

МЕТОДЫ И МЕТОДОЛОГИЯ

DOI: 10.14515/monitoring.2017.5.01

Правильная ссылка на статью:

Девятко 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. С. 1—19. DOI: 10.14515/monitoring.2017.5.01.

For citation:

Deviatko I. F., Lebedev D. V. 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. P. 1—19. DOI: 10.14515/monitoring.2017.5.01.

И. Ф. Девятко, Д. В. Лебедев ГЛАЗАМИ ИНТЕРВЬЮЕРА, ГЛАЗАМИ РЕСПОНДЕНТА: КОНТУРЫ НОВОГО ПОДХОДА К ОЦЕНКЕ КОГНИТИВНОЙ НАГРУЗКИ ПРИ ПРОВЕДЕНИИ ОПРОСА

ГЛАЗАМИ ИНТЕРВЬЮЕРА, ГЛАЗАМИ
РЕСПОНДЕНТА: КОНТУРЫ НОВОГО
ПОДХОДА К ОЦЕНКЕ КОГНИТИВНОЙ
НАГРУЗКИ ПРИ ПРОВЕДЕНИИ ОПРОСА

THROUGH THE EYES OF THE INTER-
VIEWER, THROUGH THE EYES OF THE
RESPONDENT: OUTLINING A NEW AP-
PROACH TOWARDS THE ASSESSMENT
OF COGNITIVE LOAD DURING THE
INTERVIEW

*ДЕВЯТКО Инна Феликсовна — доктор социологических наук, профессор, департамент социологии, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия; главный научный сотрудник, Институт социологии Федерального научно-исследовательского социологического центра Российской академии наук, Москва, Россия.
E-MAIL: deviatko@gmail.com
ORCID: 0000-0002-1955-7592*

*Inna F. DEVIATKO^{1,2} — Dr. Sci (Sociology), Professor; Principal Researcher
E-MAIL: deviatko@gmail.com
ORCID: 0000-0002-1955-7592*

¹ Department of Sociology, National Research University Higher School of Economics

² Institute of Sociology of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences, Moscow, Russia

ЛЕБЕДЕВ Даниил Вадимович - студент магистратуры, департамент социологии, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия; младший научный сотрудник, Институт социального анализа и прогнозирования, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Москва, Россия.

*E-MAIL: zenon-daniil@yandex.ru
ORCID: 0000-0001-7246-8022*

*Daniil V. LEBEDEV^{1,2} — Postgraduate Student; Junior Researcher
E-MAIL: zenon-daniil@yandex.ru
ORCID: 0000-0001-7246-8022*

¹ Department of Sociology, National Research University Higher School of Economics

² Institute of Sociology of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences, Moscow, Russia

Аннотация. Представление о том, что инструментарий исследования, опосредующий коммуникацию между интервьюером и респондентом, влияет на качество получаемой в процессе проведения интервью информации, давно стало общепринятым в социологической методологии, однако среди множества методов оценки качества опросного инструментария и выявления связанных с ним источников ошибки измерения в опросных данных пока отсутствуют методы количественной оценки когнитивной нагрузки на интервьюера и респондента, возникающей непосредственно в процессе проведения опроса. В случае опросов в формате личного интервью физические и, в значительно большей степени, умственные усилия на производство опросной информации затрачивают и интервьюер, и респондент. В ситуации, когда анкета заполняется интервьюером, именно последний должен распределять ограниченные индивидуальные ресурсы внимания, памяти, зрительного и моторного контроля, активного слушания и интерпретации таким образом, чтобы минимизировать ошибки в понимании вопроса на стороне респондента и ошибки

Abstract. There is a widely held opinion in sociological methodology that the survey tools intermediating between interviewer and respondent affects the quality of information obtained during the interview. However, despite multiple methods of assessment of survey tools and detection of measurement bias, there is still a lack of techniques of quantitative assessment of cognitive load on the interviewer and the respondent during the interview. In case of personal interview, both interviewer and respondent are required physical and, to a larger extent, cognitive efforts to produce the information needed. In case of interviewer-assisted questionnaire, it is the interviewer's responsibility to allocate limited individual resources related to attention, memory, visual and motor control, active listening and interpretation to minimize respondent's misunderstanding of the questions and the mistakes made by interviewer while recording the answers; thus, the interviewer is required to possess meta-cognitive and self-regulatory skills. The interviewer might experience cognitive overload when performing multiple tasks during the interview such as sustaining the contact with respondents, carrying out control over his/her own ac-

фиксации ответа на своей стороне, что требует от интервьюера немалых метакогнитивных и саморегуляторных навыков. Решение множественных параллельных задач в процессе интервьюирования, включая поддержание коммуникации с респондентом, контроль над собственными действиями по заполнению вопросника и над работой техники, фиксацию возможных нарушений хода интервью и т. п., может приводить к исчерпанию доступных когнитивных ресурсов и когнитивной перегрузке интервьюера, что, в конечном счете, ведет к снижению качества данных. Однако до сих пор внимание социологов к когнитивной нагрузке при проведении интервью было минимальным. В статье представлены результаты аналитического обзора традиционных и новых подходов к измерению когнитивной нагрузки, используемых в таких дисциплинарных областях, как когнитивная наука, эргономика, исследования процессов обучения и решения задач и др. Мы обосновываем возможность их применения для количественной оценки и оптимизации когнитивной нагрузки интервьюера (и, в перспективе, респондента), возникающей в процессе проведения интервью, а также обозначаем контуры и ближайшие перспективы исследовательской программы по разработке интегрального методологического подхода к использованию множественных индикаторов когнитивной нагрузки (как основанных на ретроспективном самоотчете, так и физиологических и поведенческих) в опросах различных типов.

Ключевые слова: когнитивная нагрузка при проведении опроса, субъективные методы измерения когнитивной

нагрузки во время выполнения опроса и контроль над работой офисного оборудования при одновременной регистрации возможных нарушений, обнаруженных во время интервью; когнитивная перегрузка может в конечном итоге привести к получению данных низкого качества. Однако внимание социологов к измерению когнитивной нагрузки во время интервью было минимальным до сих пор. Статья предоставляет результаты аналитического обзора традиционных и новых подходов к измерению когнитивной нагрузки, используемых в областях когнитивной науки, эргономики, исследований процессов обучения и решения задач. Авторы обосновывают возможность их применения в количественной оценке и оптимизации когнитивной нагрузки интервьюера (и далее респондента) и очерчивают перспективы программы исследования, направленной на разработку интегрального методологического подхода к использованию множественных индикаторов когнитивной нагрузки (основанных как на ретроспективном самоотчете, так и на физиологических и поведенческих данных) в различных опросах.

Keywords: interview-induced cognitive load, subjective cognitive load measurement methods, objective cognitive load

нагрузки, объективные методы измерения когнитивной нагрузки, измерения динамики величины зрачка, шкала Пааса, качество данных в интервью

measurement methods, task-evoked pupillary response, the Paas Scale, survey data quality

Благодарность. Статья подготовлена в рамках работы по гранту Российского научного фонда «Разработка интерактивной методической инфраструктуры для повышения качества данных лонгитюдных социологических исследований» (№ 17-78-20172).

Acknowledgment. The paper is part of the work performed under the grant of the Russian Science Foundation entitled «Development of interactive methodological infrastructure aimed to improve the quality of longitudinal sociological studies' data» (grant no. 17-78-20172).

Введение

Массовые опросы — уникальный источник информации о мнениях, установках, личностных особенностях, социально-демографических характеристиках и поведении индивидов и групп, производимой совместными усилиями исследователей и респондентов, который остается таковым даже в эпоху растущего интереса к цифровым следам поведения и масштабным поведенческим данным, возникающим как малозаметный/нереактивный побочный результат компьютеризованных транзакций индивидуальных и корпоративных акторов. В случае самозаполняемых анкет самая существенная доля когнитивной и эмоциональной нагрузки, затрачиваемой на получение этой информации, ложится на респондента. Итоговое качество ответов зависит не только от имеющихся у последнего ресурсов внимания, памяти, предшествующего опыта выполнения сходных задач и мотивации, но и от контингентных обстоятельств ситуации заполнения, например, наличия параллельных, нерелевантных основной задаче и потенциально негативно влияющих на ее решение запросов на внимание и умственные усилия респондента, фонового шума и других факторов, которые, в свою очередь, могут избирательно усиливать когнитивную нагрузку, связанную с заполнением вопросника в целом или его отдельных блоков. В случае опросов в формате личного интервью физические и, в значительно большей степени, умственные усилия на производство опросной информации затрачивают и интервьюер, и респондент. Когда анкета заполняется интервьюером, именно он должен распределять ограниченные индивидуальные ресурсы внимания, памяти, зрительного и моторного контроля, активного слушания и интерпретации таким образом, чтобы минимизировать ошибки в понимании вопроса на стороне респондента и ошибки фиксации ответа на своей стороне, что требует от него, помимо прочего, немалых метакогнитивных и саморегуляторных навыков.

Представление о том, что инструментарий исследования, опосредующий коммуникацию между интервьюером и респондентом, определяет качество получаемой в итоге информации, давно стало общим местом социологической методологии, подкрепленным поистине необозримой литературой. Возникающие в результате неудачного выбора формулировок, непродуманного дизайна инструментария,

неоднозначных инструкций по заполнению, чрезмерных запросов к когнитивным ресурсам интервьюера или, чаще, респондента и разнообразных иных причин случайные и систематические ошибки измерения исследуются, предотвращаются и контролируются с помощью множества методов и техник — от методических экспериментов и когнитивных интервью до использования множественных индикаторов и статистических методов оценки, моделирования и контроля ошибки измерения. Однако какими методами мы можем воспользоваться, чтобы прямо оценить, сопоставить и по возможности оптимизировать ментальные усилия, когнитивную нагрузку для интервьюера и респондента, влияющую на возникновение ошибок измерения и связанную с ключевой фазой проведения опроса — заполнением анкеты?

Сложившийся в социологии и родственных социальных науках подход к анализу эффектов интервьюера и эффектов респондента, а также их потенциального неаддитивного взаимодействия, асимметричен и нацелен преимущественно на облегчение когнитивного и аффективного бремени респондента [Bradburn, 1978], достигаемое иногда за счет усложнения задач интервьюера. Последний должен не только контролировать собственные действия (речь, письмо) и соотносить их с заданными инструкциями шаблоном, но и в интерактивном режиме отслеживать реакции респондента на отдельные вопросы, ситуацию интервью в целом, технические особенности записи, фиксируя возникающие нарушения хода опроса, вопросы и недоумения (ретроспективно или в интерактивном режиме), предлагать релевантные уточнения и т. д. Это возрастание когнитивной сложности решаемых интервьюером в ходе интервью задач редко систематически отслеживается и оценивается с точки зрения баланса издержек при выборе способа реализации опроса, при оценке затрат на дополнительное обучение интервьюеров и создание инфраструктуры контроля и информационной поддержки их деятельности и, наконец, при снижении или оптимизации их когнитивной нагрузки по заданному набору критериев. Когнитивная перегрузка (overload) интервьюеров, достижение неких «красных линий» в сложности решаемых ими задач, неоправданное с точки зрения приращения качества данных злоупотребление их ментальными ресурсами, как можно обоснованно предположить, влияют не только на склонность интервьюеров использовать стратегии минимизации собственных усилий, снижая их мотивацию и добросовестность, но и на точность и успешность их работы в целом.

Не менее актуальной представляется и задача сопоставимой и точной оценки когнитивной нагрузки и сложности опросных задач «на стороне» респондента. Рассуждая о риске или, что может показаться парадоксальным, даже о необходимости увеличения когнитивной нагрузки на респондента¹, мы также оказываемся перед лицом необходимости получить валидные и надежные данные о затрачиваемых респондентами умственных усилиях и, соответственно, сложности для него возникающих в ходе опроса задач на припоминание, сравнение, понимание, выбор или формулировку наиболее точного ответа и т. д. Ограничимся лишь одним примером, позволяющим оценить необходимость разработки ин-

¹ См., например, [Stodel, 2015], где обсуждаются потенциальные позитивные эффекты когнитивной перегрузки в результате использования иррелевантной второй задачи для уменьшения эффекта социальной желательности в опросах.

тегрального подхода к оценке когнитивной нагрузки: возникший в последние годы тренд к уменьшению длины анкет, их миниатюризации и снижению уровня сложности (см., например, [Пулестон, 2016; Бурдейн, 2016]) и, соответственно, к уменьшению времени заполнения стал реакцией на общую тенденцию уменьшения участия и увеличения доли разных типов не-ответов в опросах, которая нередко интерпретируется как эффект общей «опросной усталости», вызванной различными факторами, среди которых нередко упоминается чрезмерная когнитивная нагрузка, связанная с заполнением длинных, однообразных, скучных анкет. Однако предположение о том, что длинные анкеты по определению связаны с большей когнитивной нагрузкой на респондента (или интервьюера) и чаще вызывают утомление и связанные с ним эффекты — снижение внимания, степени вовлеченности в выполнение задачи и т. п., — редко проверяется прямо, поскольку для его обоснования исследователи располагают лишь теми косвенными индикаторами нагрузки, которые прямо или косвенно связаны с длиной анкеты — временем заполнения, долей покинувших опрос, субъективной оценкой длительности заполнения или средним количеством ошибок. Вместе с тем, например, исследования в области экспериментальной оценки сложности чтения и понимания компьютерного кода для программистов показывают, что длина изучаемого кода не связана со сложностью его понимания программистами линейно: во многих случаях длинный фрагмент кода, имеющий регулярную структуру, т. е. некоторое количество повторяющихся сегментов (аналогом которых в анкете могут быть блоки или отдельные вопросы со сходным форматом ответа, типом графической шкалы и т. п.), обрабатывается легче и с меньшим количеством ошибок в сравнении с равным с точки зрения функциональности нерегулярным, что вполне может быть объяснено с точки зрения иерархической организации обработки информации человеком [Jbara, Feitelson, 2017].

Однако на сегодняшний день в методологии социологических исследований не сложился единый подход к оценке когнитивной нагрузки и сложности задач², решаемых в ходе опроса интервьюерами и респондентами, сопоставимый с существующими в когнитивной науке, педагогической и социальной психологии, эргономике и исследованиях взаимодействия в системах «человек — компьютер».

В данной статье мы ставим перед собой двоякую задачу: 1) дать критический аналитический обзор существующих подходов к измерению когнитивной нагрузки (КН) и 2) описать ближайшие перспективы, возможности и трудности разработки подхода к интегральной оценке последней применительно к деятельности по заполнению опросников, преимущественно с точки зрения оценки КН интервьюеров при

² Здесь и далее мы будем рассматривать самооценку когнитивной нагрузки на индивида, связанной с заданием, и воспринимаемую им когнитивную сложность/трудность последнего как тесно связанные и эквивалентные в большинстве измерительных контекстов конструкты, поскольку усредненная величина воспринимаемой когнитивной нагрузки, т. е. ментального усилия (mental effort) выполняющего задание индивида, является естественной мерой субъективной сложности задания (последняя может быть также дополнена некими объективными показателями, подобными количеству ошибок и времени заполнения). Неслучайно непосредственным толчком и основой для разработки обсуждаемой далее девятибалльной субъективной шкалы когнитивной нагрузки [Paas, 1992] стала предложенная О. Братфишем и соавторами шкала воспринимаемой трудности задания [Bratfish, Borg, Dornić, 1972]. На эмпирическом уровне оценки индивидом умственных усилий, потребовавшихся для решения задачи, и ее воспринимаемой трудности также коррелируют [DeLeeuw, Mayer, 2008], однако см. обсуждение в [Joseph, 2013] и [Chen F. et al., 2016: 15—16].

проведении опросов с использованием разных способов заполнения (бумажная анкета, заполнение вопросников с использованием переносных компьютеров и планшетов). Этот подход применим и к другим способам проведения интервью и, с некоторыми модификациями, к оценке КН респондентов, однако ради целостности и простоты изложения мы ограничимся здесь преимущественно его иллюстрациями и возможными приложениями, относящимися к деятельности интервьюера.

Концептуализация и измерение когнитивной нагрузки в поведенческих науках: основные подходы и методики

В интересующем нас здесь широком и неспецифическом смысле термина, используемого в разных исследовательских контекстах, когнитивная нагрузка — это многомерный конструкт, характеризующий ментальные усилия, связанные с решением одной или нескольких задач, структура и содержание которых задают необходимый объем привлекаемых для их успешного решения ограниченных когнитивных ресурсов — внимания, кратковременной памяти, моторного контроля и т. д. [Raas, 1992: 429]³. Иными словами, варьируя предполагаемую конкретной задачей (или несколькими параллельно выполняемыми задачами — скажем, чтения вслух и понимания смысла предложения, запоминания последовательностей разной длины и т. п.) загрузку ограниченных ресурсов, в частности, рабочей памяти или внимания, мы варьируем величину необходимых для ее успешного решения основных усилий индивидов, интенсивность которых и выступает показателем КН⁴.

Различные методы измерения КН исторически развивались относительно автономно в нескольких областях и включают в себя: 1) опросные техники, прежде всего основанные на субъективной оценке уровня когнитивной нагрузки на некоторой шкале; 2) объективные физиологические/нейрофизиологические неинвазивные методы измерения КН, использующие показатели, которые в большинстве ситуаций не поддаются произвольному контролю со стороны испытуемых (чаще — получаемые с помощью пупиллометрии данные о динамике размера зрачка, данные о частоте сердцебиений, реже — данные айтрекеров о фиксации взгляда, результаты ЭЭГ и т. д.); 3) дополняющие методы, основанные на объективных данных об уровне исполнения, обычно с использованием параллельно выполняемой второй задачи (количество ошибок, время исполнения) и других поведенческих индикаторах. В данном разделе мы кратко охарактеризуем основные объективные и субъективные (то есть основанные на самоотчете) методы измерения КН, относящиеся к первой и второй группам, с учетом существующих свидетельств об их надежности и валидности, а также перспектив применения для интересующих нас задач оценки КН при проведении социологических опросов, проиллюстрировав их типичными примерами использования. В заключительной части данной статьи мы несколько подробнее остановимся на описании возможных направлений и ближайших шагов по разработке и реализации соответствующего интегрального подхода, включающего в себя методы из обеих групп, а также некоторые более традиционные индикаторы уровня/качества исполнения.

³ В последнее время иногда, наряду с «умственным/ментальным усилием» и «когнитивной нагрузкой», используют и синонимичный термин «когнитивное усилие» [Westbrook, Braver, 2015].

⁴ Здесь предполагается равенство прочих факторов, прежде всего уровня мотивации, для разных условий [Ibid.].

Субъективные оценки когнитивной нагрузки, основанные на самоотчете

Ретроспективный самоотчет индивидов о воспринимаемом уровне КН в ходе выполнения различных заданий — от решения задач по статистике до нахождения ошибок в тексте — очень популярный метод оценки КН, который возник поначалу в исследованиях интеллекта [Bratfisch, Borg, Dornic, 1972], но основное развитие получил в педагогической психологии и прикладных исследованиях образования [Paas, 1992; Paas et al., 2003]. Обучаемые обычно не испытывают сложности при оценке умственных усилий, затраченных ими при решении отдельной задачи или блока заданий, по девятибалльной (иногда — семибалльной) шкале ликертовского (лайкертовского) типа (в диапазоне от «очень, очень/крайне небольшие» до «очень, очень/крайне большие»). Интерпретация данных о субъективно воспринимаемой когнитивной нагрузке в указанных исследовательских областях разрабатывается преимущественно в контексте теории КН [Sweller, Ayres, Kalyuga, 2011], выделяющей в качестве аналитически независимых три вида последней: внутренне присущую задаче (intrinsic) КН, зависящую от сложности самой задачи и предшествующего запаса знаний или опыта обучаемого; внешнюю (extraneous) нагрузку, обычно связанную с отвлекающими факторами, шумом или не связанными с внутренней сутью задачи вторичными задачами — например, необходимостью отслеживать уровень заряда на устройстве, наклон ручки и т. п. и, наконец, уместную (germane), полезную с точки зрения целей научения нагрузку. Однако на практике до последнего времени КН как ключевая переменная в исследованиях эффективности подходов к обучению в большинстве случаев измерялась с помощью описанной выше одномерной шкалы. (Обоснование возможности многомерного подхода к анализу КН на основе данных самоотчета представлено, например, в работе [Lerpink et al., 2013]. Она, впрочем, также не решает окончательным и бесспорным образом вопроса о возможности разделения видов КН на основе исключительно опросных данных, без учета объективных параметров, описывающих ситуацию обучения, типы задач, индикаторы успешности и т. д.). Поскольку связанная с обучением КН встречается лишь в небольшой части ситуаций опроса (прежде всего в задачах обучения интервьюеров, например, при переходе к новому опросному инструментарию или способу заполнения — от бумажной анкеты к заполнению на планшете или ином мобильном устройстве), мы в рамках данного обзора не будем подробно останавливаться на проблемах дифференциации разных типов КН и соотношения субъективных оценок с данными об успешности исполнения заданий, ограничившись кратким анализом существующих данных о психометрических качествах и потенциально релевантных для социологической методологии приложениях самой популярной из шкал воспринимаемой КН — вышеописанной одномерной шкалы Ф. Пааса.

В нескольких исследованиях шкала Пааса применялась для оценки сравнительной эффективности использования в процессе обучения разных форм мультимедийных подсказок и учебной анимации для оптимизации КН студентов [Tabbers, Martens, Merriënboer, 2004; de Koning et al., 2007 и др.], при этом в исследовании Х. Табберса, Р. Мартенса и Е. Мерриенбоера для оценок КН в ходе выполнения разных задач было получено значение коэффициента альфа Кронбаха, равное 0,92, что свидетельствует о высоком уровне надежности-согласованности данного

инструмента измерения КН. Кроме того, высокая надежность этой субъективной шкалы была подтверждена и исследованием П. Эйрса. По итогам двух экспериментов, в рамках которых две группы студентов решали алгебраические задачи, оценивая КН после выполнения каждой из задач, также были получены высокие значения коэффициента альфа Кронбаха (0,90 и 0,97) [Ayres, 2006]. Таким образом, и в этом исследовании была подтверждена надежность субъективной шкалы Пааса для измерения связанной с задачей КН. Наконец, в исследовании, направленном на сравнение четырех стратегий обучения с использованием компьютера, Ф. Паас и Е. Ван Мерриенбоер использовали субъективную шкалу для измерения когнитивной нагрузки респондентов при решении ими геометрических задач. Для оценки надежности-согласованности между оценками КН для всех задач по используемой шкале также был рассчитан коэффициент альфа Кронбаха (то есть рассматривалось, насколько скоррелированы были субъективные оценки когнитивных нагрузок всех задач). Рассчитанный коэффициент внутрисубъектной согласованности оценок КН при решении задач оказался равен 0,82, что говорит о том, что результаты, полученные при использовании субъективной шкалы Пааса для батареи заданий, достаточно надежны [Paas, Van Merriënboer, 1994]. Интересно, что в данном исследовании было показано, что результаты использования субъективной шкалы измерения КН оказались более чувствительны к различиям в используемых инструкциях, в то время как результаты пульсометрии (ЧСС) указывают лишь на периоды умственной активности/неактивности [Paas, Van Merriënboer, 1994].

Конкурентная валидность шкалы Пааса была недавно подтверждена в исследовании А. Корбаха и коллег, которые сравнили использование различных объективных индикаторов и субъективной шкалы Пааса для измерения когнитивной нагрузки. В этом исследовании результаты, полученные с помощью субъективной оценки КН, были отрицательно скоррелированы с результатами решения задач респондентами (успешность выполнения теста из 12 вопросов разной сложности) и с измерениями движения глаз (измерение количества и длительности фиксации взгляда на зонах с отвлекающими деталями, представленных на слайдах с заданиями) [Korbach, Brünken, Park, 2017]. То есть оценки когнитивной сложности, полученные с помощью шкалы Пааса, были убедительно кросс-валидизированы с показателями успешности решения задач и объективными измерениями движения глаз.

Стоит обратить внимание на возможность разных подходов к измерению агрегированной (общей) КН и КН, связанной с выполнением отдельных заданий, что отражается обычно в разном расположении субъективных шкал для оценки когнитивной нагрузки относительно выполняемых участниками содержательных заданий в различных исследованиях. Иногда вопросы, содержащие шкалу Пааса, задают после каждого отдельного содержательного задания/задачи, и оценкой когнитивной нагрузки в таком случае становится среднее арифметическое всех оценок, полученных в ответ на эти вопросы [Tabbers, Martens, Merriënboer, 2004; Paas, Van Merriënboer, 1994]. Однако субъективная оценка КН может производиться после решения участником исследования всех задач с помощью одного итогового вопроса про сложность выполнения заданий в целом, и в таких случаях полученная итоговая оценка и является оценкой величины когнитивной нагрузки

[Schwamborn et al. 2011]. В экспериментах, посвященных этому аспекту использования субъективной шкалы Пааса, было выявлено, что оценка респондентами когнитивной нагрузки выше в случае общего вопроса относительно всех задач в целом в сравнении с результатами усреднения оценок КН, полученных для каждой из задач в отдельности. Однако эти исследования пока не дают ответа на вопрос, какой же вариант измеряет когнитивную нагрузку точнее, и не описывают причины более высокой оценки сложности заданий при использовании одного общего вопроса [Schmeck et al., 2015; Van Gog et al., 2012]. Заметим, однако, что Т. Ван Гог и соавторы советуют измерять когнитивную нагрузку после каждого вопроса по отдельности [Van Gog et al., 2012].

Исходя из проанализированных работ, можно заключить, что основанная на ретроспективной самооценке субъективная шкала Пааса для измерения когнитивной нагрузки является надежным [Ayres, 2006; Tabbers, Martens, Merriënboer, 2004; Paas, Van Merriënboer, 1994] и, возможно, валидным [Korbach, Brünken, Park, 2017] инструментом, хотя исследования, подтверждающие последний вывод, пока не столь многочисленны. Говоря о потенциальной применимости версий одномерной субъективной шкалы Пааса при оценке когнитивной нагрузки интервьюера (или респондента) при проведении опроса или сложности отдельных вопросов (или их блоков), следует подчеркнуть важность параллельного использования других поведенческих или физиологических объективных показателей КН. Следует также отметить, что существуют и потенциально интересные для применения в этом контексте многомерные версии основанных на самооценке шкал связанной с заданием (или рабочей) нагрузки, используемые преимущественно в эргономике. Прежде всего, упоминания в этом контексте заслуживают популярные и обладающие доказанными психометрическими качествами «Индекс связанной с задачей нагрузки NASA» (NASA Task Load Index — NASA-TLX) и «Техника субъективной оценки рабочей нагрузки» SWAT (subjective workload assessment technique) [Hart, Staveland, 1988; Reid, Nygren, 1988], однако они более трудоемки и требуют оценки каждого задания/вопроса по нескольким шкалам (физической нагрузки, ментальной нагрузки, стресса, требований к темпу исполнения и т. д.), а во втором случае — и использования требующей значительных временных затрат процедуры предварительной сортировки карточек. Это затрудняет адаптацию многомерных шкал к интересующим нас исследовательским задачам, хотя и оставляет открытой возможность их применения в хорошо спланированных методических экспериментах, нацеленных, например, на предварительное тестирование отдельных вопросов или форматов ответа.

Объективные физиологические методы измерения КН

Наиболее популярным и «старейшим» методом оценки связанной с заданием когнитивной нагрузки с начала 1960-х годов и до сегодняшнего дня является измерение динамики величины зрачка. Именно на нем мы остановимся подробнее. Хотя в последние годы неоднократно исследовались возможности использования других индикаторов КН — динамики частоты сердечных сокращений (ЧСС), некоторых показателей электроэнцефалографии (ЭЭГ), кожно-гальванической реакции (КГР), их значимость остается ограниченной. Основная причина такой ситуации заключается в том, что — при схожих с использованием методик пупиллометрии

основных угрозах валидности измерения, — описываемые методики отличаются пока небольшой точностью при использовании во внелабораторных условиях и, в случае нейроинтерфейсов для измерения ЭЭГ, создают более «реактивную» ситуацию измерения, не позволяющую проводить его в фоновом (заполняющем, pervasive) и малозаметном режиме (например, в силу необходимости смачивать голову для ЭЭГ). В частности, в случае измерения ЧСС, дополнительные угрозы валидности могут возникать в результате воздействия смешивающих эндогенных факторов (изменения гормонального фона, физиологического состояния и т. п.) или ситуативно обусловленных изменений положения тела, краткосрочных стрессогенных воздействий, не связанных с заданием и т. п., а в случае использования ЭЭГ — в связи с отсутствием достаточных подтверждений конструктивной валидности патентованных алгоритмов для вычисления потенциально релевантных показателей, получаемых с помощью доступных и подходящих для использования в естественных условиях беспроводных мобильных устройств типа Emotive Eroc, а также неоднозначных данных о корреляции «сырых» данных ЭЭГ с другими показателями КН [Joseph, 2013].

В когнитивной психологии не связанные с изменениями в освещенности небольшие изменения размера зрачка используются в качестве меры интереса к зрительному стимулу и меры КН с начала 1960-х годов [Hess, Polt, 1960; Hess, Polt, 1964]. Значительный рост популярности этого подхода был связан с его применением для измерения загрузки кратковременной рабочей памяти и внимания [Kahneman, Beatty, 1966; Beatty, 1982]⁵, а также КН, связанной с пониманием и запоминанием сложных предложений [Wright, Kahneman, 1971; Schluroff, 1982] и решением иных задач (см., напр., [Канеман, 2006: 33—41]).

Последние два десятилетия характеризовались заметным ростом числа прикладных исследований в областях новых информационных технологий и эргономического проектирования контекст-ориентированных или учитывающих состояние пользователя устройств и приложений, в которых пупиллометрия, а также анализ движений глаз и направления взгляда используются для интерактивного фонового мониторинга состояний пользователей — усталости (например, при вождении), эмоционального состояния, интереса (например, при использовании шлемов/очков виртуальной реальности и иных нейроинтерфейсов для игры либо в качестве управляющего элемента роботизированных протезирующих устройств) и, наконец, для отслеживания динамики КН при решении имеющих моторный компонент мультимодальных задач (включающих чтение, слушание, запоминание, письмо и т. п.). Этот связанный с цифровой революцией тренд заметно расширил спектр технических решений, доступных для исследователей, нуждающихся в пупиллометрических данных для фиксации изменений в КН в полевых (то есть внелабораторных) контекстах. Следует, впрочем, отметить, что бурный рост разработок в области доступных и удобных устройств и программного обеспечения привел и к не столь радужным последствиям — появившиеся в последнее десятилетие недорогие и легкие в использовании устройства с открытым кодом быстро привлекают внимание не только академических исследователей,

⁵ В качестве синонима КН в исследованиях этого направления иногда используют термин «нагрузка, связанная с обработкой (информации)» — processing load и др. (см., напр. [Beatty, 1982]).

но и корпораций, ведущих инженерные разработки в соответствующих областях. Последние покупают фирмы-разработчики, делая устройства недоступными для рядовых пользователей. Самый дешевый на сегодняшний день и совместимый с имеющими камеру устройствами айтрекер, разработанный небольшой датской компанией «EyeTibe», выпущенный на рынок в 2013 г. (поначалу — по цене \$ 99) и получивший первичную валидацию в качестве оптимального для решения многих исследовательских задач с использованием показателей динамики размера зрачка инструмента [Dalmaijer, 2014; Jbara, Feitelson, 2017], который был в конце 2016 г. куплен вместе с компанией-разработчиком подразделением «Facebook» «Oculus», — не единственный, но, видимо, самый яркий пример. Впрочем, быстрый технический прогресс в этой области позволяет надеяться на скорое появление потенциально адаптируемых к исследовательским нуждам новых технических решений, которые позволят отслеживать не только динамику зрачка и движения глаз, но и движения рук в синхронизированном режиме⁶.

Каковы существующие сейчас основные подходы к измерению КН с использованием пупиллометрии (то есть к измерению вызванной задачей реакции зрачка, или ВЗРЗ — Task-evoked Pupillary Response, TERP)? Избегая погружения в сложные технические детали, можно выделить два основных подхода: 1) использование высокоточных наголовных (head mounted) систем для айтрекинга и, в некоторых случаях, профессиональных пупиллометров и 2) использование удаленных айтрекеров или данных с встроенных в компьютерные устройства видеокамер (в самые последние годы — даже обычных потребительских веб-камер, в том числе встроенных в ноутбуки и планшетные компьютеры) с использованием высокоспециализированного ПО для анализа получаемых видеозаписей с целью выделения релевантных и относительно нереактивных данных о динамике зрачков (стоит отметить, что эти изменения происходят обычно в субмиллиметровом диапазоне и требуют учета множества факторов, включая положение и движения головы, цвет глаз, изменения в освещенности, вызывающие зрачковый рефлекс, и т. п.) [Klingner, Kumar, Hanrahan, 2008; Wood, Bulling, 2014; Rafiqi et al., 2015].

Второй подход становится все более популярным по мере разработки новых алгоритмов и систем обработки данных, позволяющих выделять данные о динамике размера зрачка в реальном времени и с автоматической синхронизацией с выполнением испытуемыми заданий. С точки зрения интересующей нас цели он имеет важное преимущество, так как потенциально позволяет проведение фонового и обладающего достаточной экологической валидностью измерения динамики КН по отношению к решению задач, возникающих в процессе заполнения опросника интервьюером или респондентом. Пока убедительные результаты, позволяющие говорить о сопоставимом качестве измерения ВЗРЗ с помощью наголовных айтрекеров или профессиональных пупиллометров в лабораторных условиях и с помощью камер или удаленных айтрекеров в условиях, приближенных к естественным, получены лишь для простых задач [Klingner, Kumar, Hanrahan, 2008; Rafiqi et al., 2015]. Однако можно говорить и о первых успехах в разработке

⁶ Langley H. Move over Gear VR, Samsung's standalone eye-tracking headset gets a first look: The Exynos VR III prototype revealed [Электронный ресурс] // Wareable News. July 3, 2017. URL: <https://www.wareable.com/samsung/samsung-exynos-vr-3-headset-release-date-price-specs-586> (дата обращения: 09.09.2017).

малореактивных и фоновых способов дистанционного измерения КН при исполнении более сложных множественных задач, осуществляемого без использования лабораторных методов фиксации головы или создающих определенные неудобства наголовных систем и, что немаловажно, за пределами лаборатории.

Одно из недавних тестовых приложений «дистанционного» подхода к исследованию поведения пользователей, выполняющих опросные или экспериментальные задания на краудсорсинговой платформе Amazon Mechanical Turk (см. об этом способе рекрутирования респондентов, напр. [Девятко, 2016: 24—26]), продемонстрировало обнадеживающие результаты параллельного анализа параметров видеозаписей с веб-камеры пользовательского устройства ($N=133$) и измерения КН у подгруппы дающих надежные показатели пупиллометрии испытуемых при просмотре нормальных и подвергнутых деформации видео [Kadem, Cusack, 2017]. По крайней мере, для описанной подвыборки пользователей и одного из видео было продемонстрировано статистически значимое различие в ВЗРЗ для деформированного и обычного, то есть имеющего смысл видео. Кроме того, было показано, что из влияющих на возможность такого дистанционного измерения ВЗРЗ факторов самый существенный вклад в качество результатов вносят скорость интернет-соединения, расстояние, цвет глаз испытуемого, и характеристики освещенности (дневной свет).

Все более масштабное, малозаметное и всепроникающее использование встроенных камер, как и других сенсоров цифровых устройств для мультимодального, то есть использующего разные типы сигналов о протекании деятельности (включая ЧС, КГР, ЭЭГ и поведенческие индикаторы), анализа КН [Chen et al., 2016], с одной стороны, делает их незаменимым источником данных, в том числе для исследовательских целей, с другой стороны, оно чревато серьезными угрозами приватности или даже прямыми нарушениями исследовательской этики. Здесь мы можем лишь очень кратко остановиться на тех потенциальных угрозах и этических проблемах, которые связаны с малозаметным использованием систем айтрекинга как источника данных о КН. Д. Либлинг и С. Прайбуш отмечают, что если угрозы приватности, связанные с нелегитимным доступом к веб-камере, осознаются большинством пользователей, то нарушения приватности и неавторизованное получение чувствительных сведений о пользователе в результате отслеживания взгляда, зрительных фиксаций, пупиллометрии и других динамических характеристик зрения обычно остаются за пределами осознания широкой аудитории [Liebling, Preibusch, 2014]. Таким образом, основная ответственность за этическое использование такого рода данных ложится на исследователей и разработчиков. Иллюстрируя те возможности нарушения приватности и деанонимизации, которые связаны с описанными динамическими данными об активности глаз, названные авторы отмечают их уникальность, связанную с частичной либо полной невозможностью произвольного контроля над направлением взгляда, микродвижениями глаз и изменениями размера зрачка, упоминая, в частности, исследования, продемонстрировавшие возможность получения в фоновом режиме серий данных, позволяющих идентифицировать значимые социальные и физические атрибуты индивидов. К таким атрибутам относятся: фокус внимания, степень интереса к элементам реальных сцен и отдельным стимулам, возраст (влияет на характе-

ристики движений глаз), симптомы некоторых неврологических и поведенческих расстройств, сексуальные предпочтения, степень сложности решаемой задачи для конкретного индивида и т. п. Наконец, характеристики траектории рассматривания и даже данные об изменении размера зрачка в секундном диапазоне позволяют, при использовании методов машинного обучения и серий наблюдений, с достаточно высокой точностью идентифицировать индивидов [Ibid.]. В контексте интересующих нас методов объективной оценки КН в процессе интервьюирования возникающие в результате перечисленных угроз приватности этические проблемы потребуют разработки и использования процедур получения информированного согласия не только при оценке нагрузки интервьюера в различных режимах работы, но и при проведении методических исследований, направленных на оценку КН респондентов, участвующих в онлайн-опросах (потенциально использование таких процедур может влиять на уровень отказов от участия в исследовании).

Разработка подхода к интегральной оценке КН интервьюеров при проведении опросов: ближайшие перспективы

Целью предлагаемого нами к дальнейшей разработке и эмпирической проверке методологического подхода к оценке КН, связанной с проведением опросов, является не снижение последней до нулевого уровня, а выбор оптимальных способов заполнения и дизайнов вопросников или отдельных их элементов, обладающих более низкой КН для интервьюеров и респондентов при равной информативности (или, при более широком определении, управление уровнем КН для решения задач исследования, как в случае, когда высокая КН необходима для снижения систематической ошибки ответа при прохождении опроса респондентами, как в предлагаемом в [Stodel, 2015] исследовательском плане). Исследования в упомянутых выше областях (прежде всего психологии образования и эргономике) позволяют считать доказанными преимущества параллельного использования данных самоотчетов, объективных физиологических и поведенческих данных для оценки КН при проведении опроса. Такой мультииндикаторный и мультимодальный подход, как было продемонстрировано, в частности, в исследованиях взаимодействия человека (оператора) с ориентированными на пользователя интерфейсами и цифровыми устройствами, позволяет получить более робастные интегральные оценки КН в процессе выполнения множественных задач: хотя ни одна из мер латентных переменных КН и сложности задач несвободна от шума, т. е. угроз валидности и надежности измерения, их совместное применение позволяет добиться достаточно высокого качества измерения [Chen et al., 2016, 6—8].

Более того, детализированная и основанная на совокупности эмпирических исследований картина динамики выполнения задач, предъявляющих высокий спрос на ограниченные когнитивные ресурсы акторов (в интересующем нас случае — на кратковременную память, внимание и моторный контроль, необходимые интервьюеру при коммуникации с респондентами в ходе опроса), показывает, что окрестности точек резкого снижения уровня исполнения (т. е. увеличения количество ошибок, отказов, пропусков и т. п.), именуемых иногда «красными линиями», которые разделяют области устойчивого выполнения задачи и, соответственно, перегрузки, то есть превышения связанных с задачей запросов к когнитивным

ресурсам над доступным субъекту уровнем⁷, характеризуются неодинаковой чувствительностью различных субъективных и объективных индикаторов КН. В частности, некоторые индикаторы ментальных усилий (прежде всего субъективные оценки КН), «запаздывают» и не позволяют точно определить момент достижения «красной линии» или появления конкретного чрезмерно сложного задания в последовательности подобных [ibid., 34—40].

С точки зрения технической реализуемости и организационной доступности ближайшие перспективы методологического экспериментирования в области интегрального, т. е. мультиметодного/мультимодального измерения КН интервьюеров, как нам представляется, связаны с проведением методических экспериментов, позволяющих сравнить КН интервьюеров при разных способах проведения опроса, предпочтительно — в исследованиях со смешанным способом проведения опроса (mixed-mode / multi-mode surveys), что обеспечит сопоставимость данных с точки зрения используемого инструментария и контроля возможных эффектов взаимодействия способа реализации с переменными, характеризующими попавших в выборку респондентов. С учетом описанных выше технических, организационных (а отчасти и финансовых) сложностей, связанных с использованием мобильных айтрекеров/пупиллометров для инструментального измерения КН интервьюеров, заполняющих бумажные анкеты⁸ (однако см. [Kaminska, Foulsham, 2014]), другой сравнительно доступной возможностью представляется удаленное измерение динамики КН интервьюеров (см. выше) в процессе обучения работе с новым инструментарием, новыми или модифицированными вопросами или способом заполнения (т. е., например, при переходе от заполнения бумажного вопросника к заполнению на планшете, хотя в этом случае базовый уровень КН также будет определяться ретроспективным самоотчетом). Очевидно, что сколь-нибудь масштабные полевые или лабораторные эксперименты с интервьюерами возможны лишь при согласии исследовательских организаций и коллабораций, реализующих масштабные опросные проекты. В нашей стране такие методические исследования *in situ* получили пока ограниченное распространение, однако существующий мировой опыт изучения и контроля поведения интервьюеров и респондентов с целью обеспечения стандартов качества опросов позволяет надеяться на возможности продвижения в этом направлении. (Заметим, что в области так называемых качественных методических исследований поведения интервьюеров существуют впечатляющие по своему масштабу проекты, например, проект по исследованию сексуального насилия VIRAGE Французского института демографических исследований (INED), в которых этнографическое наблюдение за поведением интервьюеров на полевой и предполевой стадии телефонного

⁷ В случае проведения опроса интервьюерами достижение «красной линии» может вести не только к случайным ошибкам исполнения (неточному произнесению вопроса, неполной или искаженной записи ответа респондента и т. п.), но и, например, к запуску стратегий «экономичного» использования когнитивных ресурсов интервьюера (например, стереотипизации коммуникации, упрощению формулировок или самостоятельному заполнению ответов на табличные вопросы), что может справедливо расцениваться организаторами опроса как плохая или недобросовестная работа (и не вполне справедливо приписываться исключительно индивидуальным моральным качествам или уровню мотивации интервьюера).

⁸ В этом случае вполне возможным и оправданным представляется использование менее точных, но более неактивных датчиков ЧСС.

опроса было включенным, явным, а иногда сопровождалось и прямой интервенцией исследователей [Markou, Bourgea, 2017].)

Последующие перспективы, как нам представляется, связаны со сравнительно доступным и технически реализуемым дистанционным измерением КН в исследованиях ошибки измерения, связанной со склонностью уже не интервьюеров, а респондентов онлайн-опросов экономить ограниченные когнитивные ресурсы, руководствуясь принципом разумной достаточности. Эта универсальная оптимизирующая затраты стратегия (или когнитивная эвристика) поведения, обозначенная в свое время Г. Саймоном термином «минимально достаточное / минимально удовлетворяющее решение» (satisficing, «сэтисфайсинг») [Simon, 1956] становится, как мы отмечали, более вероятной в случае приближения связанной с задачей КН к «красным линиям» и, при прочих равных, в случае решения множественных задач, выдвигающих параллельные запросы к когнитивным ресурсам. Методологическая традиция исследования такого оптимизирующего (и снижающего качество данных) поведения респондентов в социальных науках существует достаточно давно [Krosnick, 1991], однако популярные в этой области модели возникновения «минимально удовлетворяющего поведения» могут существенно выиграть в точности и предсказательной силе за счет уточнения вклада не связанных с мотивацией факторов неравной КН для разных типов вопросов или способов проведения опроса. Однако более подробный анализ возникающих на этом пути возможностей и трудностей потребует отдельного дальнейшего исследования.

Список литературы (References)

Бурдейн И. Не всегда короче значит лучше // Онлайн-исследования в России: тенденции и перспективы / под общ. ред.: А. В. Шашкин, И. Ф. Девятко, С. Г. Давыдов. М. : ООО «Онлайн маркет интеллидженс», 2016. С. 113—132. [Burdein I. (2016) Ne vseгда koroche znachit luchshe [Shorter is not always better]. *Online-issledovaniya v Rossii: tendentsii i perspektivy* [Online research in Russia: trends and prospects]. Ed by: A. V. Shashkin, I. F. Deviatko, S. G. Davydov. Moscow: ООО «Onlain market intellidzhens». P. 113—132.] (In Russ.)

Девятко И. Ф. От «виртуальной лаборатории» до социального телескопа: метафоры тематических и методологических инноваций в онлайн-исследованиях // Онлайн-исследования в России: тенденции и перспективы. М. : ООО «Онлайн маркет интеллидженс», 2016. С. 19—33. [Deviatko I. F. (2016) Ot «virtual'noi laboratorii» do sotsial'nogo teleskopa»: metafory tematicheskikh i metodologicheskikh innovatsii v onlain-issledovaniyakh [From «virtual laboratory» to «social telescope»: metaphors of thematic and methodological innovations on online research]. *Online-issledovaniya v Rossii: tendentsii i perspektivy* [Online research in Russia: trends and prospects]. Moscow: ООО «Onlain market intellidzhens». P. 19—33.] (In Russ.)

Канеман Д. Внимание и усилие / пер. с англ. И. С. Уточкина. М. : Смисл, 2006. [Kahneman D. (2006) Attention and Effort. Moscow: Smysl.] (In Russ.)

Пулестон Дж. Разработка опросов в стиле бонсай© // Онлайн-исследования в России: тенденции и перспективы / под общ. ред.: А. В. Шашкин, И. Ф. Девятко,

- С. Г. Давыдов. М. : ООО «Онлайн маркет интеллидженс», 2016. С. 65—112. [Puleston J. (2016) Designing Bonsai Surveys © *Online-issledovaniya v Rossii: tendentsii i perspektivy [Online research in Russia: trends and prospects]*. Moscow: ООО «Онлайн маркет интеллидженс». P. 65—112.] (In Russ.)
- Ayres P. (2006) Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and instruction*. Vol. 16. No 5. P. 389—400. <https://doi.org/10.1016/j.learninstruc.2006.09.001>.
- Beatty J. (1982) Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*. Vol. 91. No 2. P. 276—292. <https://doi.org/10.1037/0033-2909.91.2.276>.
- Bradburn N. (1978) Respondent burden. In: Proceedings of the Survey Research Methods Section of the American Statistical Association, Alexandria, VA: American Statistical Association. V. 35. P. 35—40.
- Bratfisch O., Borg G., Dornic S. (1972) Perceived item difficulty in three tests of intellectual performance capacity. Stockholm: Institute of Applied Psychology. Report No 29.
- Chen F. et al. (2016) Robust Multimodal Cognitive Load Measurement. (Human–Computer Interaction Series.) Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-31413-6_11.
- Dalmajer E. (2014) Is the low-cost EyeTribe eye tracker any good for research? *PeerJ PrePrints*. 2014. No e585v1. URL: <http://dx.doi.org/10.7287/peerj.preprints.585v1> | CC-BY4.0 Open Access | rec: 4 Nov 2014, publ: 4 Nov 2014. <https://doi.org/10.1002/acp.1346>.
- de Koning B. B. et al. (2007) Attention cueing as a means to enhance learning from an animation. *Applied cognitive psychology*. Vol. 21. No 6. P. 731—746.
- DeLeeuw K.E., Mayer R. E. (2008) A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of educational psychology*. Vol. 100. No 1. P. 223—234. <https://doi.org/10.1037/0022-0663.100.1.223>.
- Hart S. G., Staveland L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*. Vol. 52. P. 139—183.
- Hess E. H., Polt J. M. (1960) Pupil size as related to interest value of visual stimuli. *Science*. Vol. 132. No 3423. P. 349—350. <https://doi.org/10.1126/science.132.3423.349>.
- Hess E. H., Polt J. M. (1964) Pupil size in relation to mental activity during simple problem-solving. *Science*. Vol. 143. No 3611. P. 1190—1192. <https://doi.org/10.1126/science.143.3611.1190>.
- Jbara A., Feitelson D. G. (2017) How programmers read regular code: a controlled experiment using eye tracking. *Empirical software engineering*. Vol. 22. No 3. P. 1440—1477. <https://doi.org/10.1007/s10664-016-9477-x>.

- Joseph S. (2013) Measuring cognitive load: A comparison of self-report and physiological methods. PhD dissertation. Tempe, AZ: Arizona State University. P. 1—98.
- Kadem M., Cusack R. (2017) Pearls and Perils of Pupillometry Using a Webcam. *Undergraduate Honors Posters*. 22. URL: http://ir.lib.uwo.ca/psychd_posters/22/ (дата обращения: 12.09.2017).
- Kahneman D., Beatty J. (1966) Pupil diameter and load on memory. *Science*. Vol. 154. No 3756. P. 1583—1585. <https://doi.org/10.1126/science.154.3756.1583>.
- Kaminska O., Foulsham T. (2014) Real-world Eye-tracking in Face-to-face and Web Modes. *Journal of Survey Statistics and Methodology*. Vol. 2. No 3. P. 343—359. <https://doi.org/10.1093/jssam/smu010>.
- Klingner J., Kumar R., Hanrahan P. (2008) Measuring the task-evoked pupillary response with a remote eye tracker. In: Proceedings of the 2008 Symposium on eye tracking research & applications. ACM. P. 69—72. <https://doi.org/10.1145/1344471.1344489>.
- Korbach A., Brünken R., Park B. (2017) Measurement of cognitive load in multimedia learning: a comparison of different objective measures. *Instructional science*. Vol. 45. No 4. P. 515—536. <https://doi.org/10.1007/s11251-017-9413-5>.
- Krosnick J. A. (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*. Vol. 5. No 3. P. 213—236. <https://doi.org/10.1002/acp.2350050305>.
- Leppink J. et al. (2013) Development of an instrument for measuring different types of cognitive load. *Behavior research methods*. Vol. 45. No 4. P. 1058—1072. <https://doi.org/10.3758/s13428-013-0334-1>.
- Liebling D. J., Preibusch S. (2014) Privacy considerations for a pervasive eye tracking world. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. ACM. P. 1169—1177. <https://doi.org/10.1145/2638728.2641688>.
- Markou E., Bourgea E. (2017) Observing the work of interviewers: how the quality of the data collection is constructed. Oral presentation at ESA Annual Meeting (Athens, Greece, 2017). Session RN 21_02b.
- Paas F. G. (1992) Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology*. Vol. 84. No. 4. P. 429—434. <https://doi.org/10.1037/0022-0663.84.4.429>.
- Paas F. G. W. C., Van Merriënboer J. J. G. (1994) Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of educational psychology*. Vol. 86. No 1. P. 122—133.
- Paas F. et al. (2003) Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*. Vol. 38. No 1. C. 63—71. https://doi.org/10.1207/S15326985EP3801_8.

- Rafiqi S. et al. (2015) PupilWare: towards pervasive cognitive load measurement using commodity devices. Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments. ACM, Paper #42. 2015. <https://doi.org/10.1145/2769493.2769506>.
- Reid G. B., Nygren T. E. (1988) The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in psychology*. Vol. 52. P. 185—218.
- Schlurhoff M. (1982) Pupil responses to grammatical complexity of sentences. *Brain and language*. Vol. 17. No 1. P. 133—145. [https://doi.org/10.1016/0093-934X\(82\)90010-4](https://doi.org/10.1016/0093-934X(82)90010-4).
- Schmeck A. et al. (2015) Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instructional Science*. Vol. 43. No 1. P. 93—114. <https://doi.org/10.1007/s11251-014-9328-3>.
- Schwaborn A. et al. (2011) Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in human behavior*. Vol. 27. No 1. P. 89—93. <https://doi.org/10.1016/j.chb.2010.05.028>.
- Simon H. A. (1956) Rational choice and the structure of the environment. *Psychological review*. Vol. 63. No 2. P. 129—138. <https://doi.org/10.1037/h0042769>.
- Stodel M. (2015) But what will people think? Getting beyond social desirability bias by increasing cognitive load. *International journal of market research*. Vol. 57. No 2. P. 313—321.
- Sweller J., Ayres P., Kalyuga S. (2011) Cognitive load theory. Springer Science & Business Media, NY. <https://doi.org/10.1007/978-1-4419-8126-4>.
- Tabbers H. K., Martens R. L., Merriënboer J. J. G. (2004) Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British journal of educational psychology*. Vol. 74. No 1. P. 71—81. <https://doi.org/10.1348/000709904322848824>.
- Van Gog T. et al. (2012) Timing and frequency of mental effort measurement: Evidence in favour of repeated measures. *Applied cognitive psychology*. T. 26. No 6. P. 833—839. <https://doi.org/10.1002/acp.2883>.
- Westbrook A., Braver T. S. (2015) Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*. Vol. 15. No 2. P. 395—415. <https://doi.org/10.3758/s13415-015-0334-y>.
- Wood E., Bulling A. (2014) Eyetab: Model-based gaze estimation on unmodified tablet computers. *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM. P. 207—210. <https://doi.org/10.3758/s13415-015-0334-y>.
- Wright P., Kahneman D. (1971) Evidence for alternative strategies of sentence retention. *The Quarterly journal of experimental psychology*. Vol. 23. No 2. P. 197—213. <https://doi.org/10.1080/14640747108400240>.