

Педагогические науки / Pedagogical Science  
Оригинальная статья / Original Article  
УДК 371.263  
DOI: 10.31161/1995-0659-2018-12-2-43-50

## Аргументации и доказательства надежности оценок компетенций студентов

© 2018 Ефремова Н. Ф.

Донской государственный технический университет,  
Ростов-на-Дону, Россия; e-mail: nefremova61@donstu.ru

**РЕЗЮМЕ.** Целью данного исследования является новый подход к обеспечению надежного оценивания компетенций обучающихся. **Метод.** Анализ возможностей нового метода проектирования педагогических измерителей на основе паттерн-дизайна Evidence-Centered Design (ECD), обеспечивающего доказательства и аргументацию оценок испытуемых. Метод ECD нашел широкое применение в зарубежном образовании, но мало известен в отечественной системе образования. **Результаты.** Предложен алгоритм проектирования средств оценки подготовленности студентов на основе ECD путем сопряжения модели студента (знаний и компетенций) и модели значимого задания с набором доказательств и аргументации наблюдаемых результатов деятельности обучающегося. Использование доказательного дизайна при разработке оценочных средств создает надежную основу для установления связи между оценкой и корректировкой образовательной политики. **Выводы.** Для эффективной организации образовательного процесса существенную роль играет обратная связь, обеспечиваемая достоверностью оценок, что в свою очередь требует разработки надежного оценочного инструментария. Проблема связана с тем, что конструирование оценочных средств вызывает сложности у преподавателей. Такой подход может успешно применяться как при оценке знаний обучающихся, так и при оценке компетенций, обеспечивая повышение надежности результатов, связь между процессами разработки системы оценки и обеспечением условий деятельности, чтобы с применением знаний студенты показали, что и как могут делать.

**Ключевые слова:** оценка, компетенции, доказательный дизайн, оценочное средство, Evidence-Centered Design.

**Формат цитирования:** Ефремова Н. Ф. Аргументации и доказательства надежности оценок компетенций студентов // Известия Дагестанского государственного педагогического университета. Психолого-педагогические науки. 2018. Т. 12. № 2. С. 43-50. DOI: 10.31161/1995-0659-2018-12-2-43-50

## Arguments and Evidence of Reliability Assessment of Students' Competences

© 2018 Nadezhda F. Efremova

Don State Technical University,  
Rostov-on-Don, Russia; e-mail: nefremova61@donstu.ru

**ABSTRACT.** The aim of this study is a new approach to ensuring the reliable evaluation of the students' competencies. **Method.** Analysis of the possibilities of the new method of designing the pedagogical meters based on the pattern-design Evidence-Centered Design (ECD), providing the evidence and argumentation of assessments of subjects. The ECD method has found wide application in foreign education, but is little known in the domestic education system. **Results.** The algorithm for designing students' assessment tools based on ECD is proposed by connecting the model of the student (knowledge and competences) and the model of a significant task with a set of evidence and argumentation of the observed results of the student's activity. The use of evidence-based design in the development of evaluation tools creates a reliable basis for

establishing a link between the assessment and adjustment of educational policies. **Conclusions.** For effective organization of the educational process a feedback plays an important role, which is ensured by the reliability of estimates, which in turn requires the development of a reliable evaluation tool. The problem is related to the fact that the construction of evaluation tools causes difficulties for teachers. This approach can be successfully applied both in the evaluation of students' knowledge and in the evaluation of competencies, ensuring the reliability of results, the relationship between the processes of developing the assessment system and ensuring the conditions of activity, so that with the use of knowledge, students show what and how they can do.

**Keywords:** assessment, competence, evidence design, evaluative tool, Evidence-Centered Design.

---

**For citation:** Efremova N. N. Arguments and Evidence of Reliability Assessment of Students' Competences. Dagestan State Pedagogical University. Journal. Psychological and Pedagogical Sciences. 2018. Vol. 12. No. 2. Pp. 43-50. DOI: 10.31161/1995-0659-2018-12-2-43-50 (In Russian)

---

### Введение

Перед каждым преподавателем при планировании и реализации рабочей программы дисциплины встает ряд вопросов. Какой комплекс знаний, навыков или других атрибутов следует оценивать, какие оценочные средства использовать для выявления латентных (когнитивных) и глубоко латентных (компетентностных) результатов обучения. Какие действия или выступления могут раскрыть эти результаты? Какие задачи или ситуации должны вызывать такое поведение испытуемых, чтобы результат обучения стал явным? Получение ответов на эти вопросы, как показывает практика, сложная задача, а ее понимание обеспечивает разработчику оценочных средств выбор необходимых ситуаций и контролируемых рубрик, обоснование критериев и шкал оценивания результатов контроля.

Согласно С. Мессикю (Messick S., 1994), оценка – это процесс рассуждения из конкретных действий или продуктов деятельности, которые люди могут сделать, сказать или предоставить, чтобы можно было сделать обоснованные выводы об их знаниях, навыках и способностях [5]. Он отмечает, что характер конструкта задания направляет выбор или построение соответствующих задач, определяет рациональное применение критериев и показателей, основанных на конструкции задания. Природа конструкта обуславливает создание релевантных заданий, а также разработку основанной на нем системы подсчета баллов и оценочных рубрик. Особенно полезно концептуализировать оценку как процесс рассуждения из доказательств и аргументации в случае неопределенности, следующей из выполнения студентом сложного компетентностно-

ориентированного задания. Нужны такие методы оценки, которые позволили бы интегрировать весь набор атрибутов, подтверждающих действительный уровень подготовленности испытуемых путем использования доказательств и аргументации проявления трудно измеримых латентных характеристик испытуемых. Это особенно актуально при оценке навыков 21 века как сложных результатов обучения. Разработка психометрических и технологических инструментов для этих целей требует новаторских подходов, нужны методы, позволяющие интегрировать большое число качественных и количественных данных для рассуждения, доказательного вывода и аргументации результирующей оценки.

**Целью** исследования явилось изучение доказательного дизайна оценочных средств для надежного выявления сформированности компетенций студентов.

Одним из таких **методов**, получивших распространение в зарубежной образовательной практике, стал предложенный Р. Мислеви (Mislevy R., 2003) метод доказательной аргументации Evidence Centered Design (ECD) [6; 7], опирающийся на конструкцию задания, предложенную С. Мессиком. В ECD оценка определяется как цепочка рассуждений, которая связывает доказательства с утверждениями и опровержениями об уровне подготовленности обучающихся. Все принципы, структуры и инструменты ECD следуют из этой основной концептуализации. Отличительной чертой ECD является разработка системы оценки путем формулирования цепочки рассуждений, связывающих доказательства выполнения конструкций заданий в

последовательных этапах их выполнения в процессе работы со сложными комплексными заданиями, что позволяет получать основания для доказательных рассуждений относительно обоснованности оценки.

### Методика паттерна в доказательной аргументации

В своей разработке ECD Р. Мислеви использовал также идеи и конструкции С. Тулмина (Toulmin S. 1958) о структуре аргументов в виде универсального паттерна (многократно проверенного шаблона) [9], показывающего, что данные становятся свидетельством только тогда, когда их значение для выводов однозначно установлено. По логике Тулмина все эти элементы аргументации взаимодействуют согласно общей схеме (рис. 1).

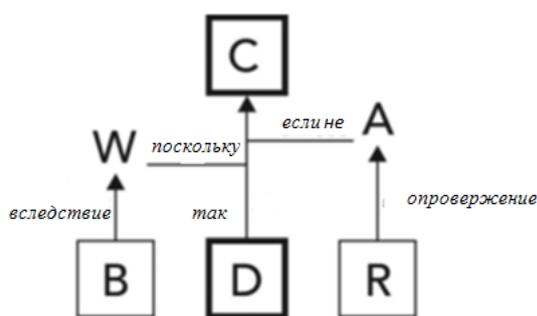


Рис. 1. Общая схема аргументации по Тулмину

Здесь D – данные (data); C – утверждение (claim); W – основание (warrant); B – поддержка (backing); R – опровержение (rebuttal); A – альтернативное объяснение (explanations). Опровержение или контраргумент в паттерне указывает на ограниченность утверждения. Альтернативное утверждение как набор слов (вероятно, возможно, поскольку, невозможно, безусловно, предположительно, всегда и др.) ограничивает достоверность утверждения и подчеркивает, при каких условиях утверждение может быть верным. Допущения и предположения служат основанием тому, как и почему из имеющихся данных следует именно сделанное утверждение об оценке. Поддержка усиливает основания за счет дополнительных данных (контрольные вопросы, результаты анкетирования и др.).

Диаграмма Тулмина адаптирует представления аргументов оценки на основе требований к выполнению заданий и возможности организовать элементы аргумента из информации о процессе получения ответов на задание. Разработка таких заданий основывается на специализированных знаниях в области педагогических измерений и использовании поддерживающих технологий, таких как цифровые симуляции, деловые и имитационные игры, решение кейсов, выполнение индивидуальных или групповых и многопрофильных проектов и др. Использование диаграммы Тулмина позволяет охватить важные элементы необходимой информации в форме, которая облегчает рассуждения и применения этой информации для аргументации оценки в соответствии с той целью, которая поставлена для контроля. Чем лучше задания отображают цели проверки, тем эффективнее процесс в контексте оценки [3].

Для применения новой модели и построения надежной аргументации оценки необходимы шаблоны проектирования оцениваемых достижений студента, психометрические модели оцениваемых характеристик и шаблоны заданий с наборами характерных задач и процессов их выполнения (рис. 2). Шаблон проектирования задает стандарт, на основе которого оцениваются все проектные решения, в том числе, связанные с заданием и спецификацией задачи, дизайном и администрированием тестовой или иной формы, элементами задач в системе оценки, а также с обработкой результатов и составлением отчета. Язык шаблонов впервые заявлен в работах С. Александра (Alexander C., 1977). Используя его, Мислеви предложил многоуровневый подход Evidence-Centered Design, основанный на моделях доказательного дизайна оценочного средства. Преимущество ECD в разработке оценки, в которой каждый шаг тщательно строится на предыдущих этапах, чтобы поддержать спецификацию и выполнение следующих шагов для обеспечения общего аргумента оценки.

В ECD последовательно выстраиваются модель компетенций обучающегося, модель сбора свидетельств (доказательств), модель оценочного средства и разработки его дизайна, описание сценария оценочного

процесса (рис. 3). Такой подход может применяться к разработке любой оценки, где априорное определение конструкций и связанные с ними переменные имеют смысл. Именно в этих контекстах значение, сложность и взаимосвязь решений, которые необходимо принять в отношении дизайна оценки, является самым сложным. Более того, поскольку такие оценки являются

трудоемкими, существует настоятельная потребность повторного использования шаблонов дизайна, структура которых может систематически связываться с конкретными наборами утверждений, которые являются естественным мостом между моделью студента и моделью конструируемого задания. Таких паттернов может быть предложено несколько.

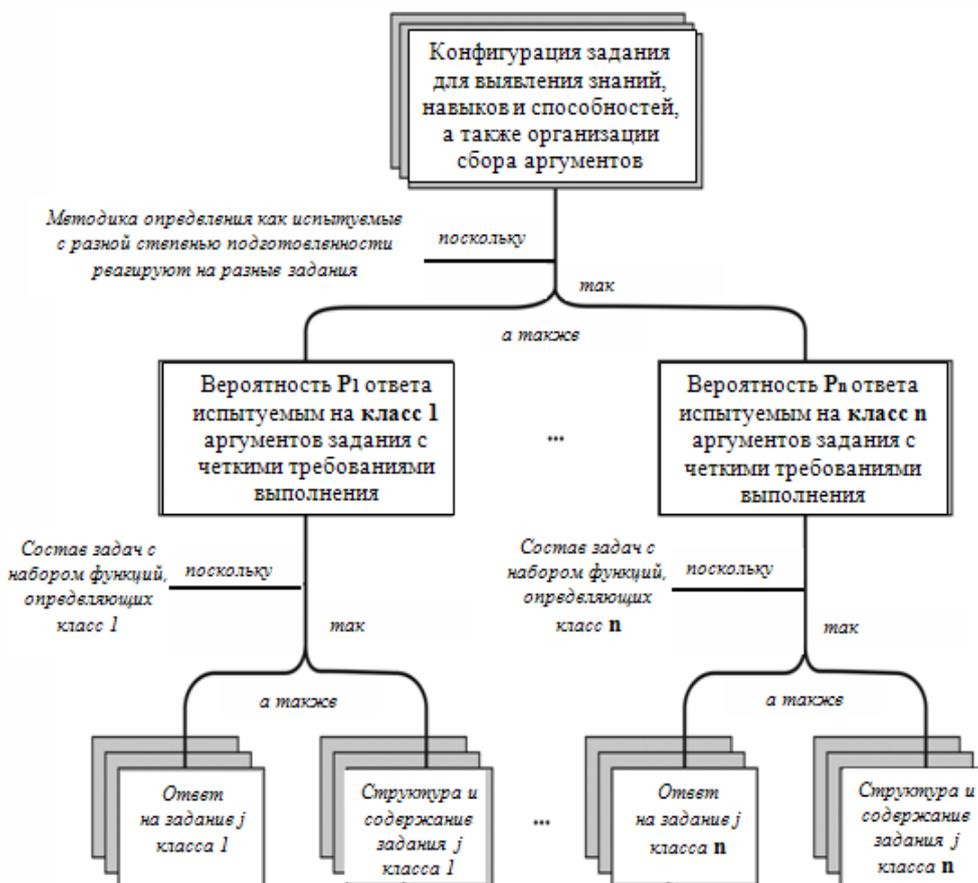


Рис. 2. Расширенная диаграмма Тулмина в контексте оценки



Рис. 3. Структура аргументов в оценке ECD

Данные о выполнении обучающимся задания становятся свидетельством его подготовленности только тогда, когда их значение для выводов установлено. Одни и те же данные могут быть хорошим доказательством для одного вывода, но плохими для доказательства другого (Schum, 1994). При аргументации оценки необходимо тщательно рассмотреть альтернативные объяснения данных, например, когда студент предоставляет неверные данные на проблему не из-за недостатка знаний математики, а из-за ограниченности понимания языка задания.

Более сложный и взаимосвязанный сбор доказательств и их интеграции полезны и необходимы для уточнения оценки. Построенная на паттерне оценка воплощает все эти основные этапы организации оценочного процесса и доказательства надежности результатов.

Для демонстрации принципа аргументации рассмотрим пример рубрик доказательств для задания множественного выбора ответов на тестовое задание (рис. 4). Новая методология оценки навыков 21-го века должна задать инструменты отслеживания работы студентов в группах и индивидуально, выявлять рефлексивность в самооценке и оценке достижений других,

оценки глобальных компетенций и множества других важных характеристик согласно заданной модели будущего специалиста [4; 8].

Алгоритм Evidence-Centered Design задает все проектные решения по разработке оценочного средства, спецификацию задачи, элементы необходимых заданий и аргументацию качества их выполнения, свидетельства для обоснования оценки студента, форматы выходных данных (рис. 5). Основные элементы паттерн-дизайна: модель студента, модель свидетельства (доказательства), модель задания.

Сопряжение модели студента, модели свидетельств и модели задания (выбор при наличии проверенных шаблонов или построение оригинального паттерн-дизайна), а также возможности перехода от баллов к выводам об уровне сформированности компетенций на уровне шкалах обеспечивает концептуальная рамка оценивания (модель сборки).

Модель студента описывает конструкт исследуемых характеристик: набор латентных переменных (количество аргументов, структура конструкта), виды деятельности, причинно-следственные связи между

ситуацией и деятельностью студента. Модель студента определяет модель задачи и задает, что и как должно быть выявлено, как представить ответы на задания и как

совместить наблюдения, доказательства и свидетельства с баллами результирующей оценки.

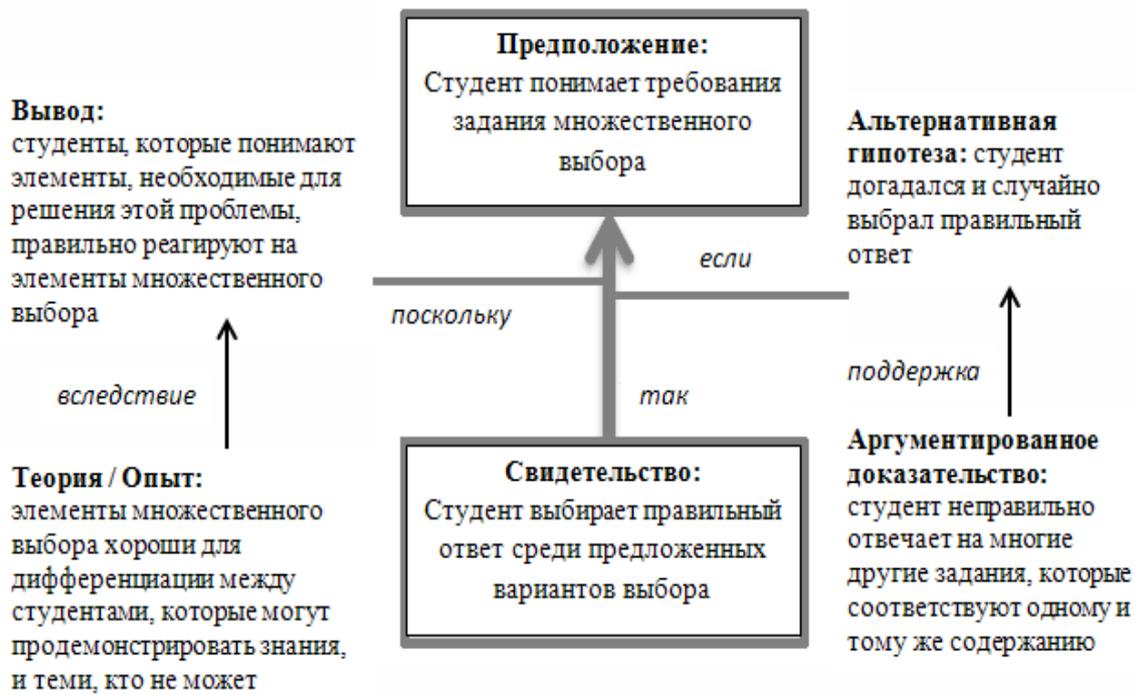


Рис. 4. Пример паттерна для задания множественного выбора

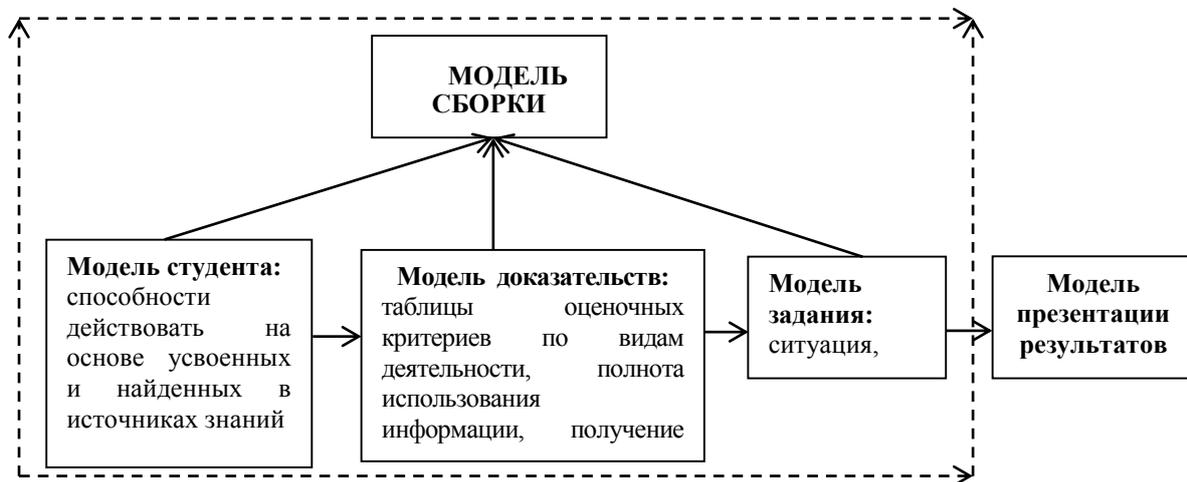


Рис. 5. Модель сборки оценочного средства (концептуальная рамка оценивания)

Модель задания: создание конструкта, отражающего структуру аргументов оценки (валидные ситуации и сценарии, акцент на наиболее важном, ключевые аргументы, принципы отбора заданий, спецификация, планируемые действия испытуемого), связь оцениваемых аргументов с функциями задания, обеспечивающие проявление компетенций при выполнении

запланированной в задании деятельности, симуляционные среды.

Модель свидетельств: оценочный компонент задания, баллы по выполнению задач, общий балл по всему заданию, продукты деятельности, сбор свидетельств и аргументов по результату выполнения задания и/или оценке продукта деятельности, правила оценки процессов или продуктов

деятельности, отображение результатов на уровневой шкале. В ECD планируются доказательства, опровержения и утверждения об уровне достижений студентов при выполнении сложных заданий, позволяющих отражать, что знает и что может сделать студент после освоения дисциплин / модулей / практик.

Анализ содержания дисциплины и поиск ситуаций для проектирования оценки являются необходимым предварительным условием для статистического моделирования оценочных средств и процессов, а формами организации деятельности студентов и оценки учебных достижений могут быть компетентностно-ориентированные задачи, решения кейсов, эпистемические игры, выполнение индивидуальных и групповых проектов, презентации, эссе и др. [1; 2]. Желаемый результат оценки может дать соблюдение последовательности этапов проектирования оценочного средства ECD:

- определение цели оценки и формирование многомерного конструкта задания;
- операционализация конструкта (представление элементов задания в удобном виде для использования и обработки результатов);
- выбор вида паттерна (шаблона) проектирования оценочного средства;
- разработка моделей студента (уровней компетенций и их профилей в терминах дескрипторов), свидетельств (как измерить или оценить) и оценочного средства (индикаторы и критерии выполнения задания);
- нахождение профессионально интересных ситуаций для конструирования сложных (комплексных) компетентностно-ориентированных заданий;
- разработка дизайна оценочного средства (каких и сколько заданий необходимо, какие предполагаемые решения могут дать студенты, спецификация);
- разработка инструкций к выполнению заданий и бланков ответов (в каком виде обучающийся дает ответ и где его фиксирует);
- обоснование шкалы оценивания компетенций и критериев обоснования достижений уровней, форм предъявления

результатов (протоколов, графиков, диаграмм и др.);

- создание комфортных условий взаимодействия студента с инструментами оценки при выполнении заданий, обеспечение условий, близких к реальным в профессиональной среде;

- проведение оценочного процесса и обеспечение его репрезентации, наблюдение и сбор свидетельств (систематический процесс, связывающий цели, задачи, обучение, оценочный процесс и данные результатов оценивания);

- обработка результатов с учетом доказательств, опровержений и аргументации свидетельств, оценка измеряемых конструктов в баллах;

- отображение полученных данных на уровневой шкале для заключения о сформированности компетенций.

#### **Заключение**

Используя ECD можно попытаться ответить на ряд вопросов обеспечения надежности оценок. Отражает ли содержание оценочного средства целевую установку проверки достижений студентов? Реагируют ли студенты на заложенные в заданиях когнитивные и компетентностные процессы и действия? Отражает ли процесс подсчета баллов способности взаимодействия студентов с заданием? Можно ли использовать оценки доказательного дизайна для прогнозирования результата обучения и повышения интереса студентов к оценочному процессу? Выполняют ли студенты аналогичные оценки в других предметных областях? Можно ли обобщить результаты таких оценок через администрирование контекстов и образцов сложных заданий? Позволяет ли интерпретация таких оценок сделать справедливые и оправданные заключения об уровнях сформированности компетенций обучающихся? Какие свидетельства компетенций или знаний студента можно непосредственно наблюдать при выполнении заданий? Как структурировать ситуацию и средство оценивания, чтобы увидеть эти свидетельства? Что еще предстоит сделать для повышения надежности результатов оценивания? Как дальше ключевые аспекты оценки компетенций концептуализировать?

Аутентичная оценка компетенций студентов имеет решающее значение в подготовке специалистов. Новые подходы к

проектированию оценочных процедур, направленных на определение компетенций и навыков 21-го века, только начинают разрабатываться и развиваться. Для этих целей методология ECD может стать одной из основных для обеспечения надежности и валидности оценок при компетентностном обучении.

Предложенные шаблоны проектирования оценочных процессов полезны при создании

инновационных задач относительно целей оценки. Использование прошедших проверку временем методик в виде паттернов (проверенных шаблонов) проектирования позволит облегчить разработку оценочных средств, их повторное применение и решение проблем большой неопределенности в оценке результатов обучения студентов.

## Литература

1. Ефремова Н. Ф. Подходы к оцениванию компетенций студентов первого курса, приступающих к освоению основных образовательных программ // Вестник Донского государственного технического университета. 2010. Т. 10. № 5 (48). С. 774-782.

2. Ефремова Н. Ф. Учебные достижения как объект тестирования и показатель качества в образовании // Вопросы тестирования в образовании. 2004. № 9. С. 39-50.

3. Assessing model-based reasoning using evidence-centered design: a suite of research-based design patterns by R. J. Mislevy, G. Haertel, M. Riconscente, D. Rutstein, C. Ziker. 2017, Springer Publ. 130 p.

4. Griffin P., McGaw B., Editors E. Assessment and Teaching of 21<sup>st</sup> Century Skills. Springer Dordrecht Heidelberg. London, New York, 2014. 362 p.

5. Messick S. The interplay of evidence and consequences in the validation of performance assessments. Education Researcher, 1994. 32 (2). Pp. 13-23.

6. Mislevy R. J., Almond R. G., & Lukas J. F. (2003). A brief introduction to evidence-centered design. (ETS Research Report RR-03-16). Princeton, NJ: Educational Testing Service. 2003. 37 p.

7. Mislevy R. J., Levy R. Bayesian psychometric modeling from an evidence-centered design perspective. In: Rao, C.R., Sinharay, S. eds. Handbook of statistics. Elsevier, Amsterdam, 2007. Pp. 839-865.

8. Sukin T., Dunn J. Validity Argument for a Statewide Alternative Assessment. Addressing Inferences with Novel Approaches. National Council for Educational Measurement. Vancouver, 2012. 33 p.

9. Toulmin S. E. The uses of argument. Cambridge, Cambridge University Press Publ., 1958. 264 p.

10. Van der Linden, W. J., & Glas, C.A.W. (2010). Elements of adaptive testing. New York, Springer Publ., 2010. 462 p.

## References

1. Efremova N. F. Approaches to assessing the competence of first-year students, embarking on the development of basic educational programs. *Vestnik Donskogo gosudarstvennogo tekhnicheskogo universiteta* [Bulletin of Don State Technical University]. 2010. Vol. 10. No. 5 (48). Pp. 774-782. (In Russian)
2. Efremova N. F. Educational achievements as an object of testing and an indicator of quality in education. *Voprosy testirovaniya v obrazovanii* [Issues of Testing in Education]. 2004. No. 9. Pp. 39-50. (In Russian)
3. Assessing model-based reasoning using evidence-centered design: a suite of research-based design patterns by R. J. Mislevy, G. Haertel, M. Riconscente, D. Rutstein, C. Ziker. 2017, Springer Publ. 130 p. (In Russian)
4. Griffin P., McGaw B., Editors E. Assessment and Teaching of 21st Century Skills. Springer Dordrecht Heidelberg. London, New York, 2014. 362 p. (In Russian)
5. Messick S. The interplay of evidence and consequences in the validation of performance assessments. *Education Researcher*, 1994. 32 (2). Pp. 13–23. (In Russian)
6. Mislevy R. J., Almond R. G., & Lukas J. F. (2003). A brief introduction to evidence-centered design. (ETS Research Report RR-03-16). Princeton, NJ: Educational Testing Service. 2003. 37 p.
7. Mislevy R. J., Levy R. Bayesian psychometric modeling from an evidence-centered design perspective. In: Rao, C.R., Sinharay, S. eds. *Handbook of statistics*. Elsevier, Amsterdam, 2007. Pp. 839-865.
8. Sukin T., Dunn J. Validity Argument for a Statewide Alternative Assessment. Addressing Inferences with Novel Approaches. National Council for Educational Measurement. Vancouver, 2012. 33 p.
9. Toulmin S. E. *The uses of argument*. Cambridge, Cambridge University Press Publ., 1958. 264 p.
10. Van der Linden, W. J., & Glas, C.A.W. (2010). *Elements of adaptive testing*. New York, Springer Publ., 2010. 462 p.

**СВЕДЕНИЯ ОБ АВТОРЕ**  
**Принадлежность к организации**

**Ефремова Надежда Федоровна**, доктор педагогических наук, профессор, заведующая кафедрой педагогических измерений, Донской государственной технической университет (ДГТУ), Ростов-на-Дону, Россия, e-mail: nefremova61@donstu.ru

*Исследование выполнено при поддержке гранта программы ЭРАЗМУС+ 575456-EPP-1-2016-1-RU-EPPJMO-CHAIR*

*Принята в печать 04.05.2018 г.*

**INFORMATION ABOUT THE AUTHOR**  
**Affiliations**

**Nadezhda F. Efremova**, Doctor of Pedagogy, professor, the head of the chair of Pedagogical Measurements, Don State Technical University (DSTU), Rostov-on-Don, Russia, e-mail: nefremova61@donstu.ru

*The study was carried out with the support of the ERASMUS+program grant 575456-EPP-1-2016-1-RU-EPPJMO-CHAIR*

*Received 04.05.2018.*