

МЕТОДИКА

А. Г. Шмелев, Д. С. Портнова, Н. Н. Страхов

МЕТОД КОНТЕКСТНОЙ ВСТАВКИ КАК ИНСТРУМЕНТ КОМПЬЮТЕРИЗИРОВАННОГО КОНТРОЛЯ КАЧЕСТВА ЗНАНИЙ

В статье представлена проблематика компьютеризированных процедур аттестационного контроля: их разработки, применения, достоинств и недостатков. Описан опыт работы с «методом контекстной вставки» как технологическим решением, позволяющим преодолеть ограничения классического тестирования и обеспечивающим комплексный подход к диагностике знаний учащихся. Приведены результаты содержательной и психометрической апробации метода на студентах факультета психологии МГУ имени М.В. Ломоносова в рамках курса «Основы психодиагностики», а также на материале проверки грамотности. Обоснована перспективность подобного метода при проведении текущего и итогового контроля, очерчены границы его применения и возможные трудности организационного характера.

Ключевые слова: компьютеризированное тестирование, система дистанционного обучения, оценка знаний, аттестация, контекстная вставка, система управления обучением, тест завершения, заполнение пропусков.

The article presents range of problems, connected with computerized procedures of attestation control, such as their elaboration, application, advantages and disadvantages. The experience of working with so called “contextual insertion method” as a technological solution, that allows overcoming the limitations of classical tests and provides complex approach to diagnostics of students’ knowledge, is described. The results of substantive and psychometric testing of this method on example of the students of department of psychology of Lomonosov Moscow State University are discussed. Participants performed tests within the bounds of “The foundations of psychodiagnostics” course. The result of literacy

Шмелев Александр Георгиевич — докт. психол. наук, профессор кафедры психологии труда и инженерной психологии, науч. рук. лаборатории компьютерной психодиагностики ф-та психологии МГУ имени М.В. Ломоносова. *E-mail:* ags06@ht.ru

Портнова Дарья Сергеевна — студентка V курса ф-та психологии МГУ имени М.В. Ломоносова. *E-mail:* maltaneba@yandex.ru

Страхов Николай Николаевич — инженер-программист лаборатории «Гуманитарные технологии». *E-mail:* nstorm@ht.ru

test is analyzed separately. The prospects of method in carrying out current and final control are proved, the limits of its use and possible organizational difficulties are delineated.

Key words: computerized testing, system of distance learning, assessment, qualifying evaluation, contextual insertion, Learning Management Systems, Closed Test, gaps filling.

Проблемный и методической контекст работы

В лаборатории компьютерной психодиагностики факультета психологии МГУ под руководством А.Г. Шмелева продолжается цикл работ по созданию и развитию инструментов компьютеризированного контроля качества образования студентов-психологов (Болсинова, Шмелев, 2010; Шмелев, 2003, 2004; Шмелев и др., 2008). Текущий и итоговый контроль с помощью компьютеризированных банков тестовых заданий с выбором ответа не только внедрен в курс «Основы психодиагностики», который уже более 30 лет ведет на факультете А.Г. Шмелев, но и применяется как первый этап государственного экзамена по психологии для студентов-выпускников (Зинченко и др., 2011; Чумаков, Шмелев, 2005). В прежних работах доказаны высокие статистические показатели валидности и надежности компьютеризированных тестовых испытаний. Причем в качестве внешних критериев для оценки валидности выступали такие объективные данные, как оценки преподавателями качества защиты студентами отчетов по выполнению практических заданий (в частности, по курсу «Основы психодиагностики»), суммарные данные успеваемости за последнюю экзаменационную сессию и за все годы обучения (Зинченко и др., 2011).

На фоне этих достижений нельзя не отметить, что метод тестов, понятый как набор заданий с выбором из готовых ответов, имеет наряду с очевидными достоинствами (объективированностью оценки) ряд серьезных ограничений и недостатков. К ним относятся следующие.

1. «Точечный фокус». Отдельные тестовые задания выхватывают из общего контекста отдельные аспекты структуры знаний и умений учащихся, не позволяя выяснить, насколько целостна и систематична та концептуальная конструкция, которая выстроена в сознании студента, т.е. не давая возможности оценить связность и интегрированность предметных знаний.

2. «Скачка идей». В тестах на скорость (с ограничением времени на ответы) быстрая смена тематического фокуса при переходе от одного задания к другому предоставляет преимущество студентам, способным лучше преодолевать экзаменационный стресс, а также

быстрее переключаться с одного проблемного контекста на другой. И наоборот, студенты с высокой экзаменационной тревожностью и слабой переключаемостью, но способные к глубокой концептуальной обработке информации за счет более длительного погружения в определенный проблемный контекст, в ситуации скоростных тестов оказываются в неудобной для себя ситуации (Болсинова, Шмелев, 2010).

3. *Репродуктивность.* Выбирая ответы из готовых вариантов, испытуемый в ходе такого тестирования не демонстрирует значимых умений и навыков, связанных с порождением собственных связанных текстов в проблемном поле изучаемой дисциплины.

В структуре комплексного многоэтапного государственного экзамена эти недостатки метода тестов компенсируются за счет более высокого весового вклада в итоговую оценку со стороны традиционных методов контроля — письменного и устного экзаменов (последние в 2010—20012 гг. дают вклад по 40%, а метод тестов — всего лишь 20%). Но такой комплексный подход неприменим для целей промежуточного, текущего контроля из-за практических ограничений. Тут возникает острая организационно-экономическая проблема, связанная с ограниченным бюджетом времени преподавателей. Преподаватели ведущих вузов страны (занятые кроме педагогической работы напряженной научно-исследовательской деятельностью, а также прикладными разработками), к сожалению, не имеют порой реальной возможности применять для текущего контроля слишком трудоемкие и индивидуализированные методы (письменные контрольные работы, устные собеседования). Отсюда возникает необходимость использования экономичных по затратам личного времени автоматизированных групповых тестовых методов: по итогам изучения какой-то темы студенты размещаются в компьютерном классе и выполняют тест с выбором ответа. Тем не менее даже такой подход требует от преподавателей значительных подготовительных усилий по разработке тестовых заданий. Понимая ограниченную дидактическую ценность этих заданий, а порой просто затрудняясь сформулировать задания с нетривиальной логико-семантической структурой, вскрывающей глубокие пласты знания или незнания, не умея сформулировать задания кейсового типа (на решение практических задач), понимая слишком короткий срок жизни этих тестов (до полного разглашения вопросов и правильных ответов среди студентов), а также осознавая уже перечисленные выше недостатки метода тестов, авторы курсов не находят в себе достаточной внутренней мотивации для создания подобных испытаний или делают это во многом лишь «для очистки совести», что превращает текущий контроль в малоосмысленное формальное мероприятие.

Существует ли реальная альтернатива методу тестов с выбором ответа, т.е. такая процедура, в которой сохранялась бы возможность автоматизированного группового контроля, но были бы снижены негативные последствия перечисленных выше проблем? В этой статье излагается первый опыт внедрения так называемого *метода контекстной вставки* (МКВ), который рассматривается авторами статьи как возможная и довольно перспективная замена тестам с выбором ответов на этапе текущего контроля знаний (речь совсем пока не идет об итоговом контроле).

Методика и программно-техническое решение *Page-test*

МКВ изобретен в позапрошлом веке, и в психологии он более известен как «тест на заполнение пропущенных слов» (его авторство приписывается Герману Эббингаузу — см. раздел «Библиография»). В педагогике он распространен под разными названиями. Наиболее широкое применение в последние десятилетия он получил, пожалуй, при обучении иностранным языкам, и в этой области именуется *Closed Test* (менее популярно название *Gaps filling* — «заполнение пропусков»). В психодиагностике широко представлен такой аналог этого метода, как «неоконченные предложения». Следует подчеркнуть, что в педагогике *Closed Test* чаще используется как обучающее упражнение (*exercise*), а не аттестационная процедура, т.е. выступает как инструмент «учебного тренинга» и самоконтроля, а не внешнего контроля с выставлением оценок. В нашем понимании преимуществ этого метода нам важно сделать акцент на работе с целостным текстом, а не с отдельными взятыми из разных текстов фразами, в которых пропущены слова. Чтобы заострить внимание именно на таком смысле, мы прибегаем в названии к слову «контекст», и получается «метод *контекстной* вставки (слов)». Не имея возможности преодолеть первую фразу (первый пропуск) или даже второй, испытуемый, справляясь с третьим пропуском, тут же упрощает для себя задачу понимания текста в целом, может вернуться к первому пропуску и успешно выполнить задание. Таким образом, МКВ предлагает испытуемому свободу и самостоятельность в выборе стратегии восстановления целостного текста. На пути к одинаково высокому результату разные испытуемые могут выбирать разную последовательность вставок, основываясь на собственном отражении тезауруса терминов изучаемой дисциплины.

Модуль *Closed Test* в настоящее время входит в состав многих развитых систем дистанционного (автоматизированного) обучения (СДО; английский термин для этого класса компьютерных прикладных систем *LMS — Learning Management Systems*). В Интернете

можно найти как англоязычные (напр., www.learnclick.com), так и русскоязычные сайты (напр., www.elearningsoft.ru), на которых любой посетитель может не только выполнить уже готовое упражнение по типу МКВ, но (после регистрации) сам создать подобное упражнение на любом собственном материале. Операционально-методический недостаток, мешающий использовать эти сайты в целях организации реальных контрольных работ в ходе занятий со студентами, заключается в отсутствии возможности реализовать аутентификацию (персональную регистрацию) отдельных студентов: каждый испытуемый может зайти и выполнить тест МКВ анонимно, а преподаватель не в состоянии увидеть, кто и как его выполнил. Для аутентификации студентов нужно осваивать более мощные СДО (*LMS*), нередко — приобретать лицензию на них и «продираться» через сложные интерфейсы, предоставляющие избыточное количество возможностей для выбора методов мониторинга работы студентов в автоматизированных режимах.

МКВ для контроля грамотности (знания русского или иностранного языка) нередко называют «интерактивным диктантом» (см.: www.totaldict.ru). Его отличительная особенность в следующем: когда пропусков в тексте очень много (так что трудно прочесть его глазами и понять смысл каждой отдельной фразы), пользователь-испытуемый может прослушать «звуковую запись», выложенную на веб-страницу, т.е. озвученный голосом текст.

Авторы данной статьи пошли по пути создания собственного модуля МКВ и включили его в состав своей системы СДО под названием *HT-LINE* (ведущий программист-разработчик — Н.Н. Страхов). Почему нас не могли устроить готовые программные решения? Причины таковы: 1) необходимость реализации определенных методических приемов внутри модуля МКВ с целью возведения его в статус инструмента аттестации, а не только тренинга и самоконтроля; 2) необходимость автоматизированного сопряжения результатов студентов, выполняющих МКВ, с результатами других оценочных процедур, прежде всего тестов с выбором ответов на вопросы из большого банка заданий, а также традиционных процедур оценки письменных работ и устных ответов студентов.

Разработанный и включенный в систему *HT-LINE* модуль МКВ получил техническое название *Page-test*. Доступ к этому модулю, как и ко всей сетевой системе *HT-LINE*, реализован в режиме онлайн в виде подключения к веб-интерфейсу особого «Личного кабинета экзаменатора», позволяющего преподавателю создавать и проводить МКВ-тесты (как и обычные тесты) на произвольной группе студентов — тех студентов, которым выслано приглашение на страницу саморегистрации в данном веб-кабинете (в роли студентов).

Модуль *Page-test* поддерживает заполнение пропусков в произвольных текстах в двух основных режимах: 1) в режиме прямого ввода, или «свободного ответа», когда текст на экране появляется с окошками для ввода пропущенных слов (см. Приложение, рис. 1); 2) в режиме меню, т.е. выбора пропущенных слов (или букв, или знаков препинания — для проверки грамотности) из выпадающего списка (см. Приложение, рис. 2).

Кроме двух указанных режимов в интерфейсе для испытуемого существует целый ряд особых режимов (веб-страниц) для преподавателя-экзаменатора — владельца «Личного веб-кабинета»:

1. *Режим создания пропусков* в произвольном тексте (см. Приложение, рис. 3), где сам преподаватель (или его ассистент с необходимым уровнем компьютерной грамотности) вставляет в окно для редактирования выбранный им связный текст с помощью известных команд «копировать-вставить» и размечает ключевые слова-пропуски с помощью прямоугольных скобок. При этом для работы с выпадающими списками в прямоугольные скобки кроме правильных ответов (помеченных звездочками) вводятся неправильные ответы-дистракторы (которые не помечаются звездочками).

2. *Режим настроек теста*, создания инструкции и словника (см. Приложение, рис. 4), в котором владелец веб-кабинета (он же администратор теста МКВ) устанавливает галочки, выбирая те или иные режимы проведения МКВ и предъявления студенту обратной связи по тесту, а также строит с помощью автоматизированной подпрограммы словник, включающий все пропущенные слова и возможные дистракторы. Этот словник помогает студентам избегать опечаток (они просто переносят-копируют на место пропусков в тексте слова из словника), а также облегчает работу со сложными текстами, включающими около полусотни пропусков и более. Для коротких текстов (с числом пропусков менее 20) словник предъявлять не рекомендуется, так как он побуждает студентов работать не на понимание текста, а методом исключения слов, представленных в словнике.

3. *Режим просмотра результатов*, полученных студентами при выполнении данного теста (см. Приложение, рис. 5): в каждой строке дается ФИО студента, далее следуют его результаты в виде процента правильных ответов, затраченного времени (с точностью до секунд), интегрального индекса (учитывает и точность, и скорость одновременно), а также гиперссылки на полный протокол — с возможностью увидеть выделенные красным шрифтом ошибки.

4. *Режим анализа заданий* (*item analysis*, как этот инструмент чаще называют в мировой тестологии). Здесь для каждого пропуска дается подробная статистика и процента правильных ответов, и процента ошибок (с указанием текста каждой ошибки), а также

демонстрируется показатель КД — «коэффициент дискриминативности», который вычисляется как разность долей правильных ответов в «высокой» и «низкой» группах по данному тесту (Шмелев и др., 2008). Эта мера давно признана более прозрачной (для понимания авторами-преподавателями) процедурой оценки внутренней валидности-надежности заданий, чем другие более изощренные меры надежности (фи-коэффициент, точноно-бисериальный коэффициент и т.п.). В данном режиме разработчик теста может легко расширить круг ключевых слов путем маркирования галочками тех новых слов-синонимов, которые он не смог учесть при первоначальном составлении *Page-test*. После расширения списка ключевых слов программа позволяет автоматически пересчитать или уточнить баллы всех студентов.

Итак, какие ноу-хау присутствуют в нашем модуле *Page-test*?

1) расчет особого индекса качества результатов испытуемых IND, а именно определенного сочетания скорости и точности, оптимизированного по результатам анализа внешней валидности МКВ через поиск максимальных коэффициентов корреляции с другими оценочными процедурами;

2) гибкая работа со словариком (включение или отключение этого вспомогательного инструмента);

3) наличие особого режима анализа пунктов с подсчетом «коэффициента дискриминативности» (см. Приложение, рис. 6);

4) полуавтоматизированная настройка МКВ на реальную статистику ответов — расширение множества ключевых слов.

Психометрическая апробация МКВ (и техническое испытание программы *Page-test*)

Первое испытание нашего технико-методического решения *Page-test* состоялось в апреле—мае 2012 г. в ходе занятий практикума «Основы психодиагностики» на факультете психологии МГУ имени М.В. Ломоносова. Наряду с основной программой практикума студентам предлагалось выполнить три факультативных задания (с интервалом в 1 неделю), основанных на использовании программного модуля *Page-test*. Каждое из трех заданий предусматривало вставку 40 слов в специальный текст, написанный А.Г. Шмелевым с целью максимального насыщения специальной терминологией курса «Краткое руководство по адаптации зарубежного теста». Первое задание включало параграфы 1—6 этого руководства, второе — параграфы 7—9, третье — параграфы 10—14. Каждый текст насчитывал около 4000—5500 знаков (с пробелами), или 400—550 слов. Это примерно 2—3 страницы или 10—15% от авторского листа. Такой размер текста был эмпирически определен нами как наиболее адекватный после

серии первичных испытаний. Количество пропусков также было определено эмпирически и учитывало несколько факторов: а) *плотность* по числу пропусков на страницу (оказалось, что введение пропусков в каждой строке излишне усложняет задание; оптимально вводить примерно по одному пропуску на каждые две строки), б) *общая продолжительность* задания: студенты должны тратить на задание 20—30 минут. Это разумное время по организационным возможностям (ресурсам компьютерного класса для текущего контроля), а также по эргономической нагрузке на внимание и усилие семантического восприятия в ходе одного сеанса. Фрагмент первого факультативного тестового задания см. на рис. 1.

Студенты были замотивированы на прохождение факультативных заданий следующим образом: за успешное выполнение (более 80% правильных ответов) они могли получить 2 бонусных очка к сумме накопленных баллов по курсу «Основы психодиагностики» (к примеру, выполнение комплексного задания «Создание и проверка психометрических свойств тест-опросника» весит максимум 25 очков в работе студента). За 60—80% правильных ответов на один МКВ-тест студенты получали по одному бонусному очку. Срок выполнения факультативных заданий с возможностью получения бонусных очков ограничивался определенным интервалом времени (с 18.00 до 24.00 по пятницам). В последующие дни задания были доступны лишь для самоподготовки, то есть в режиме «тренинг-упражнения», без начисления каких-либо очков. Студентам предоставлялась возможность свободного доступа к заданиям не только из компьютерных классов факультета психологии МГУ, но также с домашних и мобильных компьютерных устройств. Студенты не были обязаны проходить эти факультативные задания.

Результаты

Три факультативных задания в режиме «борьбы за премиальные очки» студенты выполнили следующим образом: 37 студентов (1-е задание), 60 (2-е) и 74 студента (3-е задание). В среднем это по 30% от общего числа студентов (190), зарегистрированных на потоке. Рост популярности факультативного задания по числу участников очевиден. В табл. 1 приводятся данные по расчету статистических показателей для этих трех заданий.

Таким образом, самый главный результат — это высокие (особенно для 1-го задания) и статистически значимые при данном объеме выборок линейные корреляции результатов МКВ с баллами, полученными студентами при выполнении тестового экзамена. Этот экзамен состоял из 60 заданий с выбором ответа по материалам всего курса «Основы психодиагностики». Таким образом, более успешные

Результаты психометрической апробации *Page-test* на материале отрывков из текста «Краткое руководство по адаптации зарубежного теста»

№ теста	Среднее время	Средняя точность — процент правильных ответов	Средний индекс (IND) «точность-скорость»	Линейная корреляция IND с внешним критерием (тест-экзаменом)
1	31 мин 08 сек	89	4.7	0.50
2	32 мин 23 сек	88	4.6	0.43
3	26 мин 32 сек	95	5.5	0.38

на экзамене студенты лучше выполняют методику контекстной вставки. Интересно и важно заметить, что значимые корреляции были получены нами только для сводного индекса «скорость-точность», а для каждого отдельного показателя (скорости и точности) значимые корреляции не были обнаружены (корреляции оказывались в диапазоне 0.15—0.20).

Для 2-го и 3-го заданий корреляция оказалась ниже по следующим причинам: постепенно студенты осваивали практику обмена информацией о результатах выполнения методики, пересылали друг другу по электронной почте страницы с правильными ответами (они были отмечены зелеными словами, а ошибки — красными). Об этой тенденции говорит не только рост среднего показателя точности и увеличение популярности данного факультативного задания, но также появление «фальсифицированных результатов»: трем студентам пришлось пересдавать данный факультатив на глазах у старшего преподавателя в очном режиме, так как они показали неправдоподобные результаты по индексу «скорость-точность» (выполнили тест безошибочно и в разы быстрее, чем средний студент). Отметим, что в ходе пересдачи эти студенты, несмотря на знакомство с тестом, работали медленнее, чем из дома (с опорой на страницу с подсказками, присланную по электронной почте), т.е. в обычном, доступном для человека темпе — на скорости человека, реально читающего и понимающего текст, а не переставляющего готовые ответы из другого окна с образцом.

Таким образом, кроме получения первых обнадеживающих результатов по проверке внешней валидности методики в ходе этих испытаний выявилась вполне ожидаемая проблема, затрудняющая использование методики в качестве инструмента аттестационного контроля, — техническая легкость составления «шпаргалок» путем копирования экранов с уже сделанными вставками (как правило, студенты не повторяют ошибки, и по 3—5 студенческим протоколам

можно легко восстановить полную картину правильных ответов — тех, которые повторяются у хорошо подготовленных студентов). Эта проблема ограничивает «время жизни» методики в еще большей степени, чем это имеет место для обычных тестов типа *multiple-choice* (с выбором ответов из списка), особенно если обычные тестовые задания предъявляются не на одном электронном листе-окне (с движком прокрутки экрана), а поочередно на разных экранах. Для нейтрализации этого негативного эффекта требуется отключать многие обычные настройки компьютеров в компьютерном классе (включая выход в Интернет — на почтовые сервера, на справочные сайты типа Википедии и т.п.), а главное — преподаватели должны стоять за спинами студентов и смотреть за работой каждого (препятствуя списыванию из параллельных окон, фотографированию скриншотов и т.п.). Но даже при таких расходах живого труда автоматизированный вариант МКВ (на платформе *Page-test*, в частности) требует меньших затрат, нежели проверка преподавателями контрольных работ вручную.

Впереди — более детальное исследование преимуществ МКВ по сравнению с обычными тестами в случаях со студентами, склонными к более глубокому погружению в определенный проблемный контекст. Некоторые пока лишь косвенные свидетельства конвергентной и дискриминантной (дифференцирующей) валидности МКВ принесло наше второе исследование с помощью платформы *Page-test*.

МКВ и знание русского языка

Вторым испытанием нашего программно-технического решения *Page-test* стало использование его возможностей для тестирования знания русского языка (орфографии и пунктуации) по методике так называемого «тотального диктанта». Эта методика была использована нами с разрешения ее разработчиков — авторов, в том числе создавших сайт totaldict.ru — и была адаптирована для платформы *Page-test* Д.С. Портновой. Наша версия тотального диктанта (см. рис. 2) предъявлялась участникам онлайн-клуба КИТТ («Клуба испытателей тестовых технологий») на сайте www.ht.ru, насчитывающем более 5000 зарегистрированных участников. Использованный нами вариант «тотального диктанта» включал 70 заданий (из них 25 по орфографии и 45 по пунктуации). В общей сложности в двух разных сериях-рассылках его выполнили 514 участников клуба (актуально на 30 июня — момент подготовки рукописи этой статьи). Для тех 115 участников, которые параллельно выполнили онлайн-тест «гуманитарного интеллекта», подсчиты-

**Коэффициенты линейной корреляции результатов
«Тотального диктанта» (МКВ) и субшкал традиционного
компьютерного теста интеллектуальных способностей
(n=115, уровень значимости для коэффициента 0.41 максимальный
из возможных стандартных $p < 0.001$)**

Диктант\Тесты	Лексика	Эрудиция	Абстрактная логика
Процент правильных ответов	0.41***	0.20	0.22
Сводный индекс	0.25*	0.25*	0.13

валились корреляции между обычным баллом (процент правильных ответов) и сводным показателем IND по «тотальному диктанту», с одной стороны, и баллами по трем субтестам «гуманитарного интеллекта» — с другой. Результаты корреляционного исследования представлены в табл. 2.

Почему в данном случае сводный индекс дал более низкую корреляцию (хотя тоже статистически значимую), чем более простой показатель — нормированный процент правильных ответов? Видимо, дело в мотивации испытуемых. Студенты в нашем первом эксперименте были сильнее мотивированными, а некоторые участники клуба, начиная с середины диктанта, могли сбиваться либо на тактику «прощелкивания», либо могли просто отвлекаться на какое-то время на другие домашние дела. Тем не менее полученный ожидаемый более высокий коэффициент корреляции именно с субшкалой «Лексика» (вербальный интеллект) подтверждает в целом валидность примененной технологии *Page-test*, основанной на методе контекстной вставки.

* * *

В ходе апробации технологии *Page-test* выявлены перспективные возможности и некоторые организационные и методические ограничения, связанные с методом контекстной вставки как инструментом контроля учебных достижений студентов. Ожидается рост популярности этого метода по мере освоения преподавателями компьютерной грамотности и/или прихода нового поколения молодых преподавателей. К настоящему моменту разработчики-тестологи должны изучить свойства данного метода, нормативы качественных и количественных параметров его проведения, границы применимости платформы *Page-test* для управления учебной деятельностью студентов, а также для оценки ее результатов.

Приложение

КРАТКИЕ МЕТОДИЧЕСКИЕ УКАЗАНИЯ (ПАМЯТКА) АВТОРУ АДАПТИРОВАННОЙ ВЕРСИИ ПСИХОДИАГНОСТИЧЕСКОЙ МЕТОДИКИ (часть 2, параграфы 7-9)

7. Первичная [] теста. После [] оценки тест должен быть предъявлен не менее чем 10 реальным испытуемым, чтобы уточнить, насколько хорошо [] понимают [] к тесту и смысл отдельных тестовых заданий. По замечаниям испытуемых и результатам апробации все задания и элементы инструкции, по которым получены нарекания со стороны испытуемых, должны быть переформулированы. Также следует обратить внимание на те тестовые задания, по которым все [], принявшие участие в апробации, дали одинаковые ответы. Применения особых статистических [] результаты первичной апробации не требуют. Если в результате первичной апробации в тест не внесены никакие коррективы, то результаты апробации включают в массив протоколов следующего этапа - [] исследования. Иногда именно по результатам апробации становится понятным, по каким заданиям в пилотной версии теста нужно создать избыточность (несколько разных формулировок) - с расчетом на отсеивание неудачных заданий после проверки надежности.

8. Пилотное исследование []. Чаще всего пилотное исследование надежности предшествует другим исследованиям, предпринимаемым с целью проверки [] и обеспечения репрезентативности тестовых []

Рисунок-скриншот 1. Режим респондента для прямого ввода пропущенных слов в пропуски-окошки

И нам не всё равно!

Есть тихое [?] как зуд [?] ощущение, что государство на этой земле н [?] кому [?] то не должно. Может [?] поэтому в последнее время мы так час [?] им от людей, что и я [?] мол [?] н [?] кому н [?] чего не должен. И вот [?] я не понимаю [?] как всем нам здесь выжить [?] и кто станет защищать эту страну, когда она обвалится?

Если всерьёз поверить, что Россия и [?] черпала ресурсы жизнестойкости [?] и будущего у нас нет [?] то, право слово, может [?] и переживать не стоит? Причины у нас веские [?] народ надломлен [?] все империи рано или поздно распадаются [?] и шансов у нас поэтому нет.

Российская история, не спорю, пр [?] воцировала подобные декл [?] рации. Тем не менее [?] наши предки в эти пораже [?] ые келт [?] цизмом благоглупости н [?] когда не верили. Кто решил, что у нас уже нет шансов, а, к примеру, у китайцев их больше [?] чем достаточно? У них ведь тоже многонациональная страна [?] пережившая р [?] волюции и войны.

На самом деле [?] мы живем в забавном государстве. Здесь, чтобы

Рисунок-скриншот 2. Режим респондента с выбором ответа из выпадающего списка (на примере «Тотального диктанта-2012»)

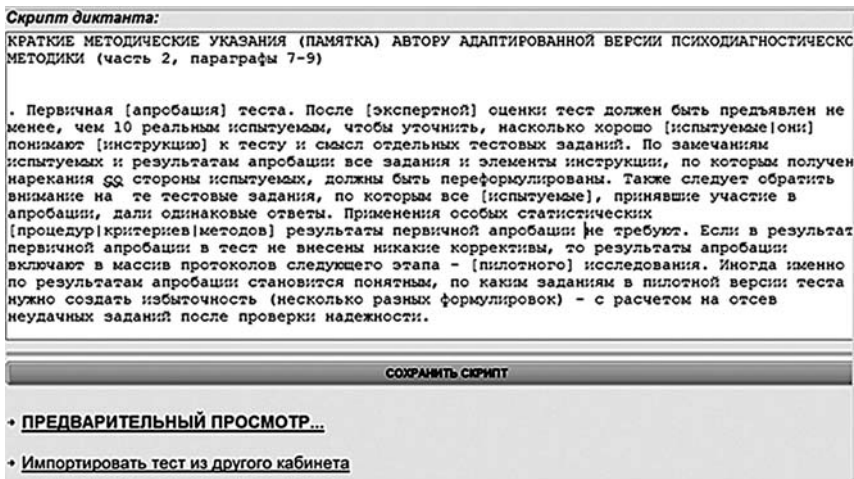


Рисунок-скриншот 3. Режим подготовки и редактирования текста преподавателем-экзаменатором: пропускаемые слова обрамляются прямоугольными скобками, возможные синонимы перечисляются через символ «|»

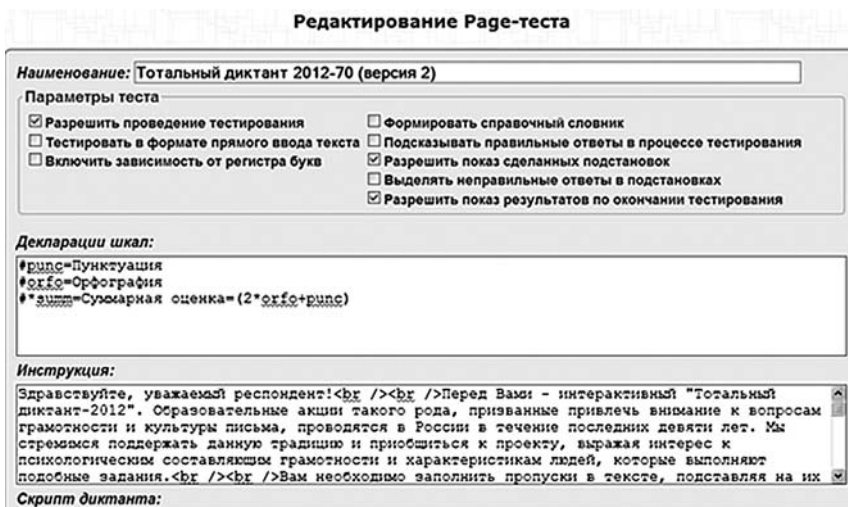


Рисунок-скриншот 4. Режим настроек теста: галочками помечаются возможные варианты настроек (опции), в отдельные окна вписываются названия субшкал и инструкция испытуемому

Список протоколов тестирования по плану «Тест № 15»

Идентификатор: 24
Способ тестирования: С прямой подстановкой
Число субтестов: 40
Общее число альтернатив: 56
Число протоколов: 74
Средний балл: 95 из 100
Среднее время выполнения: 1592 сек. (26м. 32 с.)

+ [Экспортировать групповые результаты в Excel](#)

+ [Пересчитать все результаты по тесту](#)

+ [Анализ тестовых заданий](#)

№	Фамилия Имя Отчество	Балл	Δ Дата / время	T сек	ИНДЕКС	Результаты...
1.	<input type="checkbox"/> Марина Анна Александровна	=58	(02.05.2012 /18:40)	<658>	[5.1]	ПОКАЗАТЬ
2.	<input type="checkbox"/> Тихобразова Жанна Сергеевна	=88	(02.05.2012 /18:40)	<656>	[11.8]	ПОКАЗАТЬ
3.	<input type="checkbox"/> Мартынова Елена Александровна	=100	(02.05.2012 /18:40)	<676>	[17.4]	ПОКАЗАТЬ
4.	<input type="checkbox"/> Курьянова Елизавета Сергеевна	=68	(29.04.2012 /00:09)	<1237>	[3.7]	ПОКАЗАТЬ

Рисунок-скриншот 5. Режим просмотра результатов, полученных испытуемыми (на примере третьего факультативного задания)

Анализ тестовых заданий

№	текст до	Альтернативы	послетекст	Правильно	Неправ.	КД
1.	...артизация тестовой	шкалы ×73 <input type="checkbox"/> выборы ×1	и обеспечение...	≈73 (99%)	#1 (1%)	0.04
2.	и обеспечение	репрезентативности ×73 <input checked="" type="checkbox"/> реперзентативности ×1	тестовых норм...	≈73 (99%)	#1 (1%)	0.04
3.	...товой шкалы размер	выборки ×74	должен быть н...	≈74 (100%)	#0 (0%)	0
4.	... стандартной шкалы	стенов ×74	десять град...	≈74 (100%)	#0 (0%)	0
5.	...и применения более	тонах ×11 дробных ×22 дифференцированных ×38 <input type="checkbox"/> дифференциальных ×1 <input checked="" type="checkbox"/> дифференцированных ×1	стандартных ш...	≈71 (97%)	#2 (3%)	0.07
6.	...исленность выборы	стандартизации ×68 <input type="checkbox"/> компьютеризации ×1 <input type="checkbox"/> испытуемых ×1 <input type="checkbox"/> т-баллов ×1	должна обеспе...	≈68 (96%)	#3 (4%)	0.21
7.	...и силы» на шкале	т-баллов ×15 т-баллов ×54 <input type="checkbox"/> стенов ×4	таких интерва...	≈69 (95%)	#4 (6%)	0.18
8.	...альная численность	выборки ×68 <input type="checkbox"/> интервалов ×1	для Т-шкалы –...	≈68 (93%)	#5 (7%)	0.14

Рисунок-скриншот 6. Режим анализа тестовых заданий (при пометке галочкой опечатка в слове «репрезентативности» засчитывается испытуемому как правильный ответ)

СПИСОК ЛИТЕРАТУРЫ

Болсинова М.А., Шмелев А.Г. Взаимосвязь индивидуально-личностных особенностей студентов и успешности прохождения экзамена в компьютеризированной и традиционной форме // Психологические исследования: эл. журнал. URL: <http://www.psystudy.ru>. 2010. № 4.

Зинченко Ю.П., Березанская Н.Б., Володарская И.А., Тихомандрицкая О.А., Шмелев А.Г. Опыт внедрения компьютеризированных тестовых испытаний в систему итоговой государственной аттестации студентов-психологов // Вестн. Моск. ун-та. Сер.14. Психология. 2011. № 2. С. 135—153.

Чумаков А.А., Шмелев А.Г. Опыт компьютеризированного тестового контроля знаний у студентов-психологов // Тез. Всеросс. конф. «Прикладная психология как ресурс социально-экономического развития современной России» (Москва, 17—19 ноября 2005). М., 2005. С. 103—105.

Шмелев А.Г. Компьютерное тестирование знаний у студентов-психологов // Вестн. Моск. ун-та. Сер. 14. Психология. 2003. № 1. С. 35—48.

Шмелев А.Г. «Песнь о Вещем Олеге» и профессиональные предпочтения в области психодиагностики // Вестн. Моск. ун-та. Сер. 14. Психология. 2004. № 4. С. 13—23.

Шмелев А.Г., Чумаков А.А., Ларионов А.Г., Серебряков А.Г. Методические рекомендации по разработке и внедрению системы оценки качества обучения по инновационным образовательным программам. М., Из-во Моск. ун-та, 2008.

ВЕБЛИОГРАФИЯ

- 1) http://en.wikipedia.org/wiki/Closed_testing_procedure — описание метода
- 2) <http://www.vashpsixolog.ru/psychodiagnostic-school-psychologist/61-diagnosis-of-intellectual-development/495-test-ebbingauza-filling-in-missing-words-in-the-text> — описание метода
- 3) <http://elearningsoft.ru/quizforce.html> — демоверсия инструментария по созданию упражнений.
- 4) <http://www.learnclick.com/cloze/show/> — демонстрация технологии с интерфейсом по типу «тяги-бросай» (drug-and-drop).
- 5) <http://users.ht-line.ru/page-test> — входная страница для самозаписи на дистанционный мастер-класс по технологии *Page-test* в системе *HT-LINE*.

Поступила в редакцию
03.07.12