 © Авторы статьи / Authors of the article

РУССКИЙ ЯЗЫК. ЯЗЫКИ НАРОДОВ РОССИИ/RUSSIAN LANGUAGE. LANGUAGES OF THE PEOPLES OF RUSSIA

DOI: https://doi.org/10.60797/RULB.2026.83.1 EDN: CPTYHT

ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКАЯ МОДЕЛЬ РАЗРАБОТКИ И ВАЛИДАЦИИ БАНКА ТЕСТОВЫХ ЗАДАНИЙ КОМПЕТЕНТНОСТНОЙ НАПРАВЛЕННОСТИ ДЛЯ ОЦЕНКИ ГРАММАТИКИ РУССКОГО ЯЗЫКА

Научная статья

Хоанг Ч. Т.Х.[1, *], Данг Т. Т.Х.[2], Чан Л. К.[3]
[1] ORCID : 0000-0003-1178-0893;
[1, 2, 3] Хошиминский государственный педагогический университет, Хошимин, Вьетнам

* Корреспондирующий автор (tranghth[at]hcmue.edu.vn)

Предложена: 23.06.2026; Принята: 22.10.2026; Опубликована: 09.11.2026

Аннотация

В статье предлагается теоретико-методологическая модель разработки и валидации банка тестовых заданий, ориентированного на оценку грамматической компетенции в рамках компетентностного подхода к обучению русскому языку. Работа не описывает эмпирически апробированный банк заданий: её предмет — обоснование принципов, на которых такая разработка должна строиться. На основе анализа зарубежных и отечественных источников операционализируется конструкт «грамматическая компетенция», прослеживается эволюция моделей валидности — от классической трёхкомпонентной схемы к аргументационному подходу М. Кейна, — рассматриваются психометрические методы калибровки заданий (теория моделирования и параметризации тестовых заданий, Item Response Theory, IRT; модель Раша) и условия их применимости.

Отдельное внимание уделено сложившемуся дисбалансу исследовательского внимания: оценивание грамматики и лексики оказывается статистически недопредставленным по сравнению с оцениванием рецептивных и продуктивных видов речевой деятельности. Предложена и обоснована шестиэтапная модель разработки банка, включающая спецификацию конструкта, экспертную оценку содержания, пилотирование, IRT-калибровку и анализ дифференцирующего функционирования заданий (differential item functioning, DIF); обсуждаются условия её применения, соотношение грамматической и коммуникативной компетенции, а также специфика вьетнамского контекста, в котором родной язык учащихся имеет изолирующий, а не флективный строй.

Ключевые слова: банк тестовых заданий, компетентностный подход, грамматика русского языка, валидность теста, модель Раша, тестирование русского языка как иностранного.

A THEORETICAL AND METHODOLOGICAL MODEL FOR THE DEVELOPMENT AND VALIDATION OF A BANK OF COMPETENCE-BASED TEST ASSIGNMENTS FOR ASSESSING RUSSIAN GRAMMAR

Research article

Hoang T. T.H.[1, *], Dang T. T.H.[2], Tran L. Q.[3]
[1] ORCID : 0000-0003-1178-0893;
[1, 2, 3] Ho Chi Minh City University of Education, Ho Chi Minh City, Vietnam

* Corresponding author (tranghth[at]hcmue.edu.vn)

Suggested: 23.06.2026; Accepted: 22.10.2026; Published: 09.11.2026

Abstract

The article proposes a theoretical and methodological model for the development and validation of a bank of test assignments aimed at assessing grammatical competence within the framework of a competence-based approach to teaching the Russian language. The paper does not describe an empirically tested bank of assignments: its focus is on substantiating the principles upon which such a development should be based. Based on an analysis of international and domestic sources, the construct of ‘grammatical competence’ is operationalised; the evolution of validity models is traced — from the classical three-component framework to M. Kane’s argumentative approach; and psychometric methods for calibrating items (theory of modelling and parameterisation of test items, Item Response Theory, IRT; the Rasch model) and the conditions under which they are applicable.

Particular attention is paid to the existing imbalance in research focus: the assessment of grammar and vocabulary is statistically under-represented compared with the evaluation of receptive and productive types of speech activity. A six-stage model for developing a test bank is suggested and justified, comprising construct specification, expert content evaluation, pilot testing, IRT calibration and analysis of differential item functioning (DIF); the conditions for its application, the relationship between grammatical and communicative competence, and the specifics of the Vietnamese context, in which the students' native language has an isolating rather than an inflectional system, are discussed.

Keywords: bank of test assignments, competence-based approach, Russian grammar, test validity, Rasch model, testing of Russian as a foreign language.

Введение

 © Авторы статьи / Authors of the article

Переход школьного и вузовского языкового образования на компетентностную модель обучения требует адекватного инструментария измерения. Пока образовательный стандарт формулирует результат обучения через набор компетенций, а не через сумму усвоенных правил, тестовые формы, опирающиеся на изолированное воспроизведение отдельной грамматической конструкции, перестают быть достаточными. Между тем именно грамматика как отдельный компонент языковой способности остаётся, по данным библиометрического анализа Aryadoust с соавторами, одной из наименее исследованных областей языкового тестирования: оценивание грамматики и лексики статистически значимо реже становится предметом публикаций, чем оценивание чтения, аудирования, письма и говорения [1]. Данный разрыв заметен применительно к русскому языку: методический аппарат сертификационного тестирования разрабатывался преимущественно для интегральных речевых умений, тогда как специализированные, психометрически откалиброванные банки заданий по морфологии и синтаксису остаются немногочисленными.

Целью настоящей статьи является теоретико-методологическое обоснование модели разработки и валидации банка тестовых заданий компетентностной направленности для оценки грамматики русского языка. Задачи работы включают:

1) операционализацию конструкта «грамматическая компетенция»;
2) рассмотрение психометрических моделей, применяемых при построении банков заданий;
3) обобщение отечественного и зарубежного опыта тестирования русского языка;
4) разработку поэтапного алгоритма создания и валидации искомого банка.

Необходимо сразу оговорить границы работы: она систематизирует принципы, на которых должна строиться эмпирическая апробация, но не описывает результаты уже проведённого тестирования на конкретной выборке испытуемых — соответственно, все выводы статьи носят гипотетический характер и подлежат последующей эмпирической проверке.

Методы и принципы исследования

Работа представляет собой теоретико-методологический (нарративный) обзор и синтез литературы, а не эмпирическое исследование с собственной выборкой испытуемых. Источники отбирались из баз Scopus, Web of Science, Crossref и Semantic Scholar по ключевым словам «банк тестовых заданий» / item bank, «компетентностный подход» / competency-based approach, «валидность теста» / test validity, item response theory, «модель Раша» / Rasch model, ««тестирование русского языка как иностранного (РКИ)», grammar assessment; в выборку включались публикации, содержащие процедуру разработки или валидации банка заданий средствами Item Response Theory (IRT), теоретическую модель валидности языкового теста либо описание опыта тестирования русского языка.

Синтез источников строился на трёх методологических ориентирах: компетентностной ориентации конструкта, при которой задание проверяет применение грамматической формы в коммуникативной задаче, а не изолированное знание правила; аргументационной валидации, доказывающей пригодность теста цепочкой умозаключений от балла испытуемого к принимаемому решению [4]; и психометрической калибровке заданий средствами IRT и модели Раша, обеспечивающей сопоставимость шкалы трудности при расширении банка [11].

Основные результаты

3.1. Компетентностный подход и операционализация конструкта «грамматическая компетенция»

Компетентностный подход смещает акцент с содержания обучения на результат — способность обучающегося применять знания, умения и личностные качества для решения практических задач. Makulova с соавторами определяют компетентность как интегративную характеристику личности, формируемую не суммированием отдельных знаний, а их функциональным объединением вокруг конкретной деятельности [10]. Схожим образом Kulik с соавторами, обобщая теоретические трактовки компетентностного подхода в высшем образовании, подчёркивают принципиальную контекстность компетенции: она проявляется не абстрактно, а в определённой коммуникативной или профессиональной ситуации [7].

Для того чтобы данный принцип был операционализируем в тестовом задании, а не оставался декларацией, конструкт «грамматическая компетенция» целесообразно разложить на компоненты, поддающиеся отдельному измерению. В настоящей модели выделяются четыре таких компонента.

Первый — морфологический: владение системой словоизменения (падежные формы существительных, прилагательных и местоимений; видовременные и залоговые формы глагола; формы числительных при сочетании с существительным).

Второй — синтаксический: владение моделями простого и сложного предложения, включая модели глагольного и именного управления, порядок слов и союзные средства связи частей сложного предложения.

Третий — морфосинтаксический, объединяющий первые два: согласование словоформ в пределах словосочетания и предложения — например, согласование прилагательного с существительным в роде, числе и падеже или согласование сказуемого с подлежащим.

Четвёртый — компонент грамматической уместности: способность выбирать вариант из числа формально допустимых в соответствии с коммуникативным намерением высказывания, например выбор совершенного или несовершенного вида глагола.

Первые три компонента, как правило, поддаются заданиям с однозначно определяемым правильным ответом; четвёртый требует более сложных форматов и, зачастую, экспертного шкалирования ответа. Подобный принцип содержательного разложения конструкта — переход от абстрактных правил к формам, реально используемым учащимися разных уровней владения языком в аутентичных высказываниях, — уже реализован применительно к английской грамматике в проекте English Grammar Profile [13], откуда, с необходимой адаптацией к морфологии русского языка, заимствуется общая логика описания.

 © Авторы статьи / Authors of the article

Различие между традиционным и компетентностно-ориентированным заданием удобно показать на одном и том же грамматическом явлении — глагольном управлении. Традиционное задание на трансформацию словоформы («Поставьте слово в скобках в нужную форму: Я интересуюсь (русская литература)») проверяет знание падежного окончания вне какой-либо коммуникативной задачи: правильный ответ единствен и не зависит от контекста высказывания. Компетентностно-ориентированный аналог того же грамматического явления предполагает выбор уместной формы внутри диалогической реплики: «— Над чем вы сейчас работаете? — Я работаю ____ .

а) над новую статью;

б) над новой статьёй;

в) с новой статьёй».

Здесь для успешного выполнения задания недостаточно знать падежную парадигму — необходимо также определить, какая модель управления требуется данным глаголом в данном контексте, а дистракторы включают типичную ошибку смешения моделей управления «работать над» и «работать с». Такой формат приближает задание к естественной коммуникативной ситуации, оставаясь при этом заданием закрытого типа, пригодным для последующей IRT-калибровки.

3.2. Эволюция моделей валидности языкового теста

Разработка банка заданий неотделима от проблемы валидности, поскольку именно валидность определяет корректность выводов, которые делаются на основании тестового балла. Im, Shin и Cheng прослеживают, как представления о валидности эволюционировали от трёхкомпонентной модели — содержательной, критериальной и конструктной валидности — к единой конструктной модели С. Мессика и далее к аргументационному подходу М. Кейна, в рамках которого валидация понимается не как разовая процедура, а как выстраивание цепочки доказательств, связывающих балл испытуемого с решением, принимаемым на основании этого балла [4]. Такая логика особенно уместна для компетентностно-ориентированного тестирования: недостаточно показать, что задание статистически «работает», — необходимо обосновать, что его успешное выполнение действительно свидетельствует о сформированности грамматической компетенции, а не о механическом запоминании формы.

Практическое применение аргументационного подхода к разным видам речевой деятельности демонстрирует обзор Fan и Yan, охвативший более ста публикаций по оцениванию говорения: авторы показывают, что даже для говорения — деятельности, изучаемой значительно активнее грамматики, — целостная доказательная база валидности выстраивается медленно и неравномерно, кодируясь по типам умозаключений: оценивание (scoring), обобщение (generalization), экстраполяция (extrapolation), объяснение (explanation) и использование (utilization) [3]. Показательно при этом, что рассматриваемые ниже исследования, применяющие IRT и модель Раша непосредственно к языковому материалу, как правило, ограничиваются доказательством статистического соответствия данных модели и почти не обращаются к полной аргументационной цепочке Кейна: между теоретически провозглашённым аргументационным идеалом валидации и практикой психометрической калибровки конкретных инструментов сохраняется заметный разрыв, который предлагаемая в разделе 3.4 модель пытается сократить, включая психометрическую калибровку лишь как один из шести этапов, а не как замену валидационной аргументации в целом.

3.3. IRT и модель Раша как инструмент калибровки банка заданий

Если аргументационный подход задаёт логику доказательства валидности, то технический инструментарий калибровки заданий обеспечивает IRT — теория измерения латентных (непосредственно не наблюдаемых) признаков, в рамках которой вероятность верного ответа на задание моделируется как функция уровня подготовленности испытуемого и параметров самого задания. В отличие от классической теории тестов, где трудность задания оценивается относительно конкретной выборки испытуемых, IRT помещает параметры испытуемого и задания на единую шкалу, что делает возможным сравнение результатов между разными вариантами теста и постепенное пополнение банка новыми заданиями без потери сопоставимости шкалы. Выбор именно IRT и, в частности, модели Раша в качестве основного психометрического инструментария обусловлен четырьмя обстоятельствами.

Во-первых, только IRT помещает параметры задания и испытуемого на общую шкалу, что необходимо именно для банка, предполагающего последующее расширение: без общей шкалы новые задания невозможно сопоставить со старыми без повторного тестирования всей совокупности испытуемых.

Во-вторых, применение однопараметрической модели Раша предполагает достаточно однородный, в идеале одномерный конструкт, что делает необходимой предварительную проверку размерности ещё до перехода к калибровке.

В-третьих, для устойчивой оценки параметров дихотомической модели Раша, как правило, требуется выборка в несколько сотен испытуемых на каждый пилотируемый блок заданий, что накладывает организационное ограничение на реализацию модели. В-четвёртых, если компоненты конструкта окажутся недостаточно однородными для дихотомической модели, потребуется переход к политомическим или многомерным IRT-моделям.

Вопрос о размерности конструкта остаётся дискуссионным. Систематический обзор Min и Aryadoust, охвативший шестьдесят два исследования 1985–2020 годов, показывает, что вопрос о том, является ли измеряемая языковая способность единой или распадается на несколько относительно независимых компонентов, требует эмпирической проверки для каждого нового инструмента [11]. Показательно, что практические работы по построению банков заданий не всегда решают этот вопрос эмпирически: так, Chen с соавторами при построении банка заданий по читательской грамотности исходят из унитарной модели конструкта и не приводят отдельного анализа его размерности [2], тогда как для предлагаемой в настоящей статье четырёхкомпонентной операционализации грамматической компетенции подобное допущение было бы преждевременным: прежде чем применять единую шкалу трудности ко всем четырём компонентам, необходимо эмпирически установить, образуют ли они единый континуум или требуют раздельного измерения

 © Авторы статьи / Authors of the article

Практическую процедуру построения банка заданий на основе IRT иллюстрирует исследование Chen с соавторами: авторы откалибровали девяносто девять заданий на понимание прочитанного для четвероклассников, используя схему уравнивания тестов (test equating) и сбалансированный неполный блочный план предъявления, что позволило собрать данные по всему банку, не предъявляя каждому испытуемому полный набор заданий [2]. Тот же принцип — модульное предъявление заданий с последующим уравниванием шкалы — применим и к банку заданий по грамматике русского языка: он позволяет постепенно расширять банк, не пересматривая каждый раз шкалу целиком.

Отдельного внимания заслуживает опыт применения модели Раша и смежных IRT-моделей непосредственно к русскоязычным инструментам. Logvinenko с соавторами провели IRT-анализ теста «Синонимы» из русскоязычной батареи языковых тестов ARFA-RUS, откалиброванной средствами IRT, сопоставив параметры модели для двух групп взрослых испытуемых — выросших в родной семье и имеющих опыт институционализации, — и показали, что часть заданий демонстрирует дифференцирующее функционирование заданий (differential item functioning, DIF) в зависимости от группы, что потребовало исключения одного задания из окончательной версии шкалы [8, С 35–63]. Talantseva с соавторами применили классический и IRT-анализ к русскоязычной адаптации шкал дошкольного развития языка Preschool Language Scales, Fifth Edition (PLS-5), включающей морфосинтаксические субтесты, и сопоставили результаты обоих подходов на выборке из 473 детей [14, С. 174–195]; показательно уже само по себе решение сопоставлять два подхода, а не полагаться на один из них, — оценки надёжности, получаемые классическим и IRT-методом, в принципе не обязаны совпадать, и именно поэтому двойная проверка методологически предпочтительнее. Наконец, Прихода описывает стандартизацию русскоязычного диагностического инструмента из семи субтестов, среди которых — субтест на структуру предложения и субтест на словообразование, то есть задания, непосредственно затрагивающие грамматическую компетенцию [18, С. 25–33]. Показательно, что все три работы относятся к области психологии развития и логопедии, а не к образовательному тестированию грамматики как учебного предмета, — это ещё раз подчёркивает дефицит специализированных психометрических исследований именно в сфере преподавания русского языка как учебной дисциплины.

3.4. Отечественный опыт тестирования русского языка

Показателен в этом отношении опыт локализации международного инструмента оценки ранних навыков чтения iPIPS, описанный Ivanova и Kardanova-Biryukova: авторы не просто перевели англоязычные задания, а заново смоделировали русскоязычные задания, функционально эквивалентные оригиналу, после чего Раш-анализом подтвердили соответствие структуры и иерархии заданий теоретической рамке конструкта [5, С. 93–115]. Такая процедура — не перевод, а параллельное конструирование заданий на основе общей теоретической модели конструкта — представляется методологически более корректной и для банка заданий по грамматике русского языка, если он создаётся с оглядкой на зарубежные аналоги.

Карданова и Иванова, открывая специальный выпуск журнала «Вопросы образования», посвящённый психометрике, прослеживают историю отечественной тестологии — от работ по конструированию педагогических тестов на основе современных математических моделей до сегодняшних приложений многофасеточного Раш-измерения [6]. Эта преемственность существенна: банк заданий по грамматике, разрабатываемый сегодня, методологически опирается на давнюю отечественную традицию, а не заимствует психометрический аппарат исключительно из зарубежных источников.

Применительно собственно к тестированию русского языка как иностранного показательны два направления работ. Первое — совершенствование стандартизированных экзаменационных процедур. Алимова с соавторами анализируют введение единого вступительного испытания по русскому языку для иностранных абитуриентов как новый этап образовательной практики [15, С. 4–17], тогда как Дубинина и Птюшкин, опираясь на материал субтеста «Письмо» уровней I–III Теста по русскому языку как иностранному (ТРКИ), формулируют возрастные ориентиры выбора уровня тестирования для школьников, изучающих русский язык как иностранный [17, С. 222–234]. Второе направление — автоматизация конструирования заданий: Белый, Митрофанова и Дубинина предложили и апробировали метод автоматической генерации лексико-грамматических заданий множественного выбора на основе дистрибутивно-семантических моделей и аутентичных текстов, получив при экспертной оценке точность 0,8 и полноту 0,91 [16, С. 108–118]. Схожим образом инструмент «Текстометр», разработанный Laposhina и Lebedeva, автоматически оценивает уровень сложности текста для изучающих русский язык, что может служить вспомогательным средством при подборе текстового материала для грамматических заданий, привязанных к определённому уровню владения языком [8]. Наконец, Трапицын, Поморцева и Куновски систематизируют диагностические возможности и типологию педагогических тестов, формулируя критерии оценки качества тестовой системы в целом, — эта рамка задаёт контекст, в который должен вписываться и специализированный банк грамматических заданий [19].

Обобщение проанализированной литературы в соответствии с принципами, изложенными в разделе «Методы», позволяет сформулировать поэтапную модель, включающую шесть последовательных этапов. Её научная новизна заключается не в отдельных процедурах — каждая из шести стадий по отдельности хорошо известна современной тестологии, — а в их целенаправленном соединении применительно к объекту, для которого подобное соединение прежде не описывалось: компетентностно-операционализированная грамматика русского языка. В рассмотренных выше работах IRT-калибровка применяется либо к интегральным речевым умениям (чтение — [2]), либо к инструментам вне образовательного контекста преподавания языка как учебного предмета (клиническая диагностика речевого развития — [14], [18]). Предлагаемая модель, в отличие от них, явным образом связывает этап спецификации конструкта с принципами компетентностного подхода и только затем передаёт операционализированный конструкт психометрическому аппарату IRT; именно последовательность «компетентностная спецификация конструкта → аргументационная валидация → психометрическая калибровка», а не сама калибровка как таковая, составляет предлагаемый вклад статьи.

 © Авторы статьи / Authors of the article

Первый этап — спецификация конструкта: определение того, какие из четырёх компонентов грамматической компетенции войдут в конкретный банк, и их соотнесение с уровнем владения языком и с коммуникативными задачами, а не с изолированными правилами.

Второй этап — разработка заданий и их экспертная оценка. Формулировки заданий строятся на аутентичном языковом материале; содержательная валидность обеспечивается независимой экспертной оценкой соответствия задания заявленному конструкту, а при значительном объёме банка часть заданий может конструироваться автоматически, по модели, апробированной Белым с соавторами для лексико-грамматических заданий РКИ [16, С. 108–118].

Третий этап — пилотирование. Задания предъявляются ограниченной выборке испытуемых; при большом объёме банка целесообразен сбалансированный неполный блочный план, опробованный, в частности, при построении банка заданий по читательской грамотности [2], что позволяет не перегружать одного испытуемого всем массивом заданий.

Четвёртый этап — проверка размерности конструкта и IRT-калибровка. Прежде чем применять единую шкалу трудности, проверяется, образуют ли выделенные в компоненты единый континуум или требуют раздельного измерения; затем оцениваются параметры трудности и дифференцирующей способности каждого задания в рамках модели, соответствующей установленной размерности.

Пятый этап — анализ дифференцирующего функционирования заданий (DIF). Проверяется, не работает ли задание по-разному для сопоставимых по уровню подготовки групп испытуемых — в частности, для носителей языков разного морфологического строя, что применительно к вьетнамскому контексту обсуждается отдельно в разделе «Обсуждение», — по аналогии с процедурой, применённой Logvinenko с соавторами к тесту «Синонимы» [9, С. 35–63].

Шестой этап — формирование банка и накопительное расширение шкалы. Прошедшие калибровку задания включаются в общий банк с сохранением единой шкалы трудности, что делает возможным конструирование новых эквивалентных вариантов теста и постепенное пополнение банка без повторной калибровки уже включённых заданий.

Обсуждение

Предложенная модель носит обобщающий характер и требует эмпирической проверки на конкретном материале. Ограничение настоящей работы состоит в том, что она синтезирует принципы, апробированные на смежном материале — тестах по чтению, по интеллекту, по клинической диагностике речевого развития, — но не на материале собственно грамматики русского языка как учебного предмета. Это ограничение отражает объективное состояние области: как показывает количественный анализ Aryadoust с соавторами, оценивание грамматики остаётся периферийным направлением исследований по сравнению с оцениванием других видов речевой деятельности [1]. Второе ограничение связано с тем, что предложенная модель ориентирована на достаточно крупный банк заданий и предполагает существенные организационные ресурсы, прежде всего доступ к пилотной выборке достаточного объёма, без которой устойчивая IRT-калибровка невозможна. Третье ограничение — методологическое: поскольку работа представляет собой нарративный, а не систематический обзор в строгом смысле, полнота охвата литературы не гарантирована протоколом PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses — стандарт отчётности о систематических обзорах и метаанализах) или аналогичным ему, а определяется границами предпринятого библиографического поиска.

Учитывая гипотетический характер работы, необходимо чётко разграничить, какие её положения опираются на эмпирически подтверждённые в цитируемой литературе процедуры, а какие остаются предположением автора, требующим отдельной проверки. Эмпирически подтверждены на смежном материале: применимость сбалансированного неполного блочного плана при пилотировании крупных банков заданий [2]; работоспособность DIF-анализа для выявления заданий, некорректно функционирующих в разных группах испытуемых [9]; методологическое преимущество параллельного, а не переводного конструирования заданий при локализации теста для другого языка [5]. Остаются не проверенными эмпирически именно на материале русской грамматики: предложенная в четырёхкомпонентной операционализации конструкта; принадлежность выделенных компонентов единой либо нескольким латентным шкалам; устойчивость модели Раша для заданий предложенного в 3.1 формата; и ожидаемая величина DIF-эффекта для конкретных групп испытуемых, в том числе вьетнамоязычных.

Отдельного обсуждения требует соотношение грамматической компетенции с более широкой коммуникативной компетенцией. Если, в соответствии с принципом, изложенным в 3.1, грамматическое задание строится на аутентичном коммуникативном материале, результат его выполнения потенциально зависит не только от владения грамматической системой, но и от объёма словарного запаса испытуемого, скорости чтения и общей стратегии выполнения тестового задания. Полностью исключить это влияние не позволяет ни один формат компетентностно-ориентированного задания. Минимизировать его позволяют, во-первых, контроль лексической сложности контекста — используемая в задании лексика не должна превышать уровень, на котором проверяется грамматика; во-вторых, конструирование дистракторов, различающихся исключительно грамматически, а не по значению, как показано на примере в 3.1; в-третьих, сама процедура IRT-калибровки, позволяющая на четвёртом этапе модели (3.5) обнаружить задания с аномально низкой дискриминативностью, которая нередко указывает на неучтённое влияние фактора, нерелевантного заявленному конструкту.

Применительно к вьетнамскому контексту предлагаемая модель приобретает дополнительную специфику. Вьетнамский язык принадлежит к изолирующему строю: он не имеет словоизменения по падежам, роду и числу и не обладает глагольным спряжением. Вследствие этого категории, составляющие ядро морфологического и морфосинтаксического компонентов конструкта, обозначенных в 3.1, — падежная система, согласование, глагольный вид, — как правило, оказываются для вьетнамоязычных учащихся системно более трудными и порождают иной по типу набор ошибок, чем у учащихся, чей первый язык относится к флективному строю. Это делает вероятным

 © Авторы статьи / Authors of the article

дифференцирующее функционирование части заданий именно по признаку изолирующего либо флективного строя родного языка, что необходимо проверить на пятом этапе модели (3.5, DIF-анализ). Кроме того, содержание выпускного экзамена по русскому языку во Вьетнаме регламентируется государственными требованиями — циркуляром Министерства образования и обучения Вьетнама № 19/2021/TT-BGDĐT, — которые задают распределение заданий по уровням когнитивной сложности. Сопоставление этого нормативно заданного распределения с эмпирически установленной средствами IRT трудностью заданий, по аналогии с процедурой, применённой к тесту VSTEP (Vietnamese Standardized Test of English Proficiency — вьетнамский стандартизированный тест владения английским языком) [12], могло бы стать первым этапом эмпирической апробации предложенной модели именно в вьетнамском контексте.

Заключение

Проведённый анализ показывает, что разработка банка тестовых заданий компетентностной направленности для оценки грамматики русского языка должна опираться на три группы источников: теорию компетентностного подхода, операционализированного содержание конструкта; аргументационную модель валидности, задающую логику доказательства пригодности теста; и психометрический аппарат IRT и модели Раша, обеспечивающий техническую калибровку заданий и накопительное расширение банка. Отечественный опыт — от классических работ по педагогической тестологии до современных исследований по локализации психометрических инструментов и автоматической генерации заданий — предоставляет для этого достаточную методологическую базу, тогда как специализированные исследования, посвящённые именно оцениванию грамматики русского языка средствами современной психометрики, остаются немногочисленными.

Устранение этого пробела требует эмпирической апробации предложенной шестиэтапной модели на реальном экзаменационном материале. Предложенная в статье шестиэтапная модель и четырёхкомпонентная операционализация конструкта носят гипотетический характер и представляют собой не завершённый инструмент, а программу дальнейшего эмпирического исследования; устранение обозначенного пробела требует апробации этой программы на реальном экзаменационном материале, в том числе на материале выпускного экзамена по русскому языку во Вьетнаме.

Финансирование

Это исследование финансируется Фондом науки и технологий Образовательного университета Хошимина в рамках гранта № CS.2025.19.60.

Конфликт интересов

Не указан.

Рецензия

Сообщество рецензентов журнала Russian Linguistic Bulletin.

DOI: https://doi.org/10.60797/RULB.2026.83.1.1

Funding

This research is funded by Ho Chi Minh City University of Education Foundation for Science and Technology under grant number CS.2025.19.60.

Conflict of Interest

None declared.

Review

Community of Reviewers of the Russian Linguistic Bulletin.

DOI: https://doi.org/10.60797/RULB.2026.83.1.1

Список литературы / References

1. Aryadoust V. An extensive knowledge mapping review of measurement and validity in language assessment and SLA research / V. Aryadoust, A. Zakaria, M.H. Lim et al. // Frontiers in Psychology. — 2020. — 11. — DOI: 10.3389/fpsyg.2020.01941

2. Chen Q. Construction of a reading literacy test item bank for fourth graders based on item response theory / Q. Chen, H. Zheng, H. Fan et al. // Frontiers in Psychology. — 2023. — 14. — DOI: 10.3389/fpsyg.2023.1103853

3. Fan J. Assessing speaking proficiency: A narrative review of speaking assessment research within the argument-based validation framework / J. Fan, X. Yan // Frontiers in Psychology. — 2020. — 11. — DOI: 10.3389/fpsyg.2020.00330

4. Im G.H. Critical review of validation models and practices in language testing: Their limitations and future directions for validation research / G.H. Im, D. Shin, L. Cheng // Language Testing in Asia. — 2019. — 9. — DOI: 10.1186/s40468-019-0089-4

5. Ivanova A. Constructing a Russian-language version of the international early reading assessment tool / A. Ivanova, K. Kardanova-Biryukova // Voprosy obrazovaniya / Educational Studies Moscow. — 2019. — 4. — P. 93–115. — DOI: 10.17323/1814-9545-2019-4-93-115

6. Карданова Е.Ю. Психометрические исследования: современные методы и новые возможности для образования / Е.Ю. Карданова, А.Е. Иванова // Вопросы образования. — 2023. — 3. — DOI: 10.17323/vo-2023-17951

7. Kulik A.A. Competency-based approach and competencies in higher education: A theoretical review / A.A. Kulik, N.V. Ippolitova, A.E. Egorova et al. // Propósitos y Representaciones. — 2020. — 8. — DOI: 10.20511/pyr2020.v8nspe2.645

8. Laposhina A. Textometr: An online tool for automated complexity level assessment of texts for Russian language learners / A. Laposhina, M. Lebedeva // Russian Language Studies. — 2021. — 19(3). — P. 331–345. — DOI: 10.22363/2618-8163-2021-19-3-331-345

9. Logvinenko T.I. Similar or different? An item response theory analysis of the Synonyms Test in adults with and without a history of institutionalization / T.I. Logvinenko, O.I. Talantseva, E.M. Volokhova et al. // Modelling and Data Analysis. — 2020. — 10(1). — P. 35–63. — DOI: 10.17759/mda.2020100102

10. Makulova A. Theory and practice of competency-based approach in education / A. Makulova, G.M. Alimzhanova, Z.M. Bekturganova et al. // International Education Studies. — 2015. — 8(8). — P. 183–192. — DOI: 10.5539/ies.v8n8p183

 © Авторы статьи / Authors of the article

11. Min S. A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability / S. Min, V. Aryadoust // Studies in Educational Evaluation. — 2021. — 68. — DOI: 10.1016/j.stueduc.2021.100963

12. Nguyen D.T.B. Item response theory analysis of the reading section in a mock VSTEP level 3–5 test / D.T.B. Nguyen, T.H. Nguyen, T.V.N. Bui et al. // HNUE Journal of Science: Educational Science. — 2025. — 70(4). — P. 13–21. — DOI: 10.18173/2354-1075.2025-0064

13. O'Keeffe A. The English Grammar Profile of learner competence / A. O'Keeffe, G. Mark // International Journal of Corpus Linguistics. — 2017. — 22(4). — DOI: 10.1075/ijcl.14086.oke

14. Talantseva O.I. Psychometric properties of the Preschool Language Scales, Fifth Edition (PLS-5) in Russian-speaking children: A classical and item response theory study / O.I. Talantseva, I.O. An, M.A. Zhukova et al. // Clinical Psychology and Special Education. — 2022. — 11(2). — P. 174–195. — DOI: 10.17759/cpse.2022110211

15. Алимова М.В. Единое вступительное испытание по русскому языку для иностранных абитуриентов: новая образовательная реальность / М.В. Алимова, М.Н. Куновски, А.С. Иванова и др. // Русский язык: исследования, тестирование, практика. — 2025. — 11(1). — С. 4–17. — DOI: 10.22363/3034-2090-2025-11-1-4-17

16. Белый А.В. Автоматическая генерация лексико-грамматических заданий по русскому языку как иностранному с помощью предсказывающих языковых моделей / А.В. Белый, О.А. Митрофанова, Н.А. Дубинина // Мир русского слова. — 2024. — 2. — С. 108–118. — DOI: 10.21638/spbu30.2023.212

17. Дубинина Н.А. Уровни тестирования по русскому языку как иностранному в аспекте возрастной специфики школьников / Н.А. Дубинина, Д.В. Птюшкин // Русистика. — 2021. — 19(2). — С. 222–234. — DOI: 10.22363/2618-8163-2021-19-2-222-234

18. Прихода Н.А. Оценка развития русского языка как стандартизированная методика оценки коммуникативной функции у детей 3–9 лет / Н.А. Прихода // Психологическая наука и образование. — 2016. — 21(3). — С. 25–33. — DOI: 10.17759/pse.2016210304

19. Трапицын С.Ю. Диагностические возможности и области применения педагогических тестов как инструмента оценки качества образования / С.Ю. Трапицын, Н.В. Поморцева, М.Н. Куновеки // Русский язык: исследования, тестирование, практика. — 2025. — 11(2). — С. 86–96. — DOI: 10.22363/3034-2090-2025-11-2

Список литературы на английском языке / References in English

1. Aryadoust V. An extensive knowledge mapping review of measurement and validity in language assessment and SLA research / V. Aryadoust, A. Zakaria, M.H. Lim et al. // Frontiers in Psychology. — 2020. — 11. — DOI: 10.3389/fpsyg.2020.01941

2. Chen Q. Construction of a reading literacy test item bank for fourth graders based on item response theory / Q. Chen, H. Zheng, H. Fan et al. // Frontiers in Psychology. — 2023. — 14. — DOI: 10.3389/fpsyg.2023.1103853

3. Fan J. Assessing speaking proficiency: A narrative review of speaking assessment research within the argument-based validation framework / J. Fan, X. Yan // Frontiers in Psychology. — 2020. — 11. — DOI: 10.3389/fpsyg.2020.00330

4. Im G.H. Critical review of validation models and practices in language testing: Their limitations and future directions for validation research / G.H. Im, D. Shin, L. Cheng // Language Testing in Asia. — 2019. — 9. — DOI: 10.1186/s40468-019-0089-4

5. Ivanova A. Constructing a Russian-language version of the international early reading assessment tool / A. Ivanova, K. Kardanova-Biryukova // Voprosy obrazovaniya / Educational Studies Moscow. — 2019. — 4. — P. 93–115. — DOI: 10.17323/1814-9545-2019-4-93-115

6. Kardanova E.Yu. Psixometricheskie issledovaniya: sovremenny'e metody' i novy'e vozmozhnosti dlya obrazovaniya [Psychometric Research: Modern Methods and New Opportunities for Education] / E.Yu. Kardanova, A.E. Ivanova // Educational Studies Moscow. — 2023. — 3. — DOI: 10.17323/vo-2023-17951 [in Russian]

7. Kulik A.A. Competency-based approach and competencies in higher education: A theoretical review / A.A. Kulik, N.V. Ippolitova, A.E. Egorova et al. // Propósitos y Representaciones. — 2020. — 8. — DOI: 10.20511/pyr2020.v8nspe2.645

8. Laposhina A. Textometr: An online tool for automated complexity level assessment of texts for Russian language learners / A. Laposhina, M. Lebedeva // Russian Language Studies. — 2021. — 19(3). — P. 331–345. — DOI: 10.22363/2618-8163-2021-19-3-331-345

9. Logvinenko T.I. Similar or different? An item response theory analysis of the Synonyms Test in adults with and without a history of institutionalization / T.I. Logvinenko, O.I. Talantseva, E.M. Volokhova et al. // Modelling and Data Analysis. — 2020. — 10(1). — P. 35–63. — DOI: 10.17759/mda.2020100102

10. Makulova A. Theory and practice of competency-based approach in education / A. Makulova, G.M. Alimzhanova, Z.M. Bekturganova et al. // International Education Studies. — 2015. — 8(8). — P. 183–192. — DOI: 10.5539/ies.v8n8p183

11. Min S. A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability / S. Min, V. Aryadoust // Studies in Educational Evaluation. — 2021. — 68. — DOI: 10.1016/j.stueduc.2021.100963

12. Nguyen D.T.B. Item response theory analysis of the reading section in a mock VSTEP level 3–5 test / D.T.B. Nguyen, T.H. Nguyen, T.V.N. Bui et al. // HNUE Journal of Science: Educational Science. — 2025. — 70(4). — P. 13–21. — DOI: 10.18173/2354-1075.2025-0064

13. O'Keeffe A. The English Grammar Profile of learner competence / A. O'Keeffe, G. Mark // International Journal of Corpus Linguistics. — 2017. — 22(4). — DOI: 10.1075/ijcl.14086.oke

 © Авторы статьи / Authors of the article

14. Talantseva O.I. Psychometric properties of the Preschool Language Scales, Fifth Edition (PLS-5) in Russian-speaking children: A classical and item response theory study / O.I. Talantseva, I.O. An, M.A. Zhukova et al. // Clinical Psychology and Special Education. — 2022. — 11(2). — P. 174–195. — DOI: 10.17759/cpse.2022110211

15. Alimova M.V. Edinoe vstupitel'noe ispy'tanie po russkomu yazy'ku dlya inostranny'x abiturientov: novaya obrazovatel'naya real'nost' [The unified entrance test in Russian for international applicants:a new educational reality] / M.V. Alimova, M.N. Kunovski, A.S. Ivanova et al. // Russian Language: Research, Testing and Practice. — 2025. — 11(1). — P. 4–17. — DOI: 10.22363/3034-2090-2025-11-1-4-17 [in Russian]

16. Bely'j A.V. Avtomaticheskaya generaciya leksiko-grammaticheskix zadanij po russkomu yazy'ku kak inostrannomu s pomoshh'yu predskazy'vayushhix yazy'kovy'x modelej [Automatic Generation of Lexico-Grammatical Tests for Russian as a Foreign Language Using Predictive Language Models] / A.V. Bely'j, O.A. Mitrofanova, N.A. Dubinina // The World of Russian Word. — 2024. — 2. — P. 108–118. — DOI: 10.21638/spbu30.2023.212 [in Russian]

17. Dubinina N.A. Urovni testirovaniya po russkomu yazy'ku kak inostrannomu v aspekte vozrastnoj specifiki shkol'nikov [Levels of tests of Russian as a foreign language in terms of school age characteristics] / N.A. Dubinina, D.V. Ptyushkin // Russian Language Studies. — 2021. — 19(2). — P. 222–234. — DOI: 10.22363/2618-8163-2021-19-2-222-234 [in Russian]

18. Prixoda N.A. Ocenka razvitiya russkogo yazy'ka kak standartizirovannaya metodika ocenki kommunikativnoj funkcii u detej 3–9 let [Russian language development assessment as a standardized technique for assessing communicative function in children aged 3–9 years] / N.A. Prixoda // Psychological Science and Education. — 2016. — 21(3). — P. 25–33. — DOI: 10.17759/pse.2016210304 [in Russian]

19. Trapicy'n S.Yu. Diagnosticheskie vozmozhnosti i oblasti primeneniya pedagogicheskix testov kak instrumenta ocenki kachestva obrazovaniya [Pedagogical test as a tool for assessing the quality of education] / S.Yu. Trapicy'n, N.V. Pomorceva, M.N. Kunoveki // Russian language: research, testing and practice. — 2025. — 11(2). — P. 86–96. — DOI: 10.22363/3034-2090-2025-11-2 [in Russian]