

# ДВУХЭТАПНОЕ ПРИНЯТИЕ РЕШЕНИЯ О ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ ДИСТАНЦИОННОГО ТЕСТИРОВАНИЯ (ПРОБЛЕМЫ ПРОКТОРИНГА)

*Шмелев А.Г.*

*профессор факультета психологии*

*МГУ имени М.В.Ломоносова*

*научный руководитель Центра*

*«Гуманитарные технологии»*

Аннотация. Предлагается многоэтапная схема обеспечения дистанционного контроля достоверности онлайн-тестирования знаний. Показан ее высокий потенциал в психопрофилактике читинга.

Abstract. The author offers a multi-stage scheme for providing remote control of the reliability of online knowledge testing. Its high potential in psychological prevention of cheating is shown.

Ключевые слова: прокторинг, принятие решения о достоверности, читинг.

Keywords: proctoring, decision-making about reliability, cheating.

В эпоху глобализации и непрерывной компьютерно-коммуникационной революции постоянно повышается роль процессов принятия решений о достоверности той или иной информации, которая получается с помощью современных информационных технологий. К такой информации, несомненно, следует отнести результаты дистанционного компьютерного тестирования знаний и способностей учащихся вузов – онлайн-тестирования. Техническая вооруженность специалистов по сбору разнообразной информации растет. Например, появилась сравнительно недавно (в ходе последнего десятилетия) возможность прямого или косвенного (по видеозаписи) наблюдения за поведением испытуемого в ходе такого тестирования – с помощью веб-камер, которые в большинстве случаев вмонтированы непосредственно в мониторы или в моноблоки современных компьютерных устройств (включая не только

настольные десктоп-компьютеры в учебных классах, но и нетбуки, планшеты, смартфоны и различные устройства промежуточного типа, которыми оснащены современные учащиеся непосредственно в домашних условиях). Такое наблюдение за испытуемыми получило название «прокторинг».

Вместе с тем до сих пор несовершенны сами процедуры принятия решений о том, нарушает ли испытуемый правила (пользуется ли запрещенными материалами в ходе решения заданий теста, прибегает ли к так называемому читингу) или работает честно. Крайне редко можно найти сведения о том, чтобы кто-то измерял процент ошибок, иначе говоря, точность в суждениях экспертов-прокторов о поведении испытуемых. С другой стороны, появляется все больше разработок со стороны софт-инженеров, которые пытаются целиком «вытеснить» человека-эксперта из этой цепочки принятия решения о достоверности: мы не говорим сейчас о новейших изысканиях в области ИИ (искусственного интеллекта) по распознаванию видеозаписей, но, по крайней мере, мы говорим об уже внедренной в некоторых странах процедуре автоматизированного анализа хронометрического протокола (времени реакции испытуемых на появление тестового задания и выбора ответа на него). В продвижении различных методов при этом чаще всего работает не логика сотрудничества и взаимодополнения методов, а логика конкуренции между их разработчиками.

Исследования по сравнению дистанционных и очных результатов мы начали еще 20 лет назад – в ходе так называемой «Олимпиады Телетестинг» (для старшеклассников и абитуриентов). Победить в этом конкурсе могли только те участники, которые на очном этапе подтверждали свои заочные (дистанционные) результаты. В последние годы различные конкурсы с использованием дистанционного тестирования обрели особую массовость. Кроме образовательных конкурсов, это еще и корпоративные конкурсы среди кандидатов на ту или иную привлекательную вакансию.

В 2020 году мы провели исследование, в котором была поставлена цель сравнить эффективность автоматизированных и экспертных методик контроля

достоверности онлайн-тестирования знаний, или, иными словами, методик выявления признаков «читинга» и изобличения «читеров». Методика проведенного компьютеризированного онлайн-эксперимента состояла из двух этапов, в ходе которых 2 группы участников выполняли разные роли: 1) 35 взрослых людей разного пола и возраста были испытуемыми (респондентами) – отвечали на домашнем компьютере «под видеозапись» на 30 заданий теста эрудиции с 4-мя вариантами ответов, 2) 14 экспертов выполняли задачу оценки достоверности поведения испытуемых (по видеозаписи с веб-камеры) - их мимики, голоса и моторных действий по управлению курсором на экране компьютера. В свою очередь, в первой группе подгруппа из 18 человек выполняли тест как «подставные читеры» - использовали шпаргалки с правильными ответами. В результате 8 успешных экспертов показали удовлетворительные результаты по уровню точности в обнаружении читинга (точность-ассигасу, измеренная с помощью коэффициента Карра, у них была выше 0,5). Это невысокий уровень точности, но он оказался значимо более высоким, чем автоматизированный хронометрический показатель – доля атипичных ответов (слишком медленных и слишком быстрых). Показатели точности хронометрического контроля оказались на уровне 0,38. Еще более высокое превосходство над техническим показателем обнаружено не для отдельных экспертов, а для коллективных результатов – когда усредняются оценки сразу нескольких экспертов. В случае трех и тем более пяти подготовленных экспертов коэффициент точности (ассигасу) достигает требуемых значений – в районе 0,8-0,9.

Но и хронометрическими данными, как выяснилось, не следует пренебрегать. За счет анализа четырехклеточных таблиц сопряженности (иное название для них – «матрицы ошибок», или *confusion matrix*) нами была выявлена высокая асимметричная прогностичность и хронометрического показателя, и экспертных оценок: и хронометрия, и самые успешные эксперты допускают сравнительно много ошибок типа «пропуск», но зато пренебрежимо малое (около 5 процентов) число ошибок типа «ложная тревога». Это дает

основания для следующих **рекомендаций по практическому внедрению** двухэтапной схемы обеспечения дистанционного контроля достоверности с использованием и технических инструментов, и прокторинга: этап 1 - это автоматический хронометрический анализ обычного онлайн-тестирования, этап 2 - это прокторинг (экспертное наблюдение) тестирования в условиях видеозаписи (для заподозренных на первом этапе). Наконец, на заключительном этапе целесообразно ввести еще одно - контрольное очное тестированием всех испытуемых (экзаменуемых), заподозренных в читинге на первом и втором этапах. Следует отметить, что эффективность хронометрического контроля достоверности мы измеряли не только в описанном выше камерном эксперименте (скромном по численности выборки), но и на большой выборке – на нескольких тысячах участниках конкурса «Лидеры России», которые после дистанционного этапа выполняли контрольное тестирование в аудиториях – на очных этапах конкурса (в полуфиналах и финале).

Содержательно предлагаемую двухэтапную схему можно описать таким образом. Вначале все учащиеся выполняют тест в обычных условиях (не под веб-камеру), но среди них с помощью хронометрического критерия выделяется подгруппа «подозрительных испытуемых» - это те, у кого скорость реакции оказалась атипичной. Наши данные показывают, что если эта группа насчитывает не более 10-15 процентов участников, то в ее состав, как правило, попадают в очень малом числе честные испытуемые (практически не бывает «ложных тревог»), хотя некоторые читеры остаются не выявленными (ошибки типа «пропуск»). Эту «атипичную группу» следует подвергнуть повторному дистанционному тестированию – уже с использованием видеозаписи (с помощью систем прокторинга). К анализу видеозаписей желательно привлечь не менее двух экспертов, а при расхождении этих оценок следует привлекать третьего («третейского судью»). С точки зрения экономического эффекта, следует отметить, что первый этап является дешевым и существенно сокращает затраты на привлечение квалифицированных экспертов на втором,

более дорогостоящем этапе. Последний этап сводит практически к нулю процент «ложных тревог». Оставшихся «подозреваемых» можно смело, не опасаясь «ложных обвинений», подвергать повторному очному тестированию, ибо для этого уже достаточно оснований.

Главный эффект от такой многоэтапной схемы заключается даже не в том уровне статистической надежности, который она обеспечивает («пропуски» она все-таки допускает), а в том, что угроза реального серьезного контроля достоверности работает для многих участников образовательного конкурса (образовательного экзамена) как «фактор сдерживания» - они просто опасаются быть уличенными в нечестности и избегают «читинга». Это и есть наиболее ценный эффект в методологии любой диагностики, хорошо осознанный и явно сформулированный в научной криминалистике: главное – это психопрофилактический (предупреждающий) эффект, который можно обеспечить только в случае неотвратимости наказания. Таким образом, ошибки типа «пропуск» такая многоэтапная схема сокращает по чисто психологическому механизму, который можно назвать простыми словами – «избегание дополнительных более строгих испытаний».

Что же показали многотысячные выборки участников конкурса «Лидеры России» на этапе очного контрольного тестирования (использовались тесты знаний и когнитивных способностей)?

Статистически необъяснимое падение уровня результатов (за пределами интервала, связанного с естественными колебаниями тестовых баллов) продемонстрировали не более 5-10 процентов участников. Это происходило не при повторении тестов, которые проводились на дистанционном этапе (такие тесты оказываются «раскрытыми» и теряют после проведения свою диагностическую силу), но и на новых тестовых заданиях (половина заданий в каждом сеансе были совершенно новыми заданиями, припасенными исключительно для очного этапа). Данный пренебрежимый процент «изобличенных читеров» говорит о высоком психопрофилактическом потенциале многоэтапной схемы обеспечения достоверности.

