

УДК 371.263+378.001

## СТАНДАРТИЗАЦИЯ КАК УСЛОВИЕ ОБЕСПЕЧЕНИЯ КАЧЕСТВА ФОНДОВ ОЦЕНОЧНЫХ СРЕДСТВ ВУЗОВ

Ефремова Н.Ф.

*Донской государственный технический университет,  
Ростов-на-Дону, e-mail: nefremova61@dstu.edu.ru*

Лучшие мировые образовательные практики, как известно, опираются на надежную систему оценки, развитию которой уделяется большое внимание. В отечественной системе образования этот процесс идет очень медленно и сложно. В статье описаны основные принципы построения оценки качества в системе высшего образования, показаны наиболее распространенные проблемы университетов в развитии системы оценки качества, направления научно обоснованного выхода из ситуации. Основное внимание уделено необходимости достижения качества оценочного инструментария за счет его стандартизации и определения психометрических характеристик. Показаны различия между квазитестами и психометрическими тестами, а также возможностями применимости тех и других оценочных средств. Отмечено, что опора на стандарт тестирования обеспечит надежность педагогических измерений и позволит избежать многих ошибок. Объективность оценки учебных достижений способствует повышению мотивации обучения и повышению его качества.

**Ключевые слова:** результаты обучения, системы оценивания, компетенции, оценочные средства, качество обучения, ключевые ценности оценивания достижений

## STANDARDIZATION AS A CONDITION FOR ENSURING QUALITY OF FUNDS EVALUATION TOOLS OF UNIVERSITIES

Efremova N.F.

*Don State Technical University, Rostov-on-Don, e-mail: nefremova61@dstu.edu.ru*

The world's best educational practices, as we know, rely on a reliable evaluation system, the development of which received much attention. In the national education system this process is very slow and difficult. The article describes the basic principles of quality assessment in higher education, shows the most common problems of universities in the development of quality assessment system, evidence-based direction out of the situation. The main attention is paid to the need to achieve quality of assessment tools by standardizing and defining psychometric characteristics. Shows the differences between quasi tests and psychometric tests, as well as the possibilities of the applicability of those and other assessment tools. It is noted that reliance on standard testing will ensure the reliability of pedagogical measurements and will allow to avoid many mistakes. Objectivity evaluation of educational achievement enhances learning motivation and improving its quality.

**Keywords:** learning outcomes, assessment system, competence, assessment tools, the quality of learning, key values of assessment of achievements

Текущее десятилетие характеризуется объединением усилий многих стран в разработке единых концептуальных подходов к совершенствованию образовательных стандартов, оценке качества обучения и разработке надежного оценочного инструментария. В целом модернизация контрольно-оценочной системы в образовании характеризуется смещением акцента в сторону высокой объективности, обоснованности и прогностичности результатов оценивания. В связи с этим широкое распространение получило использование современных методик оценки в практике образования, возникла потребность в формировании фондов оценочных средств (ФОС). К основным функциям ФОС можно отнести: разработку, хранение и учет (банк контрольных заданий; стандартизацию оценочных средств и процедур; обеспечение системности и независимости оценивания; формирование и предоставление статистических отчетов; гибкое управление права-

ми пользователей; оперативное изменение оценочных материалов и пользовательских форматов результатов обучения.

Несмотря на большое количество работ, посвященных этой проблеме, до настоящего времени пока еще нет единого подхода и общепринятого документа, содержащего перечень требований и критериев оценки качества самих фондов оценочных средств. Имеющиеся в вузах ФОС не в полной мере используют методологию теории и практики педагогических измерений, как правило, ограничиваясь авторскими вариантами оценочных средств, не дающими достоверных сведений о достижениях обучающихся и не позволяющих проводить сравнительный анализ полученных результатов обучения. Поскольку процесс формирования фондов оценочных средств образовательных организаций находится на начальной стадии, важно, чтобы он изначально шел в русле современных требований к оценочным средствам, педагогическим измерениям,

психометрическим тестам и процедурам тестирования.

**Анализ ситуации  
в контрольно-оценочной  
деятельности вузов**

Сегодня оценочным процессам отводится особая роль не только в образовании, но и в других отраслях экономики для сопряжения потребностей рынка труда, квалификации работников и возможностей системы образования. В учебно-методических комплексах основных профессиональных образовательных программ должны быть представлены дифференцированные по видам аттестации (текущие, рубежные, итоговые) оценочные средства, обеспечивающие высокую надежность оценивания знаний, умений, навыков, компетенций и трудовых действий/функций обучающихся, на соответствие требованиям образовательного и профессионального стандартов. Контрольно-оценочная деятельность вузов переходит в плоскость сертификации квалификаций в производственной сфере, где также широко используются средства и методы стандартизированного тестирования.

Следует отметить, что понятие «тест» в образовательной практике часто используется для любого контрольного задания, что может быть и справедливо на бытовом уровне. Так как «test» в переводе – проба, то в таком понимании любые наборы заданий могут быть названы квазитестом. К квазитестам относятся оценочные средства, состоящие из наборов некалиброванных заданий (заданий неизвестной трудности и других не обоснованных статистически психометрических характеристик) и не имеющие формализованных количественных шкал, а также допускающих различия оценок у разных экспертов при проверке свободных ответов (развернутого решения, эссе, рассуждения и др.). Вместе с тем, широко используемые в образовательной практике квазитесты должны включать в себя все те же компоненты (фиксированный набор заданий, ключи, нормы, шкалы), что и психометрический тест, но эти компоненты в отличие от психометрического теста не будут обоснованы эмпирико-статистической проверкой и сопутствующей ей калибровкой шкалы тестовых баллов. В принципе от использования квазитестов может быть определенная польза. Однако при этом следует понимать, что квазитест не дает результатов с заданной точностью, а полученные данные могут не соответствовать истинным состояниям исследуемых объектов.

Регламентация основных правил разработки и применения качественных оце-

ночных средств и процедур тестирования (психометрических тестов) недавно представлена в Российском стандарте тестирования персонала [5], разработанном по инициативе Национальной конфедерации «Развитие человеческого капитала» (НК РЧК). В стандарте показано, что такое психометрическая проверка теста, обозначены его основные психометрические свойства. Основные положения этого стандарта в полной мере могут быть применимы и при формировании ФОС вузов. Особенность стандарта заключается не в том, чтобы что-то запрещать или разрешать, а в том, чтобы разработчикам и пользователям тестов показать, какой оценочный инструментарий надо выбрать для каждой конкретной ситуации. Опора на стандарт тестирования позволит избежать многих ошибок. При оценивании достижений обучающихся важно учитывать, что тестовые измерения дают результаты с вероятностной точностью: истинный результат находится не в какой-то точке, а в определенном (доверительном) интервале на шкале, содержащей в себе заранее заданную степень погрешности (ошибку измерения). Поэтому перед проведением тестирования во избежание неправильного или излишне прямолинейного толкования результатов тестирования необходимо определиться с пониманием того, каким диапазоном точности (степенью погрешности) будет обладать полученная информация.

**Стандартизация  
психометрических тестов**

Важно понимать, как оценивать возможные погрешности результатов оценочной процедуры, когда следует применять тесты, имеющие психометрические характеристики, а когда те, которые не прошли психометрической проверки (авторские тесты или квазитесты). Стандарты оценочной процедуры требуют выполнения ряда условий: использования психометрических тестов и периодичности проведения оценки (по мере завершения изучения модуля ОПОП, дисциплин и практик); соблюдения последовательности и преемственности оценки (по возрастанию требований к уровням освоения ОПОП); многоступенчатости (оценка, самооценка, рефлексия, обсуждение результатов и комплекс мер по устранению недостатков); единства используемой технологии тестирования для всех испытуемых одной стадии обучения (заполнение оценочных листов, использования одинаковых шкал оценивания и критериев); обеспечения сопоставимости результатов оценивания

с обоснованными нормами и по однотипным выборкам [2].

Двумя базовыми требованиями к психометрическим тестам и процедурам тестирования выступают объективность и стандартизация. Теоретически объективность означает, что один и тот же испытуемый при повторном выполнении теста должен получить такой же результат, как и при первоначальном, на практике это невозможно, так как при повторном тестировании испытуемый уже знаком с содержанием теста. Стандартизация предполагает такие структуру и процедуру проведения тестирования, которые обеспечивают максимально возможное равенство условий для всех участников оценочного процесса. Тесты для аттестации студентов вуза, релативизирующие одинаковые направления подготовки, можно стандартизировать на собственных выборках в тех случаях, когда результаты аттестации используются только внутри самого вуза. В случае, если речь идет об аттестации студентов при комплексной проверке деятельности учебных заведений, стандартизация тестов должна проводиться на репрезентативной выборке, состоящей из студентов разных вузов и отражающей пропорции генеральной совокупности всех студентов, получающих подготовку по данному направлению.

В теории педагогических измерений к «педагогическому тесту» или, как его иногда называют, «психометрическому тесту», предъявляется ряд требований, указывающих на наличие: калиброванных заданий разного типа и измерительных шкал, позволяющих выразить результаты тестирования количественно; связи каждого ответа на каждое задание с одной или несколькими измерительными шкалами; наличие «ключей к тесту» и инструкций по проведению процедуры тестирования, одинаковых условий проверки результатов выполнения теста и др. Одним из наиболее важных признаков измерительных (психометрических) свойств теста является калибровка (определение уровня трудности) его тестовых заданий для получения разметки шкалы тестовых баллов. Такое тестирование предусматривает возможность автоматизированной обработки материалов (бланков ответов испытуемых), проверки правильности выполнения заданий и формирования протоколов результатов с помощью формализованной процедуры подсчета тестовых баллов по шкалам перевода первичных баллов на шкалу тестовых баллов с помощью весовых коэффициентов трудности используемых заданий, направленность результатов на индивидуальную количественную

оценку достижений испытуемого. Для этого необходимы тестовые нормы как оценочные категории или фиксированные границы перевода первичных баллов в тестовые. При тестовом контроле также важными являются методы интерпретации результатов и рекомендации для принятия по анализу результатов тех или иных решений.

К качеству педагогических тестов и методик их применения (психометрических требований) предъявляется четыре группы основных требований: достоверности, надежности, валидности и репрезентативности.

Достоверность в большей степени характеризует не сам тест, а процедуру его применения, в основном отображает имеющиеся при тестировании нарушения информационной защиты тестов и первичных результатов тестирования, мотивационные искажения, угадывание ответов, расклевывание правильных ответов и др.

Надежность связана с устойчивостью результатов, получаемых при повторных тестированиях. Учитывая сложность организации такого подхода, чаще всего для определения надежности используют метод расщепленных групп [6]. Разработчик обязан указывать размер выборки, на которой определялась надежность. Показатель надежности для тестов способностей не должен быть ниже 0,7. Если в тестовой методике используются параллельные тесты (варианты, одинаковые по структуре и трудности, но включающие разные тестовые задания), необходимы количественные подтверждения эквивалентности этих форм и валидности полученных результатов, что достигается статистической обработкой данных апробации на репрезентативной выборке испытуемых.

Различают три основных вида валидности: содержательную, конструктивную и критериальную. Содержательная валидность должна отвечать на вопросы: какой конструкт (тестируемое свойство) лежит в основе методики тестирования; какие компоненты входят в шкалу измеряемой величины; какова содержательная область, отображаемая в тесте; какие виды деятельности тестируемого оцениваются содержанием заданий теста и др.? Содержательная валидность обосновывается с помощью качественных характеристик, чаще всего путем экспертных оценок (желательно не одним экспертом). Конструктивную валидность определяют как взаимосвязь результатов тестирования с результатами других методик оценивания, содержательно близких (или отличных) путем сравнения результатов исследуемых групп. Коэффициенты корреляции (либо эквивалентные статистические

показатели) должны быть сопоставлены с изначальными теоретическими ожиданиями, при этом корреляции 0,1–0,3 расцениваются как «слабые», 0,3–0,5 как «умеренные», 0,5–0,7 как «сильные», выше 0,7 – как «настораживающие», ибо очень высокие значения валидности могут говорить о методических ошибках при ее определении, а потому требуют перепроверки. **Критериальная** или **прагматическая валидность** представляет собой взаимосвязь ее результатов с заданными научно обоснованными критериями. Данный вид валидности признается наиболее важным. Данные о критериальной валидности могут быть получены после фактического применения теста, сбора и статистической обработки данных.

Репрезентативность обосновывается свойствами выборки нормирования как концептуально (через описание состава этой выборки с точки зрения основных параметров), так и эмпирически путем апробации. Необходимый размер выборки нормирования для стандартизованных методик тестирования зависит от степени однородности предполагаемой группы участников тестирования. В общем случае размер выборки для исследования надежности не должен составлять меньше 100 чел., желательно 200 и более, что в условиях вузов не всегда возможно при проведении тестирования в малых группах. Тогда следует или накапливать данные или использовать сравнения данных с результатами, полученными другими способами. Для тестов, ориентированных на локальное применение в малых группах, выборка нормирования может быть ограничена. В этом случае участники тестирования сравниваются внутри своей группы, что определяет и специфику интерпретации результатов. Результаты подобной локальной методики при ее использовании в других группах (потоках) не могут быть признаны без проведения дополнительных исследований. В любом случае нормы должны периодически обновляться и уточняться.

В новой концептуальной модели педагогических измерений задания для психометрического теста конструируются так, чтобы обеспечить проверку различных уровней подготовленности испытуемых, выявить не только внешнее выражение умений, но и оценить латентные процессы мыслительной деятельности при формировании конечного ответа на тестовые задания. Для этого конструирование современных педагогических измерителей проводится по методике моделирования и параметризации тестов с помощью математического аппарата IRT (Item Response Theory), предполагающего

создание теста согласно предварительно заданной информационной функции. Методология IRT позволяет уровни трудности заданий и подготовленности испытуемых разместить на единую логистическую шкалу, в результате чего тест выступает в виде своеобразного измерительного инструмента определенной разрешающей способности, надежности, точности и других квалитетических характеристик. Такой подход к контролю и оцениванию обеспечивает получение более достоверной информации о подготовленности испытуемых, позволяет ее выразить количественно, дифференцировать обучающихся по уровням подготовленности и сопоставлять результаты тестирования по любым массивам испытуемых [1].

Требования к качеству тестов носят комплексный и контекстно-зависимый характер, то есть оценивать применимость методики тестирования следует для каждого конкретного оценочного мероприятия отдельно. Существует также ряд специфических требований, возникающих при решении определенного круга задач тестирования. Например, требование эффективности предполагает, что в некоторых ситуациях затраты на создание или приобретение психометрических тестов не оправдываются в связи с небольшой численностью групп потенциальных тестируемых. При больших выборках испытуемых или «высоких ставках» использования результатов предпочтительнее отдавать тестам, в большей степени удовлетворяющим психометрическим требованиям.

Объективная оценка уровня учебных достижений позволяет любому субъекту образования самоидентифицировать себя во внешней образовательной среде, определить свой рейтинг в однотипном ряду по определенным показателям, выявить и мобилизовать свои возможности для достижения планируемого результата, наметить план актуальных действий в соответствии с собственной системой ценностей и мотиваций. Выявление объективного уровня учебных достижений при самоанализе чаще всего проявляется как мотив к повышению подготовленности и как аспект личностного развития: происходят латентные процессы изменений в самооценке, самокоррекции, самовыражении и целеполагании. Воздействие результатов тестирования на личность проявляется в усилении развивающей функции обучения и контроля: приобретение навыков подготовки ответа и предъявления своих знаний; формирование общеучебных компетенций при выявлении элементов ответов на задания теста; стремление улучшить свой результат и приобрести более

устойчивые знания к последующему тестированию. Существенным признаком, обусловливающим доминирование обучающей и развивающей функций при тестировании, является оптимизация трудности тестовых заданий применительно к уровню подготовленности обучающегося, чему в значительной степени способствует технология адаптивного тестирования. Этот метод является наиболее перспективным, реализация которого обеспечивается комплексом педагогических и технологических условий [3, 4].

Следует также иметь в виду, что этика деловой коммуникации (взаимоотношений) в ситуации тестирования также способствует повышению мотивации. Она требует корректности на условиях сотрудничества. Вся информация об оценочных средствах и баллах тестирования до утверждения протоколов результатов является конфиденциальной. Предоставление тестируемому обратной связи в той или иной форме является обязательным и необходимым. Отсутствие обратной связи признается ошибочным как в силу несоответствия морально-этическим нормам, так и по причине исключения развивающего эффекта от проведенного тестирования.

### Вывод

Потребности в оценке знаний и компетенций непрерывно увеличиваются, растет запрос на качество педагогических измерителей и качественное проведение оценочных процедур, существует острая потребность в специалистах в области педагогических измерений и разработки надежных оценочных инструментов. Рассмотренные формы оценочной деятель-

ности не могут быть эффективными без специалистов тестологов, психологов, аналитиков, интерпретаторов данных образовательной статистики, подготовленных организаторов оценочных процессов и техник оценивания. Только специально подготовленные кадры могут квалифицированно обеспечить программно-инструментальную среду образовательной организации; научно-методическую помощь преподавателям разработчикам оценочных средств; возможность аккумулировать средства оценивания и проводить автоматизированную обработку результатов тестирования, определять нормы качества и стандартизацию тестов, разработку критериев и шкал оценивания; создать условия для проведения независимого внутреннего и внешнего оценивания, формировать базы образовательной статистики результатов и обеспечивать условия доступа к ней различным категориям пользователей.

### Список литературы

Ефремова Н.Ф. Тестовый контроль в образовании: учеб. пособие. – М.: Логос, 2007. – 386 с.

Ефремова Н.Ф. К вопросу о создании и функционировании фондов оценочных средств в вузе // Высшее образование в России. – 2015. – № 7. – С. 63–67.

Звонников В.И. Адаптивное тестирование в дистанционном обучении / В.И. Звонников, М.Б. Чельшкова, А.А. Малыгин // Высшее образование сегодня. – 2012. – № 6. – С. 7–10.

Малыгин А.А. Адаптивное тестирование в дистанционном обучении: монография. – Иваново: Иван. гос. хим.-технол. ун-т, 2012. – 138 с.

Российский стандарт тестирования персонала (временная версия 2015 г.) // Организационная психология. – 2015. – Т. 5. – № 2. – Режим доступа – [www.orgpsyjournal.hse.ru](http://www.orgpsyjournal.hse.ru).

Чельшкова М.Б. Теория и практика конструирования педагогических тестов: учеб. пособие. – М.: Логос, 2002. – 432 с.