

МЕТОД СЕРТИФИКАЦИОННОГО ОЦЕНИВАНИЯ

Живицкая Е.Н.

*Белорусский государственный университет информатики и радиоэлектроники,
г. Минск*

Тестовые технологии становятся существенным компонентом образования благодаря пониманию таких преимуществ тестов, как технологичность, точность, нацеленность на объективность. Эти преимущества возникают только при условии научного обоснования тестового процесса.

Тесты способствуют эффективной организации учебного процесса, без них немислима сертификация специалистов. Использование автоматизированного обучения и контроля на основе достижений новой педагогики, педагогических измерений, психологии, информатики, кибернетики и компьютерной техники постепенно становится нормой.

Сертификационные тесты позволяют получить объективные оценки уровня знаний, умений, навыков, проверить соответствие требований к подготовке выпускников заданным стандартам, выявить пробелы в знаниях, повысить качество сертификационного оценивания специалиста. В сочетании с персональными ЭВМ и программно-педагогическими средствами, тесты помогают перейти к созданию современных систем адаптивного обучения и адаптивного контроля.

В адаптивных обучающих системах сертификационное тестирование может применяться как средство идентификации личности для построения индивидуальной последовательности обучения с возможностью выдачи сертификата при подтверждении определенного уровня подготовки. При этом можно различать три вида тестирования:

1) Предварительное тестирование применяется перед началом обучения и направлено на выявление предварительных знаний обучаемого по ряду дисциплин, которые ему предстоит изучать. Сюда же могут включаться психологические тесты для определения индивидуальных характеристик личности обучаемого, которые учитываются в ходе обучения для настройки на работу с конкретным обучаемым. По результатам предварительного тестирования строится предварительная последовательность изучения учебных курсов.

2) Текущее тестирование – это контроль или самоконтроль знаний по отдельному элементу учебного курса, например, разделу или теме. По его результатам строится последовательность изучения тем и разделов внутри курса, а также может осуществляться возврат к темам, которые были изучены недостаточно хорошо.

3) Итоговое тестирование – это контроль знаний по курсу в целом или по совокупности курсов. По его результатам корректируется последовательность изучения отдельных тем или учебных курсов.

Разработка метода составления теста для сертификационного оценивания, который наиболее точно может оценить знания тестируемых, является актуаль-

ной задачей. Предлагаемый метод позволяет оценить сертификационный тест и выявить задания, которые являются неприемлемыми для оценки знаний тестируемых посредством удаления из теста заданий (вопросов), которые являются слишком легкими или слишком сложными, а также определения заданий на доработку.

На основе анализа существующих способов оценки качества тестов предлагается трехэтапный метод оценки сертификационных тестов.

На **первом этапе** определяется индекс трудности составного компонента теста. Он является показателем, характеризующим количество испытуемых, которые получили наибольшее количество баллов по определенному заданию. Индекс трудности определяется по формуле:

$$T_j = \sum_{i=1}^n \frac{x_{ij}}{x_{j\max}} \cdot n$$

T_j – индекс трудности j -го вопроса; x_{ij} – числовая оценка успешности выполнения j -го задания; $x_{j\max}$ – максимальная оценка, которую может получить испытуемый за j -ое задание; n – число испытуемых.

Величина индекса трудности должна находиться в пределах 25 – 75%, где число 25 указывает на тот факт, что данный вопрос теста под силу только одной четверти испытуемых (это трудный вопрос), а число 75 свидетельствует о возможности решения задания тремя четвертями испытуемых. Индекс позволяет выявить легкие и трудные вопросы теста и, следовательно, принять решение об удалении этих вопросов из теста.

Индекс трудности рассчитывается исходя из количества баллов, набранных каждым студентом по данному заданию, и максимально возможного по нему балла. Это связано с тем, что в расчете используется многобалльная шкала оценки. Многобалльная шкала оценки результатов тестов более эффективна, чем дихотомическая и дает реальное представление о знаниях студентов. Существуют разные способы формулирования вопросов в тестах, например, вопросы, где нужно выбрать несколько правильных ответов. Сделать вывод о незнании студентом данного вопроса нельзя, если он выбрал большинство правильных ответов, но не указал все правильные, поэтому использование многобалльной шкалы наилучшим образом отражает знания студента по конкретному вопросу.

Второй этап оценки результатов теста состоит в отборе вопросов, по которым ни один испытуемый не набрал максимальное количество баллов. Данные вопросы не удаляются из теста, а отправляются на доработку. Это связано с тем, что задания могут быть неправильно сформулированы, не полностью изложены и т.д. Отбор вопросов на обработку проходит только после прохождения первого этапа из уже оставшихся заданий, удовлетворяющих соответствующему критерию.

На **третьем этапе** исключаются задания, плохо коррелирующие с суммой баллов ($R_j < 0,15$).

Для нахождения R_j предлагается использовать коэффициент корреляции Пирсона. Таким образом определяется связь каждого j -го задания ($j=1, \dots, m$) с суммой баллов по всему тесту.

$$R_j = \frac{\sum_{i=1}^n (x_{ij} \times y_i)}{s_j \times s_y} \times \frac{\bar{x}_j \times \bar{y}}{n-1}$$

Для определения коэффициента корреляции Пирсона необходимо проделать следующие шаги:

1 шаг. Вычислить индивидуальные баллы испытуемых y_i ($i=1, \dots, n$), показывающие результат выполнения теста каждым испытуемым:

$$y_i = \sum_{j=1}^m x_{ij}$$

2 шаг. Вычислить средние результаты \bar{y} суммарных баллов испытуемых:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

3 шаг. Определить средние результаты \bar{x}_j испытуемых по каждому заданию:

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}$$

4 шаг. Вычислить дисперсию s_y^2 и стандартное отклонение s_y суммарных баллов испытуемых:

$$s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \quad s_y = \sqrt{s_y^2}$$

5 шаг. Рассчитать дисперсию s_j^2 результатов испытуемых по j -ому заданию ($j=1, \dots, m$).

$$s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}$$

Вычислив дисперсию, можно найти и стандартное отклонение $s_j = \sqrt{s_j^2}$.

Компьютерное тестирование уровня освоения материала учебных дисциплин является мощным средством повышения качества учебного процесса в различных учебных заведениях, а разработка качественных тестов сертификационного оценивания специалистов – важным инструментом определения и повышения уровня подготовки специалистов.

РЕАЛИЗАЦИЯ ЭКОЛОГИЧЕСКОГО ПОДХОДА В ИССЛЕДОВАНИЯХ ВОСПРИЯТИЯ ВРЕМЕНИ

Садов В.А., Шпагонова Н.Г.

*Государственный университет гуманитарных наук при Институте психологии Российской Академии Наук, Институт психологии Российской Академии Наук,
г. Москва*

Время является одним из самых важных условий жизни и деятельности человека. Традиционно в психофизических исследованиях восприятие времени изучается на материале интервалов определенной длительности, задаваемыми простыми сенсорными сигналами – вспышками, либо свечением световой точки или пятна, либо простыми тональными звуками. Испытуемым предъявляются либо незаполненные интервалы, где краткими сигналами отмечают начало и конец сигнала, либо заполненные длительности световых или звуковых сигналов. В повседневной жизни человек имеет дело как с искусственными сигналами определенной длительности, так и с реальными протяженными процессами, длительность которых ему требуется адекватно воспринимать. Перенесение феноменов, установленных на искусственных лабораторных интервалах, является условным, и необходимо изучать восприятие реальных временных интервалов. Экологический подход к изучению восприятия времени на сегодня имеет определенное развитие за рубежом. Однако он представлен лишь прикладными исследованиями с использованием несложных методик диагностического характера в ходе выполнения человеком различных видов практической, музыкальной или спортивной деятельности.

В данной работе развивается экологический подход к исследованию восприятия времени. Процессуальная концепция восприятия временных интервалов (Sadov, 1993) основана на понимании временных интервалов как реальных процессов. Предполагается выявление в качестве внутренних психологических механизмов субъективного отражения времени – качественного содержания воспринимаемой человеком сенсорно-перцептивной информации, а также восприятия оцениваемого им временного интервала как целостного процесса, а не как ряда последовательных дискретных событий.

В данной работе экспериментальному исследованию подвергалась гипотеза о влиянии качественного содержания звуковых сигналов на восприятие их длительности. С этой целью был создан специальный экспериментальный комплекс, позволяющий создавать, воспроизводить и трансформировать как простые психофизические сигналы, так и сложные звуковые фрагменты в свободном акустическом поле.

В работе в качестве естественных звучаний использовались следующие звуковые сигналы: мяуканье кошки – 995 мс, цокот копыт лошади – 1010 мс, лай собаки – 555 мс, крик кукушки – 612 мс, звук падающей капли – 203 мс, пение птиц в лесу – 2449 мс, удар топора по дереву – 505 мс, крик моржа – 3039 мс, бой часов – 1082 мс. Также были использованы реверсы этих звуков. При реверсе звук изменяется и может потерять свою естественность настолько, что испытуемый затрудняется в определении предмета звучания. Все естественные сигналы, выбранные для исследования, представляли собой четко выраженные законченные звучания. При реверсе естественных сигналов амплитудные спектральные характеристики не изменялись. В эксперименте приняли участие 12 человек. Они воспроизводили длительность звучания нажатием на кнопку. Длительность каждого нажатия фиксировалась и автоматически классифицировалась. Каждое звучание (9 – естественных и 9 – реверсивных) предъявлялось 20 раз. Общее количество проб в эксперименте – 180. Время проведения отдельного эксперимента – около 30 минут. Реакции предвосхищения у испытуемых исключались из обработки данных специальным программным способом. Перед проведением эксперимента каждый испытуемый имел предварительную тренировку для понимания инструкции и вхождения в эксперимент. Исследование проводилось в затемненной, звукоизолированной камере. Естественные и реверсивные сигналы предъявлялись в свободном акустическом поле со стереозвучанием. Громкость звучаний – около 60 дБ. Отношение “сигнал – шум” – 70 дБ. Были вычислены средние значения воспроизведения длительностей, временных ошибок и относительных временных ошибок.

На каждый звуковой сигнал испытуемый давал вербальное описание. Кроме того, был разработан список прилагательных, описывающих звуковые фрагменты. Этот список формировался экспертами и испытуемыми и включил 83 прилагательных, для которых были подобраны антонимы и сформированы шкалы как в семантическом дифференциале с 7 градациями. Каждый испытуемый описал звуки в этой системе прилагательных. Из 83 шкал семантического дифференциала были отобраны 37 пар прилагательных, которые коррелировали с величиной относительной временной ошибки. Применение факторного анализа (методом принципиальных компонент с вращением Varimax) позволило выявить наличие 8-ми факторов, объединяющих 67% дисперсии. На основе факторного анализа были сформированы следующие шкалы: 1. Интегральная эмоциональная оценка звука (приятный, расслабляющий, комфортный, привлекающий, неуютительный, нераздражающий, желаемый, благоприятный). 2. Естественность звука (естественный, природный, одушевленный, живой). 3. Известность (знакомый, встречаемый, известный, обычный, стандартный). 4.

Высота звука (высокий, тонкий, легкий, острый). 5. Резкость звука (ритмичный, резкий, обрывистый, жесткий). 6. Сила звука (громкий, сильный, звонкий, четкий, яркий). Анализ полученных данных показал наличие значимой корреляционной отрицательной связи величины временной ошибки со следующими шкалами: интегральная эмоциональная оценка, естественность, сила. Со шкалой резкости – значимая отрицательная связь. Таким образом, семантическое содержание звуковых фрагментов оказывает влияние на величину временной ошибки при воспроизведении длительностей естественных и реверсивных сигналов. При оценке сигнала как сильного, естественного, положительно эмоционального, нерезкого происходит более точное воспроизведение временного интервала.

ЭМПИРИЧЕСКАЯ ВАЛИДИЗАЦИЯ СТОХАСТИЧЕСКОЙ МОДЕЛИ ПСИХОФИЗИЧЕСКИХ ИЗМЕРЕНИЙ

Колюгло Н.В.

*Государственный университет гуманитарных наук при Институте психологии Российской Академии Наук,
г. Москва*

Научные достижения, которые были получены в психофизике, относятся к субъективной организации процесса измерения, в которой роль измерительного прибора играет сам субъект. С другой стороны, в рамках этой же науки были разработаны основы психологического эксперимента и принципы проведения психологических измерений поведения испытуемых, а методы психофизики приобрели самостоятельное операциональное значение безотносительно к теоретическим основаниям, лежащим в основе их разработки. В результате термин «психологическое измерение» стал использоваться в психологии в двух различных смыслах в зависимости от того, что является объектом измерения и кто организует процедуру измерения. Принято различать измерение стимулов и измерение испытуемых.

В психодиагностике существует такая модель измерения, модель Раша, которая позволяет измерять одновременно стимулы и испытуемых. Эта модель, нацеленная на оценивание латентных качеств личности и параметров заданий теста на основе математико-статистических моделей измерения (Item Response Theory – IRT), сегодня широко используется в педагогической практике. В модели Раша предполагается, что и индивидов и задания можно расположить на одной оси «способность – трудность» или «интенсивность свойства – сила пункта».

В данном исследовании модель, разработанную в рамках психодиагностики, мы попытаемся применить в рамках психофизики. Вместе с тем данная модель будет иметь несколько иное смысловое (концептуальное) наполнение. В