



AMERICAN
PSYCHOLOGICAL
ASSOCIATION

Journal of Family Psychology

Manuscript version of

Using Web-Based Technologies to Increase Reach, Inclusion, and Generalizability in Behavioral Observation Research

Nicholas S. Perry, Timothy J. Sullivan, Karena Leo, David. M. Huebner, K. Daniel O'Leary, Brian R. W. Baucom

Funded by:

- American Psychological Foundation, Division 44
- National Science Foundation
- Stony Brook University, Center for Inclusive Education
- University of Utah

© 2021, American Psychological Association. This manuscript is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final version of record is available via its DOI: <https://dx.doi.org/10.1037/fam0000856>

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Using Web-Based Technologies to Increase Reach, Inclusion, and Generalizability in
Behavioral Observation Research**

Abstract

Social scientists have long utilized observations of human behavior in research designs. For researchers studying couples, observation of romantic partners has led to important discoveries about how such behavior is associated with physical, mental, and family health. Historically, these methods have been used in in-person laboratory paradigms that place notable limitations on reach and inclusion. This has, in turn, restricted the generalizability of such research to couples who may not attend an in-person laboratory assessment. Transferring the observational laboratory into an online format has the potential to expand the capabilities of these methods to include more diverse couples. This paper presents two empirical studies that used online methods to conduct observational behavioral research with sexual and gender minority couples, populations that could be difficult to reach using traditional methods in many places. We demonstrate that we were able to reach, recruit, and enroll diverse couples that more closely resemble the population of same-sex couples in the United States than likely would have been reached in-person. Further, we show that the quality of the observational data collected via the internet allowed for over 94% of collected data to be coded, with acceptable interrater reliabilities and convergent validity. These studies provide a proof-of-concept of online observational methods, accompanied by a tutorial for using such methods. We discuss possible extensions of these online methods, their limitations, and the potential to help further the field of close relationships by reaching more diverse relationships and increasing the generalizability of our research.

Key words: observed behavior, internet, technology, generalizability, methods

Using Web-Based Technologies to Increase Reach, Inclusion, and Generalizability in Behavioral Observation Research Designs

Observational research of family interactions has valuably informed our understanding of support and conflict processes within families, as well as their associations with physical, mental, and family health. However, the methods used to conduct such research place significant restrictions on the types of families who can easily participate and therefore has limited the generalizability of subsequent findings. Translating these traditional laboratory-based methods to remote procedures via technology has the potential to significantly increase the flexibility of such research and make it more inclusive of families who cannot easily be recruited to a laboratory. The current study provides a proof-of-concept of an online observational study and maps out possible methods for conducting observational research through the internet.

The study of human social interaction has long relied on the use of the experimental laboratory. Within research on families, including romantic relationships, direct observation of dyads' and families' communication behavior has been a methodological approach that has provided rich data on how people in close relationships behave with one another (Kerig & Baucom, 2004). Such research is predicated on two assumptions: first, that individual's behavior with each other is operationally reinforced into a cohesive pattern within close relationships, and second, that self-report of these behaviors is only moderately correlated with observer ratings (Heyman, 2001). Observational research among romantic couples has answered several important research questions, including distinguishing couples who are significantly distressed in their relationships from those who are not (Heyman, 2001), how violent versus non-violent couples differ in observed behavior (Holtzworth-Munroe et al., 1998), how couples dealing with health issues interact (Johansen & Cano, 2007), and if there are differences in behavior between

heterosexual and sexual minority couples (Baucom et al., 2010; Gottman et al., 2003; Julien et al., 2003; Roisman et al., 2008).

In typical observational research designs for couples, partners are brought into a research laboratory. After providing informed consent, couples are instructed on the topic and length of the conversation they will be required to have, and that this conversation will be recorded. This paradigm is used to leverage the experimental control afforded by a controlled setting to ensure space for uninterrupted conversation, adequate audiovisual quality for reliable observational coding, and control over conversation topic and thereby enhance internal validity. However, laboratory-based research comes at the cost of reducing the ecological validity of the design (e.g., these couples do not live in a controlled setting where they are videotaped and recorded during an intimate conversation).

Further, the logistical constraints inherent to participation in couples' studies using in-laboratory observational designs may limit research access and inclusion in important ways. Requiring both partners in a relationship to rearrange their schedules to come into a laboratory in an academic environment very likely does not allow certain kinds of couples to participate (e.g., those who are not inclined to participate in academic research, those who face significant life stressors or logistic restraints [e.g., work schedules, childcare needs], or those who are not in reasonable proximity to academic institutions). These barriers may make it particularly difficult to reach minority couples, such as couples of lower socioeconomic status, racial or ethnic minority couples, and sexual and gender minority couples, among others. Indeed, much of the existing literature on couples' romantic relationships consists of samples that are predominately White, middle-class, highly educated, and heterosexual (Hartwell et al., 2017; Karney et al., 2004). Similarly, it has been suggested that study design choices imposed by researchers might

restrict demographic diversity in couples' intervention research (Rogge et al., 2006). Further, some types of couples who may have more difficulty participating in laboratory-based research, such as low-income couples, rate relationship problems differently (Jackson et al., 2016) and can have meaningfully different relationship processes (Ross et al., 2019) from the types of couples that typically participate in research. These findings suggest that the generalizability of some of the most common scientific findings in couples research is likely to be even more limited than is often acknowledged, highlighting a critical need to reach more diverse populations.

Conducting observational research using traditional laboratory-based methods also raises important considerations for the ecological validity of findings. Researchers have argued for the external validity of in-laboratory observational designs based on evidence from studies that show couples' self-rate their conversations in the lab as similar to those at home (Foster et al., 1997) and that couples do not find observation to be intrusive (Jacob et al., 1994). However, findings from other studies raise important questions about this assumption. At-home conversations may *enhance* couples' reactivity to conflict (Gottman & Krokoff, 1989). For example, observed conflict behavior in one study was less negative and more positive in the laboratory than at home (Gottman & Krokoff, 1989). Similarly, Baucom and colleagues (2018) found that physiological reactivity (i.e., heart rate) to conflict was significantly larger at home than in the laboratory and associations between conflict and relationship functioning were stronger at homes versus the laboratory. Together, these findings indicate that, while findings between the laboratory and home are often correlated, the setting of observation appears to add unique variance in behavior. This variance specific to the home setting is likely to be valuable for its ecological validity.

Thus, there is likely two-fold value in moving the observational laboratory to meet participants "where they are" as much as possible. It could enhance sample diversity and allow

researchers to reach more marginalized couples, as well as improve the ecological, real-world validity of the behaviors being observed. Indeed, one alternative used by family researchers who have had a difficult time bringing families into the laboratory is to conduct observational assessments in the home (e.g., Repetti et al., 2015). This method can be more convenient for the couple than in-laboratory assessments and allow researchers to reach more marginalized families (e.g., Conger & Conger, 2002). However, in-home methods generally shift logistic challenges (e.g., travel time) from the participants to the researchers. To date, researchers have not had an alternative to in-person assessments for conducting observational research of couples' communication behavior.

Online research methods offer an affordable, flexible, and wide-reaching strategy for conducting observational research with couples, which could be a valuable alternative to in-person methods. Indeed, this method has been suggested as a viable way of conducting observational research (Atkins & Baucom, 2016), but no empirical examples have demonstrated the feasibility of this approach or provided experience-based guidance in how to do so. The current paper aims to remedy this gap by presenting two empirical examples of online behavioral observation methods. The goals of the current paper are to address: (1) is collecting quality behavioral observation data online possible? and (2) what is the psychometric quality of behavioral data when collected online? We use our two empirical examples to offer evidence for the acceptability, efficiency, feasibility, validity, and effectiveness these methods. We also provide a tutorial in conducting online observed behavior research with romantic couples.

Data come from two similar studies using entirely online methods to recruit, enroll, and collect observational data from sexual and gender minority couples (i.e., lesbian, gay, bisexual, transgender, and queer; LGBTQ). Sexual and gender minority couples are often difficult to reach

due to being a smaller proportion of the population and because stigma-related concerns can discourage participation. As a result, historically, this population had not been well-represented in couples research (Hartwell et al., 2017). Thus, sexual and gender minority couples are a particularly apt group to utilize online methods with, both to reach wider numbers of couples more easily and to increase privacy for a stigmatized population.

The first study (Study 1), conducted by the first author, was the initial implementation of this methodological approach. The second study (Study 2), conducted by the second author, replicates the methods used in Study 1. We discuss both studies in tandem given procedural similarities and space constraints, noting key design differences where appropriate.

Empirical Examples of Online Methods in Observational Research

Both Studies 1 and 2 were cross-sectional online studies where couples completed self-report measures and video-recorded observed communication tasks in a single assessment. Study 1, conducted from March to July 2017, was a study of how minority stress processes are associated with couples' relationship functioning. It recruited and enrolled 60 couples (30 male same-sex couples, 30 female same-sex couples).

Study 2, conducted from August 2019 to March 2020, examined how sexual minority couples cope during discussions of sexual orientation-related discrimination experiences. Eighty-two sexual minority couples were recruited and enrolled. We actively monitored recruitment to ensure enrollment of relatively equal proportions of cisgender men, cisgender women, and gender diverse people (i.e., transgender, non-binary, genderqueer, gender non-conforming, or any other non-cisgender identity). Study procedures were approved by Institutional Review Boards at the University of Utah and Stony Brook University.

Procedures

Both studies recruited couples through passive advertising on social media (e.g., Facebook, Craigslist) and active outreach to LGBTQ-friendly organizations across the United States (e.g., chapters of Parents and Friends of Lesbians and Gays, LGBTQ-friendly adoption agencies). Both studies also posted in-person flyers around the local areas where they were conducted. Advertisements directed individuals to complete online screening questionnaires hosted in Qualtrics. Each partner was compensated \$40 in Study 1 and \$35 in Study 2.

Screening criteria and procedures for Study 1 and Study 2 differed. The screener for Study 1 assessed self and partner's gender identity, relationship length, relationship satisfaction (Funk & Rogge, 2007), and recent, severe physical intimate partner aggression using items from the Conflict Tactics Scales-2 (Straus et al., 1996). Individuals completed the screener and provided their partner's contact information, who was then contacted by the study team to complete their own screening. Couples were considered eligible if they and their partner reported being in a relationship for at least three months, being in a cisgender, same-sex relationship, and if both denied recent, severe intimate partner aggression.

For Study 2, individuals completed the screener and, if determined to be eligible, provided their own contact information and their partner's first and last name. Individuals were then asked to have their partner complete the online screener to determine eligibility. Couples were invited to participate in Study 2 if 1) both partners were between the ages of 18 and 35, 2) both partners identified as sexual minority (i.e., non-heterosexual), 3) both partners reported a relationship length longer than three months, 4) both partners spoke English, 5) both partners had internet access and a webcam/microphone, 6) neither partner endorsed a score over 5 on any item on the Fear of Partner Scale (O'Leary et al., 2013), 7) neither partner endorsed injury during partner conflict in the past six months, and 8) at least one partner had experienced sexual

orientation-related discrimination in the last six months.

Both studies followed a similar protocol for study procedures. Researchers scheduled eligible couples by email for a one-time “online laboratory” appointment. Researchers accommodated participants’ schedules as much as possible, which included completing study appointments in the evenings and on weekends (see Supplemental Figure 1 for data on when study visits were scheduled). Each member of the couple was sent a hyperlink via email to the online survey, hosted in Qualtrics, as well as a link to the meeting in the videoconferencing software, Adobe Connect. Couples were “met” in the Adobe Connect videoconferencing platform by study staff. The first author ran all Study 1 appointments alone, whereas trained research assistants (supervised by the second author) ran Study 2 appointments. Study staff were available in real-time over video conferencing throughout the couple’s entire participation. By using Qualtrics and Adobe Connect software in this manner, both studies were able to simulate an experience over the internet similar to in-person laboratory approaches.

In both studies, study staff first met the couple via video conferencing and oriented them to study procedures, which included ensuring the couple had appropriate privacy during the appointment. Sound and video checks were performed before proceeding with appointments. These were conducted both subjectively by study staff (i.e., could they see and hear the couple clearly), as well as using the sound check feature in Adobe Connect for participants’ microphone. Each member of the couple could view the informed consent in Qualtrics and had it read to them by study staff. Staff answered any questions couples had related to informed consent and offered to meet privately with either member of the couple if they wished. After both partners verbally consented, they then indicated consent electronically in Qualtrics and proceeded with the online survey of self-report measures.

Individual partners were able to separately complete the online surveys as long as they had two internet-capable devices (e.g., laptop, smart phone, or tablet). Prior to beginning the self-report measures, study staff instructed couples to proceed through the questionnaires until the survey told them to stop at the first conversation task. Partners were encouraged to sit apart or in separate rooms to have privacy from one another while completing questionnaires. Study staff informed the couple they would turn off their camera and mute their microphone for privacy and that the couple could do the same if they wished. Study staff remained at their computer with the video conference active so partners could ask questions verbally or through the chat function.

After both partners had completed the self-report measures, the survey would inform partners to come back together (if separated to complete the survey) and tell study staff they were ready to proceed. Study staff would turn on their camera and microphone to orient the couple to the conversation tasks. In Study 1, partners engaged in three 10-minute conversations. They first engaged in a conversation about their early relationship history designed to prompt affection (prompt: “Talk about how you first met, what attracted each of you to the other, and what your early dating life was like.”). Couples then engaged in two conflict discussions, each one focused on one partner’s lowest rated item on the Problem Area Questionnaire (Heavey, Christensen, & Malamuth, 1995) that both partners were willing to discuss.

For each of the 10-minute conversations, the first author would introduce the topic and check if either partner had questions. Couples would be asked to complete the conversations using a single device (e.g., laptop). This decision was made so that partners could be observed conversing in the same, shared space to keep the conversations in-person between partners. This also helped the online laboratory retain the non-verbal cues between partners that might be observed in an in-person laboratory setting. Then, the first author would begin to record video in

Adobe Connect, turn off his own camera and mute his microphone, and set a timer for 10 minutes. The first author would step away from his computer to allow couples privacy, while checking visually on them periodically. After the ten minutes, the first author would turn his camera and microphone on and flag the couple's attention to move to the next part of the study.

During conversation tasks, the first author periodically monitored conversation progress to insure technical or environmental difficulties were not impeding the quality of the audio and video data collection. The first author would only turn on his camera to address the couple during the conversation if audio or video quality deteriorated (e.g., a camera froze or microphone cut-out permanently). The decision made prior to the start of the study was to minimally intervene for interruptions in the environment (e.g., pets, children, etc.) to allow conversations to occur as similarly to everyday life as possible. The only other rule under which the first author would turn on his camera during the conversation to address the couple was if they had completely stopped talking before the time period had ended. If this occurred, the first author would instruct the couple, "Please do your best to continue discussing the topic for the full time."

In Study 2, couples completed three 8-minute discussions at the end of the survey: one sexual orientation-related discrimination stressor discussion, one general life stressor discussion (unrelated to sexual orientation), and one positive discussion. Using a 5-point Likert scale, partners rated up to 5 sexual orientation-related discrimination experiences and 5 general life stressor experiences from the last six months, and rank-ordered these experiences from most to least stressful. Partners were instructed to only list experiences they felt comfortable discussing with their partner. The most stressful discrimination and life stressor experiences endorsed by either partner were chosen as the topics of discussions; couples were told that topics would be chosen randomly between them. For positive discussions, partners rank-ordered up to 5 positive

aspects of their relationship and the topic that was highest rated by both partners was selected. While positive discussions were always conducted last, the order of the discrimination stressor and life stressor discussions was randomized and counterbalanced across couples. Partners were instructed to discuss related topics if they finished discussing the selected topic early.

Video-recording procedures for Study 2 were mostly identical to Study 1, although five couples completed discussions using two different devices (3 were not in the same room, 1 was in the same room but seated separately, 1 was in the same room and not seated separately). There were no interruption rules implemented in Study 2's protocol, though in the event conflict became too intense (e.g., appeared it could escalate to physical violence), standard de-escalation protocols were used to ensure participant safety.¹ In both studies, study staff documented any notable events or protocol deviations during a study visit.

For both studies, each partner completed a post-discussion questionnaire evaluating current mood and qualities of the conversation (see "Effectiveness" section). After all study procedures were concluded, couples were debriefed and paid electronically. Observational coding of the video-recorded conversation tasks was conducted after enrollment was completed. In Study 1, two independent teams of research assistants, supervised by the fourth author, were trained to reliability in the Asymmetrical Behavioral Coding System (Leo et al., in press). For Study 2, two independent teams of research assistants, supervised by the second author, were trained to reliability in the Coding System for Dyadic Coping (Bodenmann, 2000) and the Social Support Interaction Coding System (Bradbury, 1992).

Results

We organize our results to answer key questions that may be posed by researchers

¹ De-escalation was only needed for 1 couple in Study 2 and no couples in Study 1.

interested in utilizing this approach. We describe participant sample composition (i.e., what are the demographic characteristics of the couples reached through this method?), acceptability (i.e., how many couples are interested in enrolling in a study using this method?), efficiency and feasibility (i.e., how long does it take to recruit samples with this method?), validity (i.e., what is the quality of the observational data collected with this method?), and effectiveness (i.e., do conversations using this method evoke anticipated behavioral and emotional responses?).

Sample Composition

Sample demographics from both studies are presented in Table 1. Of note, couples from both studies were recruited from every major US geographic area (Northeast, South, Midwest, Mountain West, West Coast). Couples in Study 1 lived an average of 28 miles ($SD = 45$, range = 0.1 to 288 miles) from the nearest “research intensive” (i.e., R1) university as designated by the Chronicle of Higher Education (Indiana University Center for Postsecondary Research, n.d.).

Although our two empirical studies did not directly compare online data collection to in-person methods for collecting behavioral observation data, we wished to situate our findings in the context of similar prior research with sexual and gender minority populations. Thus, we compared participant demographics from our two studies to 1) laboratory-based studies of sexual and gender minority couples and individuals and 2) demographic estimates of the U.S. population of same-sex couples. First, we compared Study 1 to three samples: (1) an unpublished in-person laboratory-based study of sexual minority couples in Salt Lake City, (2) to an in-person laboratory-based study of sexual minority adults in Salt Lake City (McGarrity et al., 2020), and (3) to an in-person laboratory-based study of sexual minority female couples in Denver (Scott et al., 2019). In comparison to the study of couples in Salt Lake City, Study 1 reached participants who were similar in average age (30 ($SD=10$) vs 29 years old ($SD=8$)),

slightly longer in average relationship length ($M=5$ years ($SD=6$ years) vs $M=3$ years ($SD=3$ years)), slightly lower income (67% $< \$45,000$ vs 42% $< \$50,000$), and more racially/ethnically diverse (70% white vs 84% white). This was similar to the laboratory-based study of sexual minority individuals in Salt Lake City, where the average age was 25 years, with 85% of the sample identifying as white. In the study of sexual minority female couples, couples were older ($M=33$ years old ($SD=9$)) than those in Study 1. The sample of sexual minority female couples also had shorter relationship length ($M=4$ years ($SD=5$)) and more identified as white (73%).

We compared Study 2 to a laboratory-based study of sexual minority adults (Hatzenbuehler & McLaughlin, 2014) in the New York metropolitan area. Sexual and gender minority couples reached in Study 2 were a similar age to those individuals recruited into the lab-based study ($M=25$; $SD=4$) vs $M=24$; $SD=4$), although more sexual and gender minority couples were white than in the laboratory study (62% vs 42%).

Finally, we compared both studies to the demographics of same-sex couples nationally, based on data from the American Community Survey (ACS) (Gates, 2015). The average age of individuals in same-sex relationships in the ACS was 45 years old (for unmarried couples) and 50 (for married couples). The average household income of all same-sex couples in the U.S. was \$89,680. Twenty-three percent of individuals in same-sex couples nationally identified as a racial or ethnic minority.

Taken together, Studies 1 and 2 using online observational methods were able to reach sexual and gender minority couples that were mostly demographically comparable to laboratory-based studies of sexual minority couples and individuals in the local regions where the studies were conducted. Thus, there did not appear to be a noticeable restriction in sample demographics when comparing our online methods to similar prior studies using in-person methods. Notably,

however, Study 1 was able to reach a much more racially and ethnically diverse sample than similar laboratory-based studies in the Mountain West region. In comparison to national data on same-sex couples across the U.S., both studies using online methods more closely approximated the racial and ethnic diversity of same-sex couples in the country, although recruited younger couples than national data suggest is the average.

Acceptability

In Study 1, 541 individuals initiated the screener. After removal of incomplete and duplicate responses, 399 (74% of the total) respondents completed the screener and provided their partner's contact information. 87 individuals (21% of those screened) were ineligible or their partner could not be reached. Of the 156 intact dyads (312 individuals) who completed the screener, twenty-three couples (15% of dyads screened) were ineligible due to past-year severe physical partner aggression. Sixteen couples were otherwise eligible, but declined to schedule (10% of dyads screened). One couple did not have a webcam and declined to have one mailed to them (0.01%). Fifty-six couples were eligible, but placed on a waiting list due to stratified recruitment procedures (36% of eligible dyads). 60 couples (79% of all couples eligible and contacted to participate) completed all study procedures, which was the recruitment goal.

In Study 2, a total of 1,590 complete responses were recorded for the online screener. Three responses were ineligible due to lack of internet access (0.002%) and 50 responses due to lack of appropriate equipment (0.03%). 61 responses were ineligible due to past injury from partner or significant fear of partner (0.04%). Of the 369 intact couples screened, 92 (25% of couples screened) were ineligible based on specific study eligibility criteria, 138 (37% of couples) were waitlisted to ensure sufficient distributions of gender identities, and 139 were eligible for the study (38% of couples screened) and contacted to participate. 84 couples were

enrolled in the study. Two of these couples enrolled, but did not complete all procedures. 82 couples completed all Study 2 procedures (59% of all couples eligible and contacted).

Together, these results across both studies indicate that reasonable numbers of couples appear comfortable with online observational procedures to respond to recruitment efforts, screen for eligibility, and ultimately enroll for participation.

Efficiency

In Study 1, the entire sample ($N=60$ couples) was recruited and completed all study procedures in 16 weeks, with an average of 3.7 appointments completed weekly. Average study completion time was 90 minutes. All study appointments were run solely by the first author. In Study 2 ($N=82$ couples), study completion time was 32 weeks, with an average of 2.67 appointments completed weekly. Average study completion time was 2 hours and appointments were primarily run by 10 undergraduate research assistants. Supplemental Figure 1 displays data on the timing of assessment appointments for both studies. Thus, online methods appear capable of enrolling couples rapidly with relatively low staff demand.

Feasibility

In Study 1, 95% of recordings had adequate audio/video quality to be behaviorally coded. Domains of the behavioral codes also demonstrated good inter-rater reliability using videoconferencing-captured conversations: argumentativeness had an inter-rater reliability ranging from .76-.86, demand had an inter-rater reliability ranging from .92-.93, and withdraw had an inter-rater reliability ranging from .88-.95. In Study 2, 94% of recordings had adequate audio/video quality to be behaviorally coded. Domains of behavioral codes for dyadic coping demonstrated good reliability after consensus was reached across raters: .90 for stress expression frequency, .88 for stress expression severity, .88 for positive dyadic coping, and .89 for negative

dyadic coping. Domains of behavioral codes for social support receipt and provision also demonstrated strong reliability: .94 for positive support provision and .94 for positive support receipt. Audio and video data collected online was of sufficient quality to be behaviorally coded with acceptable reliability by coders.

Effectiveness

Both studies included questions to measure self-reported qualities of the conversations, in addition to observed behavior. Descriptive statistics of scales are displayed in Tables 2 and 3.

Individual Distress. In Study 1, distress during discussions was measured with one item: “My level of distress right now” (baseline) and “My highest level of distress during the conversation” (post-discussion) rated with a Likert scale (1 = *not at all* to 9 = *very much*).

Intimacy. In Study 1, intimacy was an average of five items assessing felt understanding, validation, acceptance, caring, and closeness from one’s partner during conversation using the above 1 to 9 Likert scale. Cronbach’s alpha for the measure ranged .96-.97 across conversations.

Negative Communication. In Study 1, negative communication was an average of five items adapted from the Communications Pattern Questionnaire (Heavey et al., 1995), which used a 1 (*not at all*) to 9 (*very much*) Likert scale. Cronbach’s alpha ranged .61-.83 across conversations.

Affect. In Study 2, participants completed measures of state affect (using the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) after conversations. Participants endorsed how much they were feeling that particular emotion in the moment on a 1 (*not at all*) to 5 (*extremely*) scale. Cronbach’s alpha for positive affect was .91 across conversations and Cronbach’s alpha for negative affect ranged from .90 to .92.

Validity. In Study 2, three items rated using 5-point Likert scales (1 = *not at all*, 5 =

exactly/extremely) assessed similarity, importance, and frequency to typical conversations.

Relationship Satisfaction. Both studies used the 16-item Couples Satisfaction Index to measure global relationship satisfaction (Funk & Rogge, 2007). The CSI is a psychometrically sound measure and widely used in couples research. Psychometric data is provided in Table 3.

Conversations in Study 1 evoked changes in self-reported communication behaviors and emotion for couples, consistent with the content of discussions. Average intimacy increased after the positive conversation and then decreased to below baseline levels following conflict conversations, while distress decreased after the positive conversation and increased after conflict discussions. Negative communication demonstrated a similar pattern.

Conversations in Study 2 also evoked expected changes in self-reported affect. Relative to both stressor discussions, positive affect increased and negative affect decreased after positive discussions. Mean ratings for conversation similarity, importance, and frequency to couples' typical conversations across all discussions were high. These results suggest that couples respond to conversations in their self-reported behavior and affect consistent with the conversation topic.

To provide additional support for the validity of the observed communication data we collected, we also present correlations of coded, observed communication with relationship satisfaction (see Table 3). In Study 1, using behaviors coded during conflict discussions, there were significant between-couple correlations among demand-withdraw behavior and positive behavior with relationship satisfaction. In Study 2, significant correlations were observed between coded communication behaviors and relationship satisfaction in the stressor discussions.

Second, our sample averages in Study 1 of observed demand-withdraw and positive behavior during conflict were within the ranges of the small number of studies of observed communication among sexual minority couples assessing similar observed behavioral constructs.

Baucom and colleagues (Baucom et al., 2010) found demand-withdraw behavior averaged between 2.3-3.6 among male and female sexual minority couples. Scott and colleagues' (Scott, Rhoades, & Markman, 2019) study of sexual minority female couples found that negative communication averaged 1.7 and positive communication averaged 6.3. Although prior research has not evaluated the dyadic coping observational system with sexual minority couples, the frequencies of coping behaviors in Study 2 appear consistent with similar work among heterosexual couples (Kuhn et al., 2017).

This body of results suggest that we can successfully screen and recruit an adequate number of couples over the internet, do so in a period of time comparable to (or perhaps shorter than) in-person studies, collect high quality audio- and video-recordings of couple conversations, and that the conversations evoke expected affective and observed behavioral reactions, which were in line with those observed in prior studies conducted in the laboratory.

Considerations When Conducting Online Observational Research

The studies presented above have outlined the methods used to undertake online observational research with couples over videoconferencing. When using an online approach, there are several important methodological decision points to consider. We highlight several of these decision points below, as well as possible directions to proceed in.

First is the question of which technology to use. Multiple platforms are now widely used for videoconferencing and many people are likely familiar with some of these platforms used socially (e.g., Google Hangouts, Zoom, FaceTime). However, there are both ethical and logistical issues to consider when using such platforms for research (e.g., Atkins & Baucom, 2016). First, is the logistical burden on participants. Many of the more popular platforms are “client-side” or client-focused technologies, requiring research participants to have installed

software or own hardware specific to the platform in order to connect with researchers. However, “server-side” platforms are hosted by the company’s server and place less demand on participants (i.e., typically no special software is required). Such platforms include Adobe Connect, Red5, and mobile applications, such as FaceTime, among others. We should also note that many of these platforms are mobile-friendly and, in our experience, participants have successfully completed observational study procedures using a smartphone or tablet.

An additional consideration for selecting a technological platform is data security. Not all videoconferencing platforms are designed for maximum user security. For research purposes, videoconferencing software should use data encryption for transmission. Researchers should also check in advance where and how recorded videos are stored. For example, recorded videos stored on the company’s cloud server, where they are transferred to the cloud over a secure encrypted connection, backed up, and use access logs, are likely safer than when stored on the researcher’s own device. Researchers can also obtain a business associate agreement from the videoconferencing company to further ensure the company will protect participants’ data.

Another consideration is how to appropriately “set the stage” for remote observational tasks with couples. The first purpose of this is to ensure couples’ comfort and privacy. Researchers should ensure that couples are home alone or in a private location where they feel comfortable speaking openly. Researchers should also make sure they themselves are in a private secure location (e.g., hanging a “do not disturb” sign on the office door, using headphones so others do not overhear the couple). This also includes having the researcher mute their microphone and turn off their camera during participants’ conversations to reduce the sensation of being observed. Making participants feel as comfortable and private as possible will likely maximize their comfort to speak freely and engage comfortably with each other. Indeed, many of

these suggestions are consistent with best practices for the delivery of therapy over the internet (i.e., teletherapy) (e.g., Wrape & McGinn, 2019).

Setting the stage is also critical to ensuring high quality audio and video data are captured. It is advisable to do a “sound check” before starting procedures to make sure each participant can be heard clearly based on their position relative to their device’s microphone. Researchers also want to ask participants to minimize ambient noise (e.g., having participants close the window to a busy street). For video, researchers want to be sure both partners are well-lit (i.e., reducing backlighting or glare) and should encourage them to remain as much in the frame as possible so that each partner is seen clearly. Additionally, some enhancement of video quality can be done after recording (in “post-production”). For example, several freely available programs (e.g., Audacity) can be used to enhance audio quality of recorded videos.

Finally, researchers may need to make decisions regarding managing participant flow through the study. It is not uncommon for problems to arise during a conversation task that might reduce the quality of the observational data. These can be technical issues (e.g., a partner moves out of frame) or distractions in the home environment. In our experience, it was not uncommon for couples to be interrupted mid-conversation by a phone call, a couple’s child, or a roommate.

This consideration returns to the original tension between the control allotted to the researcher in a laboratory versus the ecological validity of allowing couples to participate from the places that are most comfortable, familiar, and private for them. Technical issues (e.g., a microphone cutting out, a camera freezing) may be best managed by immediate intervention by the researcher and “re-starting” the conflict discussion. In contrast, researchers may choose to intervene not at all or minimally for at-home distractions to enhance the ecological validity of the discussion. We would encourage researchers to develop a protocol with their team outlining

circumstances under which they might intervene during a conversation and what the intervention might involve. Further, developing such a protocol with the research team during the design of the study will help ensure this process is systematically built into the online observational procedures. An additional consideration in the protocol is the consideration of how to code observed behavior in the context of the couples' setting. Coding teams should engage in ongoing discussions of how a behavioral code might manifest in a lived environment (e.g., how to account for demand behavior that is interrupted by a phone call) with the team lead. This will help ensure reliability is not negatively impacted by the varied settings couples were recorded in. We ourselves maintained such quality assurance logs for each participant to record any such problems that arose or deviations from the study protocol.

Extension of Methods to Other Designs

Our hope is that these online observational methods might be usefully extended to other types of research designs where direct observation is needed. This approach could easily be adapted to other types of observed analogue tasks with couples or families (e.g., Haynes, 2001) designed to elicit different interactions or behaviors (e.g., joint problem solving). Another valuable extension would be to treatment research. Clinical assessment of couples' relationship distress and observational research paradigms share common roots in the field of relationship science (Weiss & Heyman, 1997). A potentially interesting and beneficial extension of the findings from the current study to intervention research would be to combine online observation of couples' communication with clinical trials methods. Theoretically, this could make a clinical trial using observational assessment of couples' communication possible with an entirely online approach. Further, video-based telehealth has seen a massive and incredibly rapid expansion during the COVID-19 pandemic. This has likely resulted in a paradigm shift for both clinicians

and patients in their comfort and familiarity with mental and physical healthcare delivered via synchronous video telehealth. It is possible that the observational methods we describe could be used to reach couples during times of crisis (e.g., following natural disasters or other emergencies) when it would not be safe or feasible to come to the research laboratory.

Concerns and Limitations

We suggest that the use of online approaches is a useful, but not singular, option for observational research with couples. Indeed, online methods may not be preferred or viable for all couples. We believe that the decision to conduct observational research online versus in-laboratory should be guided by research questions and ethics, as this approach may not be best for all types of couples or for all empirical questions.

Concerns

One important point that bears careful consideration for these methods is that of intimate partner aggression. Among couples where there is severe intimate partner aggression (e.g., kicking, choking, burning), researchers may be understandably concerned that engagement in conflict discussions facilitated online may increase risk to participant safety because researchers are not able to physically intervene if couples' conflict escalates. Some evidence from in-laboratory paradigms suggests that engagement in conflict discussions does not result in undue risk of aggression to participants (Clements & Holtzworth-Munroe, 2009; Owen, Heyman, & Slep, 2006). We also acknowledge that intimate partner aggression occurs along a continuum of severity, with less severe aggression being more typical (Jose & O'Leary, 2009), and is relatively common: physical aggression occurs in approximately 35% of young couples and 50% of distressed couples (Jose & O'Leary, 2009). Thus, in a traditional laboratory setting, exclusion decisions based on aggression should be weighed against the prevalence of intimate partner

aggression within the population of interest.

However, the applicability of laboratory-based findings to at-home conversations among couples with histories of aggression is unknown. Further, due to the screening measures implemented in the two studies described here, we cannot speak to if and how best to conduct online observational assessments of couples who engaged in serious intimate partner aggression. More research will be needed to determine best practices for ensuring safety during conflict discussions (e.g., de-escalation, safety protocols) facilitated online as compared to in-laboratory.

The online laboratory is also not without the limitations that are common to all observational communication tasks (see Heyman, 2001). While our studies presented here provide preliminary evidence that an online approach appears to provide similar quality data to laboratory studies among sexual and gender minorities, future studies are needed to directly compare the validity of this approach against an in-person laboratory design (e.g., through randomization) and to understand participant preferences for method of data collection. Couples' willingness to engage in observational research will likely vary as a function of the data collection method and the couple (e.g., by race, ethnicity, socioeconomic status, etc.). Our intent in this paper is to demonstrate how online methods may be one more valuable tool, in addition to laboratory and at-home methods, for researchers collecting observational data on families.

An important concern also rests with the video quality obtained using online methods. Online approaches for observational designs necessarily rely on participants' technology capabilities (e.g., video/microphone quality, internet speed, user comfort). This may not be ideal for some researchers and some coding systems. For example, in our studies, we utilized global observational coding systems focused primarily on verbal behavior; other coding systems may need higher quality video data (e.g., microanalytic systems or those focused on facial

expressions) that could make using online methods more challenging. However, researchers may find this concern has reduced over time with the rapidly expanded use of video conferencing technology for many people during the COVID-19 pandemic. Further, in all of Study 1's and the vast majority of Study 2's conversations, partners were seated together and recorded with a single device to closely mimic in-laboratory procedures. However, researchers may want to consider having partners record on two devices to expand the video perspective both partners are captured from, which may enhance researchers' ability to conduct finer-grained microanalytic coding with remotely-captured data.

Limitations

Although online approaches may facilitate inclusion of diverse couples that are typically excluded from couples research, the generalizability of results may still be limited. For example, online methods for observational research do not circumvent the inherent logistical challenge of lining up partners' schedules to complete study procedures. While this can be mitigated by offering flexible scheduling options (e.g., nights, weekends) and limiting the time involved with study procedures, it may still be challenging for busy couples to participate.

Some couples may have concerns about data privacy and security when transmitting sensitive discussions over the internet, even with the data protections put in place by researchers. We also acknowledge that some couples may be more difficult to reach via the online observational methods we used. Although internet and smartphone usage continue to increase (Pew, 2019), a digital divide does persist among important populations, including older adults, adults with lower education, and Latinx and Black individuals (Pew, 2019). We also note that, compared to the U.S. population of same-sex couples, our own samples were notably younger. Further, our recruitment strategies included online advertising, which likely selected for couples

comfortable with technology. Despite these limitations, our studies were still able to recruit couples who may have been less likely or less able to attend an in-laboratory study. We base this assertion partially on our finding that many couples lived some distance from the nearest major research universities, as well as anecdotal evidence of couples who might have faced significant other barriers to in-person attendance (e.g., having a recent newborn at home).

Even with the greater opportunity for many types of couples to participate in observational research when conducted online, online methods do not replace the need for other deliberate strategies to ensure diverse and representative samples. We think that the best approach for increasing the generalizability of observational couples studies will be to combine the flexibility of easier participation via the internet with other methods specifically aimed at increasing sample diversity and generalizability (e.g., hiring diverse research teams, utilizing strategic recruitment strategies, asking scientific questions relevant to the community).

Conclusions

The studies presented here provide evidence that online observational research of romantic relationships is acceptable, efficient, feasible, valid, and effective. We also offer a tutorial in our use of these methods to be instructive to other researchers in the field of close relationships. We strongly encourage behavioral scientists to use wider-reaching methods to recruit underrepresented populations. Researchers will need to carefully weigh decisions about online versus in-person laboratory methods. Nevertheless, we urge researchers to seriously consider the benefits of using web-based observational methods, which could allow for reaching more diverse samples. These methods open up valuable opportunities for the field to reach farther and include more diverse relationships in science, particularly for those who might not otherwise come to a university laboratory to participate in such research.

References

- Atkins, D. C., & Baucom, B. R. (2016). Emerging Methodological and Statistical Techniques in Couple Research. *The Oxford handbook of relationship science and couple interventions*, 148.
- Baucom, B. R., McFarland, P. T., & Christensen, A. (2010). Gender, topic, and time in observed demand–withdraw interaction in cross-and same-sex couples. *Journal of Family Psychology*, 24, 233-242.
- Baucom, B. R. W., Baucom, K. J. W., Hogan, J. N., Crenshaw, A. O., Bourne, S. V., Crowell, S. E., . . . Goodwin, M. S. (2018). Cardiovascular reactivity during marital conflict in laboratory and naturalistic settings: differential associations with relationship and individual functioning across contexts. *Family process*, 57, 662-678.
- Bodenmann, G. (2000). *Stress und Coping bei Paaren*. Göttingen: Hogrefe.
- Bradbury, T., Pasch, P.A.,. (1992). *The social support interaction coding system. Unpublished coding manual*. University of California, Los Angeles. Los Angeles, CA.
- Clements, K., & Holtzworth-Munroe, A. (2009). Participation in laboratory research results in emotional distress with one’s partner: Comparing violent and nonviolent couples. *Journal of Family Violence*, 24, 283-295.
- Conger, R. D., & Conger, K. J. (2002). Resilience in Midwestern families: Selected findings from the first decade of a prospective, longitudinal study. *Journal of Marriage and Family*, 64, 361-373.
- Foster, D. A., Caplan, R. D., & Howe, G. W. (1997). Representativeness of observed couple interaction: Couples can tell, and it does make a difference. *Psychological Assessment*, 9, 285-294.

- Funk, J. L., & Rogge, R. D. (2007). Testing the ruler with item response theory: increasing precision of measurement for relationship satisfaction with the Couples Satisfaction Index. *Journal of Family Psychology, 21*, 572-583.
- Gates, G. (2015). *Demographics of Married and Unmarried Same-sex Couples: Analyses of the 2013 American Community Survey*. Los Angeles, CA: Williams Institute.
- Gottman, J. M., & Krokoff, L. (1989). Marital interaction and satisfaction: A longitudinal view. *Journal of Consulting and Clinical Psychology, 57*, 47-52.
- Gottman, J. M., Levenson, R. W., Swanson, C., Swanson, K., Tyson, R., & Yoshimoto, D. (2003). Observing gay, lesbian and heterosexual couples' relationships: Mathematical modeling of conflict interaction. *Journal of Homosexuality, 45*, 65-91.
- Hartwell, E. E., Serovich, J. M., Reed, S. J., Boisvert, D., & Falbo, T. (2017). A systematic review of gay, lesbian, and bisexual research samples in couple and family therapy journals. *Journal of Marital and Family Therapy, 43*, 482-501.
- Hatzenbuehler, M. L., & McLaughlin, K. A. (2014). Structural stigma and hypothalamic–pituitary–adrenocortical axis reactivity in lesbian, gay, and bisexual young adults. *Annals of Behavioral Medicine, 47*, 39-47.
- Haynes, S. N. (2001). Clinical applications of analogue behavioral observation: Dimensions of psychometric evaluation. *Psychological Assessment, 13*, 73-85.
- Heavey, C. L., Christensen, A., & Malamuth, N. M. (1995). The longitudinal impact of demand and withdrawal during marital conflict. *Journal of Consulting and Clinical Psychology, 63*, 797-801.
- Heyman, R. E. (2001). Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. *Psychological Assessment, 13*, 5-35.

- Holtzworth-Munroe, A., Smutzler, N., & Stuart, G. L. (1998). Demand and withdraw communication among couples experiencing husband violence. *Journal of Consulting and Clinical Psychology, 66*, 731-743.
- Indiana University Center for Postsecondary Research. (n.d.). *The Carnegie Classification of Institutions of Higher Education, 2018 edition*. Bloomington, IN: Author.
- Jackson, G. L., Trail, T. E., Kennedy, D. P., Williamson, H. C., Bradbury, T. N., & Karney, B. R. (2016). The salience and severity of relationship problems among low-income couples. *Journal of Family Psychology, 30*, 2-11.
- Jacob, T., Tennenbaum, D., Seilhamer, R. A., Bargiel, K., & Sharon, T. (1994). Reactivity effects during naturalistic observation of distressed and nondistressed families. *Journal of Family Psychology, 8*, 354-363.
- Johansen, A. B., & Cano, A. (2007). A preliminary investigation of affective interaction in chronic pain couples. *Pain, 132*, S86-S95.
- Jose, A., & O'Leary, K. D. (2009). Prevalence of partner aggression in representative and clinic samples. In K. D. O'Leary & E. M. Woodin (Eds.), *Psychological and physical aggression in couples: Causes and interventions* (p. 15–35). American Psychological Association.
- Julien, D., Chartrand, E., Simard, M.-C., Bouthillier, D., & Bégin, J. (2003). Conflict, social support and relationship quality: An observational study of heterosexual, gay male and lesbian couples' communication. *Journal of Family Psychology, 17*, 419-428.
- Karney, B. R., Kreitz, M. A., & Sweeney, K. E. (2004). Obstacles to ethnic diversity in marital research: On the failure of good intentions. *Journal of Social and Personal Relationships, 21*, 509-526.

- Kerig, P. K., & Baucom, D. H. (2004). *Couple observational coding systems*: Taylor & Francis.
- Kuhn, R., Milek, A., Meuwly, N., Bradbury, T. N., & Bodenmann, G. (2017). Zooming in: A microanalysis of couples' dyadic coping conversations after experimentally induced stress. *Journal of Family Psychology*, *31*, 1063-1073.
- Leo, K., Crenshaw, A. O., Hogan, J. N., Baucom, K. J. W., & Baucom, B. R. W. (In press). Affective expression moderates the association between affective experience and demand/withdraw behavior. *Journal of Family Psychology*.
- McGarrity, L. A., Huebner, D. M., Smith, T. W., & Suchy, Y. (2020). Minority stress, emotion regulation, and executive function: An experimental investigation of gay and lesbian adults. *Personality and Social Psychology Bulletin*, *46*, 365-376.
- O'Leary, K. D., Foran, H., & Cohen, S. (2013). Validation of Fear of Partner Scale. *Journal of Marital and Family Therapy*, *39*, 502-514.
- Owen, D. J., Heyman, R. E., & Slep, A. M. S. (2006). The risk of partner aggression research: Impact of laboratory couples conflict protocols on participants. *Violence and victims*, *21*, 483-497.
- Pew Research Center. (2019). Internet/Broadband Fact Sheet. Retrieved at: <https://www.pewresearch.org/internet/fact-sheet/internet-broadband/>.
- Repetti, R. L., Reynolds, B. M., & Sears, M. S. (2015). Families under the microscope: Repeated sampling of perceptions, experiences, biology, and behavior. *Journal of Marriage and Family*, *77*, 126-146.
- Rogge, R. D., Cobb, R. J., Story, L. B., Johnson, M. D., Lawrence, E. E., Rothman, A. D., & Bradbury, T. N. (2006). Recruitment and selection of couples for intervention research: Achieving developmental homogeneity at the cost of demographic diversity. *Journal of*

Consulting and Clinical Psychology, 74, 777-784.

- Roisman, G. I., Clausell, E., Holland, A., Fortuna, K., & Elieff, C. (2008). Adult romantic relationships as contexts of human development: A multimethod comparison of same-sex couples with opposite-sex dating, engaged, and married dyads. *Developmental Psychology*, 44, 91-101.
- Ross, J. M., Karney, B. R., Nguyen, T. P., & Bradbury, T. N. (2019). Communication that is maladaptive for middle-class couples is adaptive for socioeconomically disadvantaged couples. *Journal of Personality and Social Psychology*, 116, 582-597.
- Scott, S. B., Rhoades, G. K., & Markman, H. J. (2019). Observed communication and relationship quality in female same-gender couples. *Couple and Family Psychology: Research and Practice*, 8, 137-151.
- Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised conflict tactics scales (CTS2) development and preliminary psychometric data. *Journal of Family Issues*, 17, 283-316.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.
- Weiss, R. L., & Heyman, R. E. (1997). A clinical-research overview of couples interactions. In W. K. Halford & H. J. Markman (Eds.), *Clinical handbook of marriage and couples interventions* (p. 13-41). John Wiley & Sons Inc.
- Wrape, E. R., & McGinn, M. M. (2019). Clinical and ethical considerations for delivering couple and family therapy via telehealth. *Journal of Marital and Family Therapy*, 45, 296-308.