

**СОСТАВ ВЫБОРКИ. СРАВНЕНИЕ ВЕЛИЧИН
ВОСПРОИЗВОДИМОСТИ ЭФФЕКТА И КАЧЕСТВА ОТВЕТА СРЕДИ
ВЫБОРОК СТУДЕНТОВ И ПОЛЬЗОВАТЕЛЕЙ ПЛАТФОРМ
MTurk и QUALTRICS¹**

Курт Крейгер,^{a} Алисса К. МакГонагл^b и Дайана Р. Санчез^c*

^a Факультет менеджмента, Университет Мемфиса, Мемфис, Соединённые Штаты Америки;

^b Факультет психологии, Университет Северной Каролины в Шарлотт, Шарлотт, Соединённые Штаты Америки;

^c Факультет психологии, Университет штата Калифорния в Сан-Франциско, Соединённые Штаты Америки

** Факультет менеджмента, Колледж бизнеса и экономики им. Фогельмана, 3675 Централ авеню, Университет Мемфиса, Мемфис, TN 38152 USA. OF: 901.678.3159; email: Kurt.Kraiger@memphis.edu*

Аннотация. Исследователи всё чаще используют краудсорсинг, т. е. оплачиваемое привлечение участников выборок в сети Интернет в научных целях. Ключевым аспектом рассмотрения выборок является понимание того, в какой степени результаты будут обобщены для более широких генеральных совокупностей. В наше исследование включены три выборки: (1) студенты университета, (2) пользователи MTurk и (3) онлайн-панель Qualtrics. Используя данные источники, наше исследование вносит вклад в понимание того, как генерация выборки влияет на обобщаемость результатов исследования. Выявляются различия между выборками в степени их репрезентации рабочей силы в США. Исследуется, насколько хорошо результаты из каждой выборки повторяют известный мета-аналитически определенный эффект генеральной совокупности. Выборки сравниваются по четырем показателям качества ответа (*расстояние Махаланобиса, корреляции человек-целое, вариативность индивидуального ответа и психометрические синонимы*) и демонстрируется то, как исключение недобросовестных респондентов влияет на то, насколько результаты выборки повторяют

¹ Статья представлена на 11-ой Конференции Организационной психологии: Люди и риски, Саратовский Национальный Исследовательский Государственный Университет им. Н.Г. Чернышевского, 24 апреля 2020

результаты генеральной совокупности. В статье приводится обсуждение результатов генерации выборки, а также применение и анализ критериев качества ответа.

Ключевые слова: краудсорсинговые выборки, исследование-опрос, недостаточное усилие при ответе, воспроизводимость, организационная лояльность, намерение покинуть компанию.

Исследователи всё чаще используют для сбора данных краудсорсинговые выборки, полученные от Mechanical Turk (MTurk) или Qualtrics (Cheung, Burns, Sinclair, & Sliter, 2017). Участники краудсорсинга представляют собой демографически разнообразную, глобальную совокупность (Cheung et al., 2017). Однако достоверность данных, полученных от краудсорсинговых выборок, пока не является общепризнанной (например Paolacci & Chandler 2014). Центральное значение для внешней достоверности результатов, полученных с помощью краудсорсинговых выборок, заключается в том, являются ли они репрезентативными для интересующей совокупности (Cheung et al., 2017; Landers & Behrend, 2015). Исследователи задаются вопросом: «Являются ли участники, взятые из этой выборки, представителями генеральной совокупности, которую мне необходимо обобщить?» и «Могу ли я получить реальную оценку эффекта генеральной совокупности?». В последнее время возрос интерес к данным из краудсорсинговых выборок, о чем свидетельствуют многочисленные публикации о недостаточном усилии при ответе (IER - *Insufficient effort responding*) или недобросовестном ответе (Huang, Curran, Keeney, Poposki, & DeShon, 2012; Huang, Liu, & Bowling, 2015; McGonagle, Huang, & Walsh, 2016). Мы исследуем эти проблемы, изучая, повторяют ли результаты, полученные из краудсорсинговых выборок, мета-аналитические значения генеральной совокупности, и публикуя информацию о том, как IER влияет на выводы, сделанные на основе этих результатов. В частности, мы исследуем, воспроизводят ли результаты, полученные из трех разных источников (студенты колледжа, MTurk и Qualtrics), известные значения генеральной совокупности, установленные мета-анализом. Кроме того, в статье мы показываем, что включение или исключение участников, демонстрирующих IER, влияет на полученные оценки.

Научный вклад исследования состоит из двух пунктов. Во-первых, мы используем новый показатель качества данных для изучения внешней достоверности полученных из краудсорсинга результатов - сравнения

наблюдаемых конструктивных корреляций с данными, полученными в результате мета-анализа. Мета-анализы являются отличным образцом (референтом) из-за их больших размеров выборки, стабильных результатов и разнообразного состава выборки. В исследовании сравнивается мета-аналитическая оценка с корреляциями между двумя устоявшимися конструктами (организационная лояльность и намерение покинуть компанию) в трёх выборках.

Во-вторых, мы представляем и сравниваем результаты при исключении участников, демонстрирующих IER, и без исключения таких участников. При этом показывается, как сравнение мета-аналитических результатов с выборочными корреляциями может быть противоречивым в зависимости от отношения к IER.

Источники выборки: Воспроизводимость

MTurk – это набирающий в последнее время популярность краудсорсинговый источник выборок (Buhrmeister, Kwang, & Gosling, 2011; Paolacci & Chandler, 2014). Платформы-опросники, такие как Qualtrics и SurveyMonkey, также предоставляют исследователям доступ к изучению онлайн-панелей, но являются менее изученными инструментами.

Изучение достоверности выводов, сделанных на основе краудсорсинговых выборок, берет своё начало в выдающейся практической статье Бурмайстера и др. (2011) (Buhrmeister et al., 2011), вводящей MTurk в общественные науки. Исследователями были использованы личностные опросники разного объёма как для выборок из MTurk, так и для студентов, сравнена тестовая надёжность между выборками и сделан вывод, что «полученные данные, по крайней мере, так же надежны, как и данные, полученные традиционными методами» (р. 3).

Результаты могут быть надёжными, но недостоверными. Первые подходы к установлению достоверности заключались в демонстрации данных, полученных от выборок из MTurk, воспроизводящих классические эффекты поведенческой экономики и принятия решений (например Paolacci, Chandler, & Ipeirotis, 2010). Например Паолаччи и др. проверили несколько эмпирически установленных ошибок при принятии решений в выборках студентов и пользователей MTurk и обнаружили, что эти эффекты воспроизводятся в выборках.

Учёными предпринимались попытки воспроизведения известных эффектов с помощью краудсорсинговых выборок в экспериментальных исследованиях. Крамп, МакДоннел и Гурекис (2013) (Crump, McDonnell, and Gureckis, 2013) попытались воспроизвести «разнообразный набор задач экспериментальной психологии», используя выборки из MTurk, и в своей

работе опубликовали следующее: «...в основном данные кажутся соответствующими лабораторным результатам, при условии, что методы эксперимента были надёжными» (р. e57410), предполагая, что экспериментальные эффекты, доказанные традиционными способами, могут быть воспроизведены, используя краудсорсинговые выборки. Маллиникс, Липер, Дракман и Фриз (2015) (Mullinix, Leeper, Druckman, and Freese, 2015) не обнаружили существенных различий в эффектах формирования мнений участников по четырем выборкам, включая выборки MTurk и студентов.

В исследованиях, сравнивающих достоверность выводов из традиционных выборок и краудсорсинговых выборок, почти все выводы основаны на обобщаемости по статистической значимости. Данный критерий для определения эквивалентности оценки выборки является спорным по двум причинам. Во-первых, из-за удобства и более низких материальных затрат, выборки из MTurk зачастую значительно больше традиционных выборок. Любые экспериментальные манипуляции могут быть значительными в краудсорсинговой выборке просто из-за большей статистической мощности исследования. Во-вторых, среди ученых, изучающих поведение, наблюдается тенденция к отказу от тестирования значимости величин эффекта (например Cumming, 2014). Камминг и другие утверждают, что нашей основной целью для отдельных исследований должно быть определение величины эффекта для интересующих явлений.

Касательно исследований-опросов и корреляционных исследований, Уолтер, Зайберт, Гёринг и О'Бойл (2018) (Walter, Seibert, Goering, and O'Boyle, 2018) недавно был проведен мета-анализ, в котором были обнаружены 43 корреляции между двумя переменными, относящимися к организационным наукам (например, невротизм и удовлетворенность работой) в краудсорсинговой выборке, для которой была произведена мета-аналитическая оценка. 86% корреляций попадали в восьмидесятипроцентный доверительный интервал мета-анализа. Несмотря на то, что эти данные повышают уверенность в том, что оценки корреляции, полученные от краудсорсинговых выборок, повторяют известные величины эффекта, Уолтером и др. не были представлены сравнения между типами онлайн-выборок (например, MTurk против Qualtrics) или сравнения со студенческими выборками.

На сегодняшний день не существует исследований, сравнивающих множественные краудсорсинговые выборки по величинам их эффектов или сравнивающих их величины с другими, полученными из других выборок. В настоящей статье мы изучаем, насколько факторные корреляции между организационной лояльностью (OC — organizational commitment) и намерениями покинуть компанию (WI — withdrawal intention) в трёх

выборках (MTurk, Qualtrics и студенты университета) воспроизводят мета-аналитическую оценку величины эффекта генеральной совокупности. Таким образом, перед нами встаёт вопрос:

RQ1: Насколько данные из выборок студентов, MTurk и Qualtrics воспроизводят установленную посредством мета-анализа корреляцию между ОС и WI?

IER (Недостаточное усилие при ответе) и источники выборок

В настоящее время увеличилось количество исследований, посвящённых IER, в которых представляются новые критерии выявления «проблематичных» респондентов (например Curran, 2016; Meade & Craig, 2012). Исследователями были найдены доказательства, подтверждающие наличие IER в краудсорсинговых выборках (например Huang and colleagues, 2012; 2015). Кроме того, недавно были озвучены многочисленные призывы к выявлению IER (например Huang et al., 2012; Meade & Craig, 2012) и открытой публикации методов такого выявления (например McGonagle et al., 2016).

Признание существования IER имеет значение для интерпретации мета-аналитических оценок. Нам представляется вероятным, что большинство выборок в существующих мета-анализах включает в себя *нераспознанных* недобросовестных респондентов, так как в них были объединены первичные исследования, в которых данные из выборки были собраны до того, как выявление/исключение случаев IER стало привычной практикой. Следовательно, необходимо осторожно интерпретировать результаты мета-анализа как «коэффициенты генеральной совокупности». Для примера мы представляем выборочные корреляции с недобросовестными респондентами и без них и сравниваем оба набора выборочных корреляций с мета-аналитическим референтом. Так как IER завышает найденные корреляции в определённых ситуациях (например когда переменные средние значения в двумерном соотношении отклоняются от средних точек шкалы; Huang et al., 2015) и занижает их в других (из-за увеличения погрешности измерения; McGrath, Mitchell, Kim, & Hough, 2010), направленность в данном случае нами не рассматривается. Вместо этого мы предлагаем, чтобы такие изменения могли повлиять на выводы, сделанные в отношении обобщения при использовании мета-аналитических корреляций в качестве референтов:

RQ2: Повлияет ли исключение недобросовестных респондентов на выводы, сделанные о качестве данных при использовании мета-аналитических корреляций в качестве референтов?

Обнаружение IER. Попытки обнаружения и контроля за IER могут быть предприняты до или после сбора данных (см. Curran, 2016; Meade & Craig, 2012). Априорные подходы к определению IER подробно обсуждались в работах Каррана (Curran, 2016) и Мида и Крейга (Meade and Craig, 2012). Апостериорные подходы проверяют ответы после сбора данных. Мид и Крейг выделяют несколько классов апостериорных подходов, включая анализ выбросов и согласованность ответов. Нами используются четыре апостериорных подхода: расстояние Махаланобиса (*MD - Mahalanobis Distance*) и Корреляция Человек-Целое (*Ptr - Person-Total Correlations*) для анализа выбросов, а также вариативность индивидуального ответа (*IRV - Individual Response Variability*) и психометрические синонимы (*PS - Psychometric Synonyms*) для согласованности ответов. В совокупности эти четыре подхода позволяют нам сравнивать обнаружение IER по критериям и вне критериев.

Метод

Участники и методика. Все использованные данные были архивными. Выборка студентов состояла из 382 студентов-психологов из государственного университета Соединённых Штатов Америки, выполнивших требования курса, являющихся частью более обширного исследования отношений между пригодностью к работе и отношениями сотрудников.

Выборка MTurk состояла из 581 респондентов, проживающих на территории США, привлечённых посредством опубликования задания для исследования, изучающего отношения между пригодностью к работе и отношением сотрудников. Участникам MTurk заплатили по 0,75 долларов США за участие в исследовании (без подтверждения качества данных).

Выборка Qualtrics состояла из 321 респондентов. Qualtrics разослала ссылку на опрос и обратно выслала заполненные данные. Qualtrics было перечислено по 25 долларов США за каждого участника, из которых около 5 долларов США каждый участник получил лично.

Показатели

Организационная лояльность (ОС). ОС была измерена Опросником, состоящим из 15 пунктов, для измерения лояльности (ОСQ; Mowday, Steers, & Porter, 1979), состоящим из девяти утвердительных пунктов и шести отрицательных пунктов с форматом ответа, состоящим из семи пунктов. Уровни значимости (альфы): 0,87 (Студенты), 0,93 (MTurk) и 0,88 (Qualtrics).

Намерение покинуть компанию (WI). WI было измерено шкалой, состоящей из 3 пунктов, с форматом ответа, состоящим из пяти пунктов (Mobley, Horner, and Hollingsworth, 1978). Уровни значимости (альфы): 0,88 (студенты), 0,94 (MTurk) и 0,95 (Qualtrics).

Анализ

Качество ответов. Мы оценили качество ответов, основываясь на анализе выбросов (*MD*, *PTr*) и согласованности ответов (*IRV*, *PS*). Нами были подсчитаны показатели расстояния Махаланобиса - *MD* для каждого респондента (для показателей ОС и WI) и была проверена статистическая значимость для выявления потенциальных респондентов с IER (Meade & Craig, 2012). Показатель расстояния представляет собой многовариантное расстояние между векторами ответов участников и вектором среднего значения выборки. Данный показатель может быть использован для обнаружения отдельных многовариантных выбросов (Tabachnick & Fidell, 2007). Для определения IER как выбросов отсутствуют подтвержденные предельные значения. Для установления критерия для IER, мы установили $p < 0,05$ для статистической значимости для значения хи-квадрат (Tabachnick & Fidell, 2007), но использовали поправку Холма-Бонферрони (Holm, 1979) для контроля групповой вероятности ошибки. Участники с баллами статистически значимого расстояния получили “1”, что является признаком IER.

Корреляция Человек-Общее (PTr) измеряет связь между баллами респондентов по пятнадцати ОС и трём WI пунктам с итоговым или средним баллами для выборок по тем же пунктам (Curran, 2016). Карран рекомендует отрицательные корреляции в качестве «осторожного» признака IER. Респонденты с отрицательным *PTr* отмечались “1”, означающим IER; остальные были отмечены “0.”

Вариативность индивидуального ответа (IRV) является показателем согласованности ответов и находит IER путем определения низкой изменчивости в наборе ответов (Dunn, Heggestad, Shannock, & Theilgard, 2018), рассчитываемой как стандартное отклонение ответов по всем последовательным ответам для участников. Мы вычислили *IRV* по 9 положительным и 6 отрицательным (до перестановки) пунктам ОС. Низкая изменчивость по пунктам до обратной оценки отрицательных пунктов указывает на IER, а более высокая изменчивость свидетельствует о меньшей IER. Следуя рекомендациям Данна и др., мы создали предельное значение для обнаружения респондентов с IER, вычислив десятый процентиль баллов *IRV* (по выборкам). Значение 10% равно 0,798; участники с $IRV \leq 0,80$ были отмечены как IER.

Психометрические синонимы (PS) является вторым показателем согласованности ответов и относится к внутриреспондентным корреляциям векторов, состоящих из пунктов, которые сильно коррелируют друг с другом (Curran, 2016; Meade & Craig, 2012). Векторы сформированы из пар пунктов с высокой корреляцией для всей выборки ($>0,60$ по Curran, 2016). Например, значение корреляции второго пункта «Я рассказываю моим друзьям о том, как хорошо работать в этой организации» и шестого пункта «Я горжусь тем, что могу сказать, что я – часть этой организации» равняется $r = 0,80$, и эти два пункта расположены в отдельных векторах. Чем ниже корреляция между векторами, тем менее согласованно респонденты отвечали на пункты с высокой степенью взаимосвязи. Карран рекомендует отрицательные корреляции в качестве «осторожного» признака IER. Респонденты с отрицательной корреляцией отмечались “1”, означаящим IER; остальные были отмечены “0.”

Сравнение величины эффекта Были рассчитаны скрытые корреляции переменных для респондентов с IER и без IER для каждой выборки и эти оценки были сравнены с мета-аналитическим референтом. Была использована откорректированная средняя корреляция для ОС и WI со значением $-0,47$ ($K = 351$, $N = 136\ 270$) из мета-анализа Купера-Хакима и Висвесварана (Cooper-Hakim and Viswesvaran, 2005).

Результаты

Демография участников

Участники из MTurk ($M_{Age}=34,1$, $SD_{Age}=10,6$) и Qualtrics ($M_{Age}=43,0$, $SD_{Age}=22,0$) были в среднем старше, чем участники-студенты ($M_{Age}=19,1$, $SD_{Age}=1,7$). Процент женщин варьировался от 51,3% (Qualtrics) до 54,9% (MTurk). Информация о расовой и этнической принадлежности студентов была недоступна. Выборки из MTurk и Qualtrics имели демографические показатели, достаточно представительные для рабочей силы США (примерно 75% европеоидов, 8% афроамериканцев и 5-10% латиноамериканцев по выборкам). Количество рабочих часов в неделю варьировалось между студентами и двумя другими выборками. 68% студентов работали 20 часов в неделю или меньше, тогда как только 8,4% пользователей MTurk и 4,7% пользователей Qualtrics отметили такое же количество рабочих часов.²

Воспроизводимые эффекты в генеральной совокупности

В соответствии с RQ1, скрытые корреляции переменных между ОС и WI представлены в Таблице 1. Корреляции выборки студентов и пользователей MTurk сильно превысили мета-аналитическую корреляцию в -

0,47, даже превышая порог в 0,80 для дискриминантной достоверности (Brown, 2006). Полная выборка из Qualtrics ($\rho = -0,53$) наиболее близко приблизилась к мета-аналитическому референту.

Качество ответов

Таблица 2 демонстрирует IER по критериям для каждой выборки, а также изменчивость частоты IER по показателям и выборкам. Процент респондентов с IER был низким и достаточно однородным для *MD* и *PS*, а для *PTr* процент таких респондентов был выше и более изменчивым [варьируется от 10,9% (Qualtrics) до 16,8% (MTurk)]. Большие неоднородности наблюдались при значениях IRV, варьирующихся от 5,7% (MTurk) до 24,7% (Qualtrics).

В таблице 2 также показано общее количество респондентов с нулём, одним, двумя, тремя и четырьмя индикаторами IER. Как видно в таблице 2, в выборке из Qualtrics намного меньше респондентов без индикаторов IER (59,4%) чем в студенческой (71,2%) или MTurk (72,9%) выборках. Респонденты, демонстрирующие IER, обычно были отмечены только одним из четырёх критериев.

Нижняя половина Таблицы 2 отражает корреляции между исходными значениями для критериев IER по выборке. Используемые нами два индекса выброса (*PTr*, *MD*) умеренно и положительно коррелируют по выборкам. Два использованных нами индекса однородности ответов (*PS*, *IRV*) положительно коррелируют по выборками и умеренно коррелируют в двух из трёх выборок.

Влияние IER на факторную корреляцию

В соответствии с RQ2, мы изучили факторную корреляцию для респондентов с IER и без IER и сравнили их с мета-аналитической корреляцией. Была использована скрытая переменная модель и проведён анализ нескольких групп по всем параметрами, оценённым без ограничений.

Корреляции для полных выборок и для подмножеств выборки с респондентами с IER, обнаруженными каждым исключённым критерием, представлены в Таблице 2. Как и было ожидаемо, все корреляции были отрицательными, но по большей части сильнее чем использованный нами мета-аналитический референт. Корреляции были в основном схожи как при использовании *MD*, так и без него, что неудивительно, учитывая небольшой процент в каждой выборке, определённый как IER по этим критериям. Противоположно корреляции увеличились для каждой выборки, когда определённые с помощью IER респонденты с IRV были исключены. Наибольшее увеличение было отмечено в выборке из Qualtrics (-0,53 до -0,78), возможно потому, что в этой выборке была наибольшая доля исключённых с IER ($n = 79$, 25%).

Корреляции между полными выборками и выборками с IER с использованием подхода *PS* были схожими, но корреляции снизились (и были более схожими с использованным нами мета-аналитическим референтом), когда были исключены респонденты с IER при использовании подхода *PTr*. Для лучшего понимания эффекта исключения респондентов с IER при использовании подхода *PTr*, мы сравнили среднее значение при OC и WI между участниками, отмеченными “1” и “0” в каждой выборке. Средние значения для каждой переменной были значительно ниже для OC и выше для WI в каждой выборке для участников, демонстрирующий IER. Визуально, влияние исключения участников с высоким IER, используя *PTr*, заключалось в том, чтобы исключить участников, плотно сгруппированных в верхнем левом углу графика рассеяния (чрезвычайно низкое для OC, чрезвычайно высокое для WI).

Обсуждение

Наше исследование расширяет обсуждение репрезентативности и внешней достоверности краудсорсинговых выборок несколькими способами. Ниже приведено обобщение обсуждения.

Воспроизводимость величины эффекта генеральной выборки

Нами была исследована степень воспроизведения выборочными корреляциями мета-аналитически полученных корреляций между OC и WI. Ни в одной из выборок не было выявлено факторной корреляции в пределах 95% доверительных интервалов, описанных Купером-Хакимом и Висвесвараном (-.49, -.45; 2005). Стоит отметить, что скрытые корреляции переменных не совпадают с мета-аналитически полученными оценками генеральной совокупности; тогда как первое относится к ошибке измерения, второе также может учитывать ошибку выборки и ограничение диапазона.

Качество ответа

Мы сравнили выборки по четырём показателям IER. В соответствии с предыдущими исследованиями (DeSimone & Harms, 2017; Meade & Craig, 2012), процент респондентов с IER зависит от определённых критериев. Однако 28%-40% респондентов по выборкам продемонстрировало IER по крайней мере по одному показателю, что значительно выше “модального уровня” в 8-12% определённого Карраном (Curran, 2016), и может быть связано либо с использованием нами нескольких индикаторов, либо с конкретными выбранными индикаторами. Значения *IRV* были наиболее противоречивыми по типу выборки. В выборках студентов и Qualtrics наблюдалось наибольшее количество респондентов с IER выявленных с помощью *IRV*, а в MTurk - значения *PTr*.

IER и оценки факторных корреляций

В мета-анализе Купер-Хакима и Висвесварана (Cooper-Nakim and Viswesvaran, 2005) не было исследований, в которых были бы исключены участники с IER. Таким образом, их мета-анализ может содержать неверно вычисленные оценки величины эффекта в выборках без IER. Тем не менее нам известно, что исключение IER оказывает противоречивое влияние на выборочные оценки в разных исследованиях (сравните с Huang et al., 2015; McGrath et al. 2010). Пока что преждевременно прогнозировать, будет ли исключение респондентов с IER завышать или занижать корреляции, но полученные нами результаты указывают на важность открытости при нахождении и обработке IER. Мы предлагаем учёным регулярно публиковать результаты исследований с IER и без него - независимо от выборки.

Ограничения

Существует несколько ограничений для нашего исследования с точки зрения того, как качество ответа было использовано на практике. Во-первых, мы не использовали априорные или прямые показатели IER, которые являются общепринятыми методами выявления недостаточных усилий (например Huang and colleagues, 2012, 2015). Использование дополнительных методов может помочь лучше понять, на что участники реагировали недобросовестно.

Во-вторых, так как появляются новые методы выявления IER, пока что не существует чётких критериев для обозначения ответов как содержащих IER. В других исследованиях могут быть использованы разные предельные значения и быть получены разные выводы об эффектах IER; различные предельные значения могут привести к разным значениям распространённости IER. При определении приемлемых предельных значений мы руководствовались здравым смыслом, изучая предложенные рекомендации и следуя им, однако очевидно, что требуются дальнейшие психометрические исследования для сравнения и проверки предельных значений для обнаружения IER.

Теоретические выводы и дальнейшее исследование

Результаты нашего исследования вносят вклад в проводимые научные исследования и разработку теории IER. Хуанг и другие (Huang et al., 2012) определяет IER как набор ответов, в котором на элементы опроса отвечают с нежеланием соблюдать инструкции опроса, правильно интерпретировать содержание пункта опроса и/или отвечать точно. Николс, Грин и Шмлок (Nichols, Greene, and Schmolck, 1989) в своей работе предложили нужное различие между шаблонами ответов: искажение, реагирующее на содержание, и неточность, не реагирующая на содержание. Первое предполагает, что респонденты понимают, о чем их спрашивают, но намеренно дают ответы, которые искажают их «истинные оценки» по

измеряемых конструктам, что является формой саморепрезентации, включающей в себя хорошо известные поведенческие реакции, такие как фальсификация и социальная желательность.

Напротив, неточность, не реагирующая на содержание, возникает, когда участники не читают или не интерпретируют пункты опроса внимательно (Nichols et al., 1989). Главное, что поведение участника не зависит от содержимого пункта опроса. Неточность, не реагирующая на содержание, может возникать из-за класса поведений, включая IER. Несмотря на то, что понятие IER и недобросовестных ответов используется взаимозаменяемо, мы считаем, что IER более точно описывает поведение участников. Зачастую респонденты ответственно относятся в опросу, но не настолько, чтобы проявить достаточные усилия, отвечая на содержание пунктов опроса.

Поскольку учёными предлагается, тестируется и собирается все больше критериев качества ответов на регулярной основе, возможно, что в будущем будет возможность создать номологическую сеть, связывающую критерии с характеристиками и мотивацией респондентов, а также с их влиянием на результаты исследования. Как было нами показано, критерии IER частично доказывают конвергентную достоверность, тем самым предполагая, что согласованность ответов и выбросы ответов могут быть разными конструктами. Выполненный нами анализ факторных корреляций с IER и без IER показывает, что в то время как в целом исключение респондентов с IER не оказало большого влияния на выборочные оценки, исключение *IRV*, как правило, увеличивало факторные корреляции оценок, а исключение *PTr*, как правило, уменьшало их. Мы предлагаем, чтобы *IRV* оценивал форму систематической ошибки, которая имеет специфический эффект снижения расхождения наблюдаемых переменных; исключение респондентов, демонстрирующих IER на основе *IRV*, должно, как правило, повысить оценки корреляции. Напротив, находя отдельные выбросы, *PTr* может зафиксировать форму систематической ошибки, и ее устранение должно уменьшить наблюдаемые корреляции. Учёным, в дальнейшем исследующим данную проблему, рекомендуется продолжить оценивать IER по нескольким критериям и изучать их влияние на оценки параметров генеральной совокупности.

Практические выводы

Наши результаты выделяют факторы, которые необходимо учитывать при генерации выборки для исследования. Выборки MTurk и Qualtrics достаточно репрезентативны с точки зрения рабочей силы, следовательно традиционные выборки, так называемый «золотой стандарт», по своей

природе не являются наиболее представительными (Landers & Behrend, 2015). Репрезентативность и качество выборки также зависит от внимательности участников при ответе. Исследователям рекомендуется продолжить использовать несколько выборок для изучения социальных явлений, но при этом следует внимательно влиять на участников с IER, а также оценивать и контролировать их.

Исследование критериев для определения IER в опросах является довольно новым явлением, и всё ещё появляются новые типы показателей и предельных значений (например Curran, 2016; Dunn et al., 2018; Huang & colleagues 2012, 2015; Meade & Craig, 2012). Вероятно, что разные индексы IER касаются разных базовых конструкторов (сравните DeSimone & Harms, 2017). Исследователям рекомендуется использовать несколько индикаторов IER. Критерии можно выбрать в соответствии с типом показателей, например показатели согласованности ответов могут быть более диагностическими, когда на шкалах находятся элементы с обратной оценкой.

Нами были выделены несколько рекомендаций для проведения опроса-исследования. Во-первых, мы рекомендуем исследователям регулярно и открыто публиковать всю информацию о том, как они поступают с респондентами, демонстрирующими IER, то есть какие меры обнаружения IER были использованы, какие предельные значения были использованы для определения таких респондентов как недобросовестных, и какая доля выборки была таким образом классифицирована. Результаты исследования с участниками с IER и без IER должны быть опубликованы или доступны по запросу. Во-вторых, по мере того как изучение респондентов с IER становится все более распространенным, мы рекомендуем исследователям в будущих мета-анализах использовать обработку респондентов с IER в качестве методологического посредника. Следует фиксировать как тип обработки, так и процент исключенных участников, чтобы исследователи могли лучше понять, как обработка IER влияет на статистику выборки.

Выводы

Необходимо пересмотреть представления о качестве источника данных. В соответствии с результатами нашего исследования было выявлено, что наша выборка студентов нерепрезентативно представляла рабочую силу США, в отличие от выборок MTurk и Qualtrics. Участники, определённые как недобросовестные респонденты, в значительной степени зависят от критерия IER, и от 30 до 40% участников в каждой выборке демонстрировали IER хотя бы по одному критерию. При сравнении мета-аналитической оценки связи между ОС и WI, исключение недобросовестных респондентов увеличило или уменьшило корреляцию в зависимости от критерия.

Список использованной литературы

- Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- [Браун Т.А. (2006). *Подтверждающий факторный анализ для прикладных исследований*. Нью-Йорк: Гилфорд]
- Buhrmester, M., Kwang, T., & Gosling, S.D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high quality, data? *Perspectives on Psychological Science*, 6, 3–5.
- [Бурместер М., Кванг Т., и Гослинг С.Д. (2011) Mechanical Turk от Amazon: Новый источник недорогой, но высококачественной информации? *Перспективы психологической науки*, 6, 3-5]
- Cheung, J.H., Burns, D.K., Sinclair, R.R., & Sliter, M. (2017). Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology*, 32, 347-361.
- [Чонг Дж. Эйч., Бёрнс Д.К., Синклер Р.Р. и Слитер М. (2017). Mechanical Turk от Amazon в организационной психологии: Оценка и практические рекомендации. *Журнал бизнеса и психологии*, 32, 347-361]
- Cooper-Hakim, A., & Viswesvaran, C. (2005). The construct of work commitment: Testing an integrative framework. *Psychological Bulletin*, 131, 241–259.
- [Купер-Хаким, А. И Висвесваран, С. (2005). Конструкт лояльности работе: тестирование интегративной структуры. *Психологический вестник*, 131, 241-259.]
- Crump, M.J., McDonnell, J.V., & Gureckis, T.M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, 8(3), e57410.
- [Крамп, М.Дж., МакДоннелл, Дж.В. и Гурекис, Т.М. (2013). Оценка Mechanical Turk от Amazon как инструмента экспериментального поведенческого исследования. *PLoS ONE*, 8(3), e57410]
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25, 7-29.
- [Камминг, Джи. (2014). Новая статистика: Почему и как. *Психологическая наука*, 25, 7-29]
- Curran, P.G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4-19.
- [Карран, П Джи. (2016). Методы определения недобросовестных недействительных ответов в данных опроса. *Журнал экспериментальной социальной психологии*, 66, 4-19]
- DeSimone, J.A., & Harms, P.D. (2017). Dirty data: The effects of screening

- respondents who provide low-quality data in survey research. *Journal of Business and Psychology*, 1-19. 4
- [ДеСимомоне, Дж.А, и Хармс, П.Д. (2017). Недействительные данные: Результаты скрининга респондентов, предоставляющих низкокачественные данные в исследовании-опросе. *Журнал бизнеса и психологии*, 1-19. 4]
- Dunn, A.M., Heggstad, E.D., Shannock, L.R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology*, 33, 105-121.
- [Данн, А.М., Хеггестад, Е.Д., Шеннок, Л.Р. и Тайльгард, Н. (2018). Внутриндивидуальная изменчивость ответа как показатель недостаточного усилия при ответе: сравнение с другими показателями и взаимосвязь с индивидуальными различиями. *Журнал бизнеса и психологии*, 33, 105-121.]
- Holm, S. (1979). A simple sequential rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.
- [Холм, С. (1979). Процедура простого последовательного критического множественного тестирования, 6, 65-70.]
- Huang, J.L., Curran, P.G., Keeney, J., Poposki, E.M., & DeShon, R.P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27, 99–114.
- [Хуанг, Дж.Л., Карран, П.Дж., Кини, Дж., Попоски, Е.М. и ДеШон, Р.П. (2012) Обнаружение и ограничение недостаточных усилий в ответ на опросы. *Журнал бизнеса и психологии*, 100, 828–845.]
- Huang, J.L., Liu, M., & Bowling, N.A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, 100, 828–845.
- [Хуанг, Дж.Л., Лью, М. И Боулинг, Н.А. (2015). Недостаточные усилия при ответе: Изучение скрытых искажений в данных опроса. *Журнал прикладной психологии*, 100, 828–845]
- Landers, R.N., & Behrend, T.S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 8, 142-164.
- [Ландерс, Р.Н. и Беренд, Т.С. (2015). Неудобная правда: Произвольные различия между организационными, Mechanical Turk и другими случайными выборками]
- McGonagle, A.K., Huang, J.L., & Walsh, B.M. (2016). Insufficient effort survey responding: An under-appreciated problem in work and organizational health

- psychology research. *Applied Psychology: An International Review*, 65, 287–321.
- [МакГонагл, А.К., Хуанг, Дж.Л. и Уолш, Б.М. (2016). Недостаточные усилия при ответе на опрос: недооцененная проблема при работе и в исследованиях организационной психологии здоровья. *Прикладная психология: Международный обзор*, 65, 287–321.]
- McGrath, R.E., Mitchell, M., Kim, B.H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, 136, 450-470.
- [МакГрат, Р.Е., Митчелл, М., Ким, Б.Эйч., Хоу, Л. (2010). Доказательства ошибок в ответах как источник дисперсии погрешностей при прикладной оценке]
- Meade, A.W., & Craig, S.B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17, 437–455.
- [Мид, А.В. и Крейг, С.Б. (2012). Обнаружение небрежных ответов в данных опроса. *Психологические методы*, 17, 437–455]
- Mobley, W.H., Horner, S.O., & Hollingsworth, A.T. (1978). An evaluation of precursors of hospital employee turnover. *Journal of Applied Psychology*, 63, 408-414.
- [Мобли, В.Эйч., Хорнер, С.О., и Холлингсворт, А.Т. (1978). Оценка прекурсоров текучести кадров больницы. *Журнал прикладной психологии*, 63, 408-414.]
- Mowday, R.T., Steers, R.M., & Porter, L.W. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior*, 14, 224-247.
- [Моуди, Р.Т., Стирс, Р.М. и Портер, Л.В. (1979). Оценка организационной лояльности. *Журнал профессионального поведения*, 14, 224-247.]
- Mullinix, K.J., Leeper, T.J., Druckman, J.N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, 2, 109-138.
- [Маллиникс, К.Дж., Липер, Т.Дж., Дракман, Дж.Н. и Фриз, Дж. (2015). Обобщаемость опросов-экспериментов. *Журнал экспериментальной политологии*, 2, 109-138]
- Nichols, D.S., Greene, R.L., & Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology*, 45, 239-250.
- [Николс Д.С. Грин Р.Л. и Шмолк П. (1989). Критерии оценки непоследовательных моделей выбора ответа в ММПИ: обоснование, разработка и эмпирические испытания.]
- Paolacci, G., & Chandler, J. (2014). Inside the Turk understanding Mechanical

- Turk as a student sample. *Current Directions in Psychological Science*, 23, 184-188.
- [Паолаччи, Дж. и Чендлер, Дж. (2014). Внутри Turk: понимание Mechanical Turk как студенческой выборки. *Современные направления в психологии*, 23, 184-188]
- Paolacci, G., Chandler, J., & Ipeirotis, P.G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411–419.
- [Паолаччи, Дж., Чендлер, Дж. И Ипеиротис, П.Дж. (2010). Проведение испытаний на Mechanical Turk от Amazon. *Суждения и принятие решений*, 5, 411-419.]
- Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson.
- [Табачник, Б.Г. и Фиделл, Л.С., (2007). Использование мультивариативной статистики (5-е издание). Бостон, Массачусетс: Пирсон.]
- Walter, S.L., Seibert, S.E., Goering, D., & O'Boyle, E.H., Jr. (2018). A tale of two sample sources: Do results from online panel data and conventional data converge? *Journal of Business and Psychology*.
- [Уолтер, С.Л., Зайберт, С.Е., Гёринг, Д. и О'Бойл, Е.Эйч., Мл. (2018). О двух источниках выборки: сходятся ли результаты онлайн-источников и обычные результаты? *Журнал бизнеса и психологии*]

Таблица 2. Факторная корреляция с IER и без IER

Выборка	Все		Без...							
			MD		IRV		PS		PTr	
	<i>N</i>	<i>r</i>	<i>N</i>	<i>r</i>	<i>N</i>	<i>r</i>	<i>N</i>	<i>r</i>	<i>N</i>	<i>r</i>
Студенты	382	- 0,80	380	- 0,80	331	- 0,84	352	- 0,80	337	-0,71
MTurk	581	- 0,84	574	- 0,83	548	- 0,86	550	- 0,84	483	-0,70
Qualtrics	320	- 0,53	313	- 0,53	241	- 0,78	295	- 0,51	285	-0,36

Таблица 3. Расчёт и процентная величина исследования IER по выборке

Индикатор IER	Студенты (N = 382)	MTurk (N = 581)	Qualtrics (N = 320)
Расстояние Махаланобиса	2 (0.5%)	7 (1.2%)	7 (2.2%)
Вариативность индивидуального ответа	51 (13.4%)	33 (5.7%)	79 (24.7%)
Психометрические синонимы	30 (7.9%)	31 (5.3%)	25 (7.8%)
Корреляция Человек-Целое	45 (11.8%)	98 (16.8%)	35 (10.9%)
Без индикаторов IER	272 (71.2%)	424 (72.9%)	190 (59.4%)
1 индикатор	94 (24.6%)	148 (25.4%)	116 (36.3%)
2 индикатора	14 (3.7%)	7 (1.2%)	12 (3.8%)
3 индикатора	2 (0.5%)	3 (.3%)	2 (0.6%)
4 индикатора	0 (0.0%)	0 (0%)	0 (0%)
$r_{PS \square PT}$	-0,71**	-0,38**	-0,64**
$r_{PS \square IVR}$	-0,42**	-0,40**	-0,20**
$r_{PS \square MD}$	-0,69**	-0,47**	-0,55**
$r_{PT \square IVR}$	-0,36**	-0,02**	-0,02**
$r_{PT \square MD_{ceu}}$	-0,58**	0,30**	-0,46**
$r_{IVR \square MD}$	-0,09**	0,20**	-0,29**

** $p < 0,01$; * $p < 0,05$