

DOI: 10.22363/2313-2272-2021-21-1-36-49

Динамика диаметра зрачка как индикатор когнитивной нагрузки респондента: методический эксперимент по сравнению CASI и P&PSI вопросников*

И.Ф. Девятко^{1,2}, М.Б. Богданов¹, Д.В. Лебедев¹

¹Национальный исследовательский университет «Высшая школа экономики»
ул. Мясницкая, 11, Москва, 101000, Россия

²Институт социологии ФНИСЦ РАН
ул. Кржижановского, 24/35, к. 5, Москва, 117218, Россия
(e-mail: deviatko@gmail.com; bogdanovmikle@mail.ru; zenon-daniil@yandex.ru)

В последние годы наряду с ростом общего интереса к методам измерения когнитивной нагрузки и субъективно ощущаемого умственного усилия, связанных с решением различных задач и межличностной коммуникацией, наблюдается и рост специфического интереса социальных исследователей к возможностям мультимодальной оценки когнитивной нагрузки интервьюера и респондента с использованием объективных и субъективных индикаторов, включая параданные и данные веб-камер, в целях оптимизации ее воздействия на качество опросных данных. При этом все еще недооценены возможности сравнительно новых подходов к измерению когнитивной нагрузки — с помощью нейрофизиологических методов, таких как использование малозаметных и не нарушающих естественное протекание деятельности респондентов и интервьюеров современных устройств для окулографии (айтрекинга и пупиллометрии), которые обеспечивают точную временную привязку динамики измеряемых показателей (прежде всего размера зрачка) к формату вопроса, способу и фазе заполнения анкеты, наличию локализованного во времени внешнего воздействия и т.д. Количественные исследования когнитивной нагрузки в ходе социологических опросов и ее возможного влияния на качество опросных данных фокусировались преимущественно на компьютеризованном (CAPI) или использующем бумажные технологии (PAPI) интервьюировании, тогда как специфика когнитивной нагрузки в процессе самостоятельного заполнения респондентами компьютеризованных (CASI) и бумажных (P&PSI) вопросников оставалась малоизученной. В статье представлены результаты методического эксперимента, в котором использовалась модифицированная версия разработанного ранее мультимодального подхода к сравнительной оценке когнитивной нагрузки интервьюеров при заполнении бумажных и компьютеризованных версий анкеты — для оценки когнитивной нагрузки респондентов. Мы расширили диапазон методов оценки когнитивной нагрузки за счет использования окулографического устройства (айтрекера) в целях измерения динамики величины зрачка, связанной с заполнением отдельных вопросов. Результаты эксперимента подтвердили гипотезу о приблизительной эквивалентности двух способов заполнения анкеты с точки зрения когнитивной нагрузки для молодых респондентов с высокой функциональной компьютерной грамотностью, а также позволили провести первичную оценку технических и метрологических возможностей и ограничений использования показателей динамики зрачка, измеряемых с помощью окулографического устройства, для изучения когнитивной нагрузки респондента.

Ключевые слова: когнитивная нагрузка; мультимодальная оценка когнитивной нагрузки при самостоятельном заполнении анкеты; CASI и P&PSI; качество опросных данных; динамика размера зрачка; шкала Пааса; время заполнения

* © Девятко И.Ф., Богданов М.Б., Лебедев Д.В., 2021

Статья поступила 21.09.2020 г. Статья принята к публикации 26.11.2020 г.

Методы измерения когнитивной нагрузки и умственного усилия (*mental effort*) в последние годы становятся все более популярными в различных областях социальной и аффективной нейронауки, в прикладных исследованиях сравнительной эффективности методик и платформ обучения, в изучении распределения внимания при решении разных задач и использования информационных подсказок в принятии решений [см., напр.: 10; 19; 20; 29]. Возрастает и специфический запрос на мультимодальную оценку когнитивной нагрузки интервьюера и респондента с применением объективных и субъективных индикаторов, включая параданные и данные веб-камер, — для анализа ее воздействия на качество опросных данных [3; 17; 18; 22; 33]. При этом недооцененными остаются новые подходы к измерению когнитивной нагрузки с помощью нейрофизиологических методов: так, появление небольших окулографических устройств совпало с формированием более глубокого понимания связанных с динамикой зрачка нейрофизиологических процессов и их реципрокной связи с распределением когнитивной нагрузки, внимания, принятия решений и вынесения суждений [16; 27]. Изменения размера зрачка (<1 мм), управляемые связанной с голубым (голубоватым) пятном ствола мозга (*locus coeruleus*) норадренергической системой [12], отражают динамику когнитивной нагрузки с высоким временным разрешением и считаются недоступными для произвольного контроля (несмотря на требующие уточнения данные о возможностях опосредования психосенсорной и других реакций зрачка высокоуровневыми когнитивными влияниями). Эта динамика связана с задействованием ресурсов внимания и рабочей памяти и выступает в качестве «окна» в процессы переработки информации и принятия решений, позволяющего оценить когнитивную нагрузку [8; 24].

Переход от использования самостоятельно заполняемых респондентами бумажных анкет (P&PSI – *paper and pencil survey instruments*) к самостоятельному заполнению вопросников посредством компьютерных технологий (CASI — *computer-assisted survey instruments*), преимущественно онлайн, — тенденция, все более очевидная в социальных науках в последние десятилетия. Она способствовала росту интереса исследователей к сравнению качества данных, получаемых с помощью двух подходов в различных областях и для разных совокупностей [см., напр.: 6; 11; 32; 35]. Вместе с тем в центре внимания специалистов по методологии социальных наук долгое время оставались вопросы качества данных, получаемых при помощи близких подходов — основанных не на самостоятельном заполнении, а на компьютеризованном (CAPI) или использующем бумажные технологии (PAPI) интервьюировании, а также вопросы их сравнительной эффективности в масштабных лонгитюдных проектах [см., напр.: 1; 4; 5; 7; 14; 25]. В частности, отмечались преимущества компьютеризованных интервью с точки зрения снижения издержек на полевую фазу исследования, ввод данных, контроль качества заполнения анкет, упрощение маршрутизации для интервьюеров и др. Были выявлены и специфические трудности, в том числе связанные с

оптимальным выбором цифровых устройств для проведения опроса, программированием вопросников и т.п. Однако вопрос о том, различаются ли самостоятельно заполняемые бумажные анкеты (P&PSI) и компьютеризованные анкеты (CASI) в плане когнитивной нагрузки респондента, остается практически неизученным.

В исследованиях перехода от PAPI к CASI [1; 2; 4; 13] отмечалась роль различий в опыте использования цифровых технологий: этот опыт может влиять и на удобство использования электронных устройств при заполнении анкеты интервьюером (и, соответственно, на его когнитивную нагрузку), и на опасения респондентов относительно сохранности личных данных, и на психологический комфорт в процессе заполнения анкеты. Однако в этих исследованиях участвовали преимущественно интервьюеры, имеющие значительный опыт работы с бумажными технологиями и весьма варьирующую функциональную компьютерную грамотность. Если же речь идет о сравнительно молодых респондентах, имеющих значительный опыт заполнения разнообразных электронных форм, в том числе приобретенный в ходе получения среднего и высшего образования, оправдано исходить из предположения об относительной эквивалентности CASI и P&PSI с точки зрения связанной с ними когнитивной нагрузки.

Мы применили модифицированную версию использованного нами ранее мультимодального подхода к сравнительной оценке когнитивной нагрузки интервьюеров при заполнении бумажных и компьютеризованных версий анкеты [2] для оценки когнитивной нагрузки респондентов, расширив диапазон методов ее оценки за счет окулографического устройства (айтрекера) — в целях измерения динамики величины зрачка, связанной с заполнением отдельных анкетных вопросов. Таким образом, основные задачи нашего исследования: 1) проверка предположения о приблизительной эквивалентности двух способов заполнения анкеты с точки зрения когнитивной нагрузки для респондентов с высокой функциональной компьютерной грамотностью; 2) оценка технических и метрологических возможностей и ограничений использования показателей динамики зрачка, измеряемых с помощью окулографического устройства, для изучения когнитивной нагрузки респондента при ответе на вопросы анкеты.

Дизайн эксперимента предполагал две экспериментальные группы — самостоятельное заполнение анкеты на планшете (CASI) или ее бумажного варианта (P&PSI). Участниками эксперимента стали студенты разных направлений НИУ ВШЭ, которые получали дополнительные баллы по релевантным курсам за участие в эксперименте. Все испытуемые обладали нормальным или скорректированным до нормального зрением ($N = 52$, из них 90% — женщины, средний возраст — 19 лет) и были случайным образом распределены по условиям эксперимента — CASI/P&PSI. Эксперименты были проведены с 14 по 21 декабря 2019 года. Всего было осуществлено 52 экспериментальных наблюдения (28 — CASI и 24 — P&PSI). В итоговый анализ по разным

причинам не вошли 27 экспериментов: низкое качество записи размера зрачков — нижний квартиль показателей точности замеров ниже 0,5 (17 наблюдений); невозможность разметить процесс опроса на планшетах вследствие отсутствия записи видео с экрана планшета в результате технического сбоя (8); отсутствие записи замеров размера зрачка (2). В итоговом анализе представлены 25 испытуемых (15 — CASI и 10 — P&PSI).

Анкета для методического эксперимента была разработана на основе вопросника, используемого в лонгитюдном опросе «Российский мониторинг экономического положения и здоровья населения НИУ ВШЭ» (РМЭЗ), что обеспечивало валидность инструментария, и включала в себя несколько блоков. Для снижения эффекта порядка вопросов были созданы две версии анкеты — прямая и обратная: мы использовали простое контрбалансирование порядка предъявления вопросов, поскольку полная рандомизация блоков анкеты усложнила бы логику печати бумажных анкет. В анкете были представлены следующие тематические блоки (помимо «паспортички»): здоровье, образование, материальное положение, проективные ситуации, сенситивные вопросы (оценка деятельности судов и др.). В анкете каждый вопрос был представлен на отдельной странице/листе. После каждого содержательного вопроса следовало задание по оценке субъективной когнитивной нагрузки (умственного усилия), вызванной вопросом, по шкале Пааса (была представлена на отдельной странице/листе). В отличие от ряда исследований [напр.,: 2], в качестве шкалы Пааса мы использовали исходный вариант с девятью градациями [28]. После вопроса о субъективной когнитивной нагрузке шла страница/лист, на которой испытуемым предлагалось посчитать про себя или вслух до пяти, чтобы снизить контекстуальную нагрузку. Изначально анкета содержала 20 вопросов, но в ходе экспериментов была обнаружена сложность с вопросом № 4 (спортивная активность) — его представление в CASI и P&PSI различалось. На планшетах студенты пропускали вопросы-ростеры (уточнения о частоте и длительности занятий выбранными видами спорта), а на бумаге эти вопросы были представлены в одной таблице, и их было невозможно пропустить. В итоге было решено использовать только данные заполнения 19 вопросов (без № 4).

Эксперимент проходил в закрытой комнате, где присутствовал один или два исследователя (в том числе модератор). Исследователи размещались таким образом, чтобы находиться вне поля зрения испытуемого и не отвлекать его от заполнения анкеты, что было необходимо для снижения влияния внешней среды на вариации размера зрачка. Имевшееся в комнате окно было занавешено — использовался искусственный свет, который светил из-за спины испытуемого, чтобы снизить влияние колебаний естественного освещения на размер зрачка, а сам испытуемый располагался спиной к окну.

Для сбора нейрофизиологических и опросных данных использовались разные устройства. Для сбора данных о размере зрачка и видеозаписи того, как респондент заполнял опрос на планшете / в бумажной анкете (для

последующей разметки данных), использовалось монокулярное устройство для отслеживания размера зрачка и направления взгляда (*eye-tracker*) Pupil Labs с выборочной частотой в 200 Гц. Для измерения частоты сердцебиения (пульса) применялся фитнес трекер Xiaomi Mi Band 2. В качестве планшетов в экспериментальной группе CASI была задействована модель Samsung Galaxy Tab A 16 SM-T355 с общедоступным программным обеспечением Survey Solutions [подробнее см. в: 2. С. 631–632].

В качестве параметров сбора данных с помощью айтрекера в приложении Pupil Capture, которое позволяет записывать данные с устройства на компьютер и сохранять для последующего анализа, были выставлены следующие настройки, которые не менялись между экспериментами: World камера (камера сцены) записывала процесс заполнения анкеты с разрешением (*resolution*) 800×600, частотой кадров (*frame rate*) — 60, абсолютной длительностью экспозиции (*absolute exposure time*) — 157; Pupil камера (камера зрачка) записывала показатели размера зрачка и направление взгляда с разрешением 192×192, частотой кадров — 120, абсолютной длительностью экспозиции — 32.

Процедура эксперимента была выстроена следующим образом:

- Участнику на руку надевался фитнес-браслет Xiaomi MI Band 2 и включалась запись пульса. Модератор просил испытуемого сесть так, чтобы в процессе заполнения анкеты снизить возможные движения испытуемого, которые могли привести к изменению позиции относительно анкеты/планшета. Кроме того, модератор просил выключить звук на телефоне, чтобы уведомления не отвлекали испытуемого от заполнения анкеты.
- Испытуемый читал общую информацию об исследовании и подписывал форму информированного согласия на участие в нем.
- На респондента надевалось айтрекинговое устройство («очки») и проводилась его настройка, калибровка и тестовый замер.
- Модератор эксперимента объяснял испытуемому суть эксперимента: «Сейчас вы будете заполнять анкету. Пожалуйста, внимательно читайте все вопросы. Если вы не до конца понимаете вопрос, то отвечайте на него так, как считаете нужным. Здесь нет правильных или неправильных ответов. Если вы не знаете, какой вариант ответа выбрать, то выберите тот, который больше других подходит вам. Мы очень просим вас ни на что не отвлекаться во время интервью, не разговаривать, не трогать айтрекер (устройство на голове) и планшет, не задавать вопросы по анкете и другие вопросы, оставить их до завершения эксперимента. В анкете будут паузы между вопросами, чтобы ваши глаза отдохнули — не пропускайте их, пожалуйста. Вы можете считать вслух, как указано в инструкции, или про себя, как вам удобно. Просим не щуриться и не закрывать надолго глаза».

- Планшет/бумажная анкета размещалась в 60 сантиметрах от глаз испытуемого (для контроля использовалась лента соответствующей длины) на подставке с углом в 63 градуса (наименьший из доступных), и с помощью приложения на смартфоне модератор замерял показатели люксметра (уровня освещения) у глаз испытуемого.
- Включался замер размера зрачка и направления взгляда, и испытуемым предлагалось спокойно и расслабленно в течение минуты смотреть в центр пустого белого экрана планшета/листа (в CASI и P&PSI соответственно) для создания фоновых значений (*baseline*) размера зрачка.
- В случае CASI включалась запись экрана планшета с помощью приложения VRecorder — на этом заканчивалась подготовительная часть эксперимента.
- Респондентам предлагалось заполнить анкету — основная часть эксперимента.
- После заполнения анкеты выключалась запись устройства Pupil Labs и MI Band 2 и запись экрана планшета (в CASI). Кроме того, испытуемые заполняли короткую анкету с вопросами о проблемах со зрением, сердцем, а также о потреблении кофе в течение шести часов до эксперимента и лекарств, влияющих на частоту сердцебиения. Эти показатели планировалось использовать при анализе для контроля внешних факторов, потенциально влияющих на вариацию размеров зрачка и частоту сердцебиения (для контроля межиндивидуальной вариативности применялась коррекция с вычитанием фоновых значений, собранных до начала основной фазы).
- После окончания эксперимента все данные и записи сохранялись и отправлялись в облачное хранилище для последующей разметки, подготовки базы и анализа.

Для анализа полученных данных нужно было разметить показатели размера зрачка, т.е. указать, в какие моменты происходил переход от одного вопроса к другому. Эта задача выполнялась вручную с использованием записей world камеры айтрекера (P&PSI) и экрана планшета VRecorder (CASI). При разметке также отмечались значения, выбранные испытуемым по каждому вопросу по субъективной шкале когнитивной нагрузки Пааса. После этого показатели размера зрачка и пульса совмещались по временным меткам замеров.

Для работы с показателями размера зрачка в рамках CASI/P&PSI необходимо было учесть отличия в освещении и возможные индивидуальные девиации нейрофизиологического типа. Для устранения фоновых значений из данных конкретного замера (минута перед началом эксперимента) были отобраны наблюдения с точностью измерения (*confidence*) выше 0,7. Далее из данных по фоновым значениям убирались замеры, относящиеся к последним и первым пяти секундам, поскольку в начале эксперимента испытуемый мог

быть возбужден, а в конце на замеры зрачка могли влиять действия экспериментатора, который в конце замера фоновых значений давал испытуемому стимульный материал (анкету). Для оставшихся наблюдений было посчитано среднее значение размера зрачка, которое и представляло собой фоновое значение величины зрачка для конкретного респондента. Наконец, для стандартизации измерений из каждого последующего наблюдения размера зрачка каждого испытуемого вычиталось его фоновое значение. В анализе для оценки размера зрачка использовалось именно такое значение, поправленное на фоновое.

Для оценки различий в когнитивной нагрузке, измеренной с учетом среднего размера зрачка при ответе на конкретный вопрос, при заполнении анкеты на планшете (CASI) и на бумаге (P&PSI), а также различий между анкетными вопросами был проведен двухфакторный дисперсионный анализ с повторными измерениями (*two-way repeated measures ANOVA*). В качестве внутрисубъектного фактора выступал номер вопроса в анкете, а в качестве межсубъектного фактора — способ заполнения анкеты: на планшете или на бумаге. Тест на сферичность показал, что предположение о сферичности нарушается для основного эффекта номера вопроса (Mauchly's $W = 0,000$, $p < 0,001$, $\epsilon = 0,333$), поэтому степени свободы были скорректированы с помощью поправки Гринхауса-Гайссера.

Итак, наблюдается статистически значимое влияние вопроса на величину когнитивной нагрузки, измеренной через размер зрачка ($df = 5,995$, $F = 12,445$, $p < 0,001$, $\eta_p^2 = 0,351$), т.е. имеющие разные номера вопросы различаются когнитивной нагрузкой. Этот фактор объясняет 35% необъясненной другими факторами дисперсии зависимой переменной (размер зрачка). Однако основной эффект для межсубъектного фактора, как и предполагалось, оказался статистически не значим на уровне 95% доверительной вероятности ($df = 1$, $F = 2,501$, $p < 0,127$), что не позволяет говорить о различиях когнитивной нагрузки при прохождении опроса на планшете и на бумаге.

Интересно, что наблюдается статистически значимый эффект взаимодействия между номером вопроса и способом прохождения анкеты (CASI или P&PSI) ($df = 5,995$, $F = 4,492$, $p < 0,000$), т.е. эффект номера вопроса различается в зависимости от способа опроса. Поскольку привести попарные *post-hoc* сравнения для всех 19 вопросов и двух способов заполнения анкеты нецелесообразно, визуализация эффекта взаимодействия представлена на Рисунке 1.

Что касается других индикаторов когнитивной нагрузки, то эффект номера вопроса наблюдался и для самооценки сложности вопроса по шкале Параса ($df = 5,205$, $F = 3,407$, $p = 0,006$), и для времени ответа на вопрос ($df = 4,891$, $F = 15,399$, $p < 0,001$). Однако, как и в случае с размером зрачка, влияние способа прохождения анкеты (CASI или P&PSI) на когнитивную нагрузку оказалось статистически незначимым, как и эффект взаимодействия этих факторов. На Рисунках 2 и 3 в иллюстративных целях представлена динамика средних значений размера зрачка для двух испытуемых.

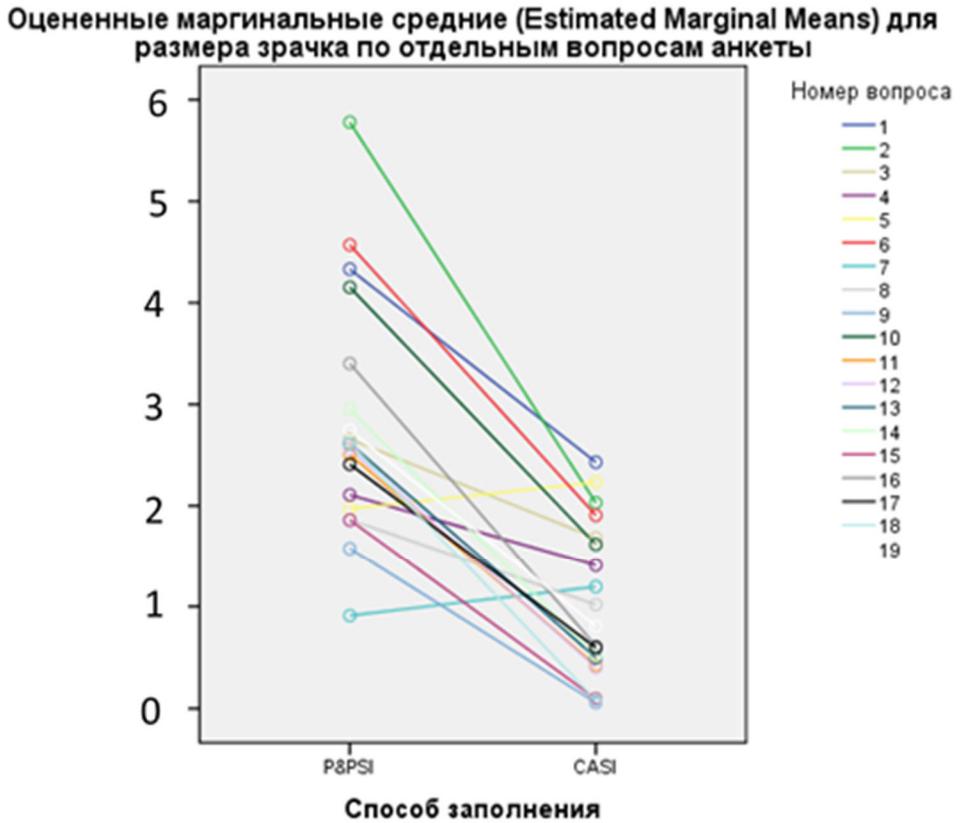


Рис. 1. Эффект взаимодействия способа заполнения анкеты и номера вопроса

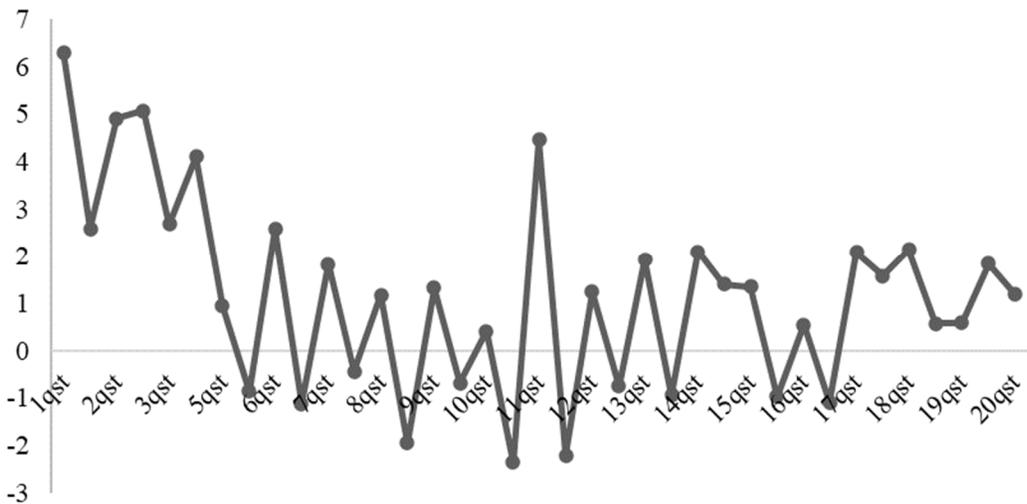


Рис. 2. Динамика среднего размера зрачка по вопросам и паузам (11-й респондент)

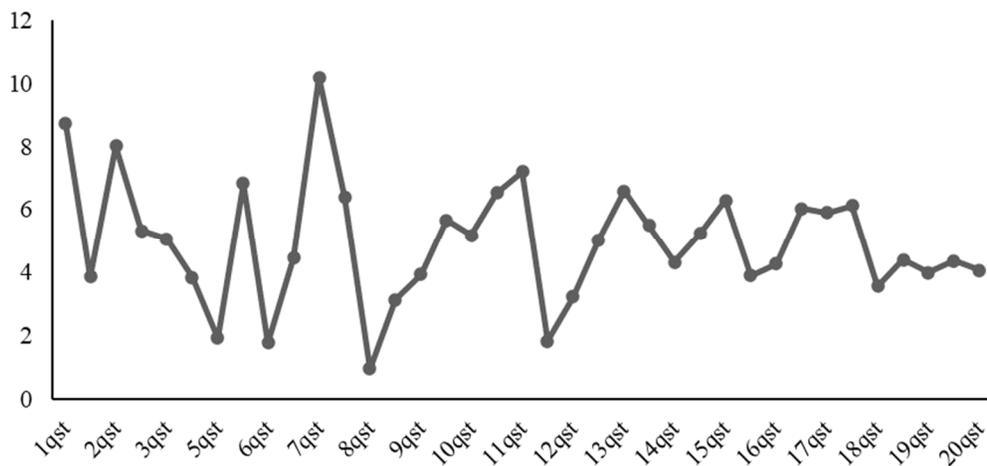


Рис. 3. Динамика среднего размера зрачка по вопросам и паузам (31-й респондент)

Первой задачей нашего исследования была проверка эквивалентности общей когнитивной нагрузки при заполнении анкеты двумя способами — CASI и P&PSI. Предположение об эквивалентности было выдвинуто с учетом специфики совокупности, к которой относились испытуемые – молодые респонденты, получающие высшее образование и обладающие высоким уровнем функциональной компьютерной грамотности, которая сформировалась в силу наличия у поколения 20–30-летних значительного опыта заполнения разнообразных электронных форм, приобретенного как в цифровых коммуникациях с индивидуальными и корпоративными акторами, так и в ходе получения среднего и высшего образования. Такого рода опыт возрастных групп, родившихся не ранее второй половины 1980-х годов, вполне сопоставим с имеющимся у них опытом заполнения бумажных форм и вопросников.

Наши результаты, подтверждающие отсутствие систематических различий между двумя способами заполнения для большинства вопросов и отражающие основные аспекты интегральной когнитивной нагрузки (динамика зрачка, субъективно воспринимаемая нагрузка, измеренная с помощью шкалы Пааса, время заполнения), соответствуют данным исследований, в которых изучался эффект способа заполнения для близких возрастных групп. Так, Д. Байер [6] показал, что в ходе основанного на самоотчете двумя способами (CASI или P&PSI) опроса учащихся старших классов о представленности в их повседневной жизни делинквентного поведения и его факторах не было обнаружено существенных различий между экспериментальными группами (N= 610 и N= 2033) [см. также: 26]. Однако Байер отмечает, что в ряде исследований делинквентного поведения, касавшихся более сенситивных тем (рискованные сексуальные практики, употребление наркотических веществ и т.п.) в конце 1990-х — начале 2000-х годов, были получены противоречивые результаты, свидетельствовавшие как о несколько большем доверии молодых респондентов компьютеризованным опросам [34; 35], так и о том, что

респонденты, описывавшие свой опыт употребления наркотиков и другие рискованные для здоровья практики, воспринимали бумажный вопросник как гарантирующий большую анонимность [9]. В нашей анкете, несмотря на вопросы о здоровье и заработке, отсутствовали столь же сенситивные вопросы, т.е. ее заполнение вряд ли могло показать существенные различия в стиле ответов и когнитивной нагрузке вследствие беспокойства респондентов об анонимности ответов. Кроме того, условия заполнения для обеих экспериментальных групп были идентичны.

Наши результаты позволяют сделать вывод, что очевидный в настоящее время тренд максимально широкого использования цифровых устройств для самостоятельного заполнения молодыми респондентами вопросников не приведет к росту когнитивной нагрузки в сравнении с самостоятельным заполнением традиционных бумажных вопросников, и наблюдаемая «индифферентность» к способу заполнения будет становиться более выраженной по мере взросления младших возрастных групп. Вместе с тем необходимо более детальное изучение отмеченных в отдельных исследованиях, но не получивших систематического рассмотрения эффектов взаимодействия способа заполнения и характерного для респондента доверия другим людям как переменной индивидуального уровня [35], а также не исследованных в достаточной мере потенциальных эффектов взаимодействия способа заполнения и общей оценки индивидом угроз анонимности в цифровых коммуникациях [15; 21; 30].

Вторая задача нашего исследования заключалась в оценке технических и метрологических возможностей и ограничений использования показателей динамики зрачка, измеряемых с помощью окулографического устройства, в изучении когнитивной нагрузки респондента при ответе на вопросы анкеты. Спецификой измерения динамики зрачка при заполнении анкеты как меры когнитивной нагрузки в сравнении, например, с психофизическими экспериментами является необходимость ручного кодирования данных с временной привязкой к отдельным эпизодам (ответам на вопросы). Поскольку мы стремились к сохранению экологической валидности методического эксперимента и к возможности использовать такие параданные, как время заполнения, мы не применяли ограничений по времени заполнения, которые могли бы превратить вопрос в стимул с четкими временными метками и помочь в частичной автоматизации выгрузки данных о динамике зрачка. Кроме того, сама ожидаемая реакция респондента представляла собой многокомпонентный процесс, предполагавший не только чтение вопроса и выбор подходящего ответа, но и возможность внесения исправлений и уточнений, заполнения полей при ответе на открытые или полузакрытые вопросы и т.д. Таким образом, экспорт и первичный анализ данных о зрачке оказались весьма трудоемкими, что отчасти компенсировалось полнотой и «естественностью» данных. Еще один вывод более частного и технического характера — необходимость использования единообразных подставок для планшетов и анкет, чтобы уменьшить долю «плохих» наблюдений (ситуаций, когда веки

респондента полуприкрыты и взгляд направлен вниз, что приводит к наблюдениям с неприемлемо низкими значениями точности измерения).

Полученные данные о динамике размера зрачка (прежде всего, статистически значимое и весьма выраженное влияние конкретного вопроса на средний размер зрачка) могут рассматриваться как подтверждение конструктивной валидности показателей размера зрачка при косвенной оценке когнитивной нагрузки при ответе на вопросы анкеты. Однако обнаруженный нами небольшой, но статистически значимый эффект взаимодействия способа заполнения с конкретным вопросом позволяет сделать предварительный вывод, что различия между средними размерами зрачка как мера когнитивной нагрузки для отдельных вопросов существенно меньше при заполнении бумажной анкеты. Впрочем, небольшой размер выборки и неравный размер групп (размер экспериментальной группы, заполнявшей бумажную анкету, в итоге был существенно меньше) требуют дальнейшего уточнения этого вывода, ограничивая возможности его содержательной интерпретации в настоящий момент.

Также требует дальнейшего изучения конвергентная валидизация показателей когнитивной нагрузки, имеющих сходные механизмы: частота сердечных сокращений и динамика зрачка. В основе этих объективных показателей лежит активация вегетативной нервной системы в ответ на требования текущей задачи, однако их временное разрешение и конкретные механизмы регуляции различаются, поэтому вопрос прямого соотнесения этих показателей друг с другом и с когнитивной нагрузкой как латентной многомерной переменной до сих пор не решен [см., напр.: 23]. Интересной перспективой для дальнейших исследований представляется и конвергентная валидизация показателей размера зрачка, с одной стороны, и субъективной оценки связанного с вопросом умственного усилия (шкала Пааса) — с другой. В нашем методическом эксперименте использовались сравнительно простые вопросы из реальной анкеты, имевшие относительно небольшую вариативность оценок по шкале Пааса, т.е. мы не можем пока сделать обобщающие выводы о связи между объективными и субъективными компонентами интегральной оценки когнитивной нагрузки. Однако в дальнейшем мы надеемся преодолеть это ограничение, создав банк анкетных вопросов с разбросом в оценках связанного с ответом на них воспринимаемого умственного усилия и используя этот банк для дальнейших методических экспериментов.

Благодарность

Авторы выражают большую благодарность Айгуль Климовой, Александру Бызову и Евгении Шмелевой за помощь в организации и проведении эксперимента.

Информация о финансировании

Статья подготовлена при поддержке Российского научного фонда. Проект № 17-78-20172.

Библиографический список / References

- [1] Богданов М.Б., Лебедев Д.В. «Главное не бояться!» — воздействие тренингов на ожидания интервьюеров при переходе от PAPI к CAPI // Социология: 4М. 2017. № 45 /

- Bogdanov M.B., Lebedev D.V. “Glavnoe ne boyatsya!” — vozdeystvie treningov na ozhidaniya intervyyueroov pri perekhode ot PAPI k CAPI [“Do not worry!” — training’s impact on interviewers’ expectations under the transition from PAPI to CAPI]. *Sotsiologiya: 4M*. 2017; 45. (In Russ.).
- [2] Девиатко И.Ф. Разработка подхода к количественной мультимодальной оценке когнитивной нагрузки интервьюеров: результаты пилотного квазиэксперимента // Вестник РУДН. Серия: Социология. 2018. Т. 18. № 4 / Deviatko I.F. Razrabotka podkhoda k kolichestvennoy multimodalnoy otsenke kognitivnoy nagruzki intervyyueroov: rezultaty pilotnogo kvaziekspierimenta [Developing an approach to the multimodal quantitative assessment of interviewers’ cognitive load: First results of a field quasi experiment]. *RUDN Journal of Sociology*. 2018; 18 (4) (In Russ.).
- [3] Девиатко И.Ф., Лебедев Д.В. Глазами интервьюера, глазами респондента: контуры нового подхода к оценке когнитивной нагрузки при проведении опроса // Мониторинг общественного мнения: Экономические и социальные перемены. 2017. № 5 / Deviatko I.F., Lebedev D.V. Glazami intervyyuera, glazami responenta: kontury novogo podkhoda k otsenke kognitivnoy nagruzki pri provedenii oprosa [Through the eyes of the interviewer, through the eyes of the respondent: Outlining a new approach towards the assessment of cognitive load during the interview]. *Monitoring of Public Opinion: Economic and Social Changes*. 2017; 5 (1). (In Russ.).
- [4] Лебедев Д.В., Богданов М.Б. Переход с PAPI на CAPI: опыт интервьюеров и характеристики, влияющие на их ожидания // Мониторинг общественного мнения: экономические и социальные перемены. 2019. № 4 / Lebedev D.V., Bogdanov M.B. Perekhod s PAPI na CAPI: opyt intervyyueroov i kharakteristiki, vliyayushchie na ikh ozhidaniya [Transition from PAPI to CAPI: Interviewers’ experience and factors influencing their expectations]. *Monitoring of Public Opinion: Economic and Social Changes*. 2019; 4. (In Russ.).
- [5] Терентьев Е.А., Мавлетова А.М., Косолапов А.М. Интервьюирование с помощью компьютерных технологий в лонгитюдных обследованиях домохозяйств // Мониторинг общественного мнения: Экономические и социальные перемены. 2018. № 3 / Terentiev E.A., Mavletova A.M., Kosolapov M.S. Intervyyuирование s pomoshchiyu kompiyuternykh tekhnologiy v longityudnykh obsledovaniyakh domokhozyaystv [Computer-assisted personal interviewing for longitudinal household studies]. *Monitoring of Public Opinion: Economic and Social Changes*. 2018; 3 (In Russ.).
- [6] Baier D. Computer-assisted versus paper-and-pencil self-report delinquency surveys: Results of an experimental study. *European Journal of Criminology*. 2018; 15 (4).
- [7] Baker R.P., Bradburn N.M., Johnson R.A. Computer-assisted personal interviewing: An experimental evaluation of data quality and cost. *Journal of Official Statistics*. 1995; 11 (4).
- [8] Beatty J., Lucero-Wagoner B. The pupillary system. *Handbook of Psychophysiology*. 2000; 2.
- [9] Beebe T.J., Harrison P.A., Park E., McRae J.A., Evams J. The effects of data collection mode and disclosure on adolescent reporting of health behavior. *Social Science Computer Review*. 2006; 24 (4).
- [10] Chen F., Zhou J., Wang Y., Yu K., Arshad S.Z., Khawaji A., Conway D. *Robust Multimodal Cognitive Load Measurement*. Cham; 2016.
- [11] Colasante E., Benedetti E., Fortunato L., Scalese M., Potente R., Cutilli A., Molinari S. Paper and-pencil versus computerized administration mode: Comparison of data quality and risk behavior prevalence estimates in the European School Survey Project on Alcohol and Other Drugs (ESPAD). *PloS One*. 2019; 14 (11).
- [12] Costa V.D., Rudebeck P.H. More than meets the eye: The relationship between pupil size and locus coeruleus activity. *Neuron*. 2016; 89 (1).
- [13] Couper M.P., Burt G. Interviewer attitudes toward computer-assisted personal interviewing (CAPI). *Social Science Computer Review*. 1994; 12 (1).
- [14] De Leeuw E.D., Hox J.J., Snijders G. The effect of computer-assisted interviewing on data quality. A review. *Market Research Society*. 1995; 37 (4).

- [15] Eynon R., Fry J., Schroeder R. The ethics of Internet research. *Sage Internet Research Methods*. Sage; 2008.
- [16] Hartmann M., Fischer M.H. Pupillometry: The eyes shed fresh light on the mind. *Current Biology*. 2014; 24 (7).
- [17] Höhne J.K. Eye-tracking methodology: Exploring the processing of question formats in web surveys. *International Journal of Social Research Methodology*. 2019; 22 (2).
- [18] Höhne J.K., Schlosser S., Krebs D. Investigating cognitive effort and response quality of question formats in web surveys using paradata. *Field Methods*. 2017; 29 (4).
- [19] Hoogerheide V., Renkl A., Fiorella L., Paas F., van Gog T. Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology*. 2019; 111 (1).
- [20] Jbara A., Feitelson D.G. How programmers read regular code: A controlled experiment using eye tracking. *Empirical Software Engineering*. 2017; 22 (3).
- [21] Joinson A.N., Ulf-Dietrich R., Buchanan T., Paine Schofield C.B. Privacy, trust, and self-disclosure online. *Human–Computer Interaction*. 2010; 25 (1).
- [22] Kaminska O., Foulsham T. Real-world eye-tracking in face-to-face and web modes. *Journal of Survey Statistics and Methodology*. 2014; 2 (3).
- [23] Kennedy D.O., Scholey A.B. Glucose administration, heart rate and cognitive performance: Effects of increasing mental effort. *Psychopharmacology*. 2000; 149 (1).
- [24] Laeng B., Sirois S., Gredebäck G. Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*. 2012; 7 (1).
- [25] Laurie H. *From PAPI to CAPI: Consequences for Data Quality on the British Household Panel Study*. ISER Working Paper Series, 2003. No. 2003-14.
- [26] Lucia S., Herrmann L., Killias M. How important are interview methods and questionnaire designs in research on self-reported juvenile delinquency? An experimental comparison of Internet vs paper-and-pencil questionnaires and different definitions of the reference period. *Journal of Experimental Criminology*. 2007; 3 (1).
- [27] Mathôt S. Pupillometry: Psychology, physiology, and function. *Journal of Cognition*. 2018; 1 (1).
- [28] Paas F., Tuovinen J.E., Tabbers H., van Gerven P.W.M. Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*. 2003; 38 (1).
- [29] Schmeck A., Opfermann M., van Gog T., Paas F., Leutner D. Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science*. 2015; 43 (1).
- [30] Schomakers E.M., Chantal L., Müllmann D., Ziefle M. Internet users' perceptions of information sensitivity — insights from Germany. *International Journal of Information Management*. 2019; 46.
- [31] Schräpler J.-P., Schupp J., Wagner G.G. Changing from PAPI to CAPI: Introducing CAPI in a longitudinal study. *Journal of Official Statistics*. 2010; 26 (2).
- [32] Spark S., Lewis D., Vaisey A., Smyth E., Wood A., Temple-Smith M., Lorch R., Guy R., Hocking J. Using computer-assisted survey instruments instead of paper and pencil increased completeness of self-administered sexual behavior questionnaires. *Journal of Clinical Epidemiology*. 2015; 68 (1).
- [33] Stodel M. But what will people think? Getting beyond social desirability bias by increasing cognitive load. *International Journal of Market Research*. 2015; 57 (2).
- [34] Turner C.F., Ku L., Rogers S.M., Lindberg L.S., Pleck J.H., Sonenstein F.L. Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*. 1998; 280.
- [35] Wright D.L., Aquilino W.S., Supple A.J. A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly*. 1998.

DOI: 10.22363/2313-2272-2021-21-1-36-49

Pupil diameter dynamics as an indicator of the respondent's cognitive load: Methodological experiment comparing CASI and P&PSI*

I.F. Deviatko^{1,2}, M.B. Bogdanov¹, D.V. Lebedev¹

¹HSE University

Myasnitskaya St., 11, Moscow, 101000, Russia

²Institute of Sociology of FCTAS RAS

Krzhizhanovskogo St., 24/35–5, Moscow, 117218, Russia

(e-mail: deviatko@gmail.com; bogdanovmikle@mail.ru; zenon-daniil@yandex.ru)

Abstract. In recent years, the increase in general interest in methods for measuring cognitive load and subjectively perceived mental effort when solving various tasks and in the interpersonal communication was accompanied by an increase in the specific interest of social researchers in the multimodal assessment of the cognitive load of interviewers and respondents based on objective and subjective indicators, including paradata and webcam data, in order to control this load's impact on the quality of survey data. The authors argue that the possibilities of relatively new approaches to measuring cognitive load with neurophysiological methods (such as the use of wearable devices for oculography — eye tracking and pupillometry — which do not disrupt the natural course of respondents and interviewers activity) are still underestimated, although they allow an accurate time linkage of measured parameters' dynamics (primarily the size of the pupil) to the question format, mode and phase of survey completion, external influences localized in time, etc. As a rule, quantitative studies of surveys' cognitive load and its possible impact on the quality of survey data focus on computer-assisted (CAPI) or paper-based (PAPI) interviewing, while the specificity of the cognitive load in the self-completed computerized (CASI) and paper (P&PSI) surveys was not studied. The article presents the results of the methodological experiment based on a modified version of the multimodal approach to the comparative assessment of the cognitive load of interviewers working with paper and computerized questionnaire. We expanded the range of methods for assessing cognitive load by using a wearable oculographic device (eye tracker) to measure the dynamics of pupil size when answering different survey questions. The results of the experiment confirmed the hypothesis about the approximate equivalence of the two modes of survey completion in terms of their cognitive load for younger respondents with a high level of functional computer literacy, and allowed an initial assessment of the technical and metrological capabilities and limitations of the use of pupil dynamics' indicators, measured with a wearable oculographic device, to study the respondents' cognitive load.

Key words: cognitive load; multimodal assessment of the cognitive load under the self-completion of the questionnaire; CASI and P&PSI; survey data quality; pupil size dynamics; Paas scale; completion time

Acknowledgments

The authors express deep gratitude to Aigul Klimova, Alexander Byzov, and Evgeniya Shmeleva for their help in organizing the experiment.

Funding

The research was supported by the Russian Science Foundation. Project No. 17-78-20172.

* © I.F. Deviatko, M.B. Bogdanov, D.V. Lebedev, 2021

The article was submitted on 13.07.2020. The article was accepted on 01.10.2020.