



К. А. ГАВРИЛОВ
(Москва)

ПЛАТФОРМА ТОЛОКА КАК ИСТОЧНИК РЕСПОНДЕНТОВ ДЛЯ ОНЛАЙН-ОПРОСА: ОПЫТ ОЦЕНКИ КАЧЕСТВА ДАННЫХ¹

В статье представлен опыт использования краудсорсинговой платформы Толока компании «Яндексе» для рекрутирования респондентов онлайн-исследования. Основываясь на методологических работах о похожей зарубежной платформе Amazon Mechanical Turk, выдвигаются гипотезы о качестве данных опроса пользователей Толоки в сравнении с результатами, полученными с применением других типов конформных выборок, – отбором участников онлайн-панели и рекрутированием респондентов через социальные сети. Дополнительно, только на данных Толоки, рассматривается показатель внимательности респондентов. Основной вывод статьи: Толока позволяет быстро и с небольшими затратами рекрутировать респондентов, а результаты сопоставимы по качеству с теми, что получены иными способами. В частности, респонденты из Толоки почти всегда проходят опрос до конца, заполняют анкеты быстрее других типов респондентов, но реже участников онлайн-панели прибегают к «стрейт-

Кирилл Андреевич Гаврилов – кандидат социологических наук, доцент кафедры анализа социальных институтов, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия; научный сотрудник, Институт социологии ФНИСЦ РАН, Москва, Россия. E-mail: gavrilov@socio.msk.ru

¹ Публикация подготовлена в ходе проведения исследования № 20-01-023 в рамках Программы «Научный фонд Национального исследовательского университета “Высшая школа экономики” (НИУ ВШЭ)» в 2020–2021 гг.

Автор благодарит М. Бутынку за первичную настройку Толоки и проведение первых волн исследования с использованием данной платформы, а также за рекрутинг участников с помощью социальных сетей.

лайнингу» (т.е. дают однотипные ответы на табличный вопрос), они так же часто дают ответы на открытый вопрос, как и респонденты из соцсетей (но реже, чем члены онлайн-панели), хотя их ответы и короче. Только 36% респондентов прошли проверку на внимательность, эти участники дольше заполняли анкету и реже являлись стрейтлайнерами. Увеличение вознаграждения не повышало долю внимательных респондентов, но уменьшало скорость заполнения анкеты, увеличивало количество ответов на открытый вопрос и снижало долю стрейтлайнеров.

Ключевые слова: Толока; краудсорсинг; онлайн-опрос; качество данных, Amazon Mechanical Turk

Постановка исследовательской задачи

Проведение онлайн-опросов давно стало привычной практикой среди исследователей. Если воспользоваться общеизвестной классификацией онлайн-опросов, предложенной М. Купером еще в 2000 г. [1], то можно отметить рост интереса к исследованиям, построенным на невероятностных выборках. Так, авторы доклада ААРОР напрямую связывают распространение неслучайных выборок с ростом количества проводимых онлайн-опросов [2].

В то же время за фразами о популярности невероятностного отбора стоят различные подходы к рекрутированию респондентов – это может быть поиск добровольцев в социальных сетях, отбор участников онлайн-панелей, посетителей веб-сайтов и др. (обзор см.: [3]). Один из сравнительно новых, но набирающих популярность подходов – привлечение респондентов с помощью краудсорсинговых платформ [4, с. 24–25]. Наиболее популярная платформа в зарубежных исследованиях – Amazon Mechanical Turk (АМТ, или Mturk)¹, которая, как отмечают М. Шмилевски и С.

¹ Amazon Mechanical Turk: [site]. URL: <https://www.mturk.com/> (date of access: 04.07.2022).

Кукер, стала «одним из самых важных исследовательских инструментов последнего десятилетия» [5, р. 464]. АМТ – это интернет-площадка, где заказчики размещают разного рода задания. Чаще всего они достаточно простые («микрозадачи») – распознать надпись на картинке, обнаружить наличие какого-либо элемента в коротком ролике, отнести комментарий в Интернете к какой-либо категории и пр. Зарегистрировавшись на платформе, исполнитель может взять любое задание и выполнить его за небольшое вознаграждение. Частным случаем «микрозадачи» может стать заполнение онлайн-анкеты.

Отечественные исследователи – по крайней мере, если говорить об опросах на русском языке, – не пользуются АМТ в своей работе, что объясняется невозможностью участников и работодателей из России официально зарегистрироваться на этой платформе. В то же время имеется российский аналог – система Толока, созданная компанией «Яндекс»¹. Толока начала работать в 2014 г. для внутренних нужд «Яндекса», спустя несколько лет внешние заказчики также получили возможность публиковать свои заказы². К настоящему времени среди представителей отечественной социальной науки этот инструмент не получил распространения: имеющиеся примеры ограничены преимущественно студенческими и аспирантскими работами [6]³. Важным исключением является работа Ф. Чапковского о проведении интерактивных экспериментов

¹ Толока: [сайт]. URL: <https://toloka.yandex.ru/> (дата обращения: 04.07.2022).

² Лекция о Толоке. Как тысячи людей помогают нам делать Яндекс: [блог компании «Яндекс»]. 13.05.2018. URL: <https://habr.com/ru/company/yandex/blog/358462/> (дата обращения: 04.07.2022).

³ В качестве примера работы, где Толока выступала не в качестве инструмента поиска информантов, можно привести исследование Е. Таран и ее коллег, использовавших данную платформу для подготовки систематизированного обзора литературы [7]. Также можно отметить исследование Е. Даниловой, изучавшей мотивацию и стратегии участников Толоки, используя отзывы, размещенные на иной платформе [8].

в Толоке, где в том числе сопоставляются результаты применения этой платформы с результатами применения платформ АМТ и Prolific¹ [9]. В то же время работа Ф. Чапковского не фокусируется на качестве собранных опросных данных и не сравнивает их с данными, собранными с помощью более привычных инструментов – онлайн-панелей или посредством рекрутинга участников в социальных сетях.

В связи с этим цель настоящей статьи – представить наш опыт использования Толоки для рекрутинга участников онлайн-опроса, обратив внимание на некоторые индикаторы качества собранных данных в сравнении теми, что получены при других способах отбора респондентов. Здесь следует оговорить, что наше исследование имело в первую очередь содержательные цели и не задумывалось как методологическое: мы не планировали проводить систематическое изучение качества данных, полученных в ходе опроса респондентов из Толоки в сравнении с другими способами рекрутирования, и поэтому не закладывали в дизайн исследования какие-либо специфические показатели. Поэтому нашу работу следует рассматривать скорее как первый, предварительный шаг на пути к более проработанным исследованиям качества опросных данных, получаемых с помощью данной платформы.

Предваряет описание нашего опыта использования Толоки краткий обзор исследований, где оценивается качество опросов, проведенных на платформе АМТ, что в силу сходств этих двух платформ позволяет сформировать ожидания относительно использования Толоки.

¹ Prolific: [site]. URL: <https://www.prolific.co/> (date of access: 04.07.2022).

Качество опросных данных, собранных с помощью платформы АМТ

Исследования результатов использования платформы АМТ для рекрутинга респондентов показывают, что качество собранных данных сопоставимо с качеством данных, полученных с применением иных «конформных» типов выборки [10, р. 234]. Анализируемые в публикациях показатели качества данных можно условно разделить на те, которые собираются «незаметно» для участников исследования, и показатели, требующие явной реакции респондентов. Первые можно обозначить как «нереактивные» (“unobtrusive”), и они, как правило, фиксируются средствами платформы, на которой проводится опрос (например, платформа определяет время заполнения анкеты). Вторые – «реактивные» (“obtrusive”), чаще всего они специально включаются исследователями для того, чтобы замерить/сравнить реакции участников (например, включение теста на внимательность) [11]¹.

Поскольку наше исследование исходно не преследовало методологических задач, то нам в первую очередь были доступны именно «нереактивные» показатели, поэтому далее рассмотрим эту категорию показателей более подробно. В табл. 1 приведены ключевые показатели и указаны примеры методологических работ по АМТ, где они фигурируют, с краткой характеристикой результатов, релевантных нашей работе.

Время заполнения анкеты – один из наиболее часто используемых показателей при оценке онлайн-опросов, проведенных с помощью АМТ. Предшествующие исследования показывают, что участники АМТ тратят столько же или меньше времени, чем респонденты, рекрутированные иным способом. Так, К. Антуан с коллегами обнаружили, что пользователи АМТ тратят примерно в два раза меньше времени на заполнение анкеты по сравнению

¹ См. также: [12].

**КЛЮЧЕВЫЕ ПОКАЗАТЕЛИ КАЧЕСТВА ДАННЫХ
ОНЛАЙН-ОПРОСОВ, ПРОВОДИМЫХ С ПОМОЩЬЮ АМТ,
И ПРИМЕРЫ ИССЛЕДОВАНИЙ**

Показатель	Пример публикации	С чем сравнивается АМТ	Особенности АМТ
Время заполнения анкеты	Антуан и соавт., 2016 [10]	Рекрутинг с помощью Google; Рекрутинг с помощью Facebook*; Онлайн-объявления	Пользователи АМТ значимо быстрее заполняют анкету, чем респонденты, рекрутированные с помощью Google
	Кис и соавт., 2017 [13]	2 выборки студентов; 2 онлайн-панели	Пользователи АМТ меньше тратили времени на заполнение анкеты, чем члены одной из двух онлайн-панелей
	Оглтри и Капц, 2020 [14]	–	Величина вознаграждения в АМТ не влияет на затраченное на заполнение анкеты время
Время заполнения отдельных блоков анкеты	Смит и соавт., 2016 [11]	Онлайн-панель населения США;	Пользователи АМТ тратят меньше времени на ответы почти по всем блокам по сравнению с участниками онлайн-панели

*Сеть Facebook принадлежит компании Meta, признанной экстремистской организацией и запрещенной на территории Российской Федерации.

Продолжение табл. 1

Показатель	Пример публикации	С чем сравнивается АМТ	Особенности АМТ
Качество ответов на открытые вопросы	Беренд и соавт., 2011 [15]	Выборка студентов	Ответы пользователей АМТ содержат столько же слов, что и ответы студентов
	Кис и соавт., 2017 [13]	2 выборки студентов; 2 онлайн-панели	Участники АМТ давали ответ из больше числа символов, чем респонденты из двух коммерческих онлайн-панелей
Количество ответов «затрудняюсь ответить»	Шмилевски и Кукер, 2020 [5]	Несколько волн респондентов из АМТ	Падение качества ответов на открытые вопросы в 2015–2019 гг.: больше бессмысленных фраз, нерелевантных ответов
	Антуан и соавт., 2016 [10]	Рекрутинг с помощью Google; Рекрутинг с помощью Facebook*; Онлайн-объявления	Пользователи АМТ реже отмечают вариант «затрудняюсь ответить», чем респонденты, рекрутированные с помощью Google

Окончание табл. 1

Показатель	Пример публикации	С чем сравнивается АМТ	Особенности АМТ
Доля недозаполненных анкет	Паолачи и соавт., 2010 [16] Чжоу и Фишбах, 2016 [17]	Выборка студентов Онлайн-объявление	<p>Пользователи АМТ реже заполняют анкеты до конца (больше частичных ответов) по сравнению со студентами.</p> <p>По сравнению с респондентами, зашедшими на опрос через онлайн-объявление, пользователи АМТ чаще заполняют анкету до конца (меньше частичных ответов)</p> <p>Пользователи АМТ реже заполняют анкеты до конца по сравнению со студентами – участниками лабораторных исследований</p>

с теми, кто пришли по ссылке, размещенной средствами рекламы Google [10, р. 240]. Дж. Кис и соавторы, сравнивая выборку из АМТ с двумя студенческими выборками (в первой студенты получали ссылку для заполнения анкеты, во второй – проходили опрос в компьютерном классе) и двумя коммерческими онлайн-панелями, выяснили, что участники из АМТ тратили меньше времени на заполнение анкеты, чем респонденты одной онлайн-панели, тогда как статистически значимых различий по времени заполнения с другими выборками обнаружено не было [13, р. 149–150].

Указанные исследования, как правило, исходят из допущения, что чем больше времени тратится на заполнение, тем более внимательно респонденты подходят к задаче. Как следствие, в ряде работ обсуждается, каким образом можно повысить время заполнения анкеты. В частности, А. Оглтри и Б. Катц показали, что время заполнения анкет не зависит от величины вознаграждения за участие в опросе [14].

Время заполнения отдельных блоков анкеты – показатель, тесно связанный с предыдущим. Здесь результаты похожи и указывают на то, что пользователи АМТ достаточно быстро заполняют анкету. Так, С. Смит и его коллеги выяснили, что участники АМТ тратят меньше времени на заполнение 6 из 7 использованных блоков анкеты по сравнению с респондентами, отобранными через онлайн-панель домохозяйств США [11, р. 3144].

Качество ответов на открытые вопросы. Обычно измеряется количеством слов или символов, но могут использоваться и качественные характеристики (например, оценка релевантности ответа на поставленный вопрос [5]). В частности, Т. Беренд с коллегами подсчитывали количество слов в ответах на открытые вопросы и не обнаружили статистически значимых различий между выборкой на основе АМТ и выборкой студентов-добровольцев [15, р. 808]. В упомянутой ранее работе Киса и его соавторов отмечается, что участники АМТ использовали больше символов для ответов на два открытых вопроса, чем в выборке из обеих

онлайн-панелей, тогда как по этому показателю не было разницы между респондентами из АМТ и студентами, заполнявшими анкету в компьютерном классе [13, р. 149–150].

Количество выбранных вариантов «затрудняюсь ответить». Данный показатель фигурирует в литературе по АМТ редко. Так, К. Антуан с коллегами обнаружили, что пользователи АМТ статистически реже выбирают вариант «затрудняюсь ответить», чем респонденты, рекрутированные с помощью рекламы Google [10, р. 240], что предположительно свидетельствует о более высоком качестве получаемых с помощью АМТ данных.

Доля недозаполненных анкет (dropout) или заполненных до конца анкет (completion rate). Эти показатели скорее характеризуют процесс отбора респондентов, нежели полученные результаты. Тем не менее для оценки эффективности использования краудсорсинговых платформ как источника участников онлайн-исследований эти показатели имеют немаловажное значение. Имеющиеся исследования АМТ дают неоднозначную картину. С одной стороны, Г. Паолачи и его коллеги пишут, что пользователи АМТ практически всегда доходили до конца анкеты (91,6%, соответствующий показатель для студенческой выборки составил 98,6%), причем это происходило гораздо чаще, чем при использовании выборки добровольцев, рекрутированных по ссылке на онлайн-форуме (69,3%) [16]. Похожие результаты были получены в исследовании Дж. Спроуса – количество не до конца заполненных анкет составило 4%, тогда как все анкеты в оффлайн-эксперименте были заполнены полностью [18, р. 160]. С другой стороны, в работе Х. Чжоу и А. Фишбах было показано, что опросы в АМТ характеризуются высоким показателем недозаполненных анкет: проанализировав 88 исследований на этой платформе, они выяснили, что в 20% из них показатель недозаполненных анкет составил более 30%, тогда как в 96% лабораторных исследованиях со студенческими выборками все анкеты были заполнены полностью [17].

В нашем исследовании мы имели возможность проанализировать все указанные показатели, за исключением времени

заполнения отдельных блоков анкеты (это не позволяла делать платформа, на которой размещался опрос) и количества выборов варианта ответа «затрудняюсь ответить» (в анкете практически не предлагался этот вариант ответа).

Кроме того, в последней волне опроса мы включили вопрос, оценивающий внимательность респондента (“attention check question”). Это пример «реактивного» показателя, который часто фигурирует в исследованиях об АМТ, затрагивающих тему повышения качества получаемых данных за счет отсеивания недостаточно вовлеченных или внимательных респондентов [19; 20, р. 6–7]. Обычно респондентам предлагается ряд вопросов: например, в исследовании И. Агли и его коллег участникам задавались вопросы, посещали ли они за последние два года Латверию (вымышленную страну из комиксов Marvel), по какому номеру вызывать службу спасения, а также ряд других вопросов [21].

Смежными с показателями невнимательности оказываются вопросы недобросовестности. Так, М. Шмилевски и С. Кукер отсеивали респондентов, которые давали неправдоподобные ответы о количестве детей (более 10 детей) или отмечали наличие у них несовместимых психических расстройств [5, р. 466]. Разумеется, недостаточная внимательность или недобросовестность – это только некоторые из проблем, влияющих на валидность получаемых результатов. Помимо них исследователи обсуждают эффект самоотбора, возможность повторного участия, эффект социальной желательности, а также вопросы нечестности и искажения личных характеристик в целях успешного прохождения вопросов-фильтров или предоставления «удовлетворительных» (satisficing) ответов (см., например: [22; 23; 24; 25]¹).

К сожалению, ограниченные возможности по увеличению анкеты позволили нам включить лишь один вопрос, оценивающий

¹ См. также обзор преимуществ и особенностей применения АМТ, представленный в статье Л. Лу и ее соавторов [26].

внимательность. Примечательно, что в цитируемых работах практически не обсуждается то, насколько внимательность участников АМТ близка соответствующему показателю у респондентов, отобранных иным способом. Имеющиеся свидетельства редки и указывают на то, что АМТ позволяет получить не менее качественные в этом смысле данные. Так, Д. Хаузер и М. Шварц обнаружили, что респонденты, отобранные с помощью АМТ, даже более внимательны при чтении инструкций к вопросам – по сравнению с выборкой, состоящей из студентов [27].

В заключение обзора отметим, что в нашей работе мы не затрагиваем тему соответствия содержательных результатов (например, обнаруженных эффектов), полученных при опросе участников краудсорсинговой платформы и респондентов, рекрутированных иным способом. В то же время это – важнейшее направление исследований АМТ, показывающее пригодность этой платформы для проведения онлайн-исследований и экспериментов. Так, было показано, что АМТ позволяет воспроизвести эффекты лабораторных экспериментов, связанных с поведением в условиях «дилеммы заключенного», реакциями на прайминг и с выбором ситуаций, предъявленных под разными описаниями (фрейминг) [28]. В задании на поиск функции знакомого и незнакомого предмета для решения некоторой задачи не было обнаружено различий между тем, выполнялось ли задание в университетской лаборатории или онлайн – через социальные сети или АМТ [29]. Также АМТ позволила воспроизвести результаты предшествующих исследований, связанных с готовностью участвовать в благотворительности [30], а характеристики, полученные по шкале оценки личности (Personality Assessment Inventory), очень похожи на результаты национального обследования [31]. Иными словами, такого рода исследования свидетельствуют в пользу использования АМТ для проведения онлайн-экспериментов в социальных науках, что позволяет надеяться на продуктивность применения Толоки отечественными исследователями для решения своих задач.

Рекрутинг респондентов с помощью Толоки

Настройка сбора данных

В рамках нашего исследования «Соотношение традиционных рисков и киберрисков: опыт использования расширенной версии “психометрической парадигмы”» (2020–2021) мы использовали Толоку в качестве одного из способов рекрутирования респондентов. Поскольку использование этой платформы не является распространенной практикой среди социологов, опишем более подробно процесс рекрутинга¹.

Как правило, задания в Толоке выполняются непосредственно на платформе, с использованием ее средств. Однако для онлайн-опросов, в том числе рассмотренных в нашем обзоре по АМТ, это чаще всего не так: краудсорсинговая платформа лишь отвечает за рекрутинг, а заинтересованные участники обычно переходят по ссылке для прохождения опроса. Так и в нашем исследовании опрос был размещен на внешнем ресурсе – в системе SurveyMonkey². В Толоке для реализации такой схемы предлагается разместить задание, содержащее ссылку на опрос, а также поле для ввода кода подтверждения. Для создания такого задания можно воспользоваться шаблоном («пресет», в терминологии Толоки), который называется «Опрос на сторонней платформе». При выборе этого шаблона будет предложен интерфейс такого задания по умолчанию, который достаточно легко отредактировать при помощи конструктора, также можно воспользоваться средствами HTML/JS/CSS. Пример того, как может выглядеть задание, представлен на рис. 1. На странице редактирования проекта также нужно разместить инструкцию – краткое описание исследования и информацию о том, что ожидается от участника.

¹ За время проведения исследования платформа претерпела некоторые изменения. Приводимое описание соответствует интерфейсу по состоянию на лето 2022 г.

² SurveyMonkey: [site]. URL: <https://www.surveymonkey.com/> (date of access: 04.07.2022).

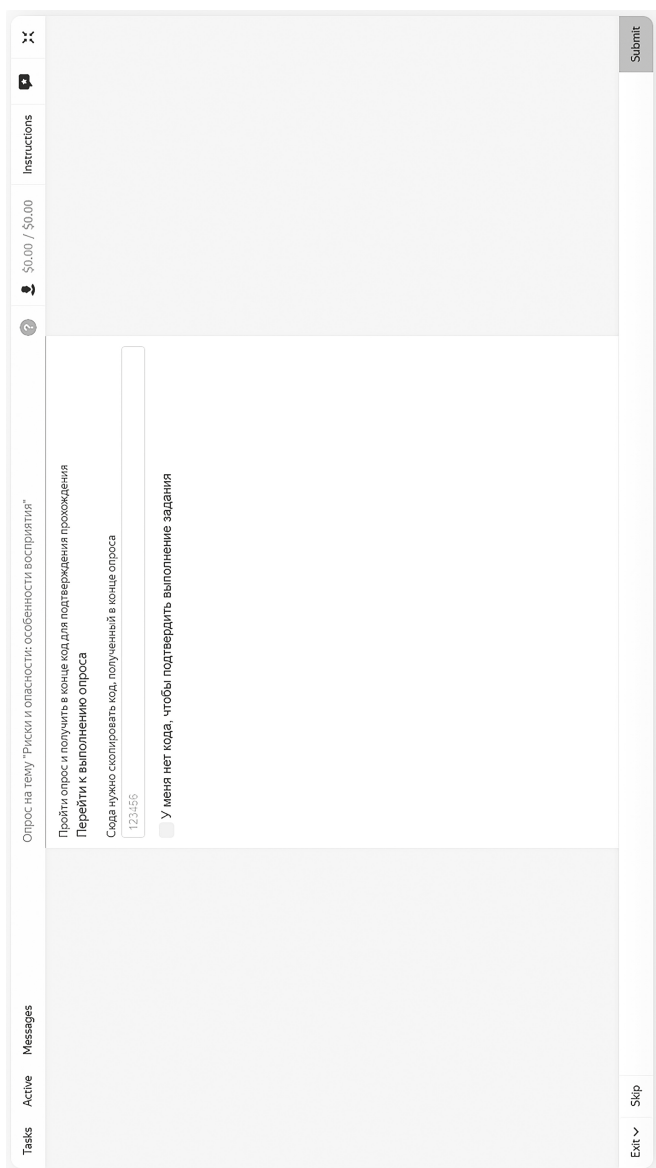


Рис. 1. Пример задания в Толоке с переходом по внешней ссылке.

После создания проекта необходимо добавить «пул» – это само задание с указанием того, кто приглашается к его выполнению. Одному проекту может соответствовать несколько заданий (пулов), что позволяет в рамках одного проекта отправлять разные ссылки на опрос, тестировать параметры поиска респондентов, предлагать разное вознаграждение и проч. Настройка пула в нашем случае включала следующие основные параметры:

- цена в долларах США: мы изначально указали 0,03 долл., но после изучения других заданий выставили 0,06 долл., эта сумма использовалась для основного массива данных;

- перекрытие (количество опрошенных) – задавалось разным, в зависимости от волны исследования;

- контент для взрослых – отмечено;

- параметры аудитории – мы использовали следующие:

- категория устройств – любая, но не включая смартфон, планшет и игровую консоль. Это было сделано по причине того, что вопросы матричного типа в SurveyMonkey на смартфонах могли отображаться некорректно;

- язык – русский;

- соотношение скорости и качества – 80% лучших исполнителей;

- параметры контроля качества:

- частота показа капчи – низкая;

- отложенная приемка – да;

- срок проверки в днях – 7 дней;

- динамическое ценообразование и динамическое перекрытие – выключены;

- дополнительные настройки:

- время на страницу заданий – 3600 секунд.

После настройки пула следует загрузить задание, в нашем случае это был tsv-файл, созданный на основе предлагаемого шаблона, со следующим содержанием (survey_link в нашем случае – это параметр, заданный в проекте):

INPUT:survey_link
<https://ru.surveymonkey.com/r/3H23812>

После этого система отправила задание на «разметку», т.е. начала предлагать участникам платформы пройти опрос. Завершив прохождение опроса, участник не получает вознаграждение автоматически: исследователь должен принять ответ, используя интерфейс «Выполненные задания».

Особенности и рекомендации

Тривиальной рекомендацией является необходимость тестирования, т.е. начинать с рекрутинга небольшого числа респондентов с отслеживанием обратной связи. Как показало наше исследование, участники Толоки активно пишут автору исследования сообщения через интерфейс Толоки при обнаружении проблемы. Также система сама направляет предупреждения, если при выполнении заданий наблюдаются аномалии (например, если пользователи каких-то устройств не могут завершить задание).

Одной из ключевых проблем, с которой мы столкнулись, – это необходимость выдать респонденту код, который он должен ввести на стороне Толоки для завершения исследования. К сожалению, платформа SurveyMonkey не позволяет генерировать числа, потому нашим основным решением стал завершающий исследование открытый вопрос, где респондент должен был сам придумать любое число и вписать его, после чего вписать этот код в интерфейсе Толоки. У этого решения имелись недостатки:

– не все респонденты осознали или обратили внимание, что придуманный ими самими код нужно ввести на стороне Толоки. Наибольшее число обращений как раз было связано с проблемами получения данного кода подтверждения;

– поскольку открытый вопрос с запросом кода был последним на странице, наряду с другими вопросами, то какое-то количество респондентов могли вписать код, ввести его на стороне Толоки,

но не завершить исследование в SurveyMonkey. Здесь мы можем рекомендовать размещать вопрос с запросом кода на отдельной (завершающей) странице.

Для решения проблемы завершения опроса можно использовать другие решения. В частности, применить функцию A/B-тестирования для выдачи респонденту одного из некоторого перечня заранее сгенерированных кодов¹.

Другой вариант состоял в том, чтобы предлагать код не в вопросе, а на странице благодарности / завершения исследования (“custom thank you message”, в терминологии SurveyMonkey). К сожалению, в этом случае респондентам будет предлагаться один и тот же код, но поскольку респонденты едва ли об этом будут догадываться, то это скорее всего не будет проблемой. Мы испытали этот вариант в заключительной волне исследования. Его дизайн включал в себя шесть версий одной и той же анкеты, но с рандомизацией блоков опросника, так что на странице благодарности участник получал один из шести возможных кодов. Отсутствие обращений по проблеме ввода кода от пользователей свидетельствует, что такой вариант предпочтительнее выбранного нами на начальном этапе.

Также в ходе заключительной волны исследования выяснилось, что заданный нами фильтр на устройство пользователя (любое устройство, но не включая смартфон, планшет и игровую консоль) работает недостаточно хорошо: администрация Толоки сообщила, что 7% исполнителей на Android не могут завершить опрос, хотя мы явно запрещали смартфон в качестве устройства. Для решения этой проблемы была дана рекомендация установить иные фильтры в настройках аудитории:

¹ Этот способ описан в гайде по подключению опроса на SurveyMonkey к AMT. См.: Tutorial: Getting great survey results from MTurk and SurveyMonkey. 03.06.2017. URL: <https://blog.mturk.com/tutorial-getting-great-survey-results-from-mturk-and-surveymonkey-f49c2891ca6f> (date of access: 04.07.2022).

- клиент: веб-версия Толоки;
- категория устройства: персональный компьютер.

Также в ходе последней волны возникла проблема со случайным предъявлением шести версий анкеты: ссылки распределяла сама Толока (для этого tsv-файл задания должен был содержать все шесть ссылок), но по умолчанию система не запрещала пользователю несколько раз выполнить задание, так как рассматривала ссылки как не исключающие друг друга. Чтобы решить эту проблему, в разделе «Контроль качества» нужно было сформулировать следующее правило: «Если отправленных страниц заданий = 1, то приостановить в пуле навсегда».

Альтернативные способы рекрутирования и результаты отбора

Помимо Толоки, рекрутинг респондентов в рамках нашего проекта осуществлялся двумя дополнительными способами.

1. Посредством распространения приглашения к участию через социальные сети – Facebook^{1*}, Instagram^{2**} и «ВКонтакте». Участники не получали какого-либо вознаграждения за заполнение анкеты.

2. С помощью онлайн-панели компании OMI³. Участники получали за заполнение анкеты вознаграждение от 20 до 50 баллов во внутренней «валюте» OMI, что не превышало 75 руб. за анкету.

^{1*} Сеть Facebook принадлежит компании Meta, признанной экстремистской организацией и запрещенной на территории Российской Федерации. Сбор данных осуществлялся до введения данного запрета.

^{2**} Сеть Instagram принадлежит компании Meta, признанной экстремистской организацией и запрещенной на территории Российской Федерации. Сбор данных осуществлялся до введения данного запрета.

³ OMI: Online Market Intelligence: [сайт]. URL: <https://www.omirussia.ru/> (дата обращения: 01.07.2022).

Все три способа рекрутирования осуществлялись в мае–июне 2020 г. Дополнительная волна сбора данных исключительно с помощью Толоки была реализована летом 2022 г. Результаты отбора представлены в табл. 2.

Таблица 2

РЕЗУЛЬТАТЫ ОТБОРА РЕСПОНДЕНТОВ

Способ отбора	Количество анкет*	Доля женщин (%)	Средний возраст (SD)	Доля с высшим образованием** (%)
Социальные сети	189	72	20,3 (7,9)	46
Онлайн-панель	338	61	39,9 (11,7)	65
Толока	1370	51	36,7 (11,5)	59
Толока (2022)***	830	43	39,5 (11,3)	51

* Указано только количество полностью заполненных анкет.

** Включает в себя респондентов, имеющих диплом бакалавра, магистра, специалиста, закончивших аспирантуру или защитивших диссертацию.

*** Дизайн анкеты отличался и поэтому результаты этой волны приводятся отдельно.

Отметим, что все выборки значимо различаются по гендерному составу. Женщин больше всего среди респондентов из соцсетей, и меньше всего – в обеих волнах из Толоки¹. Почти то же самое справедливо для возраста (средний возраст значимо различается по всем выборкам, за исключением онлайн-панели и второй волны

¹ Примечательно, что Ф. Чапковский также отмечает, что мужчин несколько больше в его выборке, чем женщин [9, р. 10]. В то же время последние исследования на основе АМТ свидетельствуют, что гендерный состав является непостоянной величиной. Так, в исследовании Шмилевски и Кукер было 57% женщин, а во всех подвыборках исследования Агли и его коллег, кроме одной, было больше мужчин (54,4–57,8%) [21].

Толоки, где он оказался равен 40 годам¹) и образования (больше всего респондентов с высшим образованием в онлайн-панели (65%), меньше всего – в выборке из социальных сетей (46%), хотя этот показатель оказался статистически незначимо отличным от второй волны Толоки (51%)²)).

Общие впечатления от рекрутинга респондентов с помощью Толоки в целом соответствуют тому, что известно про АМТ. Во-первых, стоимость одной заполненной анкеты респондентами из Толоки значительно меньше, чем стоимость рекрутинга через коммерческую онлайн-панель: с точки зрения затраченных средств исследователем Толока дешевле онлайн-панели более чем в 10 раз (ср.: [10, p. 234–235]). Во-вторых, скорость рекрутирования респондентов очень высока. Поскольку поиск осуществлялся в несколько этапов, то мы не можем точно оценить, сколько времени занял сбор 1370 анкет первой волны. Однако совершенно точно платформа позволяет собрать несколько сотен заполненных анкет за несколько часов. Так, завершающий этап последней волны включал в себя привлечение 600 исполнителей, что было выполнено Толокой за 2 часа 23 минуты. Иными словами, если по стоимости рекрутинга Толока ближе к поиску «бесплатных» респондентов через социальные сети, то по скорости достижения целевого объема выборки она ближе к онлайн-панелям. В то же время если принять во внимание усилия, которые требуются для поиска участников исследования через социальные сети³, то Толока представляется весьма эффективным инструментом.

Показатели и гипотезы

Поскольку исходно в дизайн исследования не закладывалась оценка качества собранных данных, то мы воспользовались такими

¹ $t(1166) = 0,56, p = 0,6$.

² $\chi^2(1) = 1,9, p = 0,17$.

³ См., напр., опыт рекрутирования респондентов, описанный Р. Абрамовым [32].

«нереактивными» показателями, как время заполнения анкеты и количество респондентов, не заполнивших анкету до конца.

Содержательная часть анкеты состояла из 12 однотипных таблиц, что дало возможность обратиться к такому показателю качества данных, как доля респондентов, выбиравших один и тот же ответ в каждой строке таблицы, т.е. так называемых стрейтлайнеров (straightliners). Хотя в рассмотренных исследованиях АМТ данный показатель не фигурирует, однако он достаточно часто применяется в работах по методологии онлайн-исследований [33; 34; 35; 36; 37]. Для каждого респондента мы определили количество таблиц, где он отмечал один и тот же вариант ответа в строке, и считали его стрейтлайнером при условии, что таких таблиц было не менее четырех¹.

К сожалению, единственный открытый вопрос – предложение дать комментарии и замечания по опросу – в анкете для Толоки имел другую формулировку, поэтому сравнение ответов по основному массиву данных было невозможно. Однако проведение дополнительной волны исследования в 2022 г. позволило включить открытый вопрос на той же позиции² (чтобы исключить эффекты, связанные с усталостью респондента) и с той же формулировкой с целью проведения сравнения как доли респондентов, решивших ответить на этот вопрос, так и количества слов в предложенных ответах. Поскольку после открытого вопроса были заданы дополнительные, это не позволило сопоставить время заполнения и долю завершенных анкет с исходным рекрутингом в 2020 г.

Гипотеза относительно времени, затраченного респондентами на заполнение анкеты, состояла в том, что участники из Толоки тратят меньше времени, чем респонденты, рекрутированные

¹ Такая стратегия не позволила выявить респондентов, которые применяли иной паттерн, например – выбирали ответы по диагонали.

² До этого вопроса был добавлен только вопрос на внимательность (см. далее в наст. статье), что едва ли могло повлиять на результаты исследования.

другими способами. На это указывают не только исследования АМТ, но и тривиальное соображение о том, что низкие уровни вознаграждения на платформе позволяют участникам получать значимые финансовые средства только в том случае, если они будут выполнять большое количество заданий. Как результат – максимально быстрый ответ на вопросы, это, по-видимому, рациональная стратегия работы на платформе.

Гипотеза относительно заполненных до конца анкет заключалась в том, что доля незаполненных анкет будет мала и будет ближе к уровню, который демонстрируют респонденты из онлайн-панели.

Гипотеза относительно доли однотипных ответов в таблицах аналогична гипотезе относительно времени, затраченного на заполнение, так как эти два показателя считаются связанными [35]: мы ожидали, что доля стрейтлайнеров будет максимальной именно среди респондентов из Толоки.

Гипотеза по открытым вопросам состояла в том, что участники из Толоки скорее всего будут реже отвечать на открытый вопрос для экономии времени (в отличие от предшествующих исследований, в нашей анкете вопрос был необязательным). Те же респонденты, которые решат ответить на этот вопрос, скорее всего будут формулировать ответ из такого же количества слов, как и респонденты, отобранные иными методами.

Запуск дополнительной волны позволил включить в исследование также вопрос на внимательность. Основываясь на работе И. Агли и его коллег¹, мы предложили респондентам ответить

¹ В их исследовании один из вопросов звучал так: «Когда вы были в школе, насколько усердно вы учились? Отвечая на этот вопрос, проигнорируйте все варианты и выберите последний вариант, указывающий, что вы не помните» [21, р. 888]. В цитируемой работе использовались и другие вопросы, позволяющие выявить невнимательных респондентами, но мы ограничились только одним из-за ограничений на объем анкеты, а также по причине того, что подобный вопрос органично вписывался в дизайн анкеты рядом с вопросом про образование.

на следующие вопрос: «Когда вы учились в школе в старших классах, насколько допустимо для вас было прогулять урок? Пожалуйста, не обращайтесь внимание на варианты ответов и выберите последний вариант, что вы не учились в школе»¹. Поскольку этот вопрос не задавался респондентам из онлайн-панели и социальных сетей, то мы решили проверить наличие связи внимательности с рассмотренными ранее показателями (со временем заполнения анкеты, долей ответов на открытый вопрос, числом слов в таких ответах, а также с частотой однотипных ответов). Дополнительно выяснили, влияет ли величина оплаты на внимательность: для проверки наличия такой связи часть (598) анкет заключительной волны предлагалась с повышенной оплатой за их заполнение (0,12 вместо 0,06 долл.).

Результаты

Время заполнения, доля заполненных до конца анкет и доля анкет с однотипными ответами

Результаты сравнения трех способов рекрутирования по доле заполненных анкет, времени заполнения, а также доле анкет с однотипными ответами в таблицах приведены в табл. 3. Заметим, что ряд респондентов потратили на анкету более часа, их время было решено не учитывать при вычислении среднего и медианы. Таких «аномальных» респондентов было мало при рекрутировании из соцсетей и с помощью Толоки (7 и 2 случая соответственно), однако среди участников онлайн-панели их оказалось 21, причем из них 7 человек отвечали на вопросы анкеты более 3 часов. Маловероятно, что эти респонденты действительно тратили так много времени, скорее всего они переключались на решение других задач, после чего возвращались к опросу.

¹ К сожалению, прямой вопрос о том, учились ли респондент в школе, не задавался. Но, принимая во внимание обязательность школьного образования, можно предположить, что все участники исследования так или иначе учились в школе.

Среднее время заполнения анкеты статистически не различается только для участников онлайн-панели и респондентов из социальных сетей¹. Иными словами, можно утверждать, что через Толоку рекрутируются респонденты, которые быстрее заполняют анкету, что потенциально негативно может сказываться на качестве получаемых результатов.

Одновременно участники Толока характеризуются такой же высокой мотивацией заполнить анкету до конца, как и члены онлайн-панели, – доля заполненных анкет для этих двух источников статистически не различается² и заметно выше, чем аналогичный показатель для социальных сетей. Здесь следует напомнить, что респонденты социальных сетей не получали вознаграждения за участие в опросе.

Что касается анкет с однотипными ответами, то доля респондентов-стрейтлайнеров значительно различается для всех трех типов выборки: их оказалось меньше всего среди респондентов из социальных сетей (1,6%), больше всего – из онлайн-панели (13,3%), а Толока характеризуется средним показателем (8,1%). Иными словами, наша гипотеза о том, что именно среди респондентов из Толоки будет больше всего стрейтлайнеров, не подтвердилась. Кроме того, наши результаты соответствуют выводам предыдущих онлайн-экспериментов о том, что скорость заполнения анкеты связана с наличием однотипного паттерна в ответах: тогда как стрейтлайнеры тратили на заполнение анкеты в среднем 7,5 минут, все остальные это делали в среднем за 12,5 минут³.

Качество ответов на открытый вопрос

Характеристики ответов на открытый вопрос приведены в табл. 4. Доля анкет с наличием ответа на открытый вопрос

¹ $t(500) = -0,46, p = 0,6$.

² $\chi^2(1) = 2,85, p = 0,09$.

³ $t(205) = 10,3, p = 0,000$.

Таблица 3
ДОЛЯ ЗАПОЛНЕННЫХ ДО КОНЦА АНКЕТ, ВРЕМЯ ЗАПОЛНЕНИЯ И ДОЛЯ ОДНОТИПНО ЗАПОЛНЕННЫХ ТАБЛИЦ, 2020 г.

Способ отбора	Количество анкет		Доля анкет, заполненных до конца (%)	Время заполнения*		Доля анкет с однотипными ответами** (%)
	начатых	заполненных		среднее	медианное	
Социальные сети	428	189	44,1	14:59	12:54	1,6
Онлайн-панель	358	338	94,4	14:36	11:43	13,3
Толока	1493	1370	91,8	11:14	10:20	8,1

* После исключения «выбросов» – заполнение анкеты более часа.

** Анкета считалась заполненной однотипными ответами, если не менее чем в 4 таблицах были выбраны ответы в одной колонке.

Таблица 4
ХАРАКТЕРИСТИКИ ОТВЕТОВ НА ОТКРЫТЫЙ ВОПРОС (ОВ), 2020–2022 гг.

	Количество анкет		Доля анкет с ответом на ОВ (%)	Количество слов*	
	заполненных	с ответом на ОВ		среднее	медианное
Социальные сети	189	15	7,9	21,1	13
Онлайн-панель	338	51	15,1	5,3	4
Толока (2022 г.)	830	81	9,8	7,3	2

* Среднее и медианное количество слов рассчитывается для анкет, имеющих ответ на открытый вопрос.

примерно равна при сборе данных через социальные сети (7,9%) и посредством Толоки (9,8%)¹. В то же время участники онлайн-панели гораздо чаще отвечают на открытый вопрос (15,1%)².

Респонденты, рекрутированные через социальные сети, дают более объемные ответы – в среднем 21,1 слова (при медиане 13), что больше соответствующего показателя участников онлайн-панели (5,3, медиана 4) и Толоки (7,3, медиана 2)³. Разница в среднем количестве слов в анкетах участников онлайн-панели и Толоки статистически не значима⁴.

Иными словами, паттерн ответов на открытый вопрос респондентов из Толоки отличается от паттерна, наблюдаемого при рекрутинге другими способами: участники Толоки так же редко отвечают, как и респонденты из социальных сетей, но если они дают ответ, то он по своей длине больше похож отклик участников онлайн-панели.

Также мы проверили наличие связи между наличием ответа на открытый вопрос и тем, является ли респондент стрейтлайнером: такая связь не была обнаружена⁵. Иными словами, вопреки ожиданиям, стрейтлайнеры не реже отвечают на открытый вопрос, однако они дают более короткие ответы (медиана – 1,5 слова, тогда как у остальных респондентов этот показатель составил 3⁶).

¹ $\chi^2(1) = 0,6, p = 0,44$.

² Это больше показателя в 7,9% у соцсетей ($\chi^2(1) = 6,1, p = 0,014$) и в 9,8% у Толоки ($\chi^2(1) = 7,5, p = 0,006$).

³ При сравнении обеих пар критерий *U* Манна – Уитни для независимых выборок показывает наличие различий ($p < 0,000$).

⁴ Критерий *U* Манна – Уитни для независимых выборок показывает отсутствие различий в распределениях ($p = 0,055$), хотя при доверительной вероятности 0,9 данные различия следовало бы признать значимыми.

⁵ $\chi^2(1) = 0,3, p = 0,56$.

⁶ Критерий *U* Манна – Уитни для независимых выборок показывает наличие различий в распределениях ($p = 0,015$).

Оценка внимательности и ее влияние на показатели качества

На вопрос, позволяющий оценить внимательность, ответили 830 респондентов, из них 296 (35,7%) выбрали вариант «Я не учился в школе», свидетельствующий о внимательном прохождении опроса (далее эти респонденты будут называться внимательными).

Связь внимательности при заполнении анкеты с показателями качества анкеты, использованными ранее, представлена в табл. 5.

Внимательные респонденты в целом затрачивают на заполнение анкеты значительно больше времени¹ и среди них реже встречаются стрейтлайнеры². Доля анкет с наличием ответа на открытый вопрос не зависит от внимательности³, в отличие от того, как респонденты отвечают на него. На рис. 2 представлена диаграмма с распределением количества слов в ответе, где видно, что невнимательные респонденты чаще всего дают ответ из 1–2 слов, тогда как внимательные – либо из 1–2, либо 4–5 слов. Наличие различий подтверждают и статистические тесты: согласно одновыборочному критерию Колмогорова–Смирнова, оба распределения не являются нормальными ($p = 0,000$), а критерий U Манна – Уитни для независимых выборок показывает, что перед нами – два разных распределения ($p = 0,002$).

Поскольку в заключительной волне исследования в 2022 г. участники Толоки получали разное вознаграждение за прохождение опроса – 0,06 и 0,12 долл., – это позволило проверить гипотезу о влиянии уровня вознаграждения на внимательность, а также на другие показатели качества заполнения анкеты (табл. 6).

Увеличение вознаграждения не оказывает влияния на то, насколько внимательно респонденты отвечают на вопросы⁴, однако

¹ $t(825) = -4,75, p = 0,000$.

² $\chi^2(1) = 6,7, p = 0,01$.

³ $\chi^2(1) = 0,58, p = 0,45$.

⁴ $\chi^2(1) = 1,38, p = 0,24$.

Таблица 5
ПОКАЗАТЕЛИ КАЧЕСТВА ЗАПОЛНЕНИЯ АНКЕТЫ В ЗАВИСИМОСТИ
ОТ ВНИМАТЕЛЬНОСТИ, 2022 г.

Тип респондента	Кол-во анкет	Время заполнения*		Доля анкет с однотипными ответами** (%)	Доля анкет с ответом на ОВ (%)	Количество слов***	
		среднее	медианное			среднее	медианное
Внимательный	296	13:54	12:06	6,1	10,8	11,4	4
Невнимательный	534	11:29	10:25	11,6	9,2	4,5	2
Всего	830	12:21	11:02	9,6	9,8	7,3	2

*После исключения «выбросов» – заполнение анкеты более часа.

**Анкета считалась заполненной однотипными ответами, если не менее чем в 4 таблицах были выбраны ответы в одной колонке.

***Среднее и медианное количество слов рассчитывается для анкет, имеющих ответ на открытый вопрос.

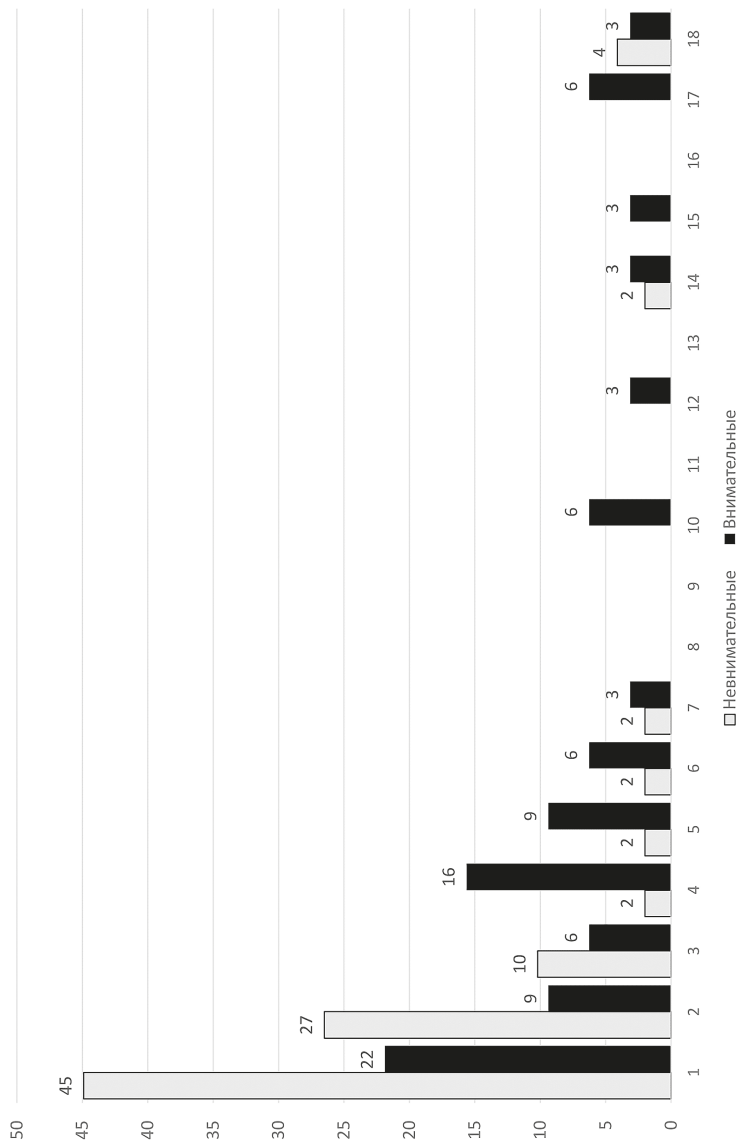


Рис. 2. Распределение количества слов в ответе на открытый вопрос (%)

Таблица 6
 ПОКАЗАТЕЛИ КАЧЕСТВА ЗАПОЛНЕНИЯ АНКЕТЫ В ЗАВИСИМОСТИ ОТ ВЕЛИЧИНЫ
 ВОЗНАГРАЖДЕНИЯ, 2022 г.

Величина вознаграж- дения	Кол-во анкет	Доля внима- тельных респонден- тов (%)	Время заполнения*		Доля анкет с однотип- ными отве- тами** (%)	Доля анкет с ответом на ОВ (%)	Количество слов***	
			среднее	медиан- ное			среднее	медиан- ное
0,06	232	30,4	11:08	9:50	14,2	4,3	21,9	3,5
0,12	598	34,4	12:49	11:28	7,9	11,9	5,2	2
Всего	830	35,7	12:21	11:02	9,6	9,8	7,3	2

*После исключения «выбросов» – заполнение анкеты более часа.

**Анкета считалась заполненной однотипными ответами, если не менее чем в 4 таблицах были выбраны ответы в одной колонке.

***Среднее и медианное количество слов рассчитывается для анкет, имеющих ответ на открытый вопрос.

увеличивает время заполнения анкеты: рост вознаграждения на 0,06 долл. приводит к тому, что респонденты затрачивают на заполнение анкеты в среднем на 1,5 минуты больше¹. Аналогично растет и доля анкет с ответом на открытый вопрос (с 4,3 до 11,9%²) и падает доля стрейтлайнеров (с 14,2 до 7,9%³). Несмотря на то что среднее и медианное количество слов в ответе выше у респондентов, получивших меньшее вознаграждение, следует учитывать, то таких респондентов в нашей выборке всего 10, как следствие – данное различие нельзя считать значимым⁴.

Таким образом, эти предварительные результаты показывают, что увеличение вознаграждения способно увеличить время заполнения анкеты и количество ответов на открытые вопросы, снизить долю стрейтлайнеров, но может не сказаться на доле внимательных респондентов и количестве слов в открытых вопросах.

Заключение

Хотя полученные результаты исследования свидетельствуют, что рекрутинг респондентов с помощью платформы Толока имеет право на существование, следует учесть ряд ограничений. Прежде всего, это касается проанализированных показателей, которые носят «нереактивный» характер и выступают лишь косвенными индикаторами собственно качества собираемых данных. Относительно времени заполнения анкеты в целом общепринята точка зрения, сформированная еще в эпоху «классических» опросов методом личного или телефонного интервью, что более медленное заполнение анкеты приводит к более вдумчивым ответам

¹ $t(825) = -3,1, p = 0,002$.

² $\chi^2(1) = 10,9, p = 0,001$.

³ $\chi^2(1) = 7,8, p = 0,005$.

⁴ Критерий U Манна – Уитни для независимых выборок показывает отсутствие различий в распределениях ($p = 0,262$).

и к снижению количества случайных ошибок (см. напр.: [38]), однако применительно к онлайн-опросам, когда отсутствует контроль за ходом заполнения анкеты, нет гарантии, что больше затраченного времени на заполнение анкеты свидетельствует о большем погружении в нее¹. Впрочем, результаты нашего исследования подтверждают наблюдение, что скорость заполнения анкеты связана с наличием однообразно заполненных табличных ответов: стрейтлайнеры действительно быстрее заполняли анкеты, т.е. давали исследователю менее качественные данные². Это позволяет сделать осторожный вывод, что, учитывая время заполнения и долю стрейтлайнеров, качество собранных данных с помощью Толоки сопоставимо с тем, что дает нам рекрутинг в соцсетях и с помощью онлайн-панелей.

Доля заполненных до конца анкет в принципе не является показателем качества данных. Безусловно, исследователи с самого начала распространения онлайн-опросов стремились к тому, чтобы снизить вероятность прерывания заполнения анкеты (dropouts), однако количество таких отказов – это скорее показатель качества инструментария и эффективности способа рекрутирования (см. напр.: [40; 41]). Иными словами, то, что рекрутинг при помощи Толоки характеризуется такой же высокой долей заполненных анкет (completion rate), что и онлайн-панель, говорит скорее об удобстве этого инструмента для исследователя, но никак не характеризует качество получаемых данных.

Что касается ответов на открытые вопросы, то количество слов гипотетически связано с погруженностью респондента и тем

¹ Также можно обсуждать вопрос о том, что время заполнения анкеты может характеризовать сложность анкеты для разных категорий респондентов. Иными словами, разница в затраченном времени может быть связана не столько с платформой, сколько с характеристиками респондентов. Например, имеются свидетельства того, что чем старше респонденты, тем медленнее они отвечают на вопросы (см. напр.: [39]).

² Отметим, что некоторые исследователи подчеркивают, что элементы стрейтлайнинга не всегда свидетельствуют о низком качестве заполнения анкеты [37].

самым соотносится с качеством данных. Исследователи отмечают различные, в том числе технические способы побудить респондента вписывать более развернутые ответы (см., напр.: [42]). И в этом смысле результаты нашего исследования, заключающиеся в том, что респонденты из Толоки с такой же вероятностью отвечают на опциональный открытый вопрос, что и респонденты из соцсетей, и что длина ответа примерно такая же, как у участников онлайн-панели, – это аргумент в пользу Толоки как способа рекрутирования. В то же время стоит помнить, что наш открытый вопрос был необязательным и содержал в себе лишь предложение дать комментарии и замечания по анкете. Как результат – очень мало респондентов в целом откликнулись на эту просьбу, кроме того, нет гарантии, что паттерн ответов на такой вопрос соответствовал бы тому, что наблюдалось бы при ответе на «полноценный» открытый вопрос.

Среди других ограничений нашей работы отметим следующие. Во-первых, отсутствие вопроса на внимательность в анкетах, предъявленных респондентам из социальных сетей и из онлайн-панели. Иными словами, тот факт, что треть респондентов оказались невнимательными, нельзя интерпретировать ни как преимущество, ни как недостаток – по сравнению с другими способами рекрутирования. Сопоставлять эти результаты с теми, что получены Агли и его коллегами, также нельзя: наша формулировка вопроса была неэквивалентной, а в указанной работе сообщается лишь итоговая доля невнимательных респондентов (31%), которые не прошли данный тип проверки (включавший, помимо нашего, еще 2 вопроса) [21].

Во-вторых, мы не анализировали разницу в содержательных ответах на вопросы анкеты представителей разных выборок, а также не изучали влияние разницы социально-демографических характеристик выборок на качество данных. Последнее представляется особенно важным: если, например, гендер (или образование) респондента влияет на скорость заполнения анкеты,

то разница между выборками по этому показателю может быть вызвана именно различиями в гендерном составе (или в уровне образования), а не спецификой рекрутирования.

В-третьих, мы не анализировали время рекрутинга разными способами. В частности, сбор данных с помощью Толоки происходил поэтапно, с отслеживанием возникающих ошибок и анализом обратной связи. Однако те сведения о скорости рекрутинга, которые показывает сама Толока, позволяют с уверенностью утверждать, что Толока – это инструмент, позволяющий очень быстро рекрутировать респондентов. Здесь стоит упомянуть также исследование Ф. Чапковского, где отмечается, что участники Толоки быстрее, чем участники других платформ (АМТ и Prolific), принимают задание и начинают его исполнять [9, р. 17–19].

В-четвертых, мы не оценивали эффекты параметров рекрутинга респондентов на качество получаемых данных. Исключением стала попытка посмотреть, связана ли величина вознаграждения с внимательностью респондентов, в результате которой мы пришли к выводу, что связь отсутствует¹, хотя рост вознаграждения снижает скорость заполнения, увеличивает долю ответов на открытые вопросы и сокращает долю стрейтлайнеров. В то же время имеются и другие параметры рекрутинга, которые настраиваются на стороне платформы. Прежде всего, это «соотношение скорости и качества», которое в нашем исследовании всегда было равно 80% (это значит, что задание предлагалось 80% «лучших» исполнителей). Справка Толоки не разъясняет, каким образом определяются эти «лучшие» исполнители, поэтому прояснение вопроса, влияет ли этот показатель на качество данных, представляет особый интерес. Применительно к АМТ было обнаружено, что рейтинг участника, основанный на доле принятых работ (НП Approval Rate), тесно связан с показателями внимательности [44],

¹ Что соответствует исследованиям АМТ, см., напр.: [43].

однако Толока не позволяет фильтровать аудиторию по такого рода параметрам¹.

В-пятых, мы не учитывали ряд аспектов, связанных со временем сбора данных. Так, Л. Кейси и его коллеги показали, что на характеристики респондентов из АМТ значимо влияет время суток, когда участник отвечал на вопросы, а также день недели [45]. Соответственно, не исключено, что время заполнения анкеты могло повлиять на качество данных, а поскольку мы его не контролировали, то это могло исказить полученные результаты. Другой аспект – это время, прошедшее между волнами исследования. Разница в ответах на открытый вопрос могла быть связана не со спецификой рекрутинга с помощью Толоки, а с тем, что заключительная волна была проведена спустя более чем год после опроса участников онлайн-панели и пользователей социальных сетей. Так, М. Шмилевски и С. Кукер показали, что качество ответов респондентов из АМТ за 4 года существенно снизилось [5, р. 469]; возможно, нечто подобное могло произойти с Толокой с 2020 по 2022 г.

Таким образом, хотя мы и считаем наш опыт использования Толоки успешным (платформа позволила собрать значительный объем данных за короткий срок и с небольшими затратами, а рассмотренные в этой статье показатели свидетельствуют о сопоставимости качества этих данных с качеством данных, полученных при опросе пользователей социальных сетей и участников онлайн-панели), но обозначенные выше ограничения указывают на то, как много еще предстоит сделать, чтобы использование Толоки в социальных исследованиях воспринималось как приемлемая

¹ Ф. Чапковский отмечает в качестве ключевого недостатка Толоки именно то, что эта платформа не предоставляет инструментов для отсеивания участников с низким качеством выполнения заданий. Обобщенный показатель качества участника, предлагаемый Толокой, «не коррелирует с каким-либо поведением, имеющим отношение к поведенческим исследованиям» [9, р. 34].

практика. Так или иначе, мы надеемся, что наша статья внесет вклад в обсуждение возможностей этой платформы для проведения опросов и экспериментальных социальных исследований в сети Интернет.

ЛИТЕРАТУРА

1. *Couper M.* Web Surveys: A review of issues and approaches // *Public Opinion Quarterly*. 2000. Vol. 64 (4). P. 464–494.

2. Отчет рабочей группы ААРОР о неслучайных выборках. Июнь 2013 г. / Американская ассоциация исследователей общественного мнения; пер. с англ. Д. Рогозина, А. Ипатовой. М.: ФОМ, 2016. 170 с. URL: https://fom.ru/uploads/files/FOM_AAPOP_book1.pdf (дата обращения: 01.07.2022).

3. *Мавлетова А. М.* Социологические опросы в сети Интернет: возможности построения типологии // *Социология: методология, методы, математическое моделирование (Социология: 4М)*. 2010. № 31. С. 115–134.

4. *Девятко И. Ф.* От «виртуальной лаборатории» до «социального телескопа»: метафоры тематических и методологических инноваций в онлайн-исследованиях // *Онлайн-исследования в России: тенденции и перспективы / Под ред. А. В. Шашкина, И. Ф. Девятко, С. Г. Давыдова*. М.: МИК, 2016. С. 19–33.

5. *Chmielewski M., Kucker S.* An MTurk crisis? Shifts in data quality and the impact on study results // *Social Psychological and Personality Science*. 2020. Vol. 11 (4). P. 464–473. DOI: 10.1177/1948550619875149

6. *Кулиева А. К.* Возможность использования Яндекс. Толока для проведения онлайн-экспериментов (на примере решения сенсомоторной задачи) // *Ананьевские чтения – 2020. Психология служебной деятельности: достижения и перспективы развития: материалы международной научной конференции, 8–11 декабря 2020 г.* СПб.: Скифия-принт, 2020. С. 195–196.

7. *Таран Е. А., Маланина В. А., Касати Ф.* Алгоритм использования краудсорсинговых инструментов для сбора и анализа данных научных исследований (на примере подготовки систематизированного обзора литературы) // *Экономика и управление инновациями*. 2020. № 4 (15). С. 39–46.

8. *Danilova E.* Worker’s motivation and planning strategies on crowdsourcing platforms. The case of Yandex Toloka // *Digital Transformation and Global Society. DTGS 2021. Communications in Computer and Information Science*. 2022. Vol. 1503. DOI: 10.1007/978-3-030-93715-7_38

9. *Chapkovski Ph.* Interactive experiments in Toloka // SSRN: [site]. 03.02.2022. 40 p. URL: <https://ssrn.com/abstract=3727227> (date of access: 04.07.2022).

10. *Antoun et al.* Comparisons of online recruitment strategies for convenience samples: Craigslist, Google AdWords, Facebook, and Amazon Mechanical Turk /

C. Antoun, C. Zhang, F. G. Conrad, M. F. Schober // *Field Methods*. 2016. Vol. 28 (3). P. 231–246. DOI: 10.1177/1525822X15603149

11. *Smith S.M. et al.* A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples / S. M. Smith, C. A. Roster, L. L. Golden, G. S. Albaum // *Journal of Business Research*. 2016. Vol. 69 (8). P. 3139–3148. DOI: 10.1016/j.jbusres.2015.12.002

12. *Девятко И. Ф.* Инструментарий онлайн-исследований: попытка каталогизации // *Онлайн исследования в России 3.0* / Под ред. А. В. Пашкина, И. Ф. Девятко, С. Г. Давыдова. М.: Кодекс, 2012. С. 17–30.

13. *Kees J. et al.* An analysis of data quality: Professional panels, student subject pools, and Amazon’s Mechanical Turk / J. Kees, B. Berry, S. Burton, K. Sheehan // *Journal of Advertising*. 2017. Vol. 46. P. 141–155.

14. *Ogletree A. M., Katz B.* How Do Older Adults Recruited Using MTurk Differ From Those in a National Probability Sample? // *International Journal of Aging and Human Development*. 2021. Vol. 93 (2). P. 700–721. DOI: 10.1177/0091415020940197

15. *Behrend T.S. et al.* The viability of crowdsourcing for survey research / T. S. Behrend, D. J. Sharek, A. W. Meade, E. N. Wiebe // *Behavior Research Methods*. 2011. Vol. 43. P. 800–813. DOI: 10.3758/s13428-011-0081-0

16. *Paolacci G., Chandler J., Ipeirotis, P. G.* Running experiments on Amazon Mechanical Turk // *Judgment and Decision Making*. 2010. Vol. 5 (5). P. 411–419.

17. *Zhou H., Fishbach A.* The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions // *Journal of Personality and Social Psychology*. 2016. Vol. 111 (4). P. 493–504. DOI: 10.1037/pspa0000056.

18. *Sprouse J.* A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory // *Behavior Research Methods*. 2011. Vol. 43 (1). P. 155–67. DOI: 10.3758/s13428-010-0039-7

19. *Fleischer A., Mead A. D., Huang J.* Inattentive responding in MTurk and other online samples // *Industrial and Organizational Psychology: Perspectives on Science and Practice*. 2015. Vol. 8 (2). P. 196–202. DOI: 10.1017/iop.2015.25

20. *Keith M. G., Tay L., Harms P. D.* Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations // *Frontiers in Psychology*. 2017. Vol. 8. DOI: 10.3389/fpsyg.2017.01359

21. *Agley J. et al.* Quality control questions on Amazon’s Mechanical Turk (MTurk): A randomized trial of impact on the USAUDIT, PHQ-9, and GAD-7 / J. Agley, Y. Xiao, R. Nolan, L. Golzarri-Arroyo // *Behavior Research Methods*. 2022. Vol. 54 (2). P. 885–897. DOI: 10.3758/s13428-021-01665-8

22. *Cheung J. H. et al.* Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations / J. H. Cheung, D. K. Burns, R. R. Sinclair, M. Sliter // *Journal of Business and Psychology*. 2017. Vol. 32. P. 347–361. DOI: 10.1007/s10869-016-9458-5

23. *Paolacci G., Chandler J.* Inside the Turk: Understanding Mechanical Turk as a participant pool // *Current Directions in Psychological Science*. 2014. Vol. 23. P. 184–188.

24. *Wessling K. S., Huber J., Netzer O.* MTurk Character misrepresentation: Assessment and solutions // *Journal of Consumer Research*. 2017. Vol. 44 (1). P. 211–230. DOI: 10.1093/jcr/ucx053

25. *Hamby T., Taylor W.* Survey satisficing inflates reliability and validity measures: An experimental comparison of college and Amazon Mechanical Turk samples // *Educational and Psychological Measurement*. 2016. Vol. 76 (6). P. 912–932. DOI: 10.1177/0013164415627349

26. *Lu L. et al.* Improving Data Quality Using Amazon Mechanical Turk Through Platform Setup / L. Lu, N. Neale, N. D. Line, M. Bonn // *Cornell Hospitality Quarterly*. 2022. Vol. 63 (2). P. 231–246. DOI: 10.1177/19389655211025475

27. *Hauser D. J., Schwarz N.* Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants // *Behavior Research Methods*. 2016. Vol. 48. P. 400–407. DOI: 10.3758/s13428-015-0578-z

28. *Horton J. J., Rand D. G., Zeckhauser R. J.* The online laboratory: conducting experiments in a real labor market // *Experimental Economics*. 2011. Vol. 14. P. 399–425. DOI: 10.1007/s10683-011-9273-9

29. *Casler K., Bickel L., Hackett E.* Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing // *Computers in Human Behavior*. 2013. Vol. 29 (6). P. 2156–2160. DOI: 10.1016/j.chb.2013.05.009.

30. *Gandullia L., Lezzi E., Parciasepe P.* Replication with MTurk of the experimental design by Gangadharan, Grossman, Jones & Leister (2018): Charitable giving across donor types // *Journal of Economic Psychology*. 2020. Vol. 78 (C). DOI: 10.1016/j.joep.2020.102268

31. *McCredie M.N., Morey L.C.* Who are the Turkers? A characterization of MTurk workers using the Personality Assessment Inventory // *Assessment*. 2019. Vol. 26 (5). P. 759–766. DOI: 10.1177/1073191118760709

32. *Абрамов Р.Н.* Практический опыт рекрутирования целевой выборки с помощью социальных сетей: кейс опроса о ностальгии по советскому // *Социология: методология, методы, математическое моделирование (Социология: 4М)*. 2019. № 48. С. 83–112.

33. *Федоровский А.М.* Качество онлайн-опросов. Методы проверок // *Мониторинг общественного мнения: экономические и социальные перемены*. 2015. № 3. С. 29–36.

34. *Корытнникова Н.В.* Параметры проверки и контроля качества онлайн-опроса с использованием параданных // *Мониторинг общественного мнения: экономические и социальные перемены*. 2018. № 3. С. 65–77. DOI: 10.14515/monitoring.2018.3.04

35. Zhang C., Conrad F. G. Speeding in web surveys: The tendency to answer very fast and its association with straightlining // *Survey Research Methods*. 2014. Vol. 8 (2). P. 127–135. DOI: 10.18148/srm/2014.v8i2.5453
36. Kim Y. et al. Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail–Web Mixed-Mode Surveys / Y. Kim, J. Dykema, J. Stevenson [et al.] // *Social Science Computer Review*. 2019. Vol. 37 (2). P. 214–233. DOI: 10.1177/0894439317752406
37. Reuning K., Plutzer E. Valid vs. Invalid Straightlining: The Complex Relationship Between Straightlining and Data Quality // *Survey Research Methods*. 2020. Vol. 14. P. 439–459. DOI: 10.18148/srm/2020.v14i5.7641
38. de Leeuw E. D. *Data Quality in Mail, Telephone, and Face-to-Face Surveys*. Amsterdam: TT-Publicaties, 1992. 170 p.
39. Yan T., Tourangeau R. Fast Times and Easy Questions: The Effects of Age, Experience, and Question Complexity on Web Survey Response Times // *Applied Cognitive Psychology*. 2008. Vol. 22 (1). P. 51–68. DOI: 10.1002/acp.1331
40. Galesic M. Dropouts on the web: Effects of interest and burden experienced during an online survey // *Journal of Official Statistics*. 2006. Vol. 22 (2). P. 313–328.
41. Galesic M., Bosnjak M. Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey // *Public Opinion Quarterly*. 2009. Vol. 73 (2). P. 349–360. DOI: 10.1093/poq/nfp031
42. Smyth J. D. et al. Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? / J. D. Smyth, D. A. Dillman, L. M. Christian, M. McBride // *Public Opinion Quarterly*. 2009. Vol. 73 (2). P. 325–337. DOI: 10.1093/poq/nfp029
43. Buhrmester M. D., Talafar S., Gosling S. D. An evaluation of Amazon’s Mechanical Turk, its rapid rise, and its effective use // *Perspectives on Psychological Science*. 2018. Vol. 13 (2). P. 149–154. DOI: 10.1177/1745691617706516
44. Peer E., Vosgerau J., Acquisti A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk // *Behavior Research Methods*. 2014. Vol. 46. P. 1023–1031. DOI: 10.3758/s13428-013-0434-y
45. Casey L. S. et al. Intertemporal Differences Among MTurk Workers: Time-Based Sample Variations and Implications for Online Data Collection / L. S. Casey, J. Chandler, A. S. Levine [et al.] // *SAGE Open*. 2017. Vol. 7 (2). DOI: 10.1177/2158244017712774

Gavrilov Kirill A.,

Assistant Professor, Department of Social Institutions Analysis, HSE University, Moscow, Russia,

Researcher, Institute of Sociology FCTAS RAS, Moscow, Russia,

gavrilov@socio.msk.ru

Toloka platform as a source of online survey participants: an experience of assessing data quality

The article presents the experience of using Yandex Toloka crowdsourcing platform to recruit respondents for an online survey. Analyzing methodological publications on a similar foreign platform Amazon Mechanical Turk we put forward hypotheses about the data quality obtained via Toloka in comparison with the results collected using other convenience sample types –online panels and recruitment of respondents through social networks. Additionally, only based on the Toloka data, we assessed the indicator of respondent’s attentiveness. The main conclusion is that Toloka allows to recruit respondents quickly and at low cost, and the results are comparable in terms of quality to those obtained by other methods. In particular, respondents from Toloka almost always complete the survey, fill out questionnaires faster than other types of respondents, but less often than participants of the online panel have a tendency to “straightline” (i.e., give the same answers in a tabular question), just as often as social media respondents give answers to the open-ended question (but less frequently than online panel participants), although their responses are shorter. Only 36% of the respondents passed the attention check question, attentive participants had a longer questionnaire complete time and were less likely to be straightliners. The increase of reward did not increase the proportion of attentive respondents, but decreased the questionnaire filling out speed, increased the number of answers to the open question, and reduced the proportion of straightliners.

Keywords: Toloka, crowdsourcing, online survey, data quality, Amazon Mechanical Turk

References

1. Couper M. Web Surveys: A review of issues and approaches, *Public Opinion Quarterly*, 2000, 64 (4), 464–494.

2. *Report of the AAPOR task force on nonprobability sampling* (transl., in Russian). Moscow: FOM publ., 2016. URL: https://fom.ru/uploads/files/FOM_AAPOR_book1.pdf (date of access: 01.07.2022).
3. Mavletova A.M. Sociological surveys on the Internet: the possibilities of building a typology (in Russian), *Sotsiologiya 4M (Sociology: methodology, methods, mathematical modeling)*, 2010, 31, 115–134.
4. Deviatko I. F. From “Virtual Lab” to “Social Telescope”: Metaphors of Theoretical and Methodological Innovations in Online Research, in: *Online research in Russia: trends and prospects* (in Russian). M.: MIK, 2016. P. 19–33.
5. Chmielewski M., Kucker S. An MTurk crisis? Shifts in data quality and the impact on study results, *Social Psychological and Personality Science*, 2020, 11 (4), 464–473. DOI: 10.1177/1948550619875149
6. Kulieva A.K. Ability to use Yandex.Toloka for conducting online experiments (on the example of solving a sensorimotor problem) (in Russian), *Ananiev readings - 2020. Psychology of performance: achievements and development prospects: materials of the international scientific conference*, December 8–11, 2020. St. Petersburg: Scythia-print, 2020. P. 195–196.
7. Taran E.A., Malanina V.A., Casati F. Algorithm for using crowdsourcing tools for collecting and analysis of scientific research data (by the example of preparation of systematized literature review), *Economics and Innovation Management*, 2020, 4 (15), 39–46.
8. Danilova E. Worker’s motivation and planning strategies on crowdsourcing platforms. The case of Yandex Toloka, in: *Digital Transformation and Global Society. DTGS 2021. Communications in Computer and Information Science*, 2022, Vol. 1503. DOI: 10.1007/978-3-030-93715-7_38
9. Chapkovski Ph. Interactive experiments in Toloka, SSRN: [site]. 03.02.2022. 40 p. URL: <https://ssrn.com/abstract=3727227> (date of access: 04.07.2022).
10. Antoun C., Zhang C., Conrad F.G., Schober M.F. Comparisons of online recruitment strategies for convenience samples: Craigslist, Google AdWords, Facebook, and Amazon Mechanical Turk, *Field Methods*, 2016, 28 (3), 231–246. DOI: 10.1177/1525822X15603149

11. Smith S. M., Roster C. A., Golden L. L., Albaum G. S. A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples, *Journal of Business Research*, 2016, 69(8), 3139–3148. DOI: 10.1016/j.jbusres.2015.12.002
12. Deviatko I. F. Online Research Tool Set: An Attempt of Cataloging (in Russian), in: *Online research in Russia 3.0*. M.: Codex, 2012. P. 17–30.
13. Kees, J., Berry, C., Burton, S., Sheehan, K. An analysis of data quality: Professional panels, student subject pools, and Amazon’s Mechanical Turk, *Journal of Advertising*, 2017, 46, 141–155.
14. Ogletree A.M., Katz B. How Do Older Adults Recruited Using MTurk Differ From Those in a National Probability Sample? *International Journal of Aging and Human Development*, 2021, 93 (2), 700–721. DOI: 10.1177/0091415020940197
15. Behrend T. S., Sharek D. J., Meade A. W., Wiebe E. N. The viability of crowdsourcing for survey research, *Behavior Research Methods*, 2011, 43, 800–813. DOI: 10.3758/s13428-011-0081-0
16. Paolacci G., Chandler J., Ipeirotis, P.G. Running experiments on Amazon Mechanical Turk, *Judgment and Decision Making*, 2010, 5 (5), 411–419.
17. Zhou H., Fishbach A. The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions, *Journal of Personality and Social Psychology*, 2016, 111 (4), 493–504. DOI: 10.1037/pspa0000056.
18. Sprouse J. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory, *Behavior Research Methods*, 2011, 43 (1), 155–67. DOI: 10.3758/s13428-010-0039-7
19. Fleischer A., Mead A.D., Huang J. Inattentive responding in MTurk and other online samples, *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2015, 8 (2), 196–202. DOI: 10.1017/iop.2015.25
20. Keith M.G., Tay L., Harms P.D. Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations, *Frontiers in Psychology*, 2017, 8. DOI: 10.3389/fpsyg.2017.01359
21. Agle J., Xiao Y., Nolan R., Golzarri-Arroyo L. Quality control questions on Amazon’s Mechanical Turk (MTurk): A randomized trial of impact

- on the USAUDIT, PHQ-9, and GAD-7, *Behavior Research Methods*, 2022, 54 (2), 885–897. DOI: 10.3758/s13428-021-01665-8
22. Cheung J. H., Burns D. K., Sinclair R. R., Sliter M. Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations, *Journal of Business and Psychology*, 2017, 32, 347–361. DOI: 10.1007/s10869-016-9458-5
 23. Paolacci G., Chandler J. Inside the Turk: Understanding Mechanical Turk as a participant pool, *Current Directions in Psychological Science*, 2014, 23, 184–188.
 24. Wessling K.S., Huber J., Netzer O. MTurk Character misrepresentation: Assessment and solutions, *Journal of Consumer Research*, 2017, 44 (1), 211–230. DOI: 10.1093/jcr/ucx053
 25. Hamby T., Taylor W. Survey satisficing inflates reliability and validity measures: An experimental comparison of college and Amazon Mechanical Turk samples, *Educational and Psychological Measurement*, 2016, 76 (6), 912–932. DOI: 10.1177/0013164415627349
 26. Lu L., Neale N., Line N. D., Bonn M. Improving Data Quality Using Amazon Mechanical Turk Through Platform Setup, *Cornell Hospitality Quarterly*, 2022, 63 (2), 231–246. DOI: 10.1177/19389655211025475
 27. Hauser D. J., Schwarz N. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants, *Behavior Research Methods*, 2016, 48, 400–407. DOI: 10.3758/s13428-015-0578-z
 28. Horton J.J., Rand D.G., Zeckhauser R.J. The online laboratory: conducting experiments in a real labor market, *Experimental Economics*, 2011, 14, 399–425. DOI: 10.1007/s10683-011-9273-9
 29. Casler K., Bickel L., Hackett E. Separate but equal? A comparison of participants and data gathered via Amazon’s MTurk, social media, and face-to-face behavioral testing, *Computers in Human Behavior*, 2013, 29 (6), 2156–2160. DOI: 10.1016/j.chb.2013.05.009.
 30. Gandullia L., Lezzi E., Parciasepe P. Replication with MTurk of the experimental design by Gangadharan, Grossman, Jones & Leister (2018): Charitable giving across donor types, *Journal of Economic Psychology*, 2020, 78 (C). DOI: 10.1016/j.joep.2020.102268
 31. McCredie M.N., Morey L.C. Who are the Turkers? A characterization of MTurk workers using the Personality Assessment Inventory, *Assessment*, 2019, 26 (5), 759–766. DOI: 10.1177/1073191118760709

32. Abramov R.N. The practical experience of recruiting the target sample using social networks: the case of the survey on the nostalgia for the Soviet past (in Russian), *Sotsiologiya 4M (Sociology: methodology, methods, mathematical modeling)*, 2019, 48, 83–112.
33. Fedorovskii A. M. Quality of Online Surveys. Methods of Control (in Russian), *Monitoring of Public Opinion: Economic and Social Changes*, 2015, 3, 28-35.
34. Korytnikova N. V. Use of paradata in quality control of online survey (in Russian), *Monitoring of Public Opinion: Economic and Social Changes*, 3, 65—77. DOI: 10.14515/monitoring.2018.3.04
35. Zhang C., Conrad F.G. Speeding in web surveys: The tendency to answer very fast and its association with straightlining, *Survey Research Methods*, 2014, 8 (2), 127–135. DOI: 10.18148/srm/2014.v8i2.5453
36. Kim Y., Dykema J., Stevenson J., Black P., Moberg D. P. Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail–Web Mixed-Mode Surveys, *Social Science Computer Review*, 2019, 37(2), 214–233. DOI: 10.1177/0894439317752406
37. Reuning K., Plutzer E. Valid vs. Invalid Straightlining: The Complex Relationship Between Straightlining and Data Quality, *Survey Research Methods*, 2020, 14, 439–459. DOI: 10.18148/srm/2020.v14i5.7641
38. de Leeuw E.D. *Data Quality in Mail, Telephone, and Face-to-Face Surveys*. Amsterdam: TT-Publicaties, 1992. 170 p.
39. Yan T., Tourangeau R. Fast Times and Easy Questions: The Effects of Age, Experience, and Question Complexity on Web Survey Response Times, *Applied Cognitive Psychology*, 2008, 22 (1), 51–68. DOI: 10.1002/acp.1331
40. Galesic M. Dropouts on the web: Effects of interest and burden experienced during an online survey, *Journal of Official Statistics*, 2006, 22 (2), 313–328.
41. Galesic M., Bosnjak M. Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey, *Public Opinion Quarterly*, 2009, 73 (2), 349–360. DOI: 10.1093/poq/nfp031
42. Smyth J. D., Dillman D. A., Christian L. M., McBride M. Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly*, 2009, 73 (2), 325–337. DOI: 10.1093/poq/nfp029

43. Buhrmester M.D., Talairar S., Gosling S.D. An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use, *Perspectives on Psychological Science*, 2018, 13 (2), 149–154. DOI: 10.1177/1745691617706516
44. Peer E., Vosgerau J., Acquisti A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk, *Behavior Research Methods*, 2014, 46, 1023–1031. DOI: 10.3758/s13428-013-0434-y
45. Casey L. S, Chandler J., Levine A. S, Proctor A., Strolovitch D. Z. Intertemporal Differences Among MTurk Workers: Time-Based Sample Variations and Implications for Online Data Collection, *SAGE Open*, 2017, 7 (2). DOI: 10.1177/2158244017712774