

УДК 371.66:004.3

**Комаринский С.М.**

## **ОСНОВНЫЕ РЕЗУЛЬТАТЫ ЭКСПЕРТНОГО ОЦЕНИВАНИЯ КАЧЕСТВА ТЕСТОВЫХ ПОДСИСТЕМ ПРОГРАММ ДИСТАНЦИОННОГО ОБУЧЕНИЯ**

*Ключевые слова:* дистанционное обучение, экспертное оценивание, тестовые подсистемы, тестовые задания.

В нашем исследовании мы намеревались выяснить тенденции разработок тестовых материалов вузовскими педагогами – разработчиками тестовых подсистем дистанционного обучения (ДО) в педвузах двух типов – обычном педагогическом вузе (Таганрогский государственный педагогический институт) и вузе, осуществляющем переподготовку педагогических кадров (Ростовский институт повышения квалификации и профессиональной переподготовки работников образования). Оба вуза занимают прочные лидерские позиции в Южно-Российском регионе, имеют сильный педагогический состав (в ТГПИ процент докторов и кандидатов наук не уступает университетским аттестационным требованиям, а в РИПК и ППРО преподают руководители ведущих научно-педагогических школ г. Ростова-на-Дону – проф. Е.В. Бондаревская, проф. В.Т. Фоменко и др.). По кадровому составу разработчиков и экспертов тестовых материалов оба вуза находились «на верхнем пределе», и по ожидаемым результатам нами предполагалось на основе анализа качества используемых на практике авторских тестовых материалов определить тенденцию их конструирования как для тестового контроля в программах ДО (оба вуза), так и для рубежного (ТГПИ); выявить возможные различия в конструировании тестовых заданий (ТЗ) для студентов педвуза (ТГПИ) и для педагогов, повышающих свою квалификацию (при том, что в обоих вузах контингент разработчиков не ниже доцента, кандидата педагогических наук с 5 и более годами стажа работы по специальности).

Как рабочие эмпирические гипотезы нами были выдвинуты следующие: – достаточно существенными могут быть различия в содержании ТЗ по доминирующим типам требуемой

- деятельности испытуемых и менее значимыми – по форме их проявления;
- всеми разработчиками будет отдаваться предпочтение закрытым ТЗ с многопозиционным выбором (4–5) верного ответа.
  - дихотомические закрытые ТЗ (ответ «да – нет») будут в явном меньшинстве во всех наборах ТЗ для всех контингентов;
  - чем «взрослее» контингент тестируемых – тем разнообразнее по форме и содержанию будут разработанные для них ТЗ;
  - в отдельных пакетах (наборах) ТЗ по различным предметам каждая группа разработчиков ТЗ будет представлять близкие по структуре, содержанию и форме ТЗ. То есть сработает общность базовых методических установок на разработку ТЗ для каждого обследуемого контингента, и пакеты ТЗ разных кафедр одного вуза (ТГПИ, РИПК и ППРО) будут по типологии ТЗ близки между собой, отличаясь все же по каким-то еще не выявленным параметрам от пакетов ТЗ другого вуза.

Изначально предполагалось, что на структуру и содержание разрабатываемых тестовых материалов могут существенно влиять такие факторы, как тип учебного заведения, в котором трудятся разработчики (преподаватели гуманитарных и технических вузов, гуманитарных и негуманитарных дисциплин); характер учебных дисциплин, которые они сами ведут и по которым разрабатывали ТЗ (традиционный стабильный курс или авторский спецкурс); стаж работы по профессии и возраст (как показатели опытности и, возможно, консервативности установок по отношению к предпочтительно конструируемым типам тестовых заданий).

В РИПК и ППРО нами при участии старшего методиста ИПК М.Е. Инькова было в совокупности подвергнуто экспериментальному анализу 212 тестовых заданий, что потенциально позволяло применять широкий спектр методов математической статистики. В ТГПИ выборочному экспертному анализу подвергались тестовые материалы учебно-методических комплексов по психолого-педагогическим дисциплинам разных факультетов – в совокупности 248 ТЗ.

Анализ общей ситуации с разработкой и применением ТЗ показал следующую картину, подтвержденную диаграммами, представленными ниже:

- тестовые задания ориентированы только на итоговый контроль усвоения курса или учебного модуля образовательной программы;
- в аспекте содержательной валидности все тестовые материалы адекватно отражают ключевые цели и задачи курсов и модулей, их разработчикам в основном (за очень редким исключением) удалось уйти от попыток изменений детализированного контроля фактологической части курсов (что является воспроизводящей ошибкой многих разработчиков тестов ЕГЭ и аттестационного вузовского тестирования);
- доминирует ориентация на небольшие по объему тестовые пакеты (от 8 до 20 тестовых заданий, в среднем по 10–12 ТЗ).

В содержательной части тестовых заданий выявлены следующие закономерности:

- явное предпочтение теоретических ТЗ (об исключениях – ниже);
- по особенностям постановки задания тестовой задачи доминируют задачи исполнения;
- доминируют четко сформулированные ТЗ;

- доминируют ТЗ с полной информацией, представляемой для решения;
- представлены преимущественно рутинные, т.е. привычные для обследуемых по содержанию задачи;
- почти все ТЗ – кадровые по форме предъявления;
- в большинстве случаев (90%) ТЗ множественные по выбору вариантов ответа либо дихотомические; открытые задания (с конструируемым вариантом ответа, более трудоемкие в обработке и позволяющие субъективно трактовать ответы, практически отсутствуют.

Отмеченное свидетельствует о достаточно высокой профессиональной солидарности всех разработчиков тестовых материалов в ТГПИ и РИПК и ППРО с разработчиками тестов для школьников – относительно предпочитаемых типов ТЗ в их содержательном аспекте.

В то же время, как показывают данные, полной унификации типов используемых ТЗ нет. Так, из 29 экспертируемых ТЗ на кафедре педагогики РИПК и ППРО 4 были не теоретической,

а практической ориентации, 3 ТЗ – с нечеткой формулировкой, требующей включения креативного мышления, 9 заданий – на использование (и выбор) процедуры

Хотя и редко, но на всех кафедрах встречались блочные независимые ТЗ (значительно более сложные и в разработке, и в решении, чем обычные кадровые).

Наряду с использованием базовой аналитической методики нами проверялась в пробном внедрении и ее модификация, заключающаяся в том, что все характеристики тестовых заданий были переведены в систему шкальных оценок в рамках традиционной пятибалльной шкалы. Это позволяло (независимо от числа ТЗ в данном конкретном пакете тестового контроля) использовать среднеарифметические величины для характеристики каждого показателя. Тем самым достигалась сопоставимость пакетов тестовых заданий по различным учебным дисциплинам.

На рис. 1 демонстрируется результат сопоставления качественных характеристик тестовых заданий на факуль-

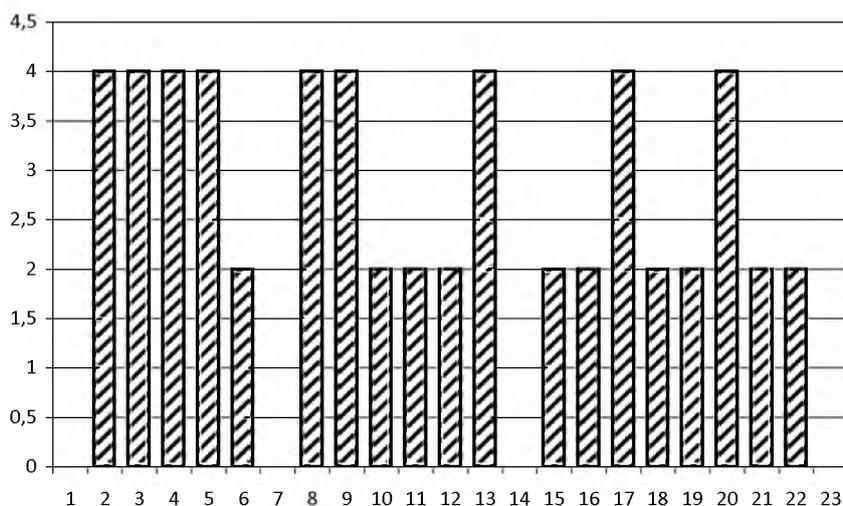


Рис. 1. Комплекс оценок тестовых материалов по пятибалльной шкале – предмет «Социальная психология»

тете социальной педагогики ТГПИ по трем учебным дисциплинам ДО.

Характеристики тестовых заданий:

1. Определение целей и задач теста.
2. Доминирующий вид деятельности при выполнении заданий (распознавание, конструирование, объяснение).
3. Способ познания при выполнении заданий.
4. Способ выполнения заданий.
5. Адекватность заданий требованиям Госстандарта.
6. Ориентированность на учебные пособия.
7. Обоснованность шкал обработки.
8. Композиция предъявления теста.
9. Количество тестовых заданий в одном варианте.
10. Расположение тестовых заданий по темам.
11. Форма тестовых заданий.
12. Количество дистракторов.
13. Количество ответов по полноте информации.
14. Охват материала в тесте.
15. Внутренняя согласованность заданий теста.
16. Трудность теста.
17. Нацеленность формулировки на адекватность восприятия.
18. Характер постановки вопроса.
19. Степень разрешимости задания.
20. Четкость формулировки задания.
21. Особенность формулировки.
22. Согласованность дистракторов.
23. Эталон ответа.

Совершенно очевиден в данном случае достаточно высокий уровень конструирования ТЗ по следующим характеристикам:

- № 2. Доминирующий вид деятельности при выполнении заданий (распознавание, конструирование, объяснение).
- № 3. Способ познания при выполнении заданий.

- № 4. Способ выполнения заданий.
- № 5. Адекватность заданий требованиям Госстандарта.
- № 8. Композиция предъявления теста.
- № 9. Количество тестовых заданий в одном варианте.
- № 13. Количество ответов по полноте информации.
- № 17. Нацеленность формулировки на адекватность восприятия.
- № 20. Четкость формулировки задания.

Методисты-разработчики достаточно четко ориентируются в локальных задачах контроля изученного учащимися материала и в аспекте обеспечения содержательной (программной) валидности тестовой подсистемы применяют задания на все указанные аспекты диагностики. Они хорошо соотносят требуемые и реально реализуемые учащимися способы выполнения заданий с учетом Госстандарта освоения учебного материала дисциплины.

В целом хорошо прослеживается в тестовой подсистеме задуманная разработчиками композиция тестовых заданий, хотя в тестовых подсистемах представляются лишь наиболее распространенные формы конструирования ТЗ – кадровые ТЗ.

Количество ТЗ в тестовых подсистемах рубежного и итогового контроля минимально достаточное, что позволяет избежать перегрузки обследуемых и больших трудозатрат диагноста на процедуры тестирования и обработки результатов. Явно предпочитают ТЗ с полной информацией – как более простые при разработке (но и нацеленные в основном на уровень распознавания, т.е. элементарного знакомства обследуемого с темой (или, точнее, с ее фрагментом).

Практически все задания даются «без подвоха» и нацелены на адекват-

ное их восприятие, обладают четкой формулировкой.

Заданий, провоцирующих проблемные ситуации и их анализ, практически нет. Отсюда и четкость формулировок, облегчающая решение задания, но не диагноз рефлексивных особенностей обследуемых.

В то же время «западают»:

- № 6. Ориентированность на учебные пособия.
- № 10. Расположение тестовых заданий по темам.
- № 11. Форма тестовых заданий.
- № 12. Количество дистракторов.
- № 16. Трудность теста.
- № 18. Характер постановки вопроса.
- № 19. Степень разрешимости задания.

Практически не обозначены в методических материалах к тестированию:

- № 7. Обоснованность шкал работы.
- № 14. Охват материала в тесте

«Западающие» аспекты экспертируемых тестовых подсистем ДО вызваны, как показывают уточняющие беседы с разработчиками, различными по характеру причинами.

Так, ориентированность ТЗ на конкретные учебные пособия оказалась сравнительно невысокой в связи с тем, что в профессионально-педагогическом образовании в большинстве случаев преподавание идет с опорой одновременно на несколько конкурирующих учебных пособий.

Из них в лекционно-семинарские курсы преподаватели выбирают разные разделы, которые им представляются наиболее соответствующими Госстандарту и учебным программам.

В низких оценках рациональности расположения ТЗ в тестовых подсистемах ДО, на наш взгляд, просвечивает дефицит тестологических знаний раз-

работчиков. Потенциально и в ТГПИ, и в РИПК и ППРО имеются возможности пополнять эти знания по литературе, изданной в Ростове-на-Дону и Таганроге по методологии и технологии разработки дидактических тестов [1–9]. Можно получить эти знания и на курсах повышения тестологической компетентности, программы которых разработаны специалистами в этой области – Е.А. Михайлычевым, С.А. Сафонцевым.

Дефицитом технолого-тестологической компетентности разработчиков тестовых подсистем ДО может быть объяснено и «западание», низкие оценки экспертов по таким аспектам тестовых подсистем, как удачность выбора формы предъявления тестовых заданий, определение количества дистракторов (отвлекающих вариантов ответа), характер постановки вопросов в ТЗ. А вот такие моменты, как неадекватная, по мнению экспертов, трудность теста (набора ТЗ в тестовой подсистеме) и степень разрешимости заданий (т.е. сложность заданий, заложенная разработчиком), более связаны с методологическими тестологическими знаниями и опытом их реализации при совершенствовании и стандартизации тестовых подсистем.

Для определения трудности и отдельных ТЗ, и тестовой подсистемы в целом требуется накопление и тестологический анализ результатов, технология которого известна [10–16]. Это трудоемкая, но более простая работа, чем проектирование рациональной степени сложности ТЗ. В стране имеются компьютерные программы тестирования, позволяющие собирать и базы данных, и персональные результаты тестирования каждого учащегося.

Для рационального определения сложности ТЗ необходим опыт проектирования конструкта – модели теста,

опыт методологической работы требуется и для обоснования шкал обработки тестовых материалов, для решения вопросов проектирования и обоснования содержательной программной валидности теста.

Фоновыми, но значимыми причинами существенной ограниченности тестологических знаний разработчиков тестовых подсистем являются их перегруженность учебной работой и слабое стимулирование освоения ими диагностических (в том числе и тестологических) технологий.

Можно сделать выводы о том, что при общем знакомстве с тестовыми формами разработчики слабо ориентируются в вопросах конструирования содержательно валидного тестового комплекса, недостаточно владеют технологиями композиции ТЗ и проектирования теста как определенной целостности.

Заметим, что по другим учебным дисциплинам ситуация с разработкой ТЗ достаточно серьезно отличается (рис. 2), хотя сильными сторонами их конструирования остаются:

- № 2. Доминирующий вид деятельности при выполнении заданий

(распознавание, конструирование, объяснение).

- № 3. Способ познания при выполнении заданий.

- № 4. Способ выполнения заданий.

При использовании приведенных выше критериев экспертного оценивания наборов тестовых заданий в тестовых подсистемах ДО, как видно из диаграмм, ряд критериев в них не представлен, поскольку описывался в качественных характеристиках:

- определение цели и задач тестовой подсистемы (1-я позиция);
- обоснованность шкал обработки (7-я позиция);
- охват учебного материала в тестовой подсистеме (14-я позиция);
- эталон ответа (23-я позиция).

Сравнивая диаграммы, построенные с помощью стандартного пакета программных средств, мы имеем основания считать достаточно высокой диагностичность предложенной нами методики анализа пакетов тестовых материалов. Она позволяет рельефно выделять сильные и слабые стороны конструирования пакетов тестовых заданий, ориентируя разработчиков на «адресный» ремонт

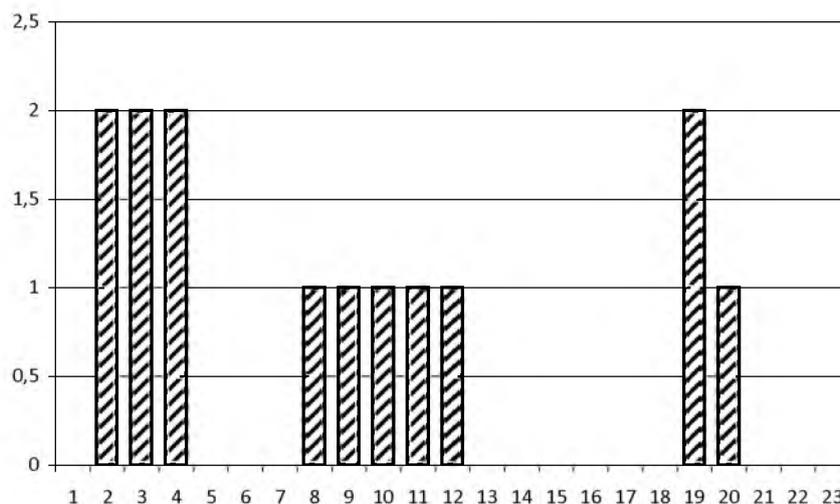


Рис. 2. Комплекс оценок тестовых материалов по пятибалльной шкале – предмет НПОО

такого пакета по четко определенным критериям.

Диаграммы показывают и такие общие слабые места конструирования комплексов тестовых материалов, как композиция предъявления теста, расположение тестовых заданий по темам, рациональная форма тестовых заданий, количество дистракторов, четкость формулировки задания. Все это вопросы общей методологии конструирования теста как диагностического инструментария, и без систематического обучения освоить тестовые технологии самостоятельно вряд ли удастся даже достаточно опытным преподавателям (что, собственно, и продемонстрировали представленные выше диаграммы).

Выводы проведенного исследования непосредственно ориентируют разработчиков программ тестологической подготовки педагогов на особое внимание к отработке тех навыков практического составления тестовых материалов, которые оказались наиболее слаборазвитыми. Проведенный экспертами анализ тестовых подсистем показал, что в нарушение известных требований к обязательной сертификации используемых в учебном процессе тестов ни одной представленной тестовой подсистеме не сопутствовали сертификация тестов, сведения об их надежности и валидности, обоснования оценочных шкал. Результаты пробного внедрения экспертной методики анализа качества тестовых подсистем ДО показали общий невысокий уровень тестологической компетентности разработчиков тестовых подсистем ДО, существенно мало отличающийся от компетентности разработчиков тестовых материалов для школьников.

Неадекватность практики разработки тестовых подсистем современным требованиям дидактической тестоло-

гии и достижениям компьютерного тестирования обусловлена, на наш взгляд, характером организации создания этих подсистем на управленческих уровнях, а именно игнорированием зарубежного опыта создания специализированных тестологических центров, проводящих высокопрофессиональную методологическую работу по конструированию и стандартизации тестовых систем и подсистем, ориентацией управленцев системы образования на самостоятельное освоение разработчиками тестологических компетенций (чего не происходит, судя по результатам диагностического эксперимента).

#### Библиография

1. Инструментарий педагогической диагностики и мониторинга образовательных процессов / под ред. Е.А. Михайлычева. Таганрог: Изд-во ТГПИ, 2000.
2. *Кадневский В.М.* Генезис тестирования в теории отечественного образования. Екатеринбург: Уральск. гос. пед. ин-т, 2006.
3. *Механцев Б.Е.* Опыт применения математических методов для тестологического анализа // Модернизация отечественного педагогического образования: проблемы, подходы, решения. Таганрог, ТГПИ, 2005. С. 316–324.
4. *Михайлычев Е.А.* Дидактическая тестология. М.: Народное образование, 2001.
5. *Михайлычев Е.А.* Современная российская дидактическая тестология: история становления и развития до конца XX века. Таганрог; Ростов н/Д, 2009.
6. *Михайлычев Е.А.* Современная российская дидактическая тестология: методология и научный статус дидактической тестологии. Таганрог; Ростов н/Д, 2010.
7. *Михайлычев Е.А., Механцев Б.Е., Иньков М.Е.* Современная российская дидактическая тестология: развитие теории и практики дидактического тестирования. Таганрог; Ростов н/Д, 2010.
8. *Переверзев В.Ю.* Критериально-ориентированные педагогические тесты для итоговой аттестации студентов. М.: Изд-во НМЦ СПО Минобразования РФ, 1998.
9. *Designing Courses for Distance Learners*, 1994. Hagerstown: Institute for Distance Education University System of Maryland.
10. *Аванесов В.С.* Тесты в социологическом исследовании. М.: Наука, 1982.

11. *Анастаси А.* Психологическое тестирование: в 2 кн. М.: Педагогика, 1982. Кн. 1.
12. *Анастаси А., Урбина С.* Психологическое тестирование. 7-е изд. СПб., 2002.
13. *Клайн П.* Справочное руководство по конструированию тестов. Киев, 1994.
14. *Михайлычев Е.А., Механцев Б.Е., Иньков М.Е.* Указ. соч.
15. *Davie, L.E. and R. Inskip,* 1992. Fantasy and Structure in Computer Mediated Courses. *Journal of Distance Education*, 2: 31–50.
16. *Flowers, L., E.T. Pascarella and C.T. Pierson,* 2000. Information technology use and cognitive outcomes in the first year of college. *The journal of higher education*. Columbus, 71 (6): 637–667.
6. *Mikhaylychev, E.A.,* 2010. Modern Russian didactic testology: methodology and scientific status of didactic testology. Taganrog; Rostov-on-Don. (rus)
7. *Mikhaylychev, E.A., B.E. Mekhantsev and M.E. Inkov,* 2010. Modern Russian didactic testology: development of the theory and practice of didactic testing. Taganrog; Rostov-on-Don. (rus)
8. *Pereverzev, V.Yu.,* 1998. Criterial pedagogical tests for final certification of students. Moscow: published by Publishing house of the Ministry of Education of the Russian Federation. (rus)
9. *Designing Courses for Distance Learners,* 1994. Hagerstown: Institute for Distance Education University System of Maryland.

#### *Bibliography*

1. *Mikhajlychev, E.A. (Ed.),* 2000. Toolkit for pedagogical diagnostics and educational processes monitoring. Taganrog: Publishing house of Taganrog State Pedagogical University. (rus)
2. *Kadnevsky, V.M.,* 2006. Genesis of testing in the theory of national education. Ekaterinburg: Ural State Pedagogical University. (rus)
3. *Mekhantsev, B.E.,* 2005. Experience of applying mathematical methods for test-based analysis. Modernization of national pedagogical education: problems, approaches, decisions. Taganrog: published by Taganrog State Pedagogical Institute: 316–324. (rus)
4. *Mikhaylychev, E.A.,* 2001. Didactic testology. Moscow: published by national education. (rus)
5. *Mikhaylychev, E.A.,* 2009. Modern Russian didactic testology: history of development up to the end of XX century. Taganrog; Rostov-on-Don. (rus)
10. *Avanesov, V.S.,* 1982. Tests in sociological researches. Moscow: published by Nauka. (rus)
11. *Anastazy, A.,* 1982. Psychological testing: in 2 books (book 1). Moscow: published by Pedagogika. (rus)
12. *Anastazy, A. and S. Urbina,* 2002. Psychological testing. 7<sup>th</sup> ed. St. Petersburg. (rus)
13. *Kline, P.,* 1994. Guidelines for designing tests. Kiev. (rus)
14. *Mikhaylychev, E.A., B.E. Mekhantsev and M.E. Inkov.* Op. cit.
15. *Davie, L.E. and R. Inskip,* 1992. Fantasy and Structure in Computer Mediated Courses. *Journal of Distance Education*, 2: 31–50.
16. *Flowers, L., Pascarella, E.T. and C.T. Pierson,* 2000. Information technology use and cognitive outcomes in the first year of college. *The journal of higher education*. Columbus, 71 (6): 637–667.