorithm. ETS Research Report Series, 1992. pp. 159–176.

9. Orlando M., Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models // *Applied Psychological Measurement*, 2000, no. 24 (1), pp. 50–64.
10. Samejima F. Graded Response Model. In W.J. van der Linden, & R.K. Hambleton (Eds.). *Handbook of Modern Item Response Theory*. New York, NY: Springer, 1997. pp. 85–100.
11. Sorokova, Marina; Odintsova, Maria; Radchikova, Nataly (2021): Scale for Assessing University Digital Educational Environment (AUDEE Scale): validation database. Psychological Research Data & Tools Repository. Dataset. <https://doi.org/10.25449/ruspsydata.14742816>

Information about the authors

Nataly P. Radchikova, PhD in Psychology, Leading Researcher of Scientific and Practical Center for Comprehensive Support of Psychological Research «PsyDATA», Moscow State University of Psychology & Education, Moscow, Russia, ORCID: <http://orcid.org/0000-0002-5139-8288>, e-mail: nataly.radchikova@gmail.com

Marina G. Sorokova, Doctor in Education, PhD in Physics and Mathematics, Professor, Chair of Applied Mathematics, Faculty of Information Technology, Head of Scientific and Practical Center for Comprehensive Support of Psychological Research «PsyDATA», Moscow State University of Psychology & Education, Moscow, Russia, ORCID: <http://orcid.org/0000-0002-1000-6487>, e-mail: sorokovamg@mgppu.ru

Maria A. Odintsova, PhD in Psychology, Professor, Chair of Psychology and Pedagogy of Distance Learning, Head of the Chair of Psychology and Pedagogy of Distance Learning, Faculty of Distance Learning, Moscow State University of Psychology & Education, Moscow, Russia, ORCID: <https://orcid.org/0000-0003-3106-4616>, e-mail: mari505@mail.ru

Elena S. Gusarova, Master in Psychology, Moderator of Scientific and Practical Center for Comprehensive Support of Psychological Research «PsyDATA», Moscow State University of Psychology & Education, Moscow, Russia, ORCID: <https://orcid.org/0000-0001-6968-5253>, e-mail: gusarovaes@mgppu.ru