

## Разработка и перспективы развития национальной интеллектуальной системы тестирования общеязыковых компетенций (ИСТОК) на базе нейросетевых решений

Научная статья

DOI: 10.31992/0869-3617-2023-32-8-9-147-166

**Базанова Елена Михайловна** – канд. пед. наук, доцент, руководитель Департамента иностранных языков, директор Центра языковой подготовки и тестирования МФТИ, ORCID: 0000-0002-6306-8892, [bazanova.em@mipt.ru](mailto:bazanova.em@mipt.ru)

**Горизонтова Анна Всеволодовна** – канд. ист. наук, старший преподаватель Департамента иностранных языков, [gorizontova.av@mipt.ru](mailto:gorizontova.av@mipt.ru)

**Грибова Наталья Николаевна** – канд. филол. наук, заместитель руководителя по административной работе Департамента иностранных языков, ORCID: 0000-0001-5444-3091, [gribova.nn@mipt.ru](mailto:gribova.nn@mipt.ru)

**Чикаке Тендай Мапунгвана** – исследователь, преподаватель-исследователь Физтех-школы прикладной математики и информатики, ORCID: 0000-0002-9512-1256, [tendaichikake@phystech.edu](mailto:tendaichikake@phystech.edu)

**Самосюк Алексей Владимирович** – инженер Лаборатории гибридных интеллектуальных систем, ORCID: 0000-0002-3030-481X, [alexeysamosyuk@phystech.edu](mailto:alexeysamosyuk@phystech.edu)

Московский физико-технический институт (Национальный исследовательский университет) (МФТИ), г. Долгопрудный, Московская область, Россия

Адрес: 141701, Россия, Московская обл., г. Долгопрудный, Институтский переулок, д. 9

*Аннотация.* В статье представлены результаты апробации интеллектуальной системы тестирования общеязыковых компетенций (ИСТОК), разработанной учёными-тестологами, лингвистами, специалистами по методике преподавания иностранных языков и искусственному интеллекту. Данная система включает в себя как тесты для подтверждения заявленного уровня владения английским языком как иностранным, так и адаптивный тест на определение этого уровня. Все тестовые материалы откалиброваны в соответствии с Общеввропейской шкалой языковых компетенций. ИСТОК представляет собой адаптивную систему тестирования, развёрнутую на базе нейросетевых решений, охватывает проверку рецептивных и продуктивных речевых навыков (чтение, аудирование, говорение и письмо) искусственным интеллектом, а именно – нейролингвистическими моделями. В процессе подготовки к тестированию были созданы базы данных, обеспечивающие тестирование проверочными материалами, разработаны критерии оценивания, алгоритмы выявления ошибок различного типа искусственным интеллектом, валидатор для определения уровня владения языком. Результаты протестированных испытуемых (более 5 000)

*показали высокую надёжность и объективированную валидность тестового контроля. Новый подход к проверке языковых компетенций может быть использован как в процессе проведения оценки в высших учебных заведениях, так и для определения или подтверждения уровня владения языком персонала, занятого в различных сферах профессиональной деятельности, а принципы обучения и практического использования нейролингвистических моделей найдут широкое применение в различных областях прикладных исследований.*

*Ключевые слова:* языковое тестирование, адаптивное тестирование, машинное обучение, нейролингвистическая модель, лингвистический корпус, базы данных, нейросетевые решения, интеллектуальная система тестирования общезыковых компетенций

*Для цитирования:* Базанова Е.М., Горизонтова А.В., Грибова Н.Н., Чикаке Т.М., Самосюк А.В. Разработка и перспективы развития национальной интеллектуальной системы тестирования общезыковых компетенций (ИСТОК) на базе нейросетевых решений // Высшее образование в России. 2023. Т. 32. № 8-9. С. 147–166. DOI: 10.31992/0869-3617-2023-32-8-9-147-166

## Development and Prospects of National Intelligent System for Testing General Language Competencies Deployed Through Neural Network Solutions

Original article

DOI: 10.31992/0869-3617-2023-32-8-9-147-166

*Elena M. Bazanova* – Cand. Sci. (Pedagogy), Associate Professor, Head of the Department of Foreign Languages, ORCID: 0000-0002-6306-8892, bazanova.em@mipt.ru

*Anna V. Gorizontova* – Cand. Sci. (History), Senior Lecturer of the Department of Foreign Languages, gorizontova.av@mipt.ru

*Natalia N. Gribova* – Cand. Sci. (Philology), Deputy Head of the Department of Foreign Languages, ORCID: 0000-0001-5444-3091, gribova.nn@mipt.ru

*Tendai M. Chikake* – Researcher and Senior Lecturer at Phystech School of Applied Mathematics and Informatics, ORCID: 0000-0002-9512-1256, tendaichikake@phystech.edu

*Alexey V. Samosyuk* – Engineer of the Laboratory of Hybrid Intelligent Systems, ORCID: 0000-0002-3030-481X, alexeysamosyuk@phystech.edu

Moscow Institute of Physics and Technology (National Research University), Moscow region, Dolgoprudny, Russia

Address: 9 Institutsky lane, Moscow region, Dolgoprudny, 141701, Russia

**Abstract.** The article presents the results of approbation of the intellectual system for testing general language competences (ISTOK) developed by testologists, linguists, specialists in methodology of teaching foreign languages and in artificial intelligence. This system includes a range of tests assessing language ability at levels from A2 to C1 of the Common European Framework of Reference (CEFR), as well as an adaptive placement test. All test materials are calibrated according to the CEFR. ISTOK is an adaptive testing system deployed through neural network solutions and providing assessment of receptive and productive language skills (reading, listening, speaking and writing) by using artificial intelligence and/or neurolinguistic models. The process of ISTOK devel-

opment implied, apart from writing test items, putting together databases of writing and speaking assignments marked by professional assessors and assessment criteria for productive skills, as well as algorithms to identify various types of mistakes with the help of artificial intelligence. The results of various testing cohorts with the total number of test takers exceeding 5,000 demonstrated high reliability and objectified test validity. The new approach to language skills testing can be used for various purposes in higher education institutions, as well and to identify and/or confirm language proficiency of personnel in different organisations and businesses, while the principles of training and practical use of neurolinguistic models will find wide application in various fields of applied research.

**Keywords:** language testing, language assessment, language skills testing, adaptive testing, machine learning, neuro-linguistic model, linguistic corpus, databases, neural network solutions, intellectual system of testing common language competencies

**Cite as:** Bazanova, E.M., Gorizontova, A.V., Gribova, N.N., Chikake, T.M., Samosyuk, A.V. (2023). Development and Prospects of National Intelligent System for Testing General Language Competencies Deployed Through Neural Network Solutions. *Vysshee obrazovanie v Rossii = Higher Education in Russia*. Vol. 32, no. 8-9, pp. 147-166, doi: 10.31992/0869-3617-2023-32-8-9-147-166 (In Russ., abstract in Eng.).

#### Отечественное и международное независимое языковое тестирование

Текущая ситуация в мире открывает новые перспективы для разработки и внедрения национальной системы оценки языковых знаний и умений, которая может стать стандартной для российских университетов и работодателей, предъявляющих требования к знанию иностранного языка при отборе кандидатов на вакантные должности, а также в процессе обучения сотрудников.

Существующие на данный момент международные системы тестирования знания английского языка следуют традиционным лингводидактическим методикам и ориентированы на Европейскую шкалу языковой компетенции, которая распределяет пользователей иностранных языков по шести уровням – от A1 до C2 (*Common European Framework of Reference: Learning, Teaching, Assessment (CEFR)*) [1]. Все тесты, используемые в международных системах языкового оценивания, подразделяются на две большие категории с точки зрения их предназначения: первая из них призвана подтвердить уровень владения языком, заявленный кандидатом, вторая – определить этот уровень.

Очевидно, что в первую группу входят экзамены и тесты, проверяющие владение

иностранном языке в пределах одного уровня по Европейской шкале языковой компетенции. К ней относятся в первую очередь такие кембриджские экзамены как *Key (A2)*, *Preliminary (B1)*, *First (B2)*, *Advanced (C1)* и *Proficiency (C2)*, а также некоторые другие. Кандидат, закончивший, например, курс английского языка уровня B2, выбирает соответствующий экзамен и, в случае успешной его сдачи, подтверждает, что владеет языком на данном уровне.

Вторую группу можно условно назвать тестами или экзаменами на определение уровня. К ней относятся такие экзамены, как, например, *IELTS*, *TOEFL* и кембриджский тест *LinguaSkill*. Как следует из названия этой группы, такие экзамены не ограничены рамками одного уровня Европейской шкалы и могут сдаваться кандидатами с разной степенью развития языковых навыков и умений.

В условиях неязыкового вуза, как представляется, необходимы оба вида тестов: входной, для правильного распределения по группам, и итоговый, проверяющий результаты определённого этапа изучения языка в конце каждого учебного года. Разный характер задач, выполняемых такими тестами, диктует необходимость использо-

вания двух отдельных типов тестов. В первом случае, когда мы имеем дело с новой для нас когортой обучающихся, очевидно, имевших разный опыт изучения английского языка, необходим тест для определения уровня, который работал бы если не на всех, то на наиболее релевантных для данной когорты уровнях Общеввропейской шкалы. В свою очередь, для подведения итогов изучения английского языка на определённом уровне Общеввропейской шкалы в конце каждого курса необходим тест на подтверждение заявленного уровня.

После приостановки работы в России всех компаний-разработчиков международных экзаменов по английскому языку появилась уникальная возможность создать и внедрить национальную систему тестирования, основанную на мировых и отечественных научных достижениях в области тестологии, педагогической теории измерений, прикладной лингвистики и искусственного интеллекта. Симбиоз этих наук позволяет совершить прорыв в решении сложных проблем, которые до сих пор, по большей части, оставались неразрешёнными в отечественной практике: в первую очередь, стандартизации процедуры определения и оценивания уровня владения иностранным языком с применением адаптивного тестирования и системы автоматизированной оценки как рецептивных, так и продуктивных навыков на базе нейросетевых решений.

Вопрос о стандартизации оценивания уровня владения иностранными языками не является новым для российской системы образования. Первые попытки были предприняты ещё в 1997 г., когда на базе МПГУ им. В.И. Ленина был открыт Центр тестирования, которым руководил профессор Владимир Хлебников. Следующим важным шагом явился переход к сдаче Единого государственного экзамена (ЕГЭ) в 2009 г. Иностранный язык, не являясь обязательным для выпускников школ, учитывается большинством вузов и факультетов гуманитарной направленности и, соответственно, является надёжным ориен-

тиром при выстраивании программ дальнейшего совершенствования навыков владения языком. Другие вузы, в частности, технические, сталкиваются с весьма разноуровневым контингентом первокурсников и вынуждены каким-то образом решать проблемы, связанные не только и не столько с распределением по группам, сколько с выстраиванием программ обучения и обеспечением прогресса в уровне владения иностранным языком для всех студентов.

Другая проблема – это определение или подтверждение уровня владения иностранным языком при окончании программ бакалавриата, поступлении в магистратуру и аспирантуру. Как правило, она решалась либо путём использования материалов или собственно экзаменов международных систем тестирования, либо самостоятельной разработкой оценочных средств, определяющих уровень сформированности языковых компетенций.

Не подвергая сомнению профессионализм вузовских специалистов, составляющих тестовые задания для бакалавров, магистрантов и аспирантов в каждом отдельном учебном заведении, нельзя не признать, что такие материалы разрабатываются вне рамок какого-либо единого стандарта, что приводит к расхождениям в интерпретации результатов тестирования, а также к фактической несопоставимости этих результатов. Для того чтобы тест по иностранному языку был качественным измерителем, он должен обладать рядом психометрических характеристик, которые включают валидность, надёжность, экономичность, аутентичность, взаимосвязь со всеми аспектами обучения и практическую применимость (*validity, reliability, impact, practicality*) [2]. Разработка и внедрение отечественной системы тестирования позволят решить вопрос о стандартизации оценки уровня владения английским языком, унифицировать требования, предъявляемые ко всем участникам языкового тестирования, а также сократить расходы на сдачу экзаменов и время ожида-

ния результатов. Кроме того, использование современных технологических решений будет способствовать перестройке методических подходов к решению задачи подготовки к единому языковому тестированию [3]. Необходимость в разработке отечественной системы языкового тестирования для вузов также, в определённой мере, обусловлена ситуацией, сложившейся после одностороннего отказа провайдеров международных экзаменов продолжить работу с российскими потребителями.

В 2022–2023 гг. Московский физико-технический институт (Национальный исследовательский университет) (МФТИ) приступил к разработке и внедрению независимого тестирования по английскому языку. Многолетний и успешный опыт преподавания иностранных языков, большая практика в проведении мероприятий по языковому оцениванию, а также уникальная научная база университета в области информационных технологий и искусственного интеллекта позволили в достаточно короткие сроки создать и внедрить в учебный процесс интеллектуальную систему тестирования общезыковых компетенций (ИСТОК). При создании системы разработчики ориентировались как на потребности высших учебных заведений, так и на передовые практики, характеризующие современный уровень развития языкового тестирования. В частности, поскольку проведение независимой оценки с помощью системы ИСТОК является частью учебного процесса, очень важное значение имеют сроки предоставления результатов. Технические решения, которые позволяют максимально сократить эти сроки, – это проведение тестирования в компьютерном варианте, а также автоматизированная проверка заданий [4]. В то время как компьютерный формат проведения тестирования обеспечивает автоматическую проверку заданий на чтение и аудирование, оценивание продуктивных навыков (письма и говорения) требует более сложных решений с использованием искусственного интеллекта. В МФТИ в настоящее

время в полном объёме работают два варианта компьютерного тестирования рецептивных навыков и ведётся финальная стадия разработки системы автоматизированного оценивания продуктивных навыков.

### Опыт разработки и проведения уровневого тестирования в МФТИ

Первым опытом в области тестирования на подтверждение заявленного уровня явилась разработка и проведение теста на уровне *B2* весной 2022 г. Выбор именно данного уровня был предопределён тем обстоятельством, что он является целевым для студентов 2-го курса, после которого студенты могут выбирать один из предложенных треков, для успешного прохождения которых необходимо владеть английским языком на указанном уровне. В то же время тестирование не было обязательным, в связи с чем коргорта тестируемых (в общей сложности более 400 человек) включала студентов разных курсов МФТИ. Тестирование было предложено апробировать также ряду неязыковых вузов, в число которых входили Тюменский государственный университет, Поволжский и Тамбовский технические университеты, а также Всероссийская академия внешней торговли. Как и студенты МФТИ, участники из этих вузов проходили тестирование в компьютерном формате. Все участники этого пилотного проекта положительно оценили опыт участия в нём. Весной 2023 г. линейка уровней инструментов оценки была расширена за счёт тестов уровнями *A2*, *B1* и *C1*. Таким образом, появилась возможность тестировать всех бакалавров, магистрантов и аспирантов МФТИ.

Тест *B2* 2022 г. включал задания по всем четырём речевым навыкам. Оценивание говорения и письма осуществлялось силами специалистов, имеющих опыт проверки данных навыков на международных экзаменах. При этом не предполагалось применять именно эту форму проверки при проведении дальнейших мероприятий по оценке уровня владения английским языком: было очевид-



но, что такой вариант является затратным, как финансово, так и с точки зрения времени, требуемого на оценивание письма и говорения большого числа кандидатов. Тем не менее было необходимо, во-первых, получить независимую оценку всех речевых навыков – как для внесения коррективов в процесс преподавания, так и с целью сбора материалов для дальнейшего использования в процессе перехода к автоматизированной оценке продуктивных навыков. В уровневые тесты 2023 г. не входили задания на говорение, поскольку задача разработки системы автоматизированного оценивания этого навыка с использованием нейросетей будет решаться по завершении проекта по разработке такой системы для оценки письма. Образцы выполнения заданий по говорению и оценки экзаменаторов, необходимые для обучения искусственного интеллекта, были в достаточном объёме получены в 2022 г., и в 2023 г. не стояла задача расширения этого массива данных.

Отдельно следует остановиться на подходе к оцениванию языковых умений и разбалловке тестовых заданий. Оценивание объективных языковых умений (чтения и аудирования), как уже упоминалось выше, не представляет особой сложности: при компьютерном формате тестирования оно проводится автоматически, и каждый правильный ответ засчитывается как один балл. В то же время оценка письма и говорения традиционно ставит перед любым разработчиком теста вопрос об определении критериев. Если проанализировать подходы, используемые для оценки письма и говорения различными международными экзаменационными комиссиями, то станет очевидным, что все эти организации фактически применяют сходные критерии оценивания. В число критериев непременно входят выполнение коммуникативной задачи, организация высказывания (устного или письменного), грамматика и лексика. При оценивании устной речи, естественно, учитываются также произношение и

интонация. Поскольку проверка письма и говорения в нашем случае проводилась силами экзаменаторов, аккредитованных *Cambridge Assessment English*, были использованы привычные этим специалистам кембриджские шкалы. Оценивание говорения и письма осуществляется по четырём основным критериям и по пятибалльной шкале. Для письма это соответствие содержания заданной теме, выполнение коммуникативной задачи, язык (то есть, лексика и грамматика) и организация текста [5]. Для говорения – управление дискурсом, лексика и грамматика, коммуникативное взаимодействие и произношение [6]. Поскольку формат устной части теста ИСТОК, в отличие от Кембриджских экзаменов, не включает диалог между кандидатами, критерий «коммуникативное взаимодействие» не учитывался при выставлении оценок.

Следующий вопрос, непосредственно относящийся к оцениванию, – это методика пересчёта т. н. «черновой» оценки (*raw score*) в окончательную. «Черновая» оценка – это количество баллов за каждый навык. Различия между традиционно применяемыми подходами обычно состоят лишь в максимальном окончательном балле. Для некоторых уровневых экзаменов это 100, для других – 200. Вне зависимости от того, на каком уровне установлен окончательный балл, «вес» отдельных навыков всегда уравнивается: если, например, общий максимальный балл за тест составляет 200, то максимальная оценка за каждый навык будет 50 баллов. Для перевода «чернового» балла в окончательный для каждого навыка высчитывается коэффициент. Например, если раздел «Аудирование» включает 30 вопросов, то коэффициент, на который нужно умножить количество правильных ответов кандидата, составляет 0,6. Кандидат, ответивший правильно, например, на 20 вопросов, таким образом, получит 12 в качестве окончательного балла за аудирование. В тестах ИСТОК общий балл рассчитывается в диапазоне от 0 до 100 баллов.

Гораздо более сложный вопрос, возникший в связи с выставлением результатов, – это определение границ уровней Общеввропейской шкалы. Этот вопрос так или иначе приходится решать разработчикам любого языкового теста. Очевидно, что, когда речь идёт о языке, о навыках его применения, практически невозможно провести однозначно чёткую границу между соседними уровнями. При использовании аналитических шкал оценивания продуктивных навыков оценки кандидатов по отдельным критериям могут существенно отличаться в рамках одного и того же навыка [7]. Например, письменное задание может быть выполнено на высоком уровне с точки зрения лексики и грамматики, но не набрать максимального балла за критерий «организация текста».

«Пограничье» между соседними уровнями всегда будет существовать, и в международной практике встречается несколько подходов к трактовке результатов, находящихся в этой зоне. Например, в экзаменах британско-греческой комиссии *LanguageCert* минимальным проходным баллом считается 50%, а за результат выше 80% даётся оценка *High Pass*. При этом в нижние границы тестируемого уровня входит некоторый процент высоких результатов предыдущего уровня, а в зону *High Pass* частично попадают результаты более высокого уровня Общеввропейской шкалы. В кембриджских экзаменах используется более сложная схема так называемой «расширенной» сертификации, когда кандидат, чьи результаты низки для получения сертификата желаемого уровня, получает сертификат предыдущего. Аналогичным образом можно получить сертификат более высокого уровня. Например, кандидаты, которые сдавали один и тот же экзамен уровня *B1*, могут получить сертификат этого уровня, а также сертификаты двух «соседних» – *A2* и *B2*. Разработчики тестов ИСТОК следовали этой модели. Было принято решение считать результаты ниже чем 50 баллов из 100 неудовлетворительными, баллы в диапазоне 50–59 засчитывались

как соответствующие уровню на один ниже заявленного, а результаты выше 90 баллов – следующему, более высокому уровню. Такая трактовка результатов не противоречит практикам, используемым профессиональными экзаменационными комиссиями.

Одной из целей проведения ежегодного уровня тестирования является сопоставление прошлых и последующих результатов студентов. Оно не просто даёт возможность отследить наличие прогресса в уровне владения английским языком, но и предоставляет ценную информацию для анализа эффективности обучения с точки зрения соответствия используемых учебных материалов целям обучения. Результаты тестов также в значительной мере позволяют оценить эффективность работы преподавателей, качество используемых учебных материалов и их соответствие задачам обучения или оценить так называемый «обратный эффект» (*washback effect*) тестирования на процессы преподавания и овладения языком [8].

На данный момент мы имеем возможность сопоставить результаты теста *B2*, который сдавался в 2022 и 2023 гг. Среди кандидатов, сдававших тест *B2* в конце 2021/22 учебного года, было 217 студентов 1-го курса, которые через год, весной 2023 г., прошли тестирование на том же уровне по окончании обучения на 2-м курсе. Было проведено сравнение результатов 2022 и 2023 гг. этой группы студентов, которое показало существенный прогресс в рамках одного уровня, либо переход на следующий уровень по Общеввропейской шкале. Также существенно сократилась группа тестируемых, выполнивших тест с результатом ниже 50%, т. е. ниже уровня *B1+*. Учитывая, что данная группа составляет 23% от общего числа обучавшихся на 2-м курсе, данные, отражённые в *рисунке 1*, представляются вполне репрезентативными и свидетельствуют о наличии у студентов, закончивших 2-й курс в 2023 г., существенного прогресса в изучении английского языка.

После тестирования всех групп изучающих английский язык в МФТИ на подтверж-

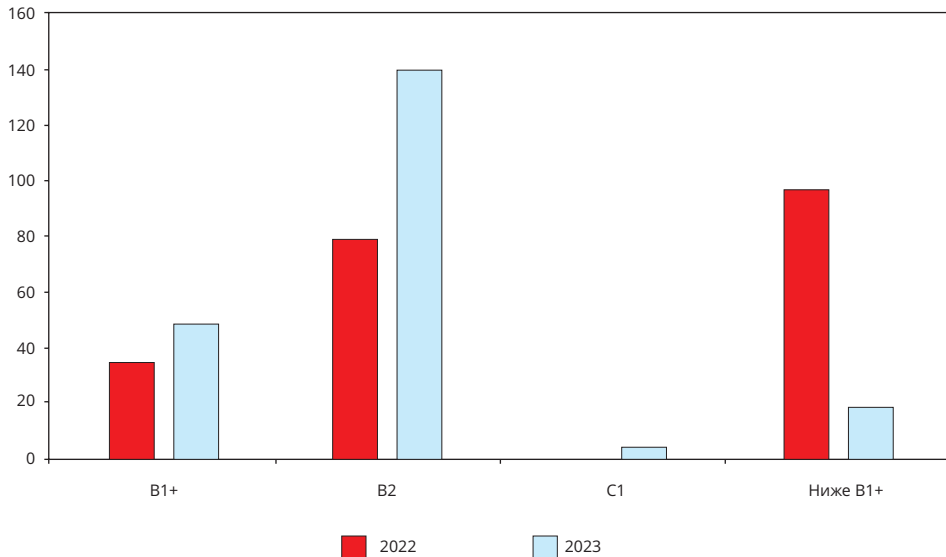


Рис. 1. Сравнение результатов студентов, сдававших тест уровня B2 в 2022 и 2023 гг.  
 Fig. 1. Results comparison of students who took the B2 level test in 2022 and 2023

дение целевого (заявленного) уровня, которое будет проводиться в конце 2023/2024 учебного года, появится возможность сопоставить полученные результаты с итогами тестирования 2022/2023 учебного года и сделать выводы о прогрессе каждой из когорт тестируемых, обучавшихся на различных программах и уровнях, а также каждого отдельного обучающегося.

#### Адаптивное компьютерное тестирование рецептивных и продуктивных навыков

Следующим тестовым продуктом, разработанным в МФТИ, явился адаптивный компьютерный тест, который проводился в начале учебного года с целью обоснованного и надёжного распределения студентов-первокурсников по группам.

Перед исследователями была поставлена цель разработать автоматизированную систему адаптивного тестирования для определения уровня владения английским языком применительно к двум речевым навыкам (чтение и письмо), для решения нескольких взаимосвязанных задач. С точки зрения содержания необходимо было разработать

корпус разноуровневых тестовых единиц (от A2 до B2 по Общеввропейской шкале языковой компетенции), составляющих банк калиброванных заданий. Техническая задача состояла в необходимости создать действующую модель на основе многошаговой варьирующе-ветвящейся стратегии адаптивного тестирования (отбор заданий непосредственно из банка калиброванных заданий по определённым алгоритмам, которые прогнозируют оптимальную трудность последующего задания по результатам выполнения испытуемым предыдущего задания адаптивного теста) [9]. Данная стратегия предполагает пошаговую переоценку уровня подготовленности кандидата, предпринимаемую после каждого выполнения очередного задания теста. Если тестируемый выполняет задание верно, то затем ему даётся более трудное задание. При неправильном выполнении задания совершается отход назад к более лёгким заданиям банка. Таким образом, в компьютерном адаптивном предъявлении число тестовых заданий и их трудность индивидуализированы с учётом ответов экзаменуемого, а инди-



видуальная совокупность заданий образует адаптивный тест. Применение адаптивного тестирования, таким образом, вносит в стандартный тест элементы индивидуализации и позволяет существенно сократить время для определения уровня [10], поскольку при прохождении теста на определение уровня в неадаптивном формате каждому кандидату приходится выполнять последовательно задания из каждого уровня Общеευропейской шкалы: от *A1* до того уровня, которому соответствуют знания данного кандидата.

В общей сложности в сентябре 2022 г. тестирование прошли более 2500 студентов первых курсов бакалавриата, магистратуры, а также аспиранты первого года обучения. Проведение адаптивного тестирования позволило корректно распределить студентов и аспирантов по группам в зависимости от установленного в результате прохождения теста уровня владения английским языком. Недостатком тестов на определение уровня является сложность соотнесения их результатов с данными, полученными по итогам тестирования на подтверждение заявленного уровня: последние гораздо глубже исследуют различные аспекты развития языковых умений каждого кандидата. Адаптивный компьютерный тест (как и любой тест на определение уровня вне зависимости от его формата), не способен в должной степени учесть «многомерность и функциональность» языка, как было отмечено ещё на ранних этапах развития этой формы тестирования [11]. В этой связи для регулярного мониторинга достижений студентов в изучении английского языка в МФТИ будут использоваться данные уровневых тестов, сдаваемых в конце каждого академического года.

#### **Особенности отечественной интеллектуальной системы тестирования общеязыковых компетенций**

В основу системы тестирования положены технологии искусственного интеллекта и междисциплинарные исследования, про-

ведённые группой учёных МФТИ: методологов, экспертов международного уровня, лингвистов, специалистов по машинному обучению и искусственному интеллекту. Развёртывание технологий по тестированию в части оценки уровня владения языком происходит на базе обучения нейросети и охватывает полный цикл, включающий оценку всех видов речевых навыков: чтение, аудирование, письмо, говорение. Это первая в России интеллектуальная система тестирования, основанная на новейших разработках и инструментах, с помощью которых создаются модели машинного обучения с возможностью переобучения и без потери, связанной с эффектом размерности, который обычно приводит к недообучению или переобучению. В сравнении с адаптивным тестированием Национальная интеллектуальная система тестирования общеязыковых компетенций отличается рядом особенностей.

1. Система развёрнута на базе нейросетевых решений, где индивидуальная траектория определяется набором признаков, основанных на лингвистических и экстралингвистических факторах.

2. Валидатор теста настроен на отслеживание речевых нарушений (ошибок), характерных для определённого уровня владения языком. В рамках проекта был разработан классификатор ошибок (перечень типов ошибок и смысловых блоков, которые могут использоваться в экспертных и алгоритмических разметках текстов эссе), учитываемых при работе валидатора.

3. Пробные тесты были проверены и детально проанализированы экспертами международного уровня с учётом нескольких перекрытий, на этой выборке проходило обучение нейросетевой модели (с результативностью около 90%).

4. В рамках проекта были разработаны базы данных: банк заданий (лексико-грамматических, текстовых в аудио-, видео- и письменном формате), база данных эссе (размеченных с учётом ошибок и неразмеченных), база данных устных монологиче-

ских и диалогических высказываний в аудиоформате.

5. Это первая система, способная охватить и проанализировать все речевые навыки для корректного определения уровня владения иностранным языком по принятой международной шкале.

В перспективе нейросеть на базе разработанного адаптивного теста сможет подбирать релевантные задания из банка и проводить исследования набора навыков, основываясь на времени выполнения задания экзаменуемым (при условии избыточности времени выполнения теста); ошибках, характерных для определённого уровня владения языком; скорости набора англоязычного текста на клавиатуре; задержки внимания на определённом фрагменте (фразе, словоформе) текста, которое имеет невербальное выражение (результат наблюдения за мимикой и движением глаз при системе проторинга) [12]. Сенситивность испытуемого будет анализироваться нейросетью также с целью определения степени уверенности на определённом уровне владения языком. В основу отбора релевантных признаков нейросетью можно положить любой из них, выявляющий лингвистические и психологические особенности экзаменуемого [13]. Эти и другие характеристики могут быть зафиксированы как дополнительные при генерации результата тестирования.

### Тестирование и программирование системы оценивания

Уровень владения языком определяется нейролингвистической моделью (НЛМ), состоящей из трёх основных модулей, работающих совместно, разработанной на основе отечественной технологии *DeepPavlov* (эта система представляет собой набор сервисов обернутых в *docker*-контейнеры под оркестрацией *Docker Compose*).

Первый модуль – объединение независимых узкоспециализированных нейросетевых моделей, которые выявляют ошибки орфографии и синтаксиса, распознают

устойчивые языковые конструкции и правильность их использования, оценивают структуру текста, связность, цельность, завершенность, модальность и наличие момента оценки в высказываниях. Основная задача данного модуля – оценка механических навыков владения языком и знание правил дистрибуции языковых единиц в потоке речи.

Второй модуль является большой языковой моделью (*LLM*) и используется для комплексной оценки текста за пределами заранее заданных правил. Он способен оценить сложность используемого словарного запаса, предсказать пропущенное слово в потоке построенной и организованной речи, оценить сложность и избыточность используемых языковых конструкций, а также качество передачи основного смысла. Основная задача данного модуля – проверка уверенности владения языком за пределами механической оценки, а также качества его использования для раскрытия темы эссе или высказывания.

Третий модуль объединяет результаты работы предыдущих двух модулей, выдавая финальную оценку как в виде статистики по отдельным категориям ошибок, так и в виде предсказания общего уровня, в том числе приведённого в соответствие со шкалой наиболее объективированной системой оценки.

Текст как результат речетворческого процесса обладает национально-культурной спецификой, его созданию предшествуют фоновые знания, осознание фактов, их осмысление (предтекстовые пресуппозиции). Система текстов, принципы текстовой организации, разные планы содержания текста, метатекстовая информация, логическая организация и другие характеристики этого сложного коммуникативного единства не могут быть описаны и заданы по всем объективным параметрам в виде алгоритмов для нейросетевой модели, но большой текстовый массив (свыше двух миллиардов словоупотреблений) представлен в виде национального языкового корпуса, оснащённого лингвистической разметкой и инструментами

<Hard working> <of course> <> one of the main <condition> for success. When you ask popular <peoply> about the recipe of success, they say that some luck also plays not <last role>. But if you think that you can achieve something only with help of <lucky>, you also will <be fail>, because success is the mix of <you> skills, <working>, ideas and chances. We know many examples of scientists and <filosophies> that <was> born earlier, that <his> ideas find popularity in minds of <another> <peoples>, I think that it <> the same thing <like> <lucky>. Everybody <have> <> friend, that was working on something, but had failed because <they> <have not> any talents for this <bussines>, like me and <> <english> language. It <> so hard to write <> text without any help of google translate ...

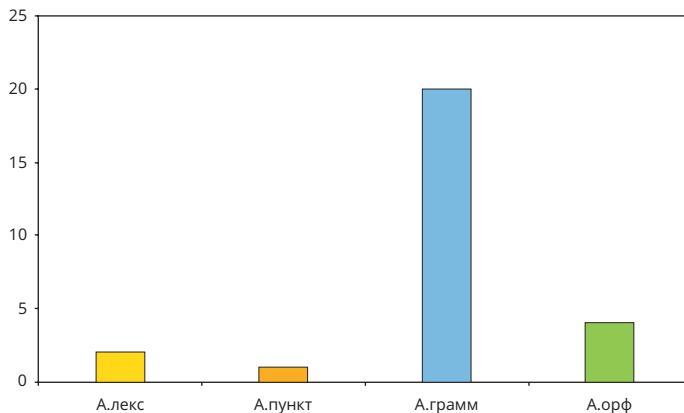


Рис. 2. Пример разметки текста и статистика ошибок (A2)

Fig. 2. Example of text markup and error statistics (A2)

поиска (программы анализа корпусов), что значительно облегчает задачу машинного обучения. Нейросетевая модель просмотрела все лингвистические корпуса английского языка, словари, изданные на их основе и способна отслеживать частотность употребления слов; изменение лексического состава; сочетаемость определённых грамматических явлений друг с другом на основе актуальных данных, полученных из живой речи; определять функциональный стиль посредством статистических характеристик текста: длину слова и предложения, характерных сочетаний слов, фиксированных фраз, идиом и т. п.; разграничивать письменный и устный дискурс. Это означает возможность использования моделью не только отдельных словоформ, но и предложений, понимание

структуры развёртывания текста, линейной последовательности, валентности, механизмов парадигматики и синтагматики [14].

Сервис, разработанный на основе искусственного интеллекта и направленный на автоматическую оценку уровня владения английским языком, обладает рядом преимущественных характеристик:

- микросервисная архитектура с базами данных;
- платформа микросервисов / оркестрация – *Docker / Docker Compose*;
- формат входных/выходных файлов с текстами на английском языке *JSON*;
- передача файлов через *Command Line Interface*;
- организация проксирования запросов на несколько параллельно запущенных

систем с помощью *Reverse Proxy* (аналог *Nginx*);

- технология *DeepPavlov* – алгоритм разметки текста (преобразование файла с неразмеченным текстом на английском языке в размеченный текст с выявлением ошибок в соответствии с сертификацией *CERF*) на основе искусственного интеллекта (рис. 2);

- технология обучения НЛМ на определение языкового уровня экзаменуемого (специально разработанный алгоритм).

Подобную разметку орфографических и грамматических ошибок можно сделать на сравнительно небольшом количестве примеров (около четырёхсот). Модель предсказывает ошибки на 7% эффективнее любого эксперта, что превосходит ожидания и так называемый «научный барьер». Изучая не правила, а массив языковых данных в процессе референции, нейросеть обрабатывает материалы корпусов английского языка, формируя обобщённую картину мира, а исследователь передаёт контекст с помощью переменных, выделяя признаки и валидируя данные.

В результате проверки письменного текста нейросетевой моделью производится оценка уровня владения языком на основе характерных для каждого уровня речевых ошибок, лексико-грамматических конструкций и показателей сформированной «правильной» речи [15].

Визуализированные данные в виде агрегированных статистик, которые выводятся с использованием нескольких значений метрики, помогают сравнить результаты оценки письменных работ нейролингвистическими моделями с перекрытием с помощью четырёх экспертов. Интересным является момент несогласованности в определении оценки уровня владения языком экспертами, которые проверяли работы независимо друг от друга. Подход модели, которая способна преодолеть эту дилемму за счёт чёткого соотношения типов ошибок по заданным критериям и калибровки по множествам лингвистических параметров и баз данных, должен быть однозначным.

На рисунке 3 приведено сравнение оценок письменного текста (эссе), выполненного НЛМ и экспертами на начальном этапе обучения модели. Языковая модель анализирует ошибки, которые допускает экзаменуемый, исходя из языковых данных и моделей построения высказывания, которыми она располагает (а это все метаданные, собранные из пространства интернета и лингвистические базы знаний). На начальном этапе обучения НЛМ продемонстрировали высокий уровень обученности, показав способности к распознаванию разных типов ошибок: стилистических, грамматических, орфографических, пунктуационных, графических, ошибок согласования (пропуск слов, неправильное употребление видовременных форм глаголов, артиклей и т. п.), ошибок аргументации, организации текста (наличие вводной части, собственного мнения, выводов), текстовых скреп (слов-связок) с присвоением среднего балла за последние три критерия.

Экспертами был разработан обособленный подход к выделению ошибок. Как видно из рисунка 4, это: применение (релевантная позиция слова в потоке речи); орфографические, лексические, грамматические ошибки и т. п. Данный пример демонстрирует более глубинное понимание природы естественной речи (функционирование языковых единиц в речевом потоке) экспертами и отсылает к образованному носителю языка, что определяется соответствующим уровнем владения *B2*. Модель анализирует текст с учётом его структурной организации, вычитывая, в том числе, такие параметры как частотность употребления слов и частей речи, размер предложений, коэффициент лексического разнообразия, что видно из рисунков 5 и 6.

Как указывалось выше, модели переобучаются посредством:

- 1) ранжирования ошибок;
- 2) создания множеств речевых моделей, свойственных определённому уровню носителей языка;
- 3) изучения оценок и комментариев экспертов;

Table 1: Averaged Expert Assessment

assessment	average rating
opinion	1.0
conclusion	0.0
introduction	1.0

### Text 211

#### Service Annotation

Some people suppose that only hard work matters in the case of achieving success, while others are sure that its impossible without a piece of luck, and even people who has used to work hard may not get the same results as lazier, but luckier men. who is right? Franclly speaking, I want fully agree with any of opponents. On the one hand, success is practically impossible without hard work. To become a professional in any sphere, its needed to have a lot of various skills and knowledge, which people can get only as a result of their work. On the other hand, we all know examples of people who tried, but didnt get a meaningful result, and people who became famous or rich because of luck. However, if we talk about average person and not about certain unfair events that always happen, the success need both luck and work. Of cause, its easier to achieve your goals if you start not from zero, you may, for example, have parents who are professionals in the sphere of your interest, of your family can just be richer then others and have more money to help your wish come true. But even in that case its impossible to become successful without personal work, bac

#### Expert Annotation

Some people suppose that only hard work matters in the case of achieving success, while others are sure that its impossible without a piece of luck, and even people who has used to work hard may not get the same results as lazier, but luckier men. who is right? Franclly speaking, I want fully agree with any of opponents. On the one hand, success is practically impossible without hard work. To become a professional in any sphere, its needed to have a lot of various skills and knowledge, which people can get only as a result of their work. On the other hand, we all know examples of people who tried, but didnt get a meaningful result, and people who became famous or rich because of luck. However, if we talk about average person and not about certain unfair events that always happen, the success need both luck and work. Of cause, its easier to achieve your goals if you start not from zero, you may, for example, have parents who are professionals in the sphere of your interest, ot your family can just be richer then others and have more money to help your wish come true. But even in that case its impossible to become successful without personal work, bac

Table 2: Expert CEFR LEVEL

expert <sub>1</sub>	expert <sub>2</sub>	expert <sub>3</sub>	expert <sub>4</sub>	expert <sub>c</sub>
B2	B1	B2-	B2 (not finished)	nan

Рис. 3. Проверка текстового материала НЛМ и экспертами  
 Fig. 3. Verification of test materials by NLM and experts



Table 2: Expert CEFR LEVEL

$expert_1$	$expert_2$	$expert_3$	$expert_4$	$expert_c$
B2	B1	B2-	B2 (not finished)	nan

Table 3: Lexical Diversity

simple	root	log	mass	msttr	mattr	hdd	mtld	mtld ma wrap	mtld ma bid
0.580	8.448	0.898	0.044	0.845	0.850	0.853	108.289	97.675	99.060

Table 4: Text Counts

sentences		words	
9		211	

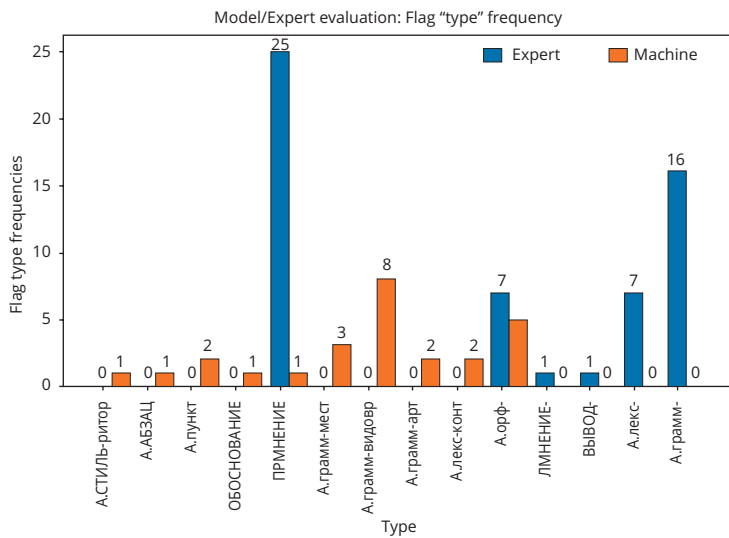


Рис. 4. Агрегированные статистики  
Fig. 4. Aggregated statistics

4) накопления правильных вариантов для определения уровней;

5) отслеживания частных и общих тенденций, характерных для определённого уровня и аккумуляции этих параметров по базам данных [16].

Интеграция технологий искусственного интеллекта в решение задач распознавания и обработки речи в перспективе может использоваться как первый этап тестирования, затем для определения уровня владе-

ния языком обученная модель анализирует развёрнутое речевое высказывание [17]. Задача проверки устного высказывания состоит в контроле умения логично и аргументированно изложить свою мысль, а также в правильности формулирования собственной точки зрения по предложенной для обсуждения проблеме и использования средств убеждающей речи. Нейросетевая модель анализирует устное речевое высказывание с позиции фонетического оформ-

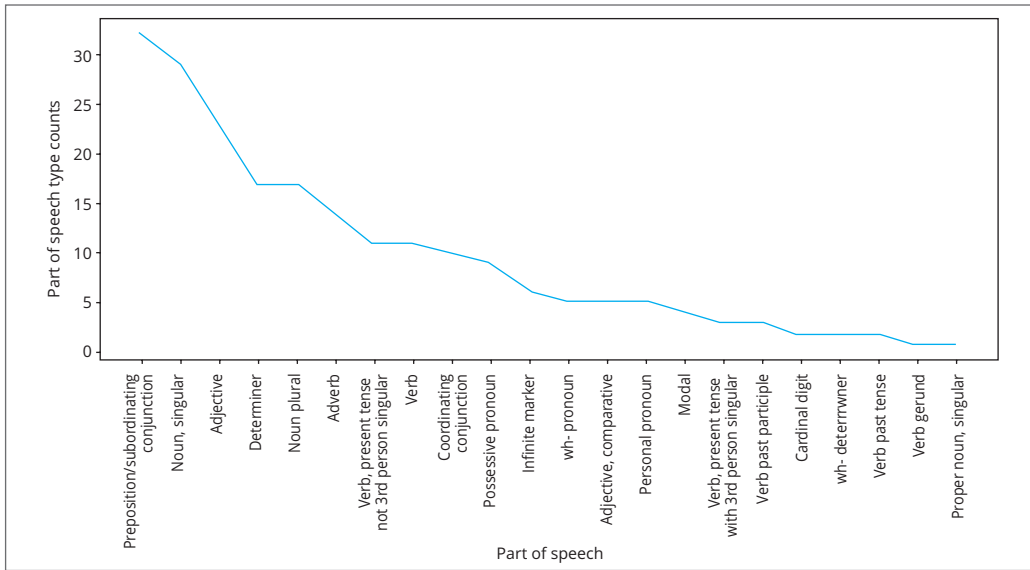


Рис. 5. Частотность употребления частей речи  
 Fig. 5. Frequency of use of parts of speech

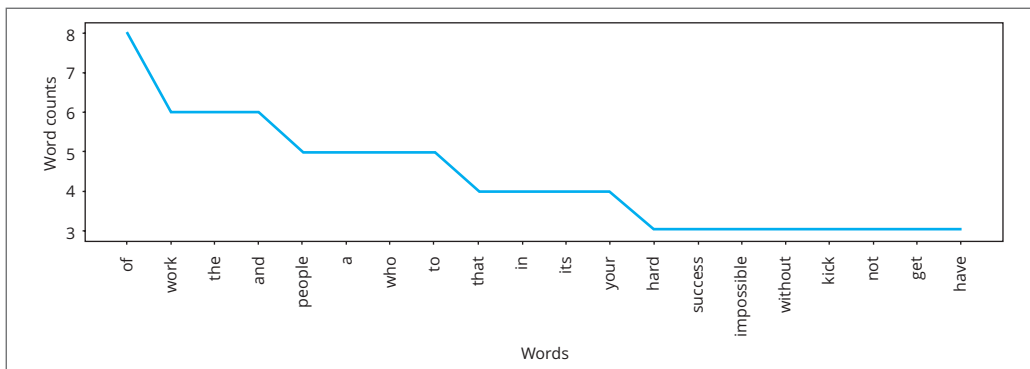


Рис. 6. Частотность употребления слов  
 Fig. 6. Frequency of use of words

ления речи (ударение, произношение, интонационный рисунок, логическое ударение и паузы, темп и громкость речи), смысловой наполненности, логичности, доказательности, умения делать выводы, убеждать и т. п. При определении уровня владения языком необходимо учитывать соответствие содержания текста теме, композицию, чёткость деления на части, удачность примеров, использование стилистических средств, усиливающих достоверность сказанного, и приёмов речевого контакта. Всесторонний

подход в оценке устного речевого высказывания может быть организован с учётом открытого метода оценки разговорного иностранного языка на основе технологии нейросети. Модель не способна анализировать синтаксическую, семантическую и прагматическую характеристики речи (для неё это абстракции), а лишь сопоставлять оформленное речевое высказывание с образцами, которыми она располагает из базы языковых данных, видеть диспропорции на основании этого сопоставления,

репродуцировать сообщение и отбирать нарушения или отклонения от речевой нормы, характерные для определённого уровня владения языком.

### Результаты тестирования и новый подход к проверке знаний

Перспектива применения результатов тестирования достаточно широка. Новый подход к проверке знаний может быть использован в различных областях прикладных исследований: в образовании – в качестве технологии *Smart Education* в аспекте цифровой трансформации образования [18; 19], для создания универсальной, объективно-валидной и адаптивной системы оценивания знаний, а также мягких (гибких) навыков, в том числе в подходе нечёткой логики в комбинации нейронных сетей и нечётких правил [20], для оценки тестовой тревожности обучающихся [21], в развитии цифровых навыков преподавателей и внедрении технологий в образовательный контент и процессы [22]; в интернет-технологиях – в интеграции технологии искусственного интеллекта и обработки естественного языка, для распознавания и синтеза речи, в качестве интеллектуальной системы принятия решений и оценивания экстралингвистических параметров, что является важным средством взаимодействия человека и компьютера, привлекает большое внимание и стимулирует исследования в промышленности и науке [23].

Нейролингвистические модели возможно задействовать в лингвистических исследованиях языковых корпусов и их интеграции в системы распознавания речи для выявления признаков, характерных для определённых речевых факторов (поток), индивидуализации речи; изучения развёртывания и анализа речевых структур [24], лингвистических и экстралингвистических факторов, оказывающих влияние на устную и письменную речь,

речепроизводство (речётворческие процессы), а также при изучении ассоциативных связей, фреймовых структур, семантических полей, формировании понятий, в гендерной лингвистике для определения гендерных характеристик речи, в теории текстообразования и при изучении текстов разных жанров, при выявлении стилистических ресурсов и потенциала языка, в переводоведении для создания системы интеллектуального перевода [25], в психологических науках для выявления особенностей и закономерностей речевых сообщений (построения высказываний) и восприятия речи.

### Заключение

Созданная и апробированная в МФТИ Интеллектуальная Система Тестирования Общеязыковых Компетенций (ИСТОК) способна не только обеспечить единый стандарт подготовки и оценивания уровня специалистов из разных областей, но и продуктивно разрешить целый ряд междисциплинарных исследовательских задач. Однако для её полноценного внедрения потребовалось переработать большой массив данных по сравнительному анализу различных систем тестирования, разработать единые требования и критерии оценки, сформировать банк заданий, апробировать компьютерное тестирование, создать на его базе адаптивное тестирование с последующим внедрением нейросетевых лингвистических решений.

Подобная система, разработанная для английского языка, впоследствии будет применима к другим языкам как иностранным – китайскому, русскому, языкам стран СНГ и т. д. Данная система является результатом внедрения новых когнитивных, мультимодальных и нейросетевых технологий, основанных на искусственном интеллекте, обработке больших данных, анализе и синтезе текстов языковых корпусов<sup>1</sup>.

<sup>1</sup> British National Corpus [Electronic Resource], URL: <http://www.natcorp.ox.ac.uk/corpus/index.xml> (дата обращения: 10.07.2023); MyCobuild.com [Electronic Resource], URL: <http://www.mycobuild.com/about-collins-corpus.aspx> (дата обращения: 10.07.2023); Oxford Dictionaries [Electronic Resource], URL: <http://oxforddictionaries.com/words/the-oxford-english-corpus> (дата обращения: 12.07.2023); The Corpus of

В настоящий момент исследовательский проект ещё не завершился, но уже близок к финальной стадии – этапу оценки устной части теста с использованием технологии распознавания речи.

### Литература

1. Common European Framework of Reference for Languages: Learning, teaching, assessment. Language Policy Unit, Strasbourg. 2001. 278 p. URL: <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4> (дата обращения: 12.07.2023).
2. *Weir C.J.* Language Testing and Validation: An Evidence-Based Approach. Houndgrange, Hampshire, UK: Palgrave-Macmillan. 2005. 316 p. URL: [https://ztcprep.com/library/tesol/Language\\_Testing\\_and\\_Validation/Language\\_Testing\\_and\\_Validation\\_\(www.ztcprep.com\).pdf](https://ztcprep.com/library/tesol/Language_Testing_and_Validation/Language_Testing_and_Validation_(www.ztcprep.com).pdf) (дата обращения: 12.07.2023).
3. *Аванесов В.С.* История педагогической теории измерений // Образовательные технологии. Москва. 2012. № 4. С. 28–38. URL: <https://www.iedtech.ru/files/journal/2012/4/pedagogical-theory-of-measurement.pdf> (дата обращения: 12.07.2023).
4. *Fernández Álvarez M.* Language Testing in the Digital Era // E. Martín-Monje, I. Elorza, B. García Riaza (Eds.), Technology-Enhanced Language Learning for Specialized Domains. Practical applications and mobility. London and New-York: Routledge: 2016. P. 61–72. ISBN: 9781138565135.
5. *Lim G.S.* Developing and Validating a Mark Scheme for Writing // Cambridge ESOL: Research Notes: No. 49. August 2012. P. 6–10. URL: <https://www.cambridgeenglish.org/images/23166-research-notes-49.pdf> (дата обращения: 12.07.2023).
6. *Taylor L.* Examining Speaking. Research and Practice in Assessing Second Language Speaking. Cambridge: University of Cambridge Press. 2011. 408 p. ISBN-10: 0521736706. ISBN-13: 978-0521736701.
7. *Маковская Л.Г.* Approaches to assessing language skills at higher educational institutions // Евразийский Союз Учёных. 2020. № 7-7 (76). С. 74–83. DOI: 10.31618/ESU.2413-9335.2020.7.76.945
8. *Shobamy E.* The Power of the Tests: the Impact of Language Tests on Teaching and Learning // NFLC Occasional Papers: June 1993. URL: [https://www.researchgate.net/publication/240324351\\_The\\_Power\\_of\\_Tests\\_A\\_Critical\\_Perspective\\_on\\_the\\_Use\\_of\\_Language\\_Tests](https://www.researchgate.net/publication/240324351_The_Power_of_Tests_A_Critical_Perspective_on_the_Use_of_Language_Tests) (дата обращения: 12.07.2023).
9. Адаптивное тестирование в системах дистанционного обучения // Наука и образование: Электронное научное издание. МГТУ им. Н.Э. Баумана. 2007. № 4. URL: <http://engineering-science.ru/doc/65577.html> (дата обращения: 12.07.2023).
10. *Гданский Н.И., Альтиментова Д.Ю.* Адаптивные методы тестирования знания при компьютерных формах обучения: монография. М.: Издательство РГСУ, 2015. 220 с. URL: [https://viperson.ru/uploads/attachment/file/951375/Gdanskiy\\_Al\\_timentova\\_Tituly\\_vvedenie\\_oglavlenie\\_bibliografiya.pdf](https://viperson.ru/uploads/attachment/file/951375/Gdanskiy_Al_timentova_Tituly_vvedenie_oglavlenie_bibliografiya.pdf) (дата обращения: 12.07.2023).
11. *Meunier L.* Computer-adaptive language tests (CALT) offer a great potential for functional testing. Yes why don't test // CALICO Journal. 1994. Vol. 11. No. 4. P. 23–39. DOI: 10.1558/cj.v11i4.23-39
12. *Laupichler M.C., Aster A., Schirch J., Raupach T.* Artificial intelligence literacy in higher and adult education: A scoping literature review // Computers and Education: Artificial Intelligence. 2022. Vol. 3. Article no. 100101. DOI: 10.1016/j.caeai.2022.100101
13. *Колдаев В.Д., Волков А.С., Павлов Р.И.* Технология применения адаптивного тестирования IT-специалистов коммерческой фирмы // Научный журнал. 2017. № 5 (18). С. 83–86. EDN: YNWSIJ.
14. *Suraworachet, W., Zhou, Q., Cukurova, M.* Impact of combining human and analytics feedback on students' engagement with, and performance in, reflective writing tasks // Int J Educ Technol High Educ. 2023. Vol. 20. No. 1. DOI: 10.1186/s41239-022-00368-0
15. *Firoozi T., Mohammadi H., Gierl M. J.* Using active learning methods to strategically select essays for automated scoring // Educational Measurement: Issues and Practice. 2022. Vol. 1. No. 10. DOI: 10.1111/emip.12537

Contemporary American English [Electronic Resource], URL: <http://corpus.byu.edu/coca/> (дата обращения: 12.07.2023).

16. Ouyang L., Wu J., Jiang X., Almeida D., Wainwright C.L., Misbin P. et al. Training language models to follow instructions with human feedback. 2022. DOI: 10.48550/arXiv.2203.02155
17. Carolus A., Augustin Y., Markus A., Wienrich C. Digital interaction literacy model – Conceptualizing competencies for literate interactions with voice-based AI systems // *Computers and Education: Artificial Intelligence*. 2023. Vol. 4. Article no. 100114. DOI: 10.1016/j.caeai.2022.100114
18. Dittrich L., Aagaard T., Hjukse H. The perceived affordances of simulation-based learning: online student teachers' perspectives // *Int J Educ Technol High Educ*. 2022. Vol. 19. No. 60. DOI: 10.1186/s41239-022-00366-2
19. Xu W., Ouyang F. The application of AI technologies in STEM education: a systematic review from 2011 to 2021 // *IJ STEM Ed*. 2022. Vol. 9. No. 59. DOI: 10.1186/s40594-022-00377-5
20. Göktepe Yıldız S., Göktepe Körpeoğlu S.. Prediction of students' perceptions of problem solving skills with a neuro-fuzzy model and hierarchical regression method: A quantitative study // *Educ Inf Technol*. 2023. Vol. 28. No. 8879–8917. DOI: 10.1007/s10639-022-11446-1
21. Immekus J.C., Jeong Ts., Yoo J.E. Machine learning procedures for predictor variable selection for schoolwork-related anxiety: evidence from PISA 2015 mathematics, reading, and science assessments // *Large-scale Assess Educ*. 2022. Vol. 10. No. 30. DOI: 10.1186/s40536-022-00150-8
22. Kaminskiene L., Järvelä S., Lehtinen E. How does technology challenge teacher education? // *Int J Educ Technol High Educ*. 2022. Vol. 19. No. 64. DOI: 10.1186/s41239-022-00375-1
23. Nan Xue. Analysis Model of Spoken English Evaluation Algorithm Based on Intelligent Algorithm of Internet of Things // *Comput Intell Neurosci*. 2022. Mar 27. Article no. 8469945. DOI: 10.1155/2022/8469945
24. Xin Wang. Research on Open Oral English Scoring System Based on Neural Network // *Comput Intell Neurosci*. 2022. Apr 23. Article no. 1346543. DOI: 10.1155/2022/1346543
25. Lei L., Wang H. Design and Analysis of English Intelligent Translation System Based on Internet of Things and Big Data Model // *Comput Intell Neurosci*. 2022. May 19. Article no. 6788813. DOI: 10.1155/2022/6788813

Статья поступила в редакцию 13.07.2023

Принята к публикации 22.08.2023

## References

1. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Language Policy Unit, Strasbourg. (2001). 278 p. Available at: <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4> (accessed 12.07.2023).
2. Weir, C.J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Houndgrange, Hampshire, UK: Palgrave-Macmillan. 316 p. Available at: [https://ztcprep.com/library/tesol/Language\\_Testing\\_and\\_Validation/Language\\_Testing\\_and\\_Validation](https://ztcprep.com/library/tesol/Language_Testing_and_Validation/Language_Testing_and_Validation) (accessed 12.07.2023).
3. Avanesov, V.S. (2012). [History of Pedagogical Theory of Measurements]. *Educational Technologies*. No. 4, pp. 28–38. Available at: <https://www.iedtech.ru/files/journal/2012/4/pedagogical-theory-of-measurement.pdf> (accessed 12.07.2023). (In Russ.).
4. Fernández Álvarez, M. (2016). Language Testing in the Digital Era. In E. Martín-Monje, I. Elorza, B. García Ríaza (Eds.), *Technology-Enhanced Language Learning for Specialized Domains. Practical applications and mobility*. London and New-York: Routledge: pp. 61-72. ISBN: 9781138565135.
5. Lim, G.S. (2012). Developing and Validating a Mark Scheme for Writing: *Cambridge ESOL: Research Notes*: No. 49 / August 2012: pp. 6-10. Available at: <https://www.cambridgeenglish.org/images/23166-research-notes-49.pdf> (accessed 12.07.2023).
6. Taylor, L. (2011). *Examining Speaking. Research and Practice in Assessing Second Language Speaking*. Cambridge: University of Cambridge Press. 408 p. ISBN-10: 0521736706. ISBN-13: 978-0521736701.



7. Makovskaya, L.G. [Approaches to Assessing Language Skills at Higher Educational Institutions] *Eurasian Union of Scientists*. (2020). No. 7-7 (76), pp. 74-83. DOI: 10.31618/ESU.2413-9335.2020.7.76.945 (In Russ.).
8. Shohamy, E. (1993). The Power of the Tests: the Impact of Language Tests on Teaching and Learning.: *NFLC Occasional Papers*: June 1993. Available at: [https://www.researchgate.net/publication/240324351\\_The\\_Power\\_of\\_Tests\\_A\\_Critical\\_Perspective\\_on\\_the\\_Use\\_of\\_Language\\_Tests](https://www.researchgate.net/publication/240324351_The_Power_of_Tests_A_Critical_Perspective_on_the_Use_of_Language_Tests) (accessed 12.07.2023).
9. Adaptive Testing in Distance Learning Systems (2007). [Science and Education: Electronic scientific edition]. Bauman Moscow State Technical University. No. 4. Available at: <http://engineering-science.ru/doc/65577.html> (accessed 12.07.2023). (In Russ.).
10. Gdanskii, N.I., Altimentova, D.Yu. (2015). *Adaptive Methods of Knowledge Testing in Computer-based Forms of Learning*: Monograph. Moscow: RGSU Publishing House. 220 p. Available at: [https://viperson.ru/uploads/attachment/file/951375/Gdanskii\\_Al\\_timentova\\_Tituly\\_vvedenie\\_oglavlenie\\_bibliografiya.pdf](https://viperson.ru/uploads/attachment/file/951375/Gdanskii_Al_timentova_Tituly_vvedenie_oglavlenie_bibliografiya.pdf) (accessed 12.07.2023). (In Russ.).
11. Meunier, L. (1994). Computer-Adaptive Language Tests (CALT) Offer a Great Potential for Functional Testing. Yes Why Don't Test: *CALICO Journal*. Vol. 11, no. 4, pp. 23-39, doi: 10.1558/cj.v11i4.23-39
12. Laupichler, M.C., Aster, A., Schirch, J., Raupach, T. (2022). Artificial Intelligence Literacy in Higher and Adult Education: A Scoping Literature Review, *Computers and Education: Artificial Intelligence*. Vol. 3, article no. 100101, doi: 10.1016/j.caeai.2022.100101
13. Koldaev, V.D., Volkov, A.S., Pavlov, R.I. (2017). [Technology of Application of the Adaptive Testing of the IT-specialists of the Commercial Firm] *Scientific Journal*. No. 5 (18), pp. 83-86. Available at: [https://www.elibrary.ru/download/elibrary\\_29149277\\_15922677.pdf](https://www.elibrary.ru/download/elibrary_29149277_15922677.pdf) (accessed 12.07.2023). (In Russ.).
14. Suraworachet, W., Zhou, Q., Cukurova, M. (2023). Impact of Combining Human and Analytics Feedback on Students' Engagement with, and Performance in, Reflective Writing Tasks. *Int J Educ Technol High Educ*. Vol. 20, no. 1, doi: 10.1186/s41239-022-00368-0
15. Firoozi, T., Mohammadi, H., Gierl, M. J. (2022). Using Active Learning Methods to Strategically Select Essays for Automated Scoring. *Educational Measurement: Issues and Practice*. Vol. 1, no. 10, doi: 10.1111/emip.12537
16. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C.L., Mishkin, P. et al. (2022). *Training Language Models to Follow Instructions with Human Feedback*, doi: 10.48550/arXiv.2203.02155.
17. Carolus, A., Augustin, Y., Markus, A., Wienrich, C. (2023). Digital Interaction Literacy Model – Conceptualizing Competencies for Literate Interactions with Voice-Based AI Systems, *Computers and Education: Artificial Intelligence*. Vol. 4, article no. 100114, doi: 10.1016/j.caeai.2022.100114
18. Dittrich, L., Aagaard, T., Hjukse, H. (2022). The Perceived Affordances of Simulation-Based Learning: Online Student Teachers' Perspectives. *Int J Educ Technol High Educ*. Vol. 19, no. 60, doi: 10.1186/s41239-022-00366-2
19. Xu, W., Ouyang, F. (2022). The Application of AI Technologies in STEM Education: A Systematic Review from 2011 to 2021. *IJ STEM Ed*. Vol. 9, no. 59, doi: 10.1186/s40594-022-00377-5
20. Göktepe Yıldız, S., Göktepe Körpeoğlu, S. (2023). Prediction of Students' Perceptions of Problem Solving Skills with a Neuro-Fuzzy Model and Hierarchical Regression Method: A Quantitative Study. *Educ Inf Technol*. Vol. 28, no. 8879–8917, doi: 10.1007/s10639-022-11446-1
21. Immekus, J.C., Jeong, Ts., Yoo, J.E. (2022). Machine Learning Procedures for Predictor Variable

- Selection for Schoolwork-Related Anxiety: Evidence from PISA 2015 Mathematics, Reading, and Science Assessments. *Large-scale Assess Educ.* Vol. 10, no. 30, doi: 10.1186/s40536-022-00150-8
22. Kaminskienė, L., Järvelä, S., Lehtinen, E. (2022). How Does Technology Challenge Teacher Education? *Int J Educ Technol High Educ.* Vol. 19, no. 64, doi: 10.1186/s41239-022-00375-1
23. Nan Xue. (2022). Analysis Model of Spoken English Evaluation Algorithm Based on Intelligent Algorithm of Internet of Things. *Comput Intell Neurosci.* Mar 27, article no. 8469945, doi: 10.1155/2022/8469945
24. Xin Wang. (2022). Research on Open Oral English Scoring System Based on Neural Network. *Comput Intell Neurosci.* Apr 23, article no. 1346543, doi: 10.1155/2022/1346543
25. Lei, L., Wang, H. (2022). Design and Analysis of English Intelligent Translation System Based on Internet of Things and Big Data Model. *Comput Intell Neurosci.* May 19, article no. 6788813, doi: 10.1155/2022/6788813

—————  
 —————  
 The paper was submitted 13.07.2023  
 Accepted for publication 22.08.2023



Science Index РИНЦ-2021

СОЦИОЛОГИЧЕСКИЕ ИССЛЕДОВАНИЯ	10,767
ВОПРОСЫ ФИЛОСОФИИ	9,978
ВОПРОСЫ ОБРАЗОВАНИЯ	9,805
<b>ВЫСШЕЕ ОБРАЗОВАНИЕ В РОССИИ</b>	<b>9,236</b>
ЭПИСТЕМОЛОГИЯ И ФИЛОСОФИЯ НАУКИ	8,065
ПСИХОЛОГИЧЕСКАЯ НАУКА И ОБРАЗОВАНИЕ	7,968
ОБРАЗОВАНИЕ И НАУКА	7,275
УНИВЕРСИТЕТСКОЕ УПРАВЛЕНИЕ: ПРАКТИКА И АНАЛИЗ	6,895
ИНТЕГРАЦИЯ ОБРАЗОВАНИЯ	6,494
АЛМА МАТЕР (ВЕСТНИК ВЫСШЕЙ ШКОЛЫ)	4,042
ВЫСШЕЕ ОБРАЗОВАНИЕ СЕГОДНЯ	3,236
ПЕДАГОГИКА	2,571