

КВАЛИМЕТРИЯ В ПРОФЕССИОНАЛЬНОМ ОБРАЗОВАНИИ

© 2009 Е.Н.Рябинова

Самарский государственный технический университет

Статья поступила в редакцию 10.09.2009

Представлен анализ работ по научным основам измерений в педагогике. Показано, что разработка заданий в тестовой форме – самый важный этап тестового процесса. Именно задания в тестовой форме, в сочетании с новым поколением компьютерных технологий, смогут обеспечить качество профессионального образования.

Ключевые слова: квалиметрия, профессиональное образование, задание в тестовой форме, компьютерные технологии.

Научные основы измерений в педагогике являются относительно новой областью педагогической науки, получившей признание только в середине XX века. Тесты представляют собой совокупность заданий, которые позволяют дать объективную сопоставимую и даже количественную оценку качества подготовки обучаемого в заданной образовательной области¹. В свою очередь объективность и измеримость качества образования открывают широчайшие возможности для управления учебным процессом – от корректировки содержания образовательных стандартов и программ до совершенствования методов преподавания и повышения эффективности стимулирования самостоятельных занятий студентов.

Тест как совокупность заданий составляется в стандартной форме, удовлетворяет требованиям надежности и валидности и дает возможность проводить сравнительную оценку знаний, умений, навыков, а также способностей и умственного развития в дидактических или психологических исследованиях. Под надежностью контрольного задания при этом понимается его способность с достаточной для практики одинаковостью характеризовать исследуемый в дидактических или психологических экспериментах показатель как заданием в целом, так и его частями, или тот же показатель одним и тем же заданием, но в разные моменты времени².

Под валидностью контрольного задания понимается его соответствие своему назначению. К критериям валидности относятся ответы на вопросы: соответствует ли задание программе или его раздела; охватывает ли задание достаточно полно весь курс или его раздел; доста-

точно ли вероятность того, что если учащийся успешно справился с рассматриваемым контрольным заданием, то он наверняка знает соответствующий материал? Другими словами, показатель валидности отвечает на вопрос: а то ли измеряется с помощью рассматриваемого задания, что сформулировано в его назначении? Очевидно, что задание может быть надежным, но не валидным. Так же очевидно, что между надежностью и валидностью имеется прямая связь: задание с низкой надежностью не может быть валидным.

Тест, имеющий достаточную надежность и приемлемую валидность, называют состоятельным³. На практике достаточно часто о валидности теста судят не по оценке с помощью коэффициента корреляции, а по аналитическим оценкам опытных преподавателей – экспертов самого содержания теста.

Интересна связь между надежностью теста и его длиной. При этом под длиной теста подразумевается количество заданий, составляющих тест. Короткий тест содержит 5 – 7 заданий, длинный может составлять до 500 заданий. Например, полный тест по проверке знания таблицы умножения содержит 100 заданий (такова его длина). Но на практике для проверки умения перемножать цифры может использоваться и более короткий тест. Количественная связь между коэффициентами надежности длинного и короткого тестов определяется формулой⁴

$$R = \frac{r \cdot n}{1 + r(n-1)}, \text{ где } R - \text{коэффициент надежности}$$

длинного теста; r – коэффициент надежности короткого теста; n – число, показывающее во сколько раз первый тест длиннее второго, то есть во сколько раз

⁰ Рябинова Елена Николаевна, кандидат технических наук, доцент кафедры высшей математики и прикладной информатики. E-mail: pnm@ssau.ru

¹ Челышкова М.Б. Теория и практика конструирования педагогических тестов: Учеб. пособие. - М.: 2002.

² Потеев М.И. Основы аналитической дидактики: Учеб. пособие. - СПб.: 1992.

³ Аванесов В.С. Композиция тестовых заданий. - М.: 1996.

⁴ Brown W. Some Experimental Results in the Correlation of Mental Abilities / British Journal of Psychology, 1910, V.3, N2. - P. 296 – 322; Spearman C. Correlation Calculated from Faulty Data / British Journal of Psychology. 1910, V.3, N2. - P. 271 – 295.

число заданий длинного теста больше числа заданий короткого теста.

Надежность теста оценивается как отличная, если $R > 0,9$; хорошая – если $R \in (0,8 \div 0,9]$; удовлетворительная – при $R \in (0,75 \div 0,8]$. Валидность теста считается отличной, если коэффициент валидности $V > 0,75$; хорошая – если $V \in (0,5 \div 0,75]$; удовлетворительная – при $V \in (0,4 \div 0,5]$.

Согласно В.П.Беспалько⁵, смысл теста и его отличие от любого другого варианта контроля знаний можно понять из его символической формулы: Тест = Задание + Эталон. Формулировка задания в соответствии с параметрами качества усвоения деятельности и наличия эталона, то есть полного и правильного решения теста, в котором выделены существенные операции, составляет основную специфику теста как средства объективного контроля качества знаний учащихся. Из эталона устанавливается общее число существенных операций теста - m , а по ответу испытуемого – число правильно выполненных им существенных операций l , после чего подсчитывается коэффициент усвоения $K_y = \frac{l}{m}$. Основным в построении тестов является создание адекватного задания и соответствующего ему эталона. Поскольку тесты создаются для проверки уровня усвоения заданного содержания обучения, то очевидно, что они также различаются по уровням усвоения.

Современное состояние контроля знаний представляет собой элективное соединение старого с новым, субъективного с объективным, ручного труда с компьютером. К основным недостаткам форм контроля в обучении можно отнести следующие: 1) низкую оперативность в использовании результатов контроля для управления учебным процессом; 2) неудовлетворительную объективность оценки знаний студентов, невозможность сопоставления оценок, полученных ими у разных преподавателей или в разных вузах; 3) невозможность использования их в дидактических исследованиях.

Если поиск решения проблемы оценки качества обучения в нашей стране идет в основном по пути теоретико-методологических исследований, то в ряде европейских стран, США, Австралии большинство исследований носит прикладной характер и сами исследования по сути своей крайне прагматичны.

Идеологи современных параметрических методов в теории педагогических измерений, получивших название *Item Response Theory*

(IRT), благодаря специальным математическим моделям и эффективному математическому аппарату, построенному на итерационных процессах, сумели перейти от наблюдаемых результатов выполнения к оценкам латентных (скрытых от поверхностного взгляда) параметров испытуемых, ассоциируемых с константами обучаемых на момент измерения.

Метод оценки индекса структурированности знаний (*method S-P lines*) основан на подсчете значений специального индекса структурированности знаний, введенного в рассмотрение зарубежными исследователями. Дело в том, что тесты на определенных условиях позволяют выявить не только уровень подготовки, но и структуру знаний учащихся, вернее степень ее отклонения от идеальной структуры, планируемой педагогом на момент начала обучения. Элементарное представление о степени отклонения позволяет составить анализ профиля ответов учащегося на различные задания теста. Так как один и тот же уровень подготовки может быть получен при ответах на различные задания, то можно сравнить ответы учащихся с одинаковым уровнем подготовки. Если задания ранжированы по нарастанию трудности и отражают по содержанию и порядку расположения идеальную, планируемую преподавателем при обучении структуру знаний, то общий вывод о качестве подготовки учащихся делается просто: чем меньше пробелов в ответах ученика на задания теста, тем лучше структура его знаний, тем выше качество его подготовки в группе других учеников, имеющих одинаковый с ним уровень подготовки по предмету.

Идеальный профиль ответов, не содержащий пробелов, в мировой литературе по тестовой проблематике получил название профиля Гуттемана⁶, которому соответствует нулевой специальный индекс. Большое значение индекса, характерное для инвертированных профилей с большим числом пробелов, порождает сомнение в системности знаний обучаемого и валидности его общей оценки по тесту.

Аутентичная оценка учебных достижений (*Authentic assessment*) была впервые предложена в 1988 году Арчибальдом и Ньюманом (*Archibald & Newmann*)⁷. Ее отличают две основные особенности: по результатам выполнения работы делается вывод об уровне овладения учебным материалом на основе специальным образом разработанных критериальных заданий;

⁶ Keeves J.P. (Ed.) Educational Research, Methodology Measurement: An International Handbook. – Oxford: Pergamon Press, 1988.

⁷ Bateson D., Nicol C., Achroeder T. Alternative Assessment and Tables of Specification for the Third International Mathematics and Science Study. ICC, 64, 1991.

⁵ Беспалько В.П. Образование и обучение с участием компьютеров (педагогика третьего тысячелетия). – М.; Воронеж: 2002.

учащимся предлагаются такие задания, которые могут вызвать у них интерес и представляют для них определенное значение. Данные задания позволяют получить эффективный конечный продукт и оценить уровень деятельности ученика. Обычно время на выполнение заданий не лимитируется.

Основной идеей «уравновешенной» оценки учебных достижений («*Balanced*» *assessment*) является необходимость уравновесить традиционные формы проверки качества подготовки обучаемых такими формами контроля, которые позволяют оценить интеллектуальные умения более высокого порядка (*higher order skills*). Впервые данный подход к оценке был предложен математиками в 1990 году⁸.

«Оценка деятельности учащихся» (*Performance assessment*) оценивает освоенные практические навыки ученика с помощью эссе или заданий, позволяющих получить в качестве результата некоторый материальный продукт⁹.

В практике создания сотен и тысяч тестов многочисленными авторами и организациями, даже такими уважаемыми и мощными гегемонами на рынке тестирования как американский *ETS (Education Testing Service)* и заполнивший всю Америку *SAT-тест (College Entrance Examination Board Scholastic Aptitude Test)*, различие тестов часто строится на основе их внешнего вида. Так, если в конце теста дано несколько ответов для выбора, то такие тесты относят к тестам различения (*multiple* или *single choice tests*), хотя задача, поставленная в тесте, может относиться ко второму или даже третьему уровню усвоения заданного содержания обучения. Таким образом, в батарее тестов сосуществуют тесты разных уровней, а оцениваются их решения одинаково, что на самом деле означает отсутствие всякой объективной оценки¹⁰. Кроме педагогической некорректности в построении неоднородных тестовых батарей по уровню усвоения знаний, тесты не различаются по трудности и сложности. Трудность теста определяется уровнем усвоения учебного материала, на диагностику которого он направлен. Сложность теста определяется числом существенных операций в нем. При контроле знаний учащихся с применением тестов можно обеспечить любую наперед заданную надежность суждения о качестве усвоения заданного содержания обучения. Для этого тесты каждого

уровня объединяют в батареи, содержащие несколько тестов одного и того же уровня. От суммарного числа операций в батарее тестов зависит надежность контроля.

Для педагогической теории измерения исходным является понятие «педагогическое задание»¹¹. Это понятие является общим, охватывающим цель и смысл всех возможных заданий. Оно включает такие средства как вопрос, задача, учебная проблема и другие. Педагогическое задание можно определить как средство интеллектуального развития, образования и обучения. При этом интеллектуальное развитие определяется как способность понимать, рассуждать, логически аргументировать, находить закономерности в явлениях и в изменениях объектов, критически оценивать мышление и деятельность. Образование понимается как целенаправленный процесс воспитания и обучения в интересах человека, общества и государства¹².

Педагогическое задание активизирует процесс собственной познавательной деятельности учащихся и повышает качество их знаний. Педагогические задания формулируются как в тестовой, так и нетестовой форме. Задания можно разделить на обучающие и контролируемые. Первые применяются в учебном процессе для развития личности, вторые – для определения уровня и структуры подготовленности. Часть заданий может использоваться одновременно для обучения и контроля, а также для самообразования и самоконтроля.

Форма тестовых заданий – это способ организации, упорядочения и существования содержания теста. Задание в тестовой форме определяется как педагогическое средство, отвечающее требованиям (признакам): цель, краткость, технологичность, логическая форма высказывания, определенность места для ответа, правильность расположения элементов задания, одинаковость инструкции для всех испытуемых, адекватность инструкции форме и содержанию задания.

Задания с выбором одного правильного ответа больше других распространены в тестовой практике, что объясняется их сравнительной простотой, традицией и удобством для автоматизированного контроля знаний. Вероятность угадывания правильного ответа зависит от общего числа ответов к каждому заданию и от степени привлекательности каждого неправильного ответа. В хорошо сделанных заданиях не-

⁸ Чельшкова М.Б. Теория и практика конструирования педагогических тестов...

⁹ Task Group on Assessment and Testing Report. National Curriculum. Department of Education and Science, 1988.

¹⁰ Беспалько В.П. Образование и обучение с участием компьютеров...

¹¹ Аванесов В.С. Форма тестовых заданий: Учебное пособие для учителей школ, лицеев, преподавателей вузов и колледжей. - М.: 2006.

¹² Никандров Н.Д. Понятийный аппарат педагогики и образования: перспективы исследования // Педагогика. 1977. - С. 112 - 114.

правильные ответы нередко кажутся правдоподобнее правильных. Их не случайно называют дистракторами (от английского глагола *to distract* - отвлекать). Хорошо подобранные неправильные ответы существенно снижают вероятность угадывания правильного ответа.

Число ответов к заданиям лучше определять по двум критериям – содержательному и формальному. Первый определяет число возможных правдоподобных ответов. Если их возможно только два или три, то на указанных цифрах и следует остановиться: придумывание большего числа ответов приведет к нарушению принципа содержательной корректности задания. Формальный критерий – вероятность угадывания правильного ответа теми, кто его не знает. Чем меньше число ответов в каждом задании, тем выше такая вероятность. Соответственно, точность и аргументированность такого измерения снижается.

Часто задаются вопросы об оптимальном числе ответов к заданиям с выбором одного правильного ответа. Одни авторы тяготеют к трем, другие – к четырем. В вузовских центрах тестирования обычно заказываются задания с пятью ответами. Однако, если есть возможность придумать шестой эффективный ответ, то он может улучшить качество измерения за счет уменьшения вероятности угадать правильный ответ теми испытуемыми, кто слабо подготовлен.

Тестовое задание – это составная единица теста, отвечающая требованиям к заданиям в тестовой форме. Составная единица означает, что каждое задание является частью неразрывного целого, теста, который представляет собой систему заданий. И каждое задание необходимо для выполнения тестом своей функции. Усвоение хотя бы одного системного задания приводит к возникновению пробела на континууме проверяемых знаний.

При применении заданий в тестовой форме получаемые баллы студентами сводятся к обычным педагогическим оценкам. Такие оценки не могут интерпретироваться как результат педагогического измерения. Скорее, это результат субъективного оценивания баллов, полученных применением заданий в тестовой форме.

В научной литературе по педагогическим измерениям известна аксиома локальной независимости¹³, которая утверждает, что если для испытуемых одного, фиксированного уровня знаний, вероятность правильного ответа на одно какое-либо задание теста зависит от вероятности правильного ответа на другое задание, то первое задание надо рассматривать как зависи-

мое от второго, что нарушает принцип статистической независимости тестовых заданий, положенный в основу создания теста как формальной системы. Эту же аксиому можно выразить проще: если для испытуемых одинакового уровня подготовленности правильный ответ на одно задание зависит от правильного ответа на другое, то такие задания, в научном понимании, тест не образуют¹⁴.

Однако, по мнению В.С.Аванесова, такие задания представляют особую ценность для организации самостоятельной учебной работы, а также для проведения итоговой аттестации выпускников профессиональных образовательных учреждений. При правильной организации автоматизированного контроля системы заданий могут включаться в учебный процесс, мотивируя ежедневную самостоятельную работу студентов.

Система заданий в тестовой форме – это содержательная система, охватывающая взаимосвязанные элементы знаний. В отличие от тестов, в системах заданий вероятность правильного ответа на последующее задание может зависеть от вероятности правильного ответа на предыдущие задания. При этом В.С.Аванесов выделяет четыре основные системы заданий в тестовой форме: цепные, тематические, текстовые и ситуационные задания. Последние можно определить как педагогически переработанный фрагмент профессиональной деятельности специалиста.

Разработка инновационной индивидуально-корректируемой технологии профессионального обучения, в основе которой лежит модель усвоения учебного материала¹⁵, непрерывный мониторинг развития учебных способностей учащихся и соответствующая этому мониторингу корректировка учебного процесса по структурированным познавательно-деятельностным составляющим¹⁶, требует соответствующей квалиметрии (тестирования). Квалиметрия (от латинского - *qualis* – какой (по качеству) + ... метрия) – отрасль науки, изучающая и реализующая методы количественной оценки качества продукции¹⁷. В нашем случае под качеством продукции понимается качество обучения, каче-

¹³ Lord F.M. Application of Item Response Theory to Practical Testing Problems. Hillsdale, N-J, Lawrence Erlbaum Ass. Publ. 1980. – 266 pp.

¹⁴ Lord F.M. & Novick M. Statistical Theories of Mental Test Scores. Addison – Wesley Publ. Cj., Reading, Mass, 1968. – 560 pp.

¹⁵ Рябинова Е.Н., Тутов Б.А. Феноменологическая модель усвоения учебного материала с учетом фактора мотивации // Вестник СГАУ. – Самара: – 2006. – №1(9). – С.246 – 258.

¹⁶ Рябинова Е.Н. Формирование познавательно-деятельностной матрицы усвоения учебного материала в высшей профессиональной школе: монография – Самара: 2008.

¹⁷ Словарь иностранных слов и выражений / Авт.-сост. Е.С.Зенович. – М.: 2002.

ство усвоения учебного материала.

Определение фактической траектории усвоения учебного материала, персонифицированной по каждому учащемуся, предполагает проведение тестирования текущей успеваемости учащихся по структурированным заранее уровням учебных задач в соответствии с познавательно-деятельностной матрицей усвоения учебного материала. Тестирование текущей успеваемости учащихся предлагается проводить в моменты t_i , $i = \overline{1, N}$, которые назовем моментами квалиметрии. В результате квалиметрии выявляется как количественная сторона процесса усвоения, так и качественная, поскольку по результатам решения задач j -го уровня, представленных в виде заданий в тестовой форме, ясно какие учебные элементы из ψd - матрицы не усвоены. Таким образом, предлагаемая технология профессионального обучения представляет собой многошаговую процедуру периодической квалиметрии текущей успеваемости учащихся и соответствующей измеренному отставанию оперативной корректировке учебного процесса путем вычисления внешней поддержки – необходимого для дополнительного изучения числа учебных элементов с тем, чтобы фактическая траектория усвоения учащегося стремилась бы к эталонной.

Задания в тестовой форме первого уровня, в соответствии с понятием первого уровня усвоения учебного материала, должны проверять качество узнавания учащимися ранее изученного учебного материала. Согласно нашей структуризации они состоят из 4-х учебных элементов (УЭ) - Y_{11} , Y_{21} , Y_{31} , Y_{41} . Задания в тестовой форме второго уровня соответствуют второму уровню усвоения учебного материала. Эти задания проверяют умение студента воспроизводить усвоенную информацию по памяти без внешней подсказки и решать на этой основе типовые задачи. Под типовой задачей при этом понимаем такую задачу, условие которой допускает непосредственное применение усвоенных алгоритмов, правил или формул для ее разрешения. Согласно нашей структуризации эти задания состоят из 8-ми УЭ - Y_{11} , Y_{12} , Y_{21} , Y_{22} , Y_{31} , Y_{32} , Y_{41} , Y_{42} . Задания в тестовой форме третьего уровня требуют от студента преобразования, трансформации имеющихся знаний и умения их применять в новой ситуации, подводя задачу под типовой алгоритм. Эти задания, согласно нашей структуризации состоят из 12-ти УЭ - $Y_{11} \rightarrow Y_{12} \rightarrow Y_{13} \rightarrow Y_{21} \rightarrow Y_{22} \rightarrow Y_{23} \rightarrow Y_{31} \rightarrow Y_{32} \rightarrow Y_{33} \rightarrow Y_{41} \rightarrow Y_{42} \rightarrow Y_{43}$. Задания в тестовой форме четвертого уровня – это проблемы, решаемые в рамках учебной научно-исследовательской работы студентов. Эти задания состоят из 16-ти УЭ - Y_{11} , Y_{12} , Y_{13} , Y_{14} ,

Y_{21} , Y_{22} , Y_{23} , Y_{24} , Y_{31} , Y_{32} , Y_{33} , Y_{34} , Y_{41} , Y_{42} , Y_{43} , Y_{44} . В приведенном описании заданий в тестовой форме их различие основано на особенностях деятельности, которую выполняет учащийся. В заданиях первого уровня это деятельность по узнаванию ранее усвоенных УЭ при повторной встрече с ними. В заданиях второго уровня – деятельность по воспроизведению по памяти ранее усвоенной информации. В заданиях третьего уровня – деятельность по реконструкции, преобразованию ранее усвоенной информации и переносу ее в новые условия. В задании четвертого уровня - деятельность по развитию усвоенной информации с известными УЭ и открытием новых УЭ.

Тестирование текущей успеваемости учащихся по структурированным заранее уровням сложности учебных задач в соответствии с предложенной познавательно-деятельностной матрицей происходит по определенному алгоритму.

Ведущим принципом в разработке заданий в тестовой форме является последовательное восхождение в уровнях усвоения учебного материала, отражающее иерархию уровней возможностей деятельности человека.

Системы тестовых заданий при квалиметрии студентам предлагается предъявлять в следующем порядке: сначала выполняется блок заданий, соответствующих первому уровню усвоения знаний, затем подсчитывается коэффициент усвоения $0 \leq K_y \leq 1$, характеризующий отношение правильно выполненных учебных элементов к общему числу учебных элементов в рассматриваемом тесте (рис. 1). В процессе усвоения знаний и действий учащийся в обязательном порядке должен усвоить деятельность на предшествующем уровне, чтобы подняться на последующий. На каждой кривой усвоения учебного материала студентом имеется узловая точка $K_y = 0,7$, которая разделяет обучающий процесс на две неравные части: участок кривой с $K_y \in [0; 0,7)$ называется кривой обучения (или научения), так как учащийся на этом этапе своего восхождения на данный уровень мастерства требует постоянного внимания преподавателя, проверяющего и корректирующего его деятельность, поскольку студент на этом этапе своего обучения еще нечувствителен к ошибкам и не может как заметить, так и исправить их; участок $K_y \in [0,7; 1,0]$ называют кривой самообучения, поскольку учащийся, достигший этого качества усвоения учебного материала, сам способен контролировать правильность своих действий и корректировать свои ошибки. Именно поэтому при $K_y < 0,7$ студенту назначается необходимая персонифицированная внешняя поддержка для исправления этой ситуации. Если

$K_y \geq 0,7$, то учащийся начинает выполнять систему заданий второго блока, соответствующую задачам второго уровня. При выполнении всех задач этого блока, опять подсчитывается коэффициент K_y , который сравнивается с величиной 0,7. И опять либо назначается необходимая персонифицированная внешняя поддержка учебной деятельности учащегося, либо студент

переходит к решению задач 3 уровня. Далее алгоритм повторяется. В результате такой квалиметрии мы имеем полную количественную оценку качества усвоения учебного материала каждым конкретным учащимся, что соответствует реальным оценкам педагогического измерения.

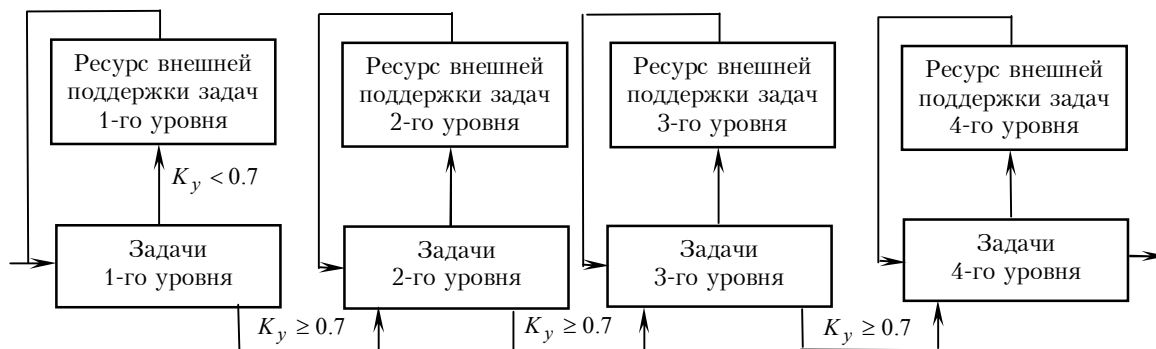


Рис. 1. Последовательность выполнения заданий разного уровня сложности в момент квалиметрии (K_y – коэффициент усвоения)

Организация опыта учебной деятельности задается постепенно в соответствии с освоением познавательных действий рассматриваемого уровня знаний. При этом учебные действия осуществляются с пониманием самого механизма формирования знаний учащегося. В монографии¹⁸ приведены примеры заданий в тестовой форме и образцы бланков ответов специально разработанной формы для текущих тестовых результатов.

Разработка заданий в тестовой форме – самый важный этап тестового процесса. Правильная форма является лишь необходимым, но недостаточным условием для создания тестов.

Нужно подобрать такое содержание заданий, которое позволяет качественно и количественно оценить уровень и структуру подготовленности обучаемых. Богатейшие возможности заданий в тестовой форме позволяют эффективно реализовать их высокий обучающий потенциал. Именно задания в тестовой форме, в сочетании с новым поколением компьютеризованных технологий, смогут обеспечить качество профессионального образования.

¹⁸Рябинова Е.Н. Разработка и реализация индивидуально-корректируемой технологии профессионального обучения: монография – Самара: 2008.

QUALITY MEASUREMENT OF VOCATIONAL EDUCATION

© 2009 E.N.Ryabinova^o

Samara State Technical University

Analysis of work on scientific bases of measurements is represented. It is shown that development of tasks in a test form is the most important stage of test process. Specifically, tasks in the test form, in combination with the new generation of computer technologies can ensure the quality of vocational education.

Key word: quality measurement, vocational education, tasks in a test form, computer technologies.

^oRyabinova Elena Nikolaevna, Candidate of Technical Sciences, Associate Professor Higher Mathematics and Applied Computer Science Department.
E-mail: pnn@ssau.ru