Detecting False Identities: A Solution to Improve Web-Based Surveys and Research on Leadership and Health/Well-Being

Jeremy B. Bernerth, Herman Aguinis, and Erik C. Taylor

Detecting False Identities: A Solution to Improve Web-Based Surveys and Research on Leadership and Health/Well-Being

Jeremy B. Bernerth1, Herman Aguinis2, and Erik C. Taylor3

1 Department of Management, San Diego State University
2 Department of Management, The George Washington University
3 Department of Management, East Carolina University

A challenge for leadership and health/well-being research and applications relying on web-based data collection is false identities—cases where participants are not members of the targeted population. To address this challenge, we investigated the effectiveness of a new approach consisting of using internet protocol (IP) address analysis to enhance the validity of web-based research involving constructs relevant in leadership and health/well-being research (e.g., leader–member exchange [LMX], physical [health] symptoms, job satisfaction, workplace stressors, and task performance). Specifically, we used study participants’ IP addresses to gather information on their IP threat scores and internet service providers (ISPs). We then used IP threat scores and ISPs to distinguish between two types of respondents: (a) targeted and (b) nontargeted. Results of an empirical study involving nearly 1,000 participants showed that using information obtained from IP addresses to distinguish targeted from nontargeted participants resulted in data with fewer missed instructed-response items, higher within-person reliability, and a higher completion rate of open-ended questions. Comparing the entire sample against targeted participants showed different mean scores, factor structures, scale reliability estimates, and estimated size of substantive relationships among constructs. Differences in scale reliability and construct mean scores remained even after implementing existing procedures typically used to compare web-based and nonweb-based respondents, providing evidence that our proposed approach offers clear benefits not found in data-cleaning methodologies currently in use. Finally, we offer best-practice recommendations in the form of a decision-making tree for improving the validity of future web-based surveys and research in leadership and health/well-being and other domains.

*Keywords:* online research, methodology, surveys, quantitative methodology

The technology used by occupational health psychology researchers and practitioners to effectively monitor and address the health and well-being of employees has changed substantially in the past few decades. One such game-changing tool is web-based data collection. Some estimates indicate that up to 75% of organizations administer web-based workplace surveys to measure health and culture, leadership, engagement, and other types of employee perceptions and attitudes (Saari & Scherbaum, 2011; Van Rooy & Oehler, 2013). In addition, organizational leaders routinely gather web-based behavioral information on the number of visits and depth of exploration spent by employees on a firm’s health/benefits page and the amount of time and the number of clicks website visitors spend on the recruiting page of their organization (Aguinis & Lawal, 2012, 2013; Chandler & Paolacci, 2017; Goodman & Paolacci, 2017; Harms & DeSimone, 2015). The use of web-based employee surveys and research in the form of, for example, MTurk, StudyResponse, and Qualtrics, has increased 10-fold over just the last decade (Aguinis et al., 2021; Walter et al., 2019). In fact, the use of web-based data collection methods is so pervasive that they are relied upon in many of the most frequently studied domains in occupational health psychology including leadership (Breevaart & Bakker, 2018; Kim & Beehr, 2020), stress and burnout (French et al., 2019; Leunissen et al., 2018), well-being (Kuykendall et al., 2020), and workplace recovery (Barber & Santuzzi, 2015; Smit & Barber, 2016), just to name a few. Interestingly, despite of the much-lamented science-practice gap, some studies in health/well-being research (e.g., leader–member exchange [LMX], physical [health] symptoms, job satisfaction, workplace stressors, and task performance) (e.g., Aguinis et al., 2021; Chandler & Paolacci, 2017; Fleischer et al., 2015; Goodman & Paolacci, 2017; Harms & DeSimone, 2015; Karim et al., 2014; Lovett et al., 2018; Sharpe Wessling et al., 2017; Walter et al., 2019; Ward & Meade, 2018). Collectively, this stream of research has been informative in terms of identifying potential validity threats and solutions to address them such as the...
use of instructed-response items, strongly worded survey instructions, psychometric synonyms and/or antonyms, and the recording of total response time (i.e., the time spent completing a web-based study, or the time spent viewing a recruiting web page; Aguinis et al., 2021; Huang et al., 2012; Ward & Meade, 2018).

The Present Study

While existing methods are certainly helpful in safeguarding the integrity of data, one important but overlooked aspect of existing research with implications for all types of web-based surveys and research is who actually completes the task: *Is it the person whom researchers and practitioners intended to include in their study, or is it perhaps someone else who is not a member of the targeted population?* Clearly, this is not a validity threat when collecting data in-person using a paper-and-pencil instrument, but it is highly relevant to contemporary web-based surveys and research that is hosted and oftentimes sourced online as false identities are common on the internet (Lovett et al., 2018; Simsek & Veiga, 2001). To this point, the majority of existing research examining the quality of web-based data has focused on comparing web-based versus nonweb-based participants, or examining the quality of web-based data against absolute standards. Our study builds upon but goes beyond existing research by focusing on the implications of a new methodological tool aimed at enhancing the validity of conclusions about constructs and their relations that are of particular interest to occupational health psychologists. We specifically focus on two types of web-based participants that may provide data on leadership and health/well-being: Those intentionally targeted for study inclusion versus nontargeted participants (i.e., false identities). False identities can take various forms including participants misrepresenting their background (e.g., giving false information), duplicating their participation (i.e., taking an online survey more than once), and being someone other than a targeted participant.

Accumulated evidence suggests that the identity of web-based participants is often in doubt, which poses a serious threat to the validity of conclusions for theory as well as the effectiveness of practices derived from those theories (Chandler & Paolacci, 2017; Fleischer et al., 2015; Goodman & Paolacci, 2017; Harms & DeSimone, 2015; Marcus et al., 2017; Sharpe Wessling et al., 2017). As an illustration, Marcus et al. (2017, p. 656, Table 4, categories 3 and 4) found that over 50% of recruited participants were suspected of false identities. Likewise, Chandler and Paolacci (2017) undertook a series of experimental investigations of MTurk participants and discovered that when a demographic screening question mentioned sexual orientation, 45.3% of participants identified as lesbian, gay, or bisexual (LGB) [sic]. Only 3.8% of participants identified as LGB [sic] in a duplicate study posted without the screening question. In an even more striking illustration, Sharpe Wessling and colleagues reported that of the 900 MTurkers who began a study “. . . all but 33 dropped out when asked to provide a screenshot that verified their qualifications” (2017, p. 212). The threat of misrepresented or outright false identities means existing data-quality controls that address other types of validity threats are less relevant. This is because the integrity of web-based research rests on the assumption that the data collected accurately represent the targeted population (DeSimone et al., 2018).

In response to the challenges raised in the preceding paragraphs, we investigate the effectiveness of a novel approach consisting of using information obtained from internet protocol (IP) addresses to improve the validity and usefulness of web-based data in research and applications in leadership and health/well-being and related domains. More precisely, we use IP addresses to (a) calculate IP threat scores—a value representing the likelihood that a given IP address is deliberately masked and/or associated with prior malicious web activity and (b) obtain information about participants’ internet service providers (ISPs). Using these two distinct but complementary pieces of information, we distinguish between different types of study participants: Targeted and nontargeted. We use IP threat scores to objectively assign a probability of group membership (i.e., targeted vs. nontargeted) and ISP information to classify participants as either consumer-focused (likely to be targeted participants) or proxy-based (likely to be nontargeted participants). We then use the information on participants’ identities to assess aspects of data-quality regarding constructs particularly relevant to leadership and health/well-being: LMX, antecedents of well-being (e.g., employee neuroticism), and various manifestations of employee health/well-being (e.g., physical [health] symptoms, job satisfaction; see Danna & Griffin, 1999, for a review of key components of health and well-being). In this way, our proposed IP analysis provides a two-part solution to identifying problematic web-based participants that may need to be excluded from a study due to false identities. As understanding and advancing theory on leadership and health/well-being is dependent on valid data, and valid data are obviously also critical for the implementation of effective organizational interventions, our methodological innovation is useful not only for researchers but also for practitioners who rely on health/well-being data collected over the internet.

For the remainder of the article, we use the generic term targeted to refer to those participants with low IP threat scores and/or consumer-focused ISPs. In addition, we use nontargeted to refer to those participants with high IP threat scores and/or proxy-based ISPs (i.e., those users who have an intermediary server separating them from the websites they browse).

Internet Protocol Addresses, Internet Protocol Threat Scores, Internet Service Providers, and Participants’ Identities

An IP address is a unique identifier assigned to a particular machine accessing a given network. Similar to how each vehicle on the interstate has a unique license plate, all computers, mobile phones, and other devices accessing a network or the internet have unique IP addresses. The distinction between networks and the broader internet is important because devices generally do not connect directly to the internet. Rather, a device first connects to a network such as that of an ISP, an organization’s private network, or a wireless hotspot. This hosting network then assigns an IP address to each connected device. Just as a vehicle carries its license plate wherever it travels, every request for information that is sent through the internet—be it the form of an email, a text message, or a request to access a web page by clicking a uniform resource locator (URL)—carries with it an IP address that uniquely identifies the machine initiating the request. Consequently, IP addresses serve as
unique identifiers for each machine accessing the internet by provid-
ing an “address” to which information accessed via the internet is sent.

While unbeknownst to some, IP addresses can be traced and attrib-
ted to a specific location. Websites such as whatismyip.com and
dnsleaktest.com reveal different types of information that can be
obtained from an IP address, including the host network or organi-
zation, approximate longitude and latitude of the requesting
machine, and in some cases even the name of the host’s customer.
Common hosts include consumer-focused companies such as Com-
cast Xfinity, Charter Spectrum, and Virgin Mobile, as well as many
businesses, universities, and governmental entities that maintain
their own networks. In addition to protecting resources such as
subscription databases and proprietary software, it is not uncommon
for hosts like universities and businesses to establish virtual private
networks (VPNs) that allow users remote access. In this way, VPNs
work as an encryption tunnel—information exchanged between the
local host and the remote user is generally not visible or interpretable
to other machines or users on the internet. Put differently, a VPN or
other proxy service (e.g., the onion router) cloaks the data trans-
ferred in the last leg of the data exchange—the communication
between the local host and end-user (i.e., the individual using the
internet-accessing device). An employee who is using a laptop on a
Wi-Fi hotspot at a coffee shop can therefore use a VPN to ensure that
data transferred between the institution’s network servers and the
user’s laptop are not observable by the hotspot provider or anyone
else monitoring data transferred through the Wi-Fi hotspot.

Increasingly, though, individuals use VPNs and other similar
proxy services outside of institutional contexts to mask or hide their
personal internet traffic (Longworth, 2018; Normile, 2017; Volz &
Tau, 2020; Zhipeng et al., 2018). Many personal VPN and other
proxy services are available for a modest subscription fee or even
free of charge, which provides end-users an encrypted connection to
a network separate from their ISP. The reasons individuals use proxy
services to encrypt their data vary greatly. Some may merely wish
to prevent their personal data from being accessible by big tech
corporations (e.g., Google and Facebook) or government entities
that track web traffic (Normile, 2017), others may have concerns
about the ability of their ISP to protect their information (Ikram et al.,
2016; Zhipeng et al., 2018), and yet others are engaged in activities
that are unethical or illegitimate (e.g., hacking and spamming) and
purposefully wish to maintain web anonymity (Van der Wagen &
Pieters, 2015; Volz & Tau, 2020). It is also common to use a VPN to
make it appear as though one’s web traffic originates from a different
geographic location (e.g., a user located in China may use a VPN to
falsely present an IP address associated with a U.S.-based location;
Chandel et al., 2019; Normile, 2017). Regardless of the reason,
accessing the internet through such a proxy changes the device’s IP
address and the information that can be attained from it because the
user now accesses the broader internet from the proxy’s intranet
rather than the ISP’s intranet.

By its very nature, interpreting information from these proxied
IPs is a difficult and often-precise process because IP addresses
may change based on where the user accesses the internet (e.g., a
home network vs. a public hotspot) or based on the host’s network
settings. 1 Regardless, when it is important to verify that the person is
in fact an authorized user as opposed to a robot (i.e., “bot”), hacker,
spammer, someone intentionally misrepresenting their geographic
location, or other entity of potential ill-intent, it is critical to assess
the information that can be obtained from a user’s IP address. While
host often utilize preemptive measures to ensure that only autho-
rized users access data or a web application (e.g., an outright block
of any proxied internet traffic; Davis & Zboralska, 2017), such
imprecise practices may erroneously exclude targeted participants
who are using a proxy service for legitimate reasons. In such case, an
effective alternative method for evaluating IP addresses is to
establish the extent to which web traffic originated from a particular
IP address poses a threat.

Network administrators scrutinize IP addresses to minimize web
threats routinely, although early systems focused more on eliminating
viruses, worms, and denial of service (DoS) attacks (Ohnof et al.,
2005). A more recent innovation involves the use of probability
theory and machine learning to check IP addresses against dynamic
lists of known proxies or hosts associated with malicious internet
activity (i.e., hosts from which nefarious internet activity has origi-
nated previously or that have notable similarities to such hosts). This
process generates an IP threat score in the form of a probability
ranging from zero to one. Scores approaching zero indicate increas-
ingly “safe” or nonsuspicious web traffic. In contrast, scores ap-
proaching one indicate that the traffic originates from a proxy server
and suggests that participants have taken deliberate action to mask
their identity, and/or that the server is associated with prior malicious
web activities. Although these systems are proprietary, the checking
process involves comparing a given IP address against both static lists
(e.g., IPs with established patterns of malicious web activity) and
dynamic lists, as well as utilizing machine learning to compare
patterns of activity to infer if web traffic from a given IP resembles
patterns of known malicious IPs. In other words, a variety of
information is utilized in these determinations, including the nature
of data sent and requested, prior instances of DoS or similar attacks
originating from the host, the number of activities deemed to be
suspicious or nefarious, and the recency of those occurrences (Sander
et al., 2018; Visbal, 2014, 2015, 2017). The combination of this
information is then utilized to produce a probability score represent-
ing the likelihood that web traffic originating from a given IP address
poses a threat.

An IP threat score is not the same as simply identifying IP
addresses originating from consumer-focused or proxy-based
ISPs in that the latter reveals which users have taken deliberate steps
to mask their identities and/or locations, but may also raise red flags
when using VPNs benignly as a means of protecting their personal
data. IP threat scores flag those IP addresses known to engage in
n nefarious activity and/or IP addresses whose behavior and web
traffic closely resemble nefarious activities. Consequently, if the
web traffic originates from a public hotspot that has been abused by
one or more nefarious actors in the past, but the majority of web
traffic originating from the hotspot is not malicious in nature, the
associated IP address will likely register a value less than one
(the highest threat level).

1 IP addresses can either be static (i.e., permanent), or dynamic (i.e., based
on IP assignment from the host). For dynamic IP addresses, the numbers
composing the address typically remain constant for weeks at a time, but may
change, though rarely within a 24-hr period (Balakrishnan et al., 2009;
Hildén, 2017; Xie et al., 2007). Also, while the numbers may change, the
information that can be attained from the IP address (e.g., consumer or
proxy-based ISP) does not change unless the individual uses a VPN or the
host changes its network information.
Using an Unobtrusive Innovation to Improve the Validity and Usefulness of Web-Based Data Collection

A measure’s properties fundamentally include reliability and validity. But, understanding a measure’s quality also includes understanding factors that influence these two psychometric characteristics. For example, Nunnally (1978) emphasized participant-related considerations that influence reliability and validity, including the speed of survey completion and carelessness in responding (pp. 675–676). We, therefore, conducted an empirical study to critically examine the effectiveness of using information garnered from IP addresses to confront the problem of fake identities in web-based data collection. In particular, we used IP addresses to (a) calculate IP threat scores and (b) identify research participants’ ISPs, which together provide information on whether each participant belongs in a targeted population. We then use this information to examine differences between targeted and nontargeted respondents regarding: (a) answers to instructed-response questions, (b) within-person reliability, (c) completion of open-ended (i.e., nonmultiple-choice) questions, (d) amount of time spent completing a web-based study, (e) construct means, (f) factor structure (i.e., configural invariance), (g) scale reliability, and (h) predictive validity (i.e., substantive relations between constructs particularly relevant to leadership and health/well-being). We chose these particular criteria because they are key indicators of the validity and usefulness of conclusions using web-based data (cf. Dunn et al., 2018; Johnson, 2005; Kung et al., 2018; Ward & Meade, 2018).

Potential Effects on Data Quality Criteria

Our rationale for undertaking an IP analysis as a potential unobtrusive tool for improving web-based data is the following. Those individuals who have a low IP threat score (i.e., a lower probability of being a proxy/bot/spammer) and/or use a consumer-focused ISP represent genuine users—genuine in the sense that these individuals are more likely to be members of the targeted population and not participating in a web-based activity (e.g., organizational health surveys) for nefarious reasons such as a means to obtain an unwarranted monetary incentive. We also expect participants with low IP threat scores and consumer-focused ISPs to have more knowledge and insight into the study’s concepts and organizational leaders as they are more likely to be members of the targeted population. Principles of social exchange and the norm of reciprocity are likewise applicable to these individuals to the extent that they receive a resource (e.g., a gift card) in exchange for a specific behavior and therefore feel compelled to give in equivalence (Blau, 1964; Gouldner, 1960). Equivalence in the context of web-based research means reading and carefully responding to survey questions (Simsek & Veiga, 2001; Ward & Meade, 2018). In short, we pose the following questions:

Research Question 1: Do participants’ IP threat scores correlate with (a) the number of instructed response questions missed, (b) within-person reliability, (c) the number of open-ended questions voluntarily answered, and (d) the amount of time spent on an online survey?

Research Question 2: Do data provided by respondents using consumer-focused ISPs and those using proxy-based ISPs differ with respect to (a) the number of instructed response questions missed on an online survey, (b) within-person reliability on an online survey, (c) the number of open-ended questions voluntarily answered on an online survey, (d) the amount of time spent on an online survey, (e) mean scores across common health/well-being constructs, and (f) factor structures across common leadership and health/well-being constructs?

If the aforementioned data-quality indicators differ between targeted users (i.e., consumer-focused ISPs) and nontargeted users (i.e., proxy-based ISPs), then it likely that there will be differences regarding internal consistency reliability and predictive validity. We note this possibility because individuals rushing through a study for illegitimate reasons could miss the nuances built into multi-item measures or simply randomly select answers—the combination of which would likely inflate error variance substantially. Moreover, as noted earlier, participants using consumer-focused ISPs likely represent genuine individuals specifically targeted for a research study. Having been targeted for a reason, their input is based on knowledge and experience not likely possessed by others. Participants using proxy-based ISPs take deliberate attempts to mask their identity because they may not possess some or all of the characteristics (e.g., department membership, work experience, sexual orientation, and ethnicity) desired by the researcher. As a result, the data they provide may have different psychometric properties.

Importantly for theory and practice, differences in psychometric properties between targeted and nontargeted participants are likely to result in differences in substantive conclusions regarding hypothesized models. Of particular interest to leadership and health/well-being are models involving the prediction of individual performance based on such antecedents as LMX, well-being, and workplace stressors. Also of particular interest to leadership and health/well-being are models involving the prediction of different aspects of well-being based on antecedents such as LMX and organizational justice (e.g., Eib et al., 2015, 2018; Wang et al., 2019). Models such as these are particularly relevant given a growing body of research showing that effective leadership helps enhance the health and well-being of both leaders and their direct-report employees (Arnold, 2017; Bernerth & Hirschfeld, 2016; Kaluza et al., 2020). This is notable for employees, leaders, and organizations as healthy leaders who provide support to employees set the stage for improved employee and organizational performance (e.g., Mao et al., 2019; Ott-Holland et al., 2019; Rocanin et al., 2017). Accordingly, we assessed implications of our methodological innovation for (a) leadership variables such as LMX and organizational justice (the lens through which many employees view leadership; Mackey et al., 2017) that directly and indirectly influence health and well-being (Eib et al., 2015, 2018; Giacalone & Promislo, 2010; Wang et al., 2019); (b) other antecedents to health/well-being such as neuroticism, conscientiousness, autonomy, and interpersonal conflict (Danna & Griffin, 1999; Giacalone & Promislo, 2010; Levy et al., 2012); and (c) multiple manifestations of health/well-being including job satisfaction, perceived stress, and physical (health) symptoms (Danna & Griffin, 1999; Hakanen et al., 2018; Rahimnia & Sharifirad, 2015). Formally, we pose the following question:

Research Question 3: Do internal consistency reliability and predictive validity differ when using datasets including (a) targeted versus (b) non-targeted versus; and (c) all participants combined?
Method

Participants and Procedure

We recruited study participants from an online social media forum as part of a larger study examining workplace stressors, downtime, subjective well-being, and work-related behavior. We offered a $10 Amazon gift card and the recruitment message explicitly noted that no compensation would be paid to those individuals who randomly selected answers and/or who did not fit the study criteria. We posted a link to the study directly on the message board directing potential participants to a survey hosted by Qualtrics. As is the default setting in many web-based platforms, all surveys recorded the participants’ IP addresses.

Measures of Participants’ Identities

IP Threat Score

We used an analytic tool that has been publicly available since 2015 (http://getipintel.net/) to determine each participants’ IP threat score. This tool, which is dynamic and constantly updates in real time, allows one to enter the numeric IP address and receive a probability ranging from 0 to 1 with numbers increasingly close to 1 indicating that the electronic device used to participate in the study has been flagged on lists of known masking or malicious hosts.

Internet Service Provider Designation (Targeted or Nontargeted)

After participants submitted their responses, we used the publicly available IP address lookup tool https://www.infobyip.com/ to ascertain their ISP provider. Tools such as this allow one to enter a numeric IP address and receive detailed information including the ISP for each participant. The first and third authors independently assessed the ISP provider and designated each respondent as either 0 (targeted) or 1 (nontargeted). Example respondent ISPs designated as targeted include Comcast, Time Warner, Quest Communications, AT&T, and the University of Washington. Example respondent ISPs designated as nontargeted include Psychz Networks, Vultr Holdings, Sharktech, CyberGhost, and Leaseweb. The initial inter-rater agreement was 86% and we addressed and resolved all discrepancies through discussion.

Measures of Data Quality Criteria

Instructed Response Items

We included two instructed-response items (Please select strongly disagree/strongly agree for this statement). We scored both items as either 0 (answered correctly) or 1 (answered incorrectly) and summed them to form a total score.

Within-Person Reliability

We used two estimates of within-person reliability, which provide information on individual reliability based on individual scores and complement the reliability of a measure based on scores from an entire sample. First, we used a half-scale subset method (Johnson, 2005) that splits multi-item measures into even and odd subscales (e.g., the average of LMX items 1, 3, 5, and 7 becomes the first measure of LMX and the average of LMX items 2, 4, 6, and 8 becomes the second measure of LMX). Once we did this for each measure, we then calculated the correlation between each of the split measures across an individual—giving one a measure of individual reliability. Higher scores indicate that individual participants provided reliable (i.e., consistent) data. Second, we used a psychometric antonym method in which forward and reversed-scored items are correlated across individual participants (Johnson, 2005). Our study included eight such pairs (e.g., efficient/inefficient). We reverse-scored this measure of reliability to match the half-scale method (i.e., positive scores represent greater reliability), giving us two measures of individual reliability.

Completion of Open-Ended Questions

Following the completion of Likert scale questions, we presented participants with three optional open-ended demographic questions: (a) age in years, (b) organizational tenure in years and months, and (c) number of jobs held over the last 5 years. We scored age 0 (did not respond) or 1 (responded), organizational tenure 0 (did not respond), .5 (responded with a single number), or 1 (responded with text indicating years and months), and a number of jobs 0 (did not respond) or 1 (responded). Then, we summed the three indicators to give each participant a total score.

Total Time Spent

The web-based data collection platform we used recorded the total amount of time in minutes and seconds participants spent completing our study.

Measures of Substantive Leadership and Health/Well-Being Constructs

Organizational Justice

We assessed overall organizational justice perceptions with a three-item measure by Ambrose and Schminke (2009) and a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). A sample item is “In general, the treatment I receive around here is fair.” Coefficient alphas for the measure including all participants combined and only targeted participants (i.e., consumer-focused ISP) were .61 and .77, respectively.

Leader–Member Exchange

We used the eight-item measure by Bernerth et al. (2007), with a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). A sample item is “If I do something for my manager, he or she will eventually repay me.” Coefficient alphas for all participants combined and only targeted participants (i.e., consumer-focused ISP) were .74 and .86, respectively.

Physical (Health) Symptoms

We assessed participants physical health with 12 items by Spector and Jex (1998) using a 5-point scale ranging from 1 (never) to 5 (very often) and a stem of “Over the past month, how often have you experienced the following symptoms . . .” Sample items include “trouble sleeping,” “headaches,” and “an upset stomach.
or nausea.” Coefficient alphas for all participants combined and only targeted participants (i.e., consumer-focused ISP) were .81 and .86, respectively.

**Job Satisfaction**

We used a single-item measure to capture overall job satisfaction: “All in all, I am satisfied with my job.” We deemed this appropriate because meta-analytic research on job satisfaction provides evidence that single-item measures are equally valid as multiple-item measures (Dolbier et al., 2005; Wanous et al., 1997).

**Conscientiousness and Neuroticism**

Using a 5-point response scale ranging from 1 (extremely inaccurate) to 5 (extremely accurate), we assessed conscientiousness and neuroticism with Saucier’s (1994) Big Five markers (eight items per construct). For conscientiousness, coefficient alphas for all participants combined and only participants with a consumer-focused ISP were .69 and .74, respectively. Sample items include “organized,” “systematic,” and “careless” (reversed-scored). For neuroticism, coefficient alphas for all participants combined and only targeted participants (i.e., consumer-focused ISP) were .60 and .73, respectively. Sample items include “temperamental,” “irritable,” and “touchy.”

**Job Autonomy**

We assessed job autonomy with a three-item measure by Spreitzer (1995) using a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). A sample item is “I have significant autonomy in determining how I do my job.” Coefficient alphas for all participants combined and only targeted participants (i.e., consumer-focused ISP) were .64 and .75, respectively.

**Interpersonal Conflict**

We assessed the amount of interpersonal conflict participants experienced at work with four items by Spector and Jex (1998) using a 1 (never) to 5 (very often) response format. Example items include “How often . . . do you get into arguments with others at work,” and “. . . are people rude to you at work.” Coefficient alphas for all participants combined and only targeted participants (i.e., consumer-focused ISP) were .64 and .74, respectively.

**Workplace Stressors**

We assessed workplace stressors with a five-item measure by Spector and Jex (1998) and a 5-point scale ranging from 1 (never) to 5 (very often). Example items include “How often . . . does your job leave you with little time to get things done,” and “. . . do you have to do more work than you can do well.” Coefficient alphas for all participants combined and only targeted participants (i.e., consumer-focused ISP) were .52 and .63, respectively.

**Task Performance**

We asked participants to assess their own work-related task performance using a referent-shift (“My supervisor would say I . . . ” e.g., “. . . adequately complete assigned duties;” see Schoorman & Mayer, 2008) with seven items by Williams and Anderson (1991). We recorded responses with a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Coefficient alphas for all participants combined and only targeted participants (i.e., consumer-focused ISP) were .77 and .83, respectively.

**Results**

Before answering our three research questions, we compared the demographic variables of participants designated as targeted and those designated as nontargeted (based on their ISP, a dichotomous variable) and found nontrivial differences regarding gender, education, and age. Values for targeted versus nontargeted participants were as follows: (a) Gender: 71% versus 54% male ($\chi^2(1) = 27.74$, $p < .001$), (b) education: 3% versus 11% some high school, 12% versus 17% high school grad/GED, 24% versus 16% some college, 18% versus 20% 2-year college degree, 30% versus 20% 4-year college degree, and 12% versus 16% graduate degree ($\chi^2(5) = 39.90, p < .001$), and (c) age: 28.2 ($SD = 7.5$) versus 27.0 ($SD = 7.4$) ($t(851) = 2.34, p = .02$).

Table 1 provides the means, standard deviations, and correlations among all study variables. Turning to our research questions, because an IP threat score is a continuous variable, we answered **Research Question 1** using regression analysis in which we regressed IP threat scores on the data quality criteria. Results indicate that IP threat scores are related to psychometric antonyms ($b = -.18, p < .001$) and the number of open-ended questions completed ($b = -.33, p < .001$), but not the amount of time participants spent completing an online study ($b = -.76, p = .90$). IP threat scores are also related to the number of instructed-response questions missed but in the opposite direction of what might be expected of nontargeted individuals ($b = -.49, p < .001$). These results suggest the answer to **Research Question 1** is that IP threat scores are associated with some indicators of the psychometric soundness of web-based data.

**Research Question 2** asked about potential differences between targeted and nontargeted participants across data-quality indicators based on ISP designation. Because participants’ ISP is a dichotomous variable, we answered this question using univariate analysis of variance (ANOVA; see Table 2), multivariate analysis of variance (MANOVA) with follow-up ANOVAs (see Table 3), and Kline’s (2011) factor structure comparison procedure. Results in Table 2 indicate that ISP designation was related to the total number of missed instructed-response questions such that targeted participants missed fewer instructed responses questions than nontargeted participants, $F(1, 984) = 136.18, p < .001, d = .78$. Results in Table 2 also indicate targeted participants resulted in greater within-person reliability than nontargeted participants for the overall half-scale measure, $F(1, 984) = 227.76, p < .001, d = 1.0$, and for

---

2 This negative coefficient suggests that as the probability of a proxied IP address increases, the less likely the participant is to miss an instructed response item. Given the newness of our methodological innovation and a lack of prior investigations, we are left to speculate about the reason why participants with a high IP threat score missed fewer data-quality checks. One possibility is that an IP threat score does a better job of detecting individuals that routinely try to “game” financially incentivized web-based labor activities (e.g., online surveys and eLancing). If true, such individuals might be cognizant of survey checks, and in kind pay closer attention than those simply trying to receive an unwarranted reward (see Hauser et al., 2019). Future research could examine this possibility.
### Table 1

**Descriptive Statistics and Correlations Among Study Variables**

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data quality predictors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Internet protocol threat score</td>
<td>.43</td>
<td>.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Internet service provider (ISP)</td>
<td>.59</td>
<td>.49</td>
<td>.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data quality criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Instructed response items missed</td>
<td>.79</td>
<td>.91</td>
<td>−.23</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Within-person reliability (half-scale method)</td>
<td>.54</td>
<td>.46</td>
<td>.01</td>
<td>−.43</td>
<td>−.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Within-person reliability (antonyms)</td>
<td>.01</td>
<td>.46</td>
<td>−.17</td>
<td>−.16</td>
<td>.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Open-ended questions completed</td>
<td>1.51</td>
<td>1.00</td>
<td>−.14</td>
<td>−.47</td>
<td>−.40</td>
<td>.51</td>
<td>.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Total time spent on study (in minutes)</td>
<td>12.18</td>
<td>81.74</td>
<td>.00</td>
<td>−.03</td>
<td>−.11</td>
<td>.09</td>
<td>.07</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Substantive leadership and health/well-being constructs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Organizational justice</td>
<td>4.75</td>
<td>1.39</td>
<td>−.03</td>
<td>−.33</td>
<td>−.42</td>
<td>.58</td>
<td>.14</td>
<td>.36</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Leader–member exchange</td>
<td>4.65</td>
<td>1.08</td>
<td>−.31</td>
<td>−.44</td>
<td>.54</td>
<td>.10</td>
<td>.38</td>
<td>.07</td>
<td>.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Physical (health) symptoms</td>
<td>2.48</td>
<td>.77</td>
<td>−.22</td>
<td>.57</td>
<td>−.53</td>
<td>−.11</td>
<td>−.25</td>
<td>−.09</td>
<td>−.39</td>
<td>−.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Job satisfaction</td>
<td>4.66</td>
<td>1.89</td>
<td>−.25</td>
<td>−.31</td>
<td>.37</td>
<td>.15</td>
<td>.30</td>
<td>.04</td>
<td>.46</td>
<td>.46</td>
<td>−.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Conscientiousness</td>
<td>3.48</td>
<td>.73</td>
<td>−.52</td>
<td>.55</td>
<td>.13</td>
<td>.43</td>
<td>.06</td>
<td>.47</td>
<td>.51</td>
<td>−.60</td>
<td>.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Neuroticism</td>
<td>2.63</td>
<td>.67</td>
<td>.27</td>
<td>.43</td>
<td>−.45</td>
<td>−.15</td>
<td>−.35</td>
<td>−.35</td>
<td>−.39</td>
<td>.53</td>
<td>−.30</td>
<td>−.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Job autonomy</td>
<td>4.57</td>
<td>1.38</td>
<td>−.39</td>
<td>.47</td>
<td>.12</td>
<td>.37</td>
<td>.07</td>
<td>.42</td>
<td>.51</td>
<td>−.34</td>
<td>.33</td>
<td>.36</td>
<td>−.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Interpersonal conflict</td>
<td>2.50</td>
<td>.93</td>
<td>.37</td>
<td>.47</td>
<td>−.53</td>
<td>−.25</td>
<td>−.44</td>
<td>−.44</td>
<td>−.47</td>
<td>.51</td>
<td>−.33</td>
<td>−.53</td>
<td>.45</td>
<td>−.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Workplace stressors</td>
<td>3.06</td>
<td>.75</td>
<td>.06</td>
<td>.07</td>
<td>.06</td>
<td>.08</td>
<td>.09</td>
<td>.04</td>
<td>.02</td>
<td>.07</td>
<td>.04</td>
<td>.06</td>
<td>.07</td>
<td>.06</td>
<td>.08</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Task performance</td>
<td>4.86</td>
<td>1.24</td>
<td>−.07</td>
<td>−.49</td>
<td>−.56</td>
<td>.65</td>
<td>.36</td>
<td>.54</td>
<td>.08</td>
<td>.51</td>
<td>.55</td>
<td>.51</td>
<td>.41</td>
<td>.64</td>
<td>−.50</td>
<td>.43</td>
<td>−.61</td>
<td>.20</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** N = 986. For ISP, 0 = targeted participants (i.e., consumer-focused ISPs); 1 = non-targeted participants (i.e., proxy-based ISPs). Internet protocol threat score ranges from 0 (i.e., increasingly “safe” or nonsuspicious web traffic) to 1 (i.e., participant is a likely a bot, hacker, spammer, or other non-targeted individual). Correlations of |.07| or greater are statistically significant at p < .05.
psychometric antonyms, $F(1, 984) = 18.87, p < .001, d = .26$. Moreover, while targeted participants answered more open-ended questions than nontargeted participants, $F(1, 984) = 285.91, p \leq .001, d = 1.07$, we found no differences ($F(1, 984) = 1.16, p = .28$) in the amount of time participants spent completing the study.\(^3\) Means reported in Table 2 show that targeted respondents missed fewer instructed-response questions missed by nontargeted respondents by an order of about 2.5 times, had a within-person reliability more than twice that of nontargeted respondents, and answered nearly twice as many open-ended questions as nontargeted respondents. In terms of Research Question 2e (i.e., possible differences regarding construct means), MANOVA results indicate a significant omnibus mean difference across constructs: Wilks’ $\lambda = .73, F(1, 985) = 35.25, p < .001, \eta^2 = .27$. Follow-up ANOVAs summarized in Table 3 show that the type of participant resulted in mean differences across all 10 substantive constructs and the average $d$ across the 10 constructs was a nontrivial .63.

To answer Research Question 2f (i.e., possible differences regarding factor structures), we fitted a unified measurement model for all multi-item latent constructs across both groups of participants (i.e., targeted and nontargeted). Given that there are several fit indexes available, as well as recommended cutoffs (Schermelleh-Engel et al., 2003), and an ongoing debate among methodologists regarding those cutoffs (e.g., Lai & Green, 2016; see also Nye & Drasgow, 2011, for a discussion of why rules of psychometric antonyms are not metrically equivalent. In sum, combined with results from Research Questions 2a–2e, the answer to Research Question 2 is yes, the data provided by targeted and nontargeted participants differ in terms of psychometric properties.

Regarding Research Question 3 (i.e., possible differences across reliability and predictive validity), results summarized in Table 4 show that reliability estimates were higher for the subsample of targeted participants compared to the entire sample for each of the nine multi-item measures.\(^5\) Across the nine measures, the average internal consistency reliability improved from .66 to .77 when considering only targeted participants. Differences were even more striking when comparing targeted to nontargeted participants as the average internal consistency reliability improved from .53 to .77, a reduction in error variance of over 50%. In addition, we investigated differences between the two types of participants using Item Response Theory (IRT) procedures (Lang & Tay, 2021). Specifically, we calculated empirical reliability for each multi-item latent construct using the “mirt” (Chalmers, 2012) and “ltm” (Rizopoulos, 2006) packages in R for the full sample (i.e., all available data), targeted sample, and nontargeted sample. Results indicated a similar pattern to that found in Table 4: The targeted sample demonstrated superior (and acceptable; average = .77) reliability and predictive validity, results from Research Questions 2a–2e, the answer to Research Question 2 is yes, the data provided by targeted and nontargeted participants differ in terms of psychometric properties.

\(^3\) There were no differences between the type of respondents when using all available data, but the time it took for participants to complete the survey varied widely. Although the platform we used recorded the total amount of time from starting a survey to completing a survey, it did not distinguish participants who completed the survey in a single versus multiple sessions. Therefore, we investigated the potential impact of removing outliers (using difference in fit, DFFIT, and difference in beta, DFBETA statistics). Removing a single outlier did not change study results, but removing multiple cases based on the heuristics described by Aguinis et al. (2013) did reveal differences in groups—with targeted respondents taking nearly two more minutes on average to complete the survey, providing additional evidence regarding the higher quality of data gathered from targeted respondents.

\(^4\) The reported CFI values are not as strong as the other fit indices reported, but we note CFI is unique from other indicators in that it “punishes” those who estimate a number of parameters—especially when correlations among variables are low to moderate (as was the case with our data; cf. Lai & Green, 2016; see also Nye & Drasgow, 2011, for a discussion of why rules of thumb can be problematic with fit indices). Given the other fit indices (SRMR = .07, RMSEA = .03) were within acceptable ranges, we conclude the overall fit was adequate.

\(^5\) Cronbach’s alpha provides a measure of reliability, but there are instances in which it may overestimate the reliability of a measure (Cho & Kim, 2015; Cortina et al., 2020). Our measure of physical symptoms may be one such instance given this measure is a more of an index, and as such, not intended to have tau equivalence.
Table 3

Analysis of Variance of the Effects of Internet Service Provider (ISP) Information on Construct Mean Scores

<table>
<thead>
<tr>
<th>Group</th>
<th>Physical symptoms</th>
<th>Conscientiousness</th>
<th>Neuroticism</th>
<th>Task performance</th>
<th>Job satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Targeted participants</td>
<td>2.29</td>
<td>.75</td>
<td>3.77</td>
<td>.65</td>
<td>2.41</td>
</tr>
<tr>
<td>Nontargeted participants</td>
<td>2.62</td>
<td>.75</td>
<td>3.28</td>
<td>.72</td>
<td>2.78</td>
</tr>
<tr>
<td>F(1, 984)</td>
<td>48.46</td>
<td>* *</td>
<td>118.45</td>
<td>* *</td>
<td>78.44</td>
</tr>
<tr>
<td>d</td>
<td>.44</td>
<td>.71</td>
<td>.56</td>
<td>.14</td>
<td>.54</td>
</tr>
</tbody>
</table>

Note. N = 986. Targeted participants include those using consumer-focused ISPs. Nontargeted participants include those using proxy-based ISPs.

Table 4

Scale Reliabilities and Regression Results Across All Participants Combined and Two Subsamples (i.e., Targeted vs. Nontargeted Participants)

<table>
<thead>
<tr>
<th>Predictors</th>
<th>αFull</th>
<th>αTargeted</th>
<th>αNontargeted</th>
<th>βFull, Targeted, Nontargeted</th>
<th>SEFull, Targeted, Nontargeted</th>
<th>βFull, Targeted, Nontargeted</th>
<th>SEFull, Targeted, Nontargeted</th>
<th>βFull, Targeted, Nontargeted</th>
<th>SEFull, Targeted, Nontargeted</th>
<th>R²Full, Targeted, Nontargeted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organizational justice</td>
<td>.61</td>
<td>.77</td>
<td>.41</td>
<td>.46, .33, .36a, b</td>
<td>.03, .04, .03</td>
<td>18.5, 8.9, 11.4</td>
<td>.24, .16, .18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leader–member exchange</td>
<td>.74</td>
<td>.86</td>
<td>.57</td>
<td>.64, .37, .64a, c</td>
<td>.03, .05, .04</td>
<td>20.6, 8.2, 16.3</td>
<td>.30, .14, .32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical (health) symptoms</td>
<td>.81</td>
<td>.86</td>
<td>.77</td>
<td>.83, .71, .68a, b</td>
<td>.04, .06, .05</td>
<td>.18.8, .11.3, .12.9</td>
<td>.26, .24, .22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job satisfaction</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.27, .28, .17b, c</td>
<td>.02, .03, .02</td>
<td>14.2, .91, .77</td>
<td>.17, .17, .09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.69</td>
<td>.74</td>
<td>.61</td>
<td>.10.8, 1.01, .85a, c</td>
<td>.04, .07, .05</td>
<td>25.8, 15.2, 16.3</td>
<td>.40, .36, .32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>.60</td>
<td>.73</td>
<td>.43</td>
<td>.92, .70, .75a, b</td>
<td>.05, .07, .07</td>
<td>18.0, .9.9, .11.5</td>
<td>.25, .19, .19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job autonomy</td>
<td>.60</td>
<td>.75</td>
<td>.44</td>
<td>.39, .24, .32a, b</td>
<td>.03, .04, .03</td>
<td>14.9, 6.2, 10.2</td>
<td>.19, .09, .15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpersonal conflict</td>
<td>.64</td>
<td>.74</td>
<td>.49</td>
<td>.81, .83, .56a, c</td>
<td>.03, .05, .05</td>
<td>.24.1, .16.7, .12.4</td>
<td>.37, .41, .21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workplace stressors</td>
<td>.52</td>
<td>.63</td>
<td>.46</td>
<td>.33, .20, .33</td>
<td>.05, .08, .06</td>
<td>6.4, 2.6, 5.9</td>
<td>.04, .02, .06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task performance</td>
<td>.77</td>
<td>.83</td>
<td>.60</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
</tbody>
</table>

Note. Overall N = 986; targeted participants N = 405; nontargeted participants N = 581. Full = data set composed of all available data. Targeted = data set composed of those participants with a consumer-focused ISP. Nontargeted = data set composed of those participants with a proxy-based ISP. B is the unstandardized regression coefficient. α = internal consistency reliability coefficient.

a Statistically significant difference between regression coefficients for full versus targeted samples (p < .05). b Statistically significant difference between regression coefficients for full versus nontargeted samples (p < .05). c Statistically significant difference between regression coefficients for targeted versus nontargeted samples (p < .05).
predictors: LMX \( (z = -4.22, p < .001) \), job satisfaction \( (z = 3.05, p = .002) \), conscientiousness \( (z = 1.86, p = .063) \), and interpersonal conflict \( (z = -3.81, p < .001) \).

Clearly, occupational health psychologists are concerned with more than employees’ performance. Accordingly, we also explored the influence of our proposed methodology on substantive conclusions about the relation between two popular leadership constructs (LMX and organizational justice) and three forms of employee well-being (physical [health] symptoms, job satisfaction, and perceptions of workplace stressors). As noted in the introduction, respondents using consumer-focused ISPs likely represent targeted individuals who possess certain characteristics (e.g., department membership, supervision by a particular manager) not possessed by others. Respondents using proxy-based ISPs take deliberate steps to mask their identity because they likely do not have the characteristics desired by the hosting entity. If such data are included in analyses, the data they provide are likely unique. Results summarized in Table 5 indicate an average difference of 6% in explained variance when using a data set with only targeted participants (as identified through their ISP) versus using all available data. Three of six regression coefficients differed \( (p < .10) \) and, perhaps more strikingly, two substantive conclusions differed when predicting perceptions of workplace stressors. That is, when using only targeted participants, LMX was not a significant predictor of stressors whereas organizational justice was \( (b = -.06, p = .25) \). When using all available data, not only was LMX significantly related to perceptions of workplace stressors \( (b = .05, p = .29) \), but it was related in an opposite direction to what theory would suggest. Furthermore, organizational justice was not related to perceptions of workplace stressors when using all available data. Combined, these findings provide evidence that results and substantive conclusions about relations between constructs differ depending on whether targeted versus nontargeted participants are included in the analyses. These results together with the differences in reliability estimates indicate that the answer to Research Question 3 is also yes, the data provided by targeted and nontargeted participants differ in both reliability and predictive validity.

### Additional Analyses to Assess the Value-Added Contribution of Our Proposed Solution

A legitimate question is whether our methodological innovation to detect false identities provides advantages and a value-added contribution above and beyond data-cleaning approaches already available. In addition to advantages regarding the unobtrusive nature as well as ease of implementation of our proposed solution, it would be useful to know whether potential malicious responders may remain undetected after using currently available data-cleaning and quality-check approaches. For instance, a “professional survey taker” seeking to receive a financial incentive for completing a survey may pay close attention to instructed-response items, similarly worded items, and the amount of time spent on the study (Hauser et al., 2019). As our proposed solution is unknown to even the savviest online participant, and nearly impossible to hide, it may identify nontargeted participants who might otherwise be undetected by currently available methodological tools.

To investigate this possibility empirically, we compared reliability estimates and construct means using ANOVAs across targeted and nontargeted participants after cleaning the full data set using the following five currently available procedures: (a) low within-person reliability, (b) missing instructed-response items, (c) answering open-ended questions, (d) removing duplicate IP addresses, and (e) time-to-complete survey. Results presented in Table 6 indicate that using our proposed solution reduced measurement error by an average of 25% across constructs. Results presented in Table 6 also indicate that substantive conclusions based on data from targeted participants differed significantly from those based on data from nontargeted participants across the substantive variables even after first cleaning the data by using existing methodologies. This was particularly true when using existing methodologies in isolation, but

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Simple Regression Comparison Between All Available Data and Targeted Participants for Leadership and Well-Being Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictor</td>
<td>Full data set (using all available data)</td>
</tr>
<tr>
<td></td>
<td>( R^2 )</td>
</tr>
<tr>
<td>Leader-member exchange</td>
<td>0.17</td>
</tr>
<tr>
<td>Organizational justice</td>
<td>0.5</td>
</tr>
<tr>
<td>Predictor</td>
<td>Criterion: Job satisfaction</td>
</tr>
<tr>
<td></td>
<td>( R^2 )</td>
</tr>
<tr>
<td>Leader-member exchange</td>
<td>0.21</td>
</tr>
<tr>
<td>Organizational justice</td>
<td>0.21</td>
</tr>
<tr>
<td>Predictor</td>
<td>Criterion: Stressors</td>
</tr>
<tr>
<td></td>
<td>( R^2 )</td>
</tr>
<tr>
<td>Leader-member exchange</td>
<td>0.01</td>
</tr>
<tr>
<td>Organizational justice</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. Overall \( N = 986 \); targeted participants \( N = 405 \). Targeted = participants using a consumer-focused internet service provider (ISP). \( z = z \)-scores for differences between regression coefficients across the two data sets. \( b \) is the unstandardized regression coefficient. \( ^* p < .05 \).
Table 6
Analysis of Variance of the Effects of Internet Service Provider (ISP) Information on Reliabilities and Construct Mean Scores After Cleaning Data Using Five Existing Quality-Check Methodologies

| After cleansing based on . . . | LMX | | | | | | | Organizational justice | Conflict | | | | | | | Stressors | Physical (health) symptoms | | | | | | | Conscientiousness | Neuroticism | | | | | | | | | | | (table continues) |
we also found significant differences for physical (health) symptoms \((F(1, 366) = 24.00, p < .001, \eta^2 = .06)\), conscientiousness \((F(1, 366) = 10.07, p = .002, \eta^2 = .03)\), and task performance \((F(1, 366) = 6.55, p = .011, \eta^2 = .02)\) even after using all five methodologies simultaneously (i.e., removing any participant who had low within-person reliability or missed one or both instructed-response question or did not answer at least one open-end question or took less than 3 min to complete the survey or shared an IP address with another participant).

### Discussion

Even before the recent global pandemic put a spotlight on employee health and well-being, scholars in the occupational health psychology field emphasized the important role leaders play in fostering the health and well-being of organizational stakeholders (e.g., Arnold, 2017; Bernerth & Hirschfeld, 2016; Incceoglu et al., 2018; Kaluza et al., 2020). The present special issue builds on this by promoting research aimed at better understanding the relationship between a leader’s behavior and their well-being and the well-being of their followers. At the same time, researchers across diverse fields now rely on web-based samples of participants. But, justifiably, there is skepticism about data quality and the validity of resulting conclusions (Walter et al., 2019). Accordingly, methodologists have responded by offering tools that help address a number of challenges (e.g., Aguinis et al., 2021; Harms & DeSimone, 2015; Lovett et al., 2018; Ward & Meade, 2018). We build upon this research stream by supplementing existing tools to address a particularly pernicious challenge for which no solution has yet been offered: Are participants in a study those whom researchers intended to include, or does the sample perhaps include individuals who are not members of the targeted population?

Our results provide evidence that using IP addresses to gather information on IP threat scores and ISPs offer an effective unobtrusive addition to address the challenge of false identities in web-based data collection in leadership and health/well-being and related domains. IP threat scores predicted within-person reliability and open-ended survey question completion. Targeted users—those hailing from consumer-focused ISPs (e.g., AT&T, Comcast, Verizon)—provided data that were more reliable as assessed by internal consistency reliability estimates and IRT procedures and missed significantly fewer instructed-response items. Within-person reliability and completion of open-ended survey questions were also of higher quality for targeted participants versus proxy-based participants based on their ISPs. Moreover, construct means as well as their factor structure varied depending on whether participants were targeted or nontargeted. Relations between substantive constructs and size of effects also varied depending on participants’ status as targeted versus nontargeted. We similarly found differences between participants even after cleaning the data using five existing procedures. We also found nontrivial differences between targeted and nontargeted participants regarding gender, education, and age. Altogether, we found that an IP analysis provides rich and useful information about web-based study participants and its use results in data of higher quality with better psychometric properties, differences in construct mean scores and factor structures, and differences in substantive relations among leadership and health/well-being constructs when we removed nontargeted participants from the analyses.

### Table 6 (continued)

<table>
<thead>
<tr>
<th>Consistencies</th>
<th>Neuroticism</th>
<th>After Targeted</th>
<th>Nontargeted</th>
<th>After Targeted</th>
<th>Nontargeted</th>
</tr>
</thead>
<tbody>
<tr>
<td>M (SD)</td>
<td>α</td>
<td>F</td>
<td>η²</td>
<td>M (SD)</td>
<td>α</td>
</tr>
<tr>
<td>Low within-person reliability (n = 697; 358, 339)</td>
<td>.82</td>
<td>5.80 (1.95)</td>
<td>.83</td>
<td>4.68 (1.20)</td>
<td>.73</td>
</tr>
<tr>
<td>Missed one or both instructed responses (n = 530, 301, 229)</td>
<td>.86</td>
<td>5.96 (1.28)</td>
<td>.84</td>
<td>4.92 (1.11)</td>
<td>.62</td>
</tr>
<tr>
<td>Those who took less than 3 min to complete the survey</td>
<td>.85</td>
<td>5.73 (1.06)</td>
<td>.84</td>
<td>4.93 (1.30)</td>
<td>.62</td>
</tr>
<tr>
<td>Duplicate IP addresses (n = 579, 358, 221)</td>
<td>.80</td>
<td>3.70 (1.81)</td>
<td>.77</td>
<td>3.31 (1.73)</td>
<td>.62</td>
</tr>
</tbody>
</table>

Note: Targeted = participants using a consumer-focused internet service provider (ISP); Nontargeted = participants using a proxy-based ISP. Sample sizes shown in parentheses. \(p < .05\).
Implications for Leadership and Health/Well-Being Research and Practice

Identifying threats to the validity of web-based data collection, proposing a possible solution, providing evidence that the proposed solution is effective in detecting false identities, and showing the proposed solution offers valuable information above and beyond established approaches have clear implications for substantive leadership and health/well-being research (and occupational health psychology more generally). In addition, our results are particularly relevant for theory because we found differences in both construct means and underlying factor structures when including (or excluding) participants from proxied IP addresses. These differences subsequently resulted in different effect-size estimates linking leadership with health/well-being constructs as well as self-reported task performance. That is, when our proposed solution was not used to remove data from suspicious ISPs, we reached different substantive conclusions about models involving the prediction of (a) individual performance from LMX, physical (health) symptoms, and workplace stressors, and (b) individual well-being (i.e., physical [health] symptoms, job satisfaction, and perceptions of workplace stressors) from LMX and organizational justice. Consequently, conclusions and any subsequent interventions that would be drawn from them differ based on whether all observations, or only those from targeted participants, are included.

Our results also have implications for occupational health psychologists concerned with practical aspects of leadership and health/well-being. Case in point, an organization considering developing a training program around the principles of LMX (e.g., Graen et al., 1982) might wish to investigate the impact of high-quality relationships on aspects of current employees’ well-being. If our solution is not implemented and instead we trust results based on the entire data set, results in Table 5 suggest that LMX harms employees’ perceptions of stressors (i.e., the positive regression coefficient of \( b = .05 \) indicates more stress is associated with higher LMX scores). Organizational leaders would, therefore, be rightly concerned about investing financial resources into any proposed program that improves the relationship between leaders and their direct reports. But, results, substantive conclusions, and implications for practice are just the opposite if steps are taken to ensure that only targeted participants are included in the analyses such that the regression coefficient based on targeted participants only was negative and different in size (i.e., \( z = 2.34, p = .01 \)). In addition, results displayed in Table 4 show that internal reliability estimates of prominent leadership and well-being measures differ depending on who is included in the analyses. Thus, organizational decisions could be improved by our proposed solution. Consider an organization that invests a significant amount of time and resources developing their own organizational health index (see, e.g., McKinsey & Company, 2021), one that includes assessments of organizational leaders and aspects of subjective well-being like job satisfaction. If such an organization failed to screen participants using our proposed methodological innovation, they would likely incorrectly conclude the survey measures do not possess adequate internal reliability and therefore should either be scrapped altogether or refined significantly.

In light of our results regarding construct means, scores on a 360°-feedback survey of leaders or an organizational health check-up might differ depending on who provided the web-based data. For example, if an employee forwards a survey intended for current employees to a colleague who was fired or who voluntarily left an organization, results could easily give a false impression of a leader’s ability or of where the organization’s employees stand on important issues (e.g., see differences in construct means in Table 3). This is especially problematic as surveys on leaders and well-being are increasingly collected worldwide, and those data can have direct implications for leaders—especially in organizations where surveys on leadership have taken the place of systematic performance data and/or traditional 360°-feedback. In sum, our research and proposed method bridge theory and practice, providing researchers a tool to improve the validity of conclusions and allowing practitioners to apply leading-edge methodological tools to their day-to-day operations.

Best-Practice Recommendations for Implementing an IP Address Analysis for Improving the Validity and Usefulness of Web-Based Data Collection

We endorse the continued use of existing methodologies for helping ensure high-quality web-based data with the additional suggestion that researchers and practitioners supplement these efforts with an IP analysis to help identify and include appropriate participants. Figure 1 offers a summary of our recommendations described as sequential steps. We suggest beginning the process by checking for the presence of duplicate IP addresses, which indicate that more than a single survey was completed by someone using the same computer-network configuration or potentially the same proxied or spoofed IP address. It could also mean someone took a survey more than once, perhaps as a means to sway the organization’s opinion of a particular manager or as a means to influence organizational interventions/proposals favored by that specific employee. While either scenario is plausible, sometimes a network uses a single public IP address for multiple devices within its network; when an organization’s network is structured in this way, multiple targeted individuals may appear with the same IP address. For this reason, if the IP address appears multiple times and the researcher did not invite multiple individuals from the same organization or group, we suggest these data be excluded from final analyses. If duplicates exist and the researcher invited multiple individuals from the same organization or group, or they do not know who is in their sample (e.g., they used a snowball recruiting technique), there is a need to follow-up to ascertain whether multiple individuals used the same computer to complete the survey, such as a shared work computer or unit in a computer lab. If attempts to verify the appropriateness of participants sharing IP addresses are unsuccessful, researchers should exclude these data from their analysis.

We offer this recommendation based on the evidence we gathered in follow-up analyses with a subsample of 200 participants by investigating whether participants would respond to a direct email requesting verification information. We recorded whether or not the person replied to our email (0 = no response, 1 = response), and subsequently conducted two separate binary logistic regression analyses with IP threat score and ISP designation as the predictor variable, respectively. A high IP threat score reduced the odds of a participant responding to an email (\( B = -2.06, \text{ Wald statistic} = 25.94, p < .001, \text{ odds ratio} = .13 \), Cox & Snell \( R^2 = .14 \)), and an ISP designation of proxy host also decreased the odds of responding to an email (\( B = -2.29, \text{ Wald statistic} = 21.16, p < .001 \), odds
ratio = .10, Cox & Snell $R^2 = .15$). We argue the likelihood of the data being tainted is higher if the user does not respond to a direct email. If email addresses were not collected, the researcher must use all available information (e.g., how many data-cleaning techniques raise flags) in combination with their best judgment in deciding whether or not to retain or eliminate their data.

After ensuring no duplicates (or following up with any duplicate responses), researchers can employ the IP analysis described in our research. If this analysis indicates a participant’s ISP was consumer-based and their IP threat score was less than .90, researchers should feel confident the participant is a member of the targeted population. If the analysis reveals a proxy-based ISP in combination with an IP threat score of greater than .90, it is likely this respondent is not a member of the targeted population and can, therefore, be removed from final analyses unless a follow-up inquiry provides additional information that justifies their inclusion. If an IP analysis reveals a mixed identity (e.g., a consumer-based ISP with a high IP threat score), researchers should turn to other validity checks for help in making ultimate inclusion or exclusion decisions (similar to best practices in defining and identifying outliers; Aguinis et al., 2013).

A logical conclusion generated from our results is that many nontargeted individuals will likely disguise their true location; it is also likely participants with a high IP threat score or those using a proxy-based ISP may not even know with which location their IP address is associated. Thus, researchers could ask a location demographic question on their survey and compare it with what an IP lookup indicates. When discrepancies exist, or when participants missed other checks such as forced response items, the validity of conclusions would be improved by removing those data.

Finally, as a result of the COVID-19 pandemic that began in 2020, the number of employees working from home has skyrocketed which likely resulted in an increase in the number of individuals using a VPN. Because using a VPN may increase the chance of an individual being classified as nontargeted, we emphasize that our proposed methodology is one of several tools that can help ensure high-quality data. As such, it is important to consider multiple checks prior to removing a participant from a data set as suggested in the decision-making tree displayed in Figure 1.

**Potential Limitations and Future Research Directions**

The use of unobtrusive and unique predictors represents a strength of our research, yet we are nevertheless mindful of certain study features that warrant recognition including using the terms “targeted” and “nontargeted” to distinguish between study participants even though the true identity of participants is not known with complete certainty. Our results provided empirical evidence that using information from IP addresses can help improve the validity of web-based data. However, other techniques such as individualized survey links sent to specific individual email addresses might help clarify participant classification and reduce the rate of proxy-based participants or high IP threat scores.

Second, an additional potential limitation surrounds our predictors of data quality. The second predictor, dichotomized ISPs, involved some degree of subjectivity in that we classified each ISP as either targeted (i.e., consumer-focused) or nontargeted (i.e., proxy-based) based on basic information related to the nature of the host or ISP (i.e., the services provided by the company, be it home or business internet service or proxy hosting). In the case of our study, we have some familiarity with ISPs as our backgrounds

---

6 The .90 value is based on a recommendation from threat prevention experts and is quite conservative. Regardless of the specific threshold used, it is critical that it be reported explicitly so results can be replicated.
include a technology component; that said, the rating of ISPs is something that most occupational health psychology researchers and practitioners could complete equally as well. One reason is that the majority of consumer-focused ISPs such as Verizon and AT&T are easily recognizable. If questions arise, a simple internet search of the ISP typically provides enough insight to rate the company as either consumer-focused or proxy-based. Apart from that, our recommendations do not suggest automatically disregarding data that are provided through a nontargeted proxy-based ISP. Rather, we propose following up more closely with such data (as is similarly recommended for addressing potential outliers; Aguinis et al., 2013). We additionally recognize we used a publicly available IP threat score calculator whose mathematical computations are proprietary. It is not clear the extent to which other commercially available tools correlate with the one used in this study.

Third, as alluded to earlier, some IP addresses are dynamic, meaning the actual numbers that compose the address might change. This feature has no impact on organizational health or leadership assessments that include a single survey at one point in time, but it might have implications for projects that ask individuals to complete multiple surveys (e.g., experience sampling or longitudinal studies). Fortunately, even if an IP address changes from one assessment to the next, the ISP provider (which is one of the two components of our proposed methodology) would not typically change unless the user switched internet providers or used a VPN. If the ISP does change or if the IP threat score changes from one administration to the next, the researcher would have multiple data points to help in their data inclusion versus exclusion decision. Relatedly, if a participant passes traditional validity checks (e.g., forced response questions) and appears to be a targeted participant based on our proposed methodology in an initial survey, then running our analysis on any subsequent surveys would only help corroborate or clarify initial decisions.

Finally, we acknowledge a potential ethical challenge such that researchers and Institutional Review Boards (IRBs) must balance the benefits associated with collecting and using IP addresses with the rights and safety of participants. This is particularly important as authoritative bodies concerned with research involving human participants such as professional organizations (e.g., American Psychological Association) have yet to issue any formal guidelines on how best to address such concerns. Paralleling ethical considerations are legal considerations such that some countries, especially those in the European Union, require formal consent to collect IP addresses (Hildén, 2017). Whereas the United States does not require consent, some may feel it is nevertheless a wise practice. For those hesitant to do so for fear of deterring targeted participants, we offer the suggestion that it might also prevent nontargeted participants from claiming anonymity to confidentiality (Teitcher et al., 2015). Until IRBs or other governing bodies offer more definitive guidelines, this seems like a reasonable compromise as is the suggestion that survey administrators delete IP addresses from all data files after checking for duplicates and using our proposed method.

Conclusion

As the prevalence of web-based data collection is anticipated to continue to increase for the foreseeable future in leadership and health/well-being and other occupational health psychology domains (Bowling & Huang, 2018), there is a need to understand ways in which researchers and practitioners can address challenges that threaten the accuracy of substantive conclusions. Our study provided evidence that unobtrusive information sourced from a web-based participant’s IP address reflects systematic differences in the quality and properties of the resulting data. Properties of measures and constructs including reliability, scale mean scores, and factor structures differed between targeted and nontargeted participants (as identified by their IP addresses). In addition, using targeted versus nontargeted participants resulted in different substantive conclusions about models involving the prediction of individual performance (based on such antecedents as LMX, physical [health] symptoms, and workplace stressors) and different aspects of well-being (based on antecedents such as LMX and organizational justice). We also gathered evidence that our proposed solution for detecting possible false identities is effective—and offers value-added information above and beyond existing methodologies for data cleaning. Because many web-based data collection providers already collect IP addresses and there are publicly available and free IP lookup tools, we hope future web-based data collections will implement and benefit from the sequential steps summarized in Figure 1.

References


Volz, D., & Tau, B. (2020, February 17). Two cyberattacks were linked to one web-services provider. The Wall Street Journal. https://www.wsj.com/articles/two-cyberattacks-were-linked-to-one-web-services-provider-11581944460

Received November 14, 2020
Revision received February 12, 2021
Accepted February 12, 2021