



Stepping Up: An Evaluation of Social Comparison of Physical Activity During Fitbit Challenges

Jessica A. Nastasi¹ · Erin M. Curry² · Romina E. Martinez² · Danielle Arigo² · Bethany R. Raiff²

Received: 10 May 2021 / Revised: 17 December 2021 / Accepted: 2 February 2022 / Published online: 18 February 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

The current set of within-subject, single-case design studies examined how exposure to social comparison information may impact physical activity in sedentary individuals. In Study 1, participants ($N=6$) were exposed to two Fitbit challenges, one with a physically active confederate and another with a sedentary confederate. Each challenge phase lasted 7 days, during which participants were able to compare their daily steps to the assigned confederate on a ranked leaderboard, received notifications if their cumulative steps were surpassed by the confederate, and a notification indicating if they won at the conclusion of each challenge (i.e., active confederate in challenge one then sedentary confederate in challenge two, or vice versa). Study 2 replicated the procedures used in Study 1 but controlled for the distance between confederate and participant daily steps ($N=4$). In Study 3, participants ($N=4$) were exposed to the same confederate twice to evaluate potential order effects. Results showed that physical activity increased for most participants, but the direction and magnitude of effects differed across participants, challenge type, and order of confederate exposure. The factors producing differential responding to the Fitbit challenges, and the implications for future research on the effects of competition and social comparison on behavior, are discussed.

Keywords Competition · Fitbit · Physical activity · Social comparison

The Physical Activity Guidelines for Americans (PAGA) recommends that adults engage in 10,000 steps per day (US Department of Health & Human Services [USDA], 2018). Walking is a primary method that is recommended for increasing physical activity and improving overall health. In 2018, the USDA reported that only half of Americans met these guidelines, contributing to 10% of premature mortality and \$117 billion dollars in health care costs. A number of barriers may affect compliance with PAGA guidelines, including physical limitations, chronic illness, and socioeconomic status. Thus, efficient and easily accessible interventions for increasing physical activity are needed (Gordon-Larsen et al., 2006; Salmon et al., 2003), and technology-based interventions are of particular interest for the potential to meet this need.

Opportunities for social comparison, or self-evaluation relative to others, may be a powerful component in technology-based interventions for increasing physical activity (Chapman et al., 2016; Zhang et al., 2016; Zuckerman & Gal-Oz, 2014). Specifically, social comparison processes can be activated via exposure to other people's physical activity ranked against one's own (Arigo et al., 2020). Traditionally, the effects of physical activity-based social comparisons have been evaluated using group designs. For example, Chapman et al. (2016) evaluated the use of general feedback on the participant's own performance, compared to the delivery of feedback on how walking performance compared to another participants' performance. Participants who received the latter social comparison feedback ($n=31$) walked an average of 1120 more steps per day than participants in the control group ($n=31$; Chapman et al., 2016). Further investigation is needed to understand how contextual factors associated with instances of social comparison may differentially impact physical activity.

✉ Jessica A. Nastasi
jnastasi@ufl.edu

¹ Department of Psychology, University of Florida, Gainesville, FL 32603, USA

² Department of Psychology, Rowan University, Glassboro, NJ 08028, USA

Social Comparison and Competition

Social comparison may facilitate healthy competition between users of technology-based interventions (Arigo et al., 2020). In the present report, the word “competition” is used as a summary term describing conditions under which the availability of a reward is limited or unavailable to all subjects who respond, depending on individual performance (Schmitt, 1976). A review by Cariveau et al. (2020) noted that differences between competitor performance may serve as an additional cue under lean schedules of reinforcement and that “nearly winning” might produce more persistent responding than conditions where levels of performance are vastly different. In addition, competitive arrangements have been associated with higher rates of responding than cooperative arrangements, but the use of competitive arrangements in applied contexts warrants further research (Cariveau et al., 2020; Schmitt, 1981).

Whether the competitor is above or below their counterpart’s performance may also impact responding. In the framework of social comparison theory, “upward comparison” occurs when a person is exposed to another individual’s performance that is perceived as better than their own, whereas “downward comparison” occurs when a person is exposed to performance that is perceived as worse than their own (Wills, 1981; Wood et al., 1985). Garcia et al. (2013) further defined social comparison in the context of competition as involving a combination of individual and situational factors, with individual factors including variables such as performance relevance (i.e., value of perceived performance in a particular activity) and situational factors such as incentive structure, proximity to a standard, or perceived similarity to another competitor. Garcia et al. (2013) also stated that “comparison concerns” may manifest when one’s level of performance could be surpassed. In the remainder of the current report, we use “social comparison” to describe instances of direct exposure (i.e., behavior) or indirect exposure (i.e., permanent products or some information on competitor performance) to another individual’s performance.

Commercially available wearable accelerometers such as the Fitbit can be used to track daily steps. Several platforms associated with these tools engage social comparison processes through challenges (competitions) and group/dyadic leaderboards that visually rank a user’s steps against that of other users (Arigo et al., 2020). However, further research is needed to establish if, how, and with whom social comparison is effective for increasing physical activity among sedentary adults. Specifically, Arigo and Suls (2018) called for researchers to examine the mechanisms impacting user engagement and behavioral outcomes associated with using applications that employ social comparison features. A better

understanding of social comparison processes may inform the personalization of treatment options across users and social comparison contexts (Arigo & Suls, 2018). Within the Fitbit platform, the Fitbit challenge is a feature that allows users to engage in a competition for the highest step count, measured by the Fitbit accelerometer and displayed via a leaderboard and through notifications. To our knowledge, no other research has been conducted evaluating the impact of exposure to social comparison information during Fitbit challenges on physical activity.

In response to calls for additional work in this area, the current report presents three exploratory studies that use within-subject, single-case experimental designs to examine the effects of engaging in Fitbit challenges on the daily steps of sedentary adults. Single-case research designs utilize repeated measurement to demonstrate experimental control within subjects, in which functional relations are evident if the dependent variable changes across phases (e.g., steps are low during baseline, steps increase during intervention, steps decrease when the intervention is removed, etc.). After a baseline period in which each participant served as their own baseline, each participant was exposed to challenges prompting upward comparison (competitor engaged in higher daily steps) or downward comparison (competitor engaged in lower daily steps).

In Study 1, participants competed with an active confederate (> 10,000 average steps per day) and a sedentary confederate (< 6,000 average steps per day) in two separate challenges, to evaluate potential differences in physical activity response to each type of comparison target. In Study 2, each participant competed with an active confederate and a sedentary confederate in two separate challenges; here, the active confederate’s steps were approximately 3,000 steps above (active confederate) or below (sedentary confederate) the participants’ previous daily steps, to control for distance between competitors. Study 3 was identical to Study 1, except that each participant was exposed to the active confederate or sedentary confederate twice, rather than to one of each, to examine the potential role of order effects.

Method

Study 1

All procedures were reviewed and approved by the Institutional Review Board at the supporting institution. A rigorous within-subject, single-case reversal design with counter-balanced exposure was used to measure each participant’s steps during baseline compared to their own steps during Fitbit challenges (Dallery et al., 2013, 2015; Parker & Vannest, 2012). In other words, daily steps were repeatedly measured

across phases with and without exposure to Fitbit challenges, with some participants exposed to one type of challenge (e.g., active) first and others exposed to the other type of challenge (e.g., sedentary) first. Sample size calculations to reach a particular level of power were not necessary; in single-case designs, the manipulation is evaluated within rather than across subjects, such that each subject serves as their own control (See Dallery et al., 2013 for additional information on single subject research designs and associated analytic approaches). Participants were recruited via email announcements distributed to staff, students, and faculty at a large state university in the Northeastern US. The eligibility criteria included that participants (a) be over 18 years of age, (b) report being sedentary (i.e., not exercising more than once a week), (c) be willing to wear and sync the Fitbit device daily, and (d) own a smartphone with an active data plan for at least 7 weeks. Exclusion criteria included having owned a Fitbit prior to the study or being diagnosed with health conditions that would limit ability to engage in physical activity.

Outcome of Interest: Steps per Day

Participant steps served as the primary outcome and were recorded objectively by the Fitbit device. Data points below 1,000 steps were excluded from analysis because extremely low daily steps (i.e., 500 steps) were frequently correlated with reports of interference with measurement across participants (i.e., forgetting to charge or wear the device). Single-case designs use repeated assessment, which violates many of the assumptions of inferential statistics. For this reason, the primary methods of evaluating the effects of our experimental manipulation were visual analysis and descriptive statistics. Visual analysis involves looking for trends in the data (i.e., steps increasing or decreasing), changes in level, and variability (i.e., bounce) across phases, and prioritizing clinically significant changes within-subject over statistical significance (Dallery et al., 2013).

Tau-U was also calculated to supplement visual analysis and analyze overlap between baseline and challenge phases for each participant while accounting for intervention phase trends and correcting for baseline trends (Parker et al., 2011). Tau-U, denoted by τ , serves as an index of nonoverlapping data points (PND), but does not account for magnitude of effects. Because single-case designs involve repeated measures and the evaluation of behavior across phases, which violate the assumptions of traditional inferential statistics, Tau-U was developed to account for those deviations from traditional group designs and is intended to serve as a proxy for other effect size measures. Values closer to zero indicate a greater amount of overlap and values closer to 1 indicate

the least amount of overlap. A positive value indicates that data points in the treatment condition were higher than baseline, and a negative value indicates that data in the treatment condition were lower. Variance of the sample used for Tau-U calculations, or the square of standard deviation, is denoted by s^2 .

Procedure

Two confederates competed with participants during challenge phases. The active confederate (last author) exceeded 10,000 steps per day on average. The sedentary confederate (first author) engaged in less than 6,000 steps per day on average. Participants exposed to the sedentary confederate during the first challenge competed with the active confederate during the second challenge, and vice versa, to counterbalance order of exposure. Fitbit challenges used in the current study were the 5-day Workweek Hustle (i.e., Monday through Friday with the winner having the most steps during the 5-day period) and the 2-day Weekend Warrior (i.e., Saturday and Sunday, with the winner having the most steps during the 2-day period); thus, each challenge phase included one Workweek Hustle and one Weekend Warrior challenge. Because physical activity levels can differ for adults from weekdays to weekends, we differentiated the Workweek Hustle and Weekend Warrior data points for visual analysis (Miller & Brown, 2004). The Workweek Hustle and Weekend Warrior challenges were identical, other than challenge duration and temporal location (i.e., weekdays versus weekends). Data were collected in two rounds, with Darien (Age: 63, Race/Ethnicity: White/Caucasian), Amy (Age: 28, Race/Ethnicity: White/Caucasian), and Raye (Age: 32, Race/Ethnicity: Black/African American) in the first round and Mina (Age: 45, Race/Ethnicity: White/Caucasian), Lita (Age: 21, Race/Ethnicity: Hispanic/Latine), and Artemis (Age: 75, Race/Ethnicity: White/Caucasian) in the second round (pseudonyms are used to maintain participant anonymity). See Table 1 for additional demographic information for all participants.

Intake

Each participant provided written informed consent and completed a series of questionnaires assessing basic demographic information (e.g., gender, age, ethnicity, race) at the start of the study, before the experiment began. Participants from round one were permitted to keep the Fitbit loaned to them, whereas participants from round 2 were entered into a lottery for \$50 in gift cards of their choosing at the conclusion of the study. Participants were made aware that compensation was in no way tied to their performance during the study.

Table 1 Participant demographics

Participant	Sex/gender	Age	BMI	Race/Ethnicity	Marital status	Education	Employment status	Housing	Income
Study 1									
Darien	Male	63	36.8	White/Caucasian	Married	Associate's degree	Employed for wages	Owned (free of loan)	\$100,000–\$149,999
Mina	Female	45	35.9	White/Caucasian	Married	Doctorate degree	Employed for wages	Owned (free of loan)	\$100,000–\$149,999
Amy	Female	28	30.4	White/Caucasian	Not married	Bachelor's degree	Employed for wages	Owned (free of loan)	\$100,000–\$149,999
Raye	Female	32	32.7	Black/African American	Married	Bachelor's degree	Employed for wages	Rented	\$80,000–\$89,000
Lita	Female	21	42.3	Hispanic/Latin	Single	Some college	Student	Rented	Less than \$10,000
Artemis	Male	75	23.1	White/Caucasian	N/A	Master's degree	Retired	Owned (free of loan)	\$50,000–\$59,999
Study 2									
Lilian	Female	29	52.4	Black/African American, White/Caucasian	Single	Associate's degree	Employed for wages	Rented	Less than \$10,000
Angelica	Female	35	30.7	White/Caucasian	Married	Master's degree	Employed for wages	Owned (mortgage or loan)	\$100,000–\$149,999
Suzie	Female	21	25.2	Asian, White/Caucasian	Long-term relationship	Associate's degree	Employed for wages	Owned (free of loan)	\$90,000–\$99,999
Thomas	Male	59	33.1	White/Caucasian	Married	Some college	Employed for wages	Owned (mortgage or loan)	\$40,000–\$49,000
Stage 3									
Trent	Male	32	35.1	Not specified	Married	Doctorate degree	Employed for wages	Owned (mortgage or loan)	\$50,000–\$59,000
Quinn	Female	59	22.8	White/Caucasian	Married	Bachelor's degree	Employed for wages	Owned (mortgage or loan)	\$100,000–\$149,999
Jane	Female	57	29.9	White/Caucasian	Married	Bachelor's degree	Employed for wages	Owned (mortgage or loan)	\$100,000–\$149,999
Daria	Female	21	26.6	White/Caucasian	Never married	Associate's degree	Student	Rent	\$80,000–\$89,999

N/A indicates items where participants elected not to respond

Each participant received a Fitbit Charge HR device and charger and was trained to use the Fitbit application on their personal smartphone. All participants were instructed to charge, wear, and sync their Fitbit Charge HR device daily. Participants were also informed of physical activity guidelines (i.e., recommended 10,000 steps per day) and briefly discussed common ways for increasing daily steps (i.e., taking the stairs, parking further away from their destination). All researchers were trained on intake procedures to ensure that conversations were consistent across participants. Fitbit profiles were customized to maintain anonymity for the duration of the study (e.g., profile picture set as the university logo for both confederates and actual participants, profile name set as study ID to avoid name-gender association). Information on gender, age,

and Fitbit friends were all set to private. Participants were instructed not to engage in challenges with other Fitbit users during the study and were asked if they engaged in other challenges at the end of the study. In addition, participants accepted friend invitations from an administrative account (i.e., no device paired or observable user history) managed by the first author. The Fitbit application allowed users to observe friend's steps whenever they chose, so an administrative account was used to facilitate challenges with confederates and avoid exposure to confederate steps during return to baseline or subsequent challenge phases (i.e., observing the active confederates' steps by checking their profile while in a challenge with the sedentary confederate, or seeing confederates' steps during return to baseline).

Baseline

The duration of baseline for each participant lasted up to two weeks, depending on stability of the data. The stability criteria required that the final data point during baseline (i.e., Sunday) be no more than 10% from the mean of the last three consecutive data points. If the stability criteria were not met after two weeks, challenge phases began due to time limitations.

Active Challenge Phase

Participants began a 5-day Workweek Hustle followed by a 2-day Weekend Warrior challenge against the active confederate. Challenge invitations were sent by the first author to the participant and confederate to facilitate both challenges, accompanied by a brief email explaining that the challenge would soon begin. The first author was removed from the challenge after both users accepted the invitation, and prior to the beginning of the challenge, to control for potential reactivity to researcher presence in the challenge. The Fitbit device automatically uploaded step data upon opening the Fitbit application, so participants could observe the confederate's daily and cumulative steps taken and their own rank on the leaderboard. The active confederate engaged in 10,000 steps or higher per day during the challenges ($M = 13,290$; range: 9,929–18,123), thus participants could win the Workweek Hustle or Weekend Warrior if they had a higher cumulative step count than the active confederate. Participants could also see the confederate's weekly average based on the previous week, and received periodic, automatic push notifications triggered by changes in the competitor's step count relative to the user's rank on the leaderboard (e.g., "Look out, user two is catching up with 5,000 steps!"). Users that engaged in the highest cumulative step count (i.e., steps accumulated across the entire challenge duration) also received a celebratory notification from the Fitbit application indicating that they won the challenge upon challenge completion.

Return to Baseline

Participants returned to the previous baseline conditions for 7 days. No challenge was in place, but participants were required to wear the Fitbit daily. Participants were unable to observe confederate steps during this phase.

Sedentary Challenge Phase

The sedentary challenge phase was identical to the active challenge phase, but the participants were paired with the sedentary confederate. The sedentary confederate engaged in less than 6,000 steps per day on average during challenges

($M = 2,812$, range: 896 to 7,294), thus participants could win either challenge if they had a higher cumulative step count than the sedentary confederate.

Exit Interview

Participants completed a Treatment Acceptability Questionnaire which included opened-ended questions allowing participants to provide feedback on the challenges and whether they would participate again, then were compensated with the Fitbit device (round one) or entered into a \$50 gift card lottery (round two) and thanked for their participation. The first author debriefed participants regarding the use of confederates after data collection was complete.

Study 2

Study 2 was identical to Study 1, except that the confederate's steps were approximately 3,000 steps above (active confederate) or below (sedentary confederate) the participants' previous daily steps during challenge phases. Previous research indicates that responding under competitive arrangements may differ according to the difference in competitor performance levels (Muller & Fayant, 2010; Schmitt, 2000). Thus, a 3,000 step difference was utilized to ensure a smaller distance between participant and confederate performance than in Study 1. Three participants, Lilian (Age: 29, Race/Ethnicity: Black/African American), Angelica (Age: 35, Race/Ethnicity: White/Caucasian), and Suzie (Age: 21, Race/Ethnicity: Asian, White/Caucasian), were each exposed to the sedentary challenge first and one participant, Thomas (Age: 59, Race/Ethnicity: White/Caucasian), was exposed to the active confederate first. Efforts were made to recruit more participants who would be exposed to the active confederate first, but recruitment was terminated due to the onset of the COVID-19 pandemic. One participant (Suzie) was exposed to three challenge phases after experiencing technical difficulties during her active challenge phase. Specifically, Suzie's Fitbit failed to sync to her mobile phone frequently, which affected the active confederate's ability to base their daily steps on the participant's last synced day. For this reason, Suzie requested to re-do the challenge phase (active) and those data were included for analysis.

Participant steps were checked by the investigator each morning at approximately 7:00 AM. The investigator logged into the participants' Fitbit account and exported the data from the previous day. Confederates' steps were, on average, approximately 3,000 steps above (active; $M = 2,828.62$, range: 896–7,294) or below (sedentary; $M = 3,366.81$, range: 687–8,543) the participants' steps from the previous day. The second author served as both the active and sedentary confederate, wearing two different Fitbits (i.e., removed one

Fitbit upon reaching \times steps above for the active confederate or below for the sedentary confederate).

Study 3

Jane (Age: 57, Race/Ethnicity: White/Caucasian) and Daria (Age: 21, Race/Ethnicity: White/Caucasian) competed the sedentary confederate and Trent (Age: 32, Race/Ethnicity: Not specified) and Quinn (Age: 59, Race/Ethnicity: White/Caucasian) competed with the active confederate for both challenge phases to evaluate the impact of repeated exposure to similar confederate performance and further examine potential order effects observed in Studies 1 and 2. Confederate usernames were changed between the first and second challenge to make it appear as though participants were competing against two different individuals with similar levels of physical activity, to closely match the experience of participants in Studies 1 and 2. All other study procedures were identical to Study 1, with the exception that the confederate in this case was the third author.

Results

Study 1

Darien and Mina participated in the active challenge first and the sedentary challenge second. Amy, Raye, Lita, and Artemis participated in the sedentary challenge first and the active challenge second. The battery on Lita's Fitbit died during the final 2 days of baseline and she reported losing the Fitbit charger on the first return to baseline day, limiting data collection to the final 3 days of that phase. One data point was omitted for Artemis during return to baseline because his Fitbit did not reset from the prior day (i.e., the Fitbit application incorrectly counted steps from Monday into Tuesday for Tuesday's daily count) after he synced at midnight.

Daily steps for all participants in Study 1, and Tau-U scores across phases, are displayed in Fig. 1. Within-person variability in daily step counts was high across participants. Of particular note, Darien and Mina showed an increasing trend during the active challenge phase and engaged in their highest daily steps during that condition, followed by a decreasing trend in no challenge for Darien. There was another increase in steps during the sedentary challenge phase for Darien but a lower level relative to the active challenge phase. Most participants (Amy, Raye, Lita) exposed to the sedentary challenge first also showed higher levels of steps compared to their own baseline steps, but the difference in level was smaller compared to those exposed to the active challenge first. Together, these findings show that participants exposed to the active confederate first

displayed a larger increase in steps across challenge phases one (median step difference range: 1625–7511; Tau-U range: 0.755–1) and two (median step difference range: 774–2557; Tau-U range: 0.633–0.694) relative to baseline and return to baseline (no challenge) phases, compared to participants exposed to the sedentary confederate first in challenge one (median step difference range: 143–1731; Tau-U range: –0.607–0.270) and two (median step difference range: –295–3273; Tau-U range: –0.383 to –0.122).

TAQ results indicated that the large distance between participant and confederate performance levels were apparent to Amy, Raye, and Darien, which they perceived as unlikely and lacked believability. All but one participant (Amy) said they would engage in a Fitbit challenge again in the future.

Study 2

Thomas competed in the active challenge first. Angelica, Lilian, and Susie competed in the sedentary challenge first. For Angelica, one data point from the sedentary challenge phase (day 13 = 773), one data point from return to baseline (day 21 = 392). One data point from Suzie's second active challenge phase (day 42 = 0) was omitted from analyses for meeting exclusion criteria. Daily steps for all participants in Study 2, as well as Tau-U scores across phases, are displayed in Fig. 2. As in Study 1, the participant exposed to the active confederate first displayed a larger percent increase in median daily steps from baseline to challenge one (median step difference = 2,805; Tau-U = 0.755) and return to baseline to challenge two (median step difference = 1,158; Tau-U = 0.633) than those exposed to the sedentary confederate first in challenge one (median step difference range: –2,178 – 135; Tau-U range: –0.257 – 0) and the active confederate in challenge two (median step difference range: –2,071 – 1,827; Tau-U range: –0.510 – 0.2381). TAQ results indicated that three out of four participants (i.e., Angelica, Suzie, and Thomas) reported that they would participate in the study again.

Study 3

Trent and Quinn competed with the active confederate for both challenge phases and Jane and Daria competed with the sedentary confederate for both challenge phases. For Trent, one baseline data point (day 1 = 840) was omitted from analyses for meeting exclusion criteria. For Quinn, one return to baseline data point (day 15 = 711), and two data points from the second active challenge phase (day 26 = 817, day 27 = 800) were omitted from analysis for meeting exclusion criteria. Jane and Daria competed with a sedentary confederate for both challenge phases. Two of Daria's data points from the second challenge phase (day 27 = 177, day 28 = 0) were omitted from analyses for meeting exclusion criteria.

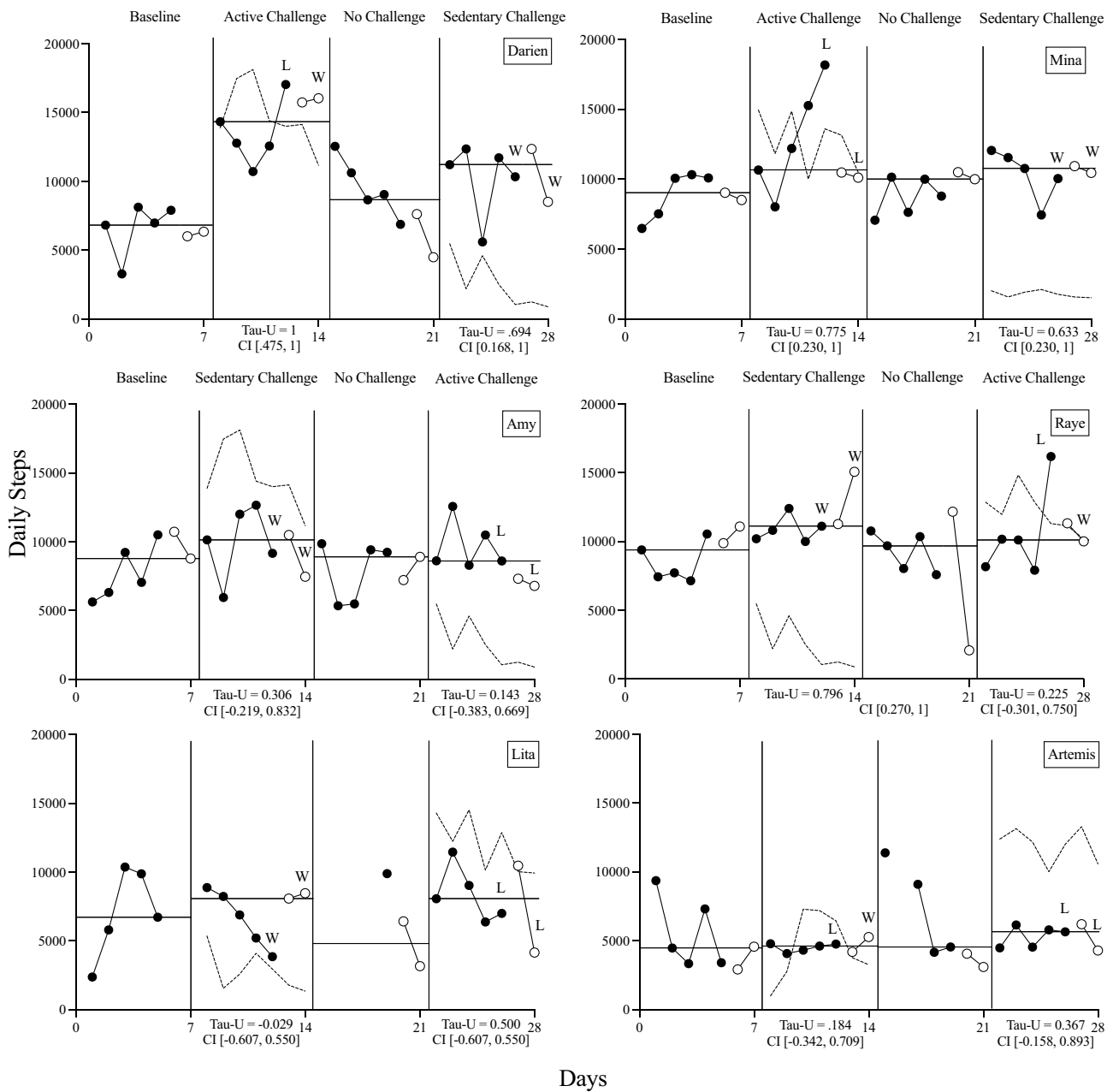


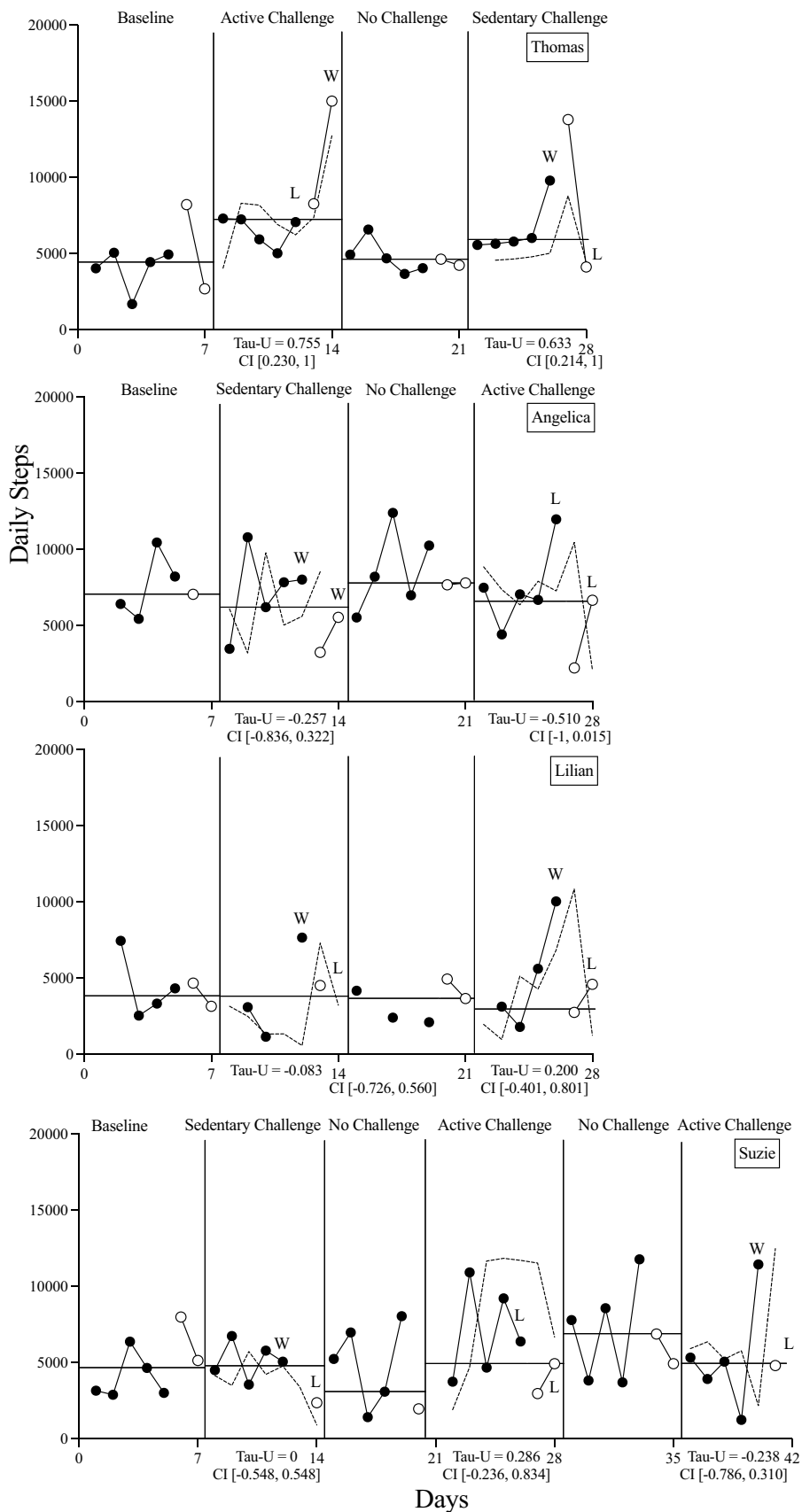
Fig. 1 Study 1 participant daily steps. *Note:* Dashed data paths indicate confederate daily steps. Closed circles indicate weekdays and open circles indicate weekends. “L” indicates when participants received a losing notification at the end of the challenge, and “W”

indicates when they received a winning notification at the end of the challenge. Horizontal lines indicate median participant steps per phase. Confidence intervals are reported at 90%

Daily steps for all participants in Study 3, as well as Tau-U scores across phases, are displayed in Fig. 3. Results were variable across participants despite controlling for order of exposure to confederates. Participants exposed to the active confederate twice showed little to no change in daily steps from baseline to challenge one (median step difference range: -397 – 70; Tau-U range: 0.095 to -0.225) and little to no change from return to baseline to challenge two for one participant (median step

difference range: 416 – 474; Tau-U range: 0.061 – 0.567). Participants exposed to the sedentary confederate twice also showed little to no changes from baseline to challenge one (median step difference range: -481 – 229; Tau-U range: 0.061 – 0.184) and some change from return to baseline to challenge two (median step difference range: 1,090 – 2,919; Tau-U range: 0.674 – 0.829). TAQ results indicated that Jane and Daria would participate in a challenge again, but Trent and Quinn would not.

Fig. 2 Study 2 daily steps. *Note:* Dashed data paths indicate confederate daily steps. Closed circles indicate weekdays and open circles indicate weekends. “L” indicates when participants received a losing notification at the end of the challenge, and “W” indicates when they received a winning notification at the end of the challenge. Horizontal lines indicate median participant steps per phase. Confidence intervals are reported at 90%



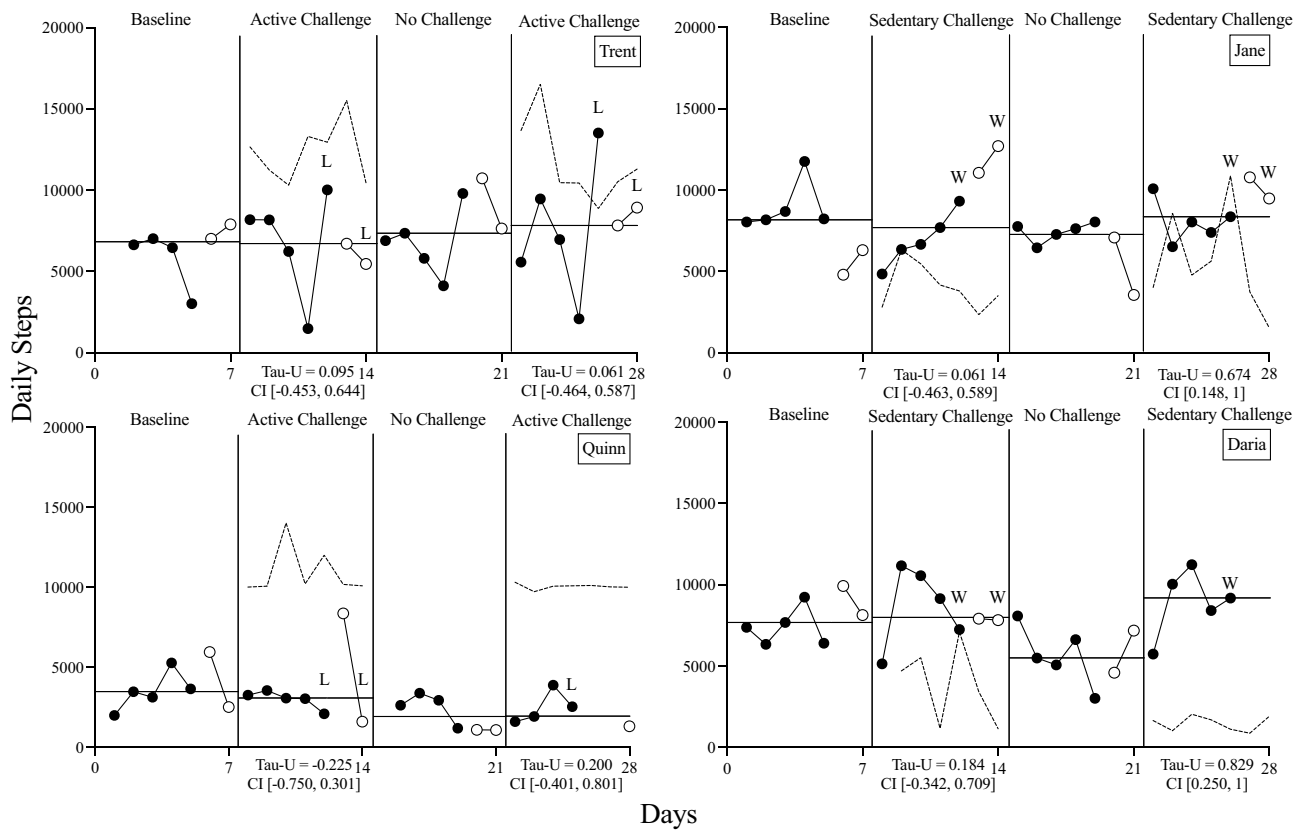


Fig. 3 Study 3 daily steps. *Note:* Dashed data paths indicate confederate daily steps. Closed circles indicate weekdays and open circles indicate weekends. L indicates when participants received a losing notification at the end of the challenge, and W indicates when they

received a winning notification at the end of the challenge. Horizontal lines indicate median participant steps per phase. Confidence intervals are reported at 90%

Discussion

The present studies were designed to evaluate change in an individual's physical activity when exposed to a confederate's physical activity progress (i.e., social comparison information in the context of a competition) in a Fitbit challenge. In Study 1, all participants (Darien, Mina, Amy, Raye, Lita, Artemis) displayed some percent increase in median daily steps from baseline to the challenge phases, although the magnitude of change and the variability of steps within phases varied greatly across participants. In Study 2, Thomas displayed some percent increase in median daily steps from baseline to challenge phases, but Lilian, Suzie, and Angelica did not. Participants exposed to the active confederate first in Studies 1 and 2 showed the largest percent increase in median daily steps taken. In Study 3, Trent, Jane, and Daria displayed increases in median daily steps from baseline to the final challenge phase, but Quinn did not. Thus, overall, the Fitbit challenge did appear to have some impact on individuals' steps, but the direction and magnitude of those effects differed across individuals, challenge type, and order of exposure.

Individual and situational factors impacting responding to social comparison may have contributed to the variability in effects observed across participants (Garcia et al., 2013; Gibbons & Buunk, 1999; Suls et al., 2002). For example, perceived similarity (e.g., age, gender), psychological closeness (i.e., perceiving a connection with the other person, such as a friend or family member), and perceived ability for the performance in question may impact outcomes (Gerber et al., 2018; Suls et al., 2002). The number of competitors in a study may also differentially impact responding (Garcia & Tor, 2009). In all three studies, participants competed with one other person, but Fitbit challenges can include up to 10 users; thus, future research may manipulate the number of competitors. Furthermore, the value of winning a challenge likely differs according to the importance of physical activity as a dimension of performance and the individual's learning history (Garcia et al., 2013; Locke, 2007).

Social comparison theory suggests that individuals may respond differently when exposed to performance that is relatively better (upward comparison) than when exposed to performance that is relatively worse (downward comparison; Buunk et al., 1990), and makes the distinction

between the reinforcing or punishing effects of that exposure based on comparison direction and the type of responding observed (Garcia et al., 2013). A variety of behaviors may produce avoidance of aversive comparison concerns in the context of a Fitbit challenge. By increasing daily step count beyond the active confederate's steps, aversive comparison concerns may have been avoided while simultaneously achieving winning outcomes such as those discussed by Zenger et al. (2017).

For example, in Study 1, TAQ results showed that Mina walked further after seeing she was behind the active confederate and reported being encouraged by being ahead during their challenge with the sedentary confederate. In Study 2, Angelica reported that she was determined to “come out on top” during the active challenge. Alternatively, exposure to high levels of performance may have a punishing effect, where an individual may choose to stop participating entirely. For example, in Study 1, Amy reported that she saw the active confederate's steps as unbelievably high on the TAQ, and displayed a decreasing trend following exposure to the active confederate during her second challenge. In Study 3, Quinn reported that there was no way she could compete with the active confederate because her steps were so low on the TAQ. These results suggest that individuals may respond differently to distinct levels of performance as one of multiple variables impacting participant steps in a Fitbit challenge.

It is unclear why participants in Studies 1 and 2 who were exposed to the active challenge first displayed a larger increase in steps, compared to those exposed to the sedentary challenge first. In Study 1, Darien displayed the largest percent increase in steps, but also had a low baseline average. In Study 2, only one participant, Thomas, was exposed to the active confederate first, but he also showed the highest percent increase in steps during the active challenge phase compared to the other three participants. Several factors could have contributed to the larger increase in steps for the active first group. One possible explanation for the order effects observed is that participants exposed to the sedentary challenge first developed a lower standard for performance than those exposed to the active confederate first. In other words, the criteria for winning or avoiding losing were vastly different during the sedentary challenge compared to the active challenge. Thus, the ratio requirement greatly decreased for the active first group when moving to the second challenge, whereas the ratio requirement greatly increased for the sedentary first group. From a social comparison perspective, having the initial sense that one is already “doing well” (i.e., better than others) with respect to steps may have lowered the perceived necessity or value of efforts to change behavior (Wills, 1981), which carried over to the active phase.

In Study 1, the abrupt increase in criteria for accessing reinforcement (i.e., the large distance between the participant and the confederate) may have been too large for participants exposed to the sedentary confederate during the first challenge. In Study 2, the distance between confederate and participant performance levels was set to approximately 3,000 steps above or below, but participants exposed to the sedentary challenge first (Lilian, Angelica, and Suzie) still displayed a smaller increase in steps from baseline to the final treatment phase than Thomas. Future research might involve a parametric analysis of distance to evaluate whether repeated exposure to varying levels of participant performance (e.g., 1,000 steps above vs. 3,000 steps above) differentially impacts the participant's steps. In Study 3, participants were exposed to the same challenge twice (active-active or sedentary-sedentary). There was no reliable difference in steps as a function of the challenges for Trent and Quinn, even though they were also exposed to the active challenge first. Further research is needed to determine whether these observed differences were unique to this study or whether they would be replicated.

Limitations

Some limitations to the current series of studies are worth noting. First, although most participants reported checking the leaderboard at least once per day during challenge phases, quantitative data were not collected on syncing frequency, frequency of checking the leaderboard, or the exact time and frequency of syncing by confederates. Although participants were not explicitly provided with a personal goal of 10,000 steps per day, they were made aware of the current physical activity guidelines during consent procedures; thus, it is possible that these guidelines served as a standard for comparison in combination with, or in place of, confederate steps. It is also unclear whether participants attended to the automatic push notifications sent through the application.

In addition, our inclusion criteria required participants to report being sedentary; thus, it is possible that participants reported being sedentary while still regularly engaging in medium to high levels of physical activity. We also excluded people who already owned a Fitbit, but did not inquire about other step measuring device, and there may have been some varying levels of experience with step measuring devices; however, any effect of those experiences would have held constant because of the within-subject nature of the design. There are also limitations associated with any intervention using a Fitbit device and measuring across an extended period of time. Users could forget to charge or put on their watch. Users could also attempt to increase their step count without walking (e.g., shaking wrist repeatedly could increase steps). We also did not measure daily steps beyond

the 7-day period, so the long-term effects of engaging in Fitbit Challenges are unknown.

There are some limitations specific to Study 2. First, it was difficult to control for “real time” similarity between participant and confederate steps. For the purposes of this study, the independent variable was approximately 3,000 steps above or below the participant’s step count on the previous day. However, it was difficult to ensure that the previous day’s step count would be similar to the actual number of steps the participant would take that day. This often resulted in a large range of differences between participant and confederate steps. The 3,000 step distance was used to increase believability, but it is possible that a larger difference in steps may have produced different outcomes. In addition, three of the four participants were exposed to the sedentary confederate during the first challenge phase and the active confederate during the second challenge phase; only Thomas was exposed to the active confederate first. We were also unable to recruit additional participants due to the onset of the COVID-19 pandemic.

Furthermore, there is a certain amount of natural variability that occurs when measuring daily steps, which poses some unique considerations for the analysis and presentation of data when steps are the primary dependent variable (Valbuena et al., 2017). In the current study, we provided mean lines and Tau-U calculations to supplement visual analysis, but it is difficult to determine all sources of variability and change in daily steps given the present arrangement. Future research might ask participants to keep a daily log of activities to capture additional influences on daily steps, to provide further insight into the additional factors that impact changes in daily steps. These may include the number of competitors in a challenge, familiarity with other competitors, frequency of exposure to challenge-related stimuli (e.g., checking the leaderboard), and the value of winning per participant, as discussed.

Conclusion

The current study offers a useful starting point for research evaluating the impact of introducing exposure to social comparison information on individuals’ physical activity. Competitive arrangements such as Fitbit challenges may be an effective method for increasing daily steps taken for some individuals, but these effects are likely to vary depending on the learning history of the participant and different contexts present during challenges. Future research should seek to isolate the environmental variables that establish winning as valuable to maximize the efficacy of interventions using competition or social comparison and identify the circumstances under which competition or exposure to social comparison information may or may not be an appropriate or effective intervention across individuals.

Funding Support for the fourth author’s time was provided by the US National Institutes of Health/National Heart, Lung, and Blood Institute (K23 HL136657; PI: Arigo).

Declarations

Informed Consent Informed consent was obtained from all individual participants included in the study. All participants consented to the publication of their data.

Conflict of Interest The authors have no financial interests or non-financial interests to disclose.

References

- Arigo, D., Brown, M. M., Pasko, K., & Suls, J. (2020). Social comparison features in physical activity promotion apps: Scoping meta-review. *Journal of Medical Internet Research*, 22(3). <https://doi.org/10.2196/15642>
- Arigo, D., & Suls, J. M. (2018). Smartphone apps providing social comparison for health behavior change: A need for better tailoring to person and context. *Mhealth*, 4. <https://doi.org/10.21037/mhealth.2018.09.10>
- Buunk, B. P., Collins, R. L., Taylor, S. E., VanYperen, N. W., & Dakof, G. A. (1990). The affective consequences of social comparison: Either direction has its ups and downs. *Journal of Personality and Social Psychology*, 59(6), 1238–1249. <https://doi.org/10.1037/0022-3514.59.6.1238>
- Chapman, G. B., Colby, H., Convery, K., & Coups, E. J. (2016). Goals and social comparisons promote walking behavior. *Medical Decision Making*, 36(4), 472–478. <https://doi.org/10.1177/0272989X15592156>
- Cariveau, T., Muething, C. S., & Trapp, W. (2020). Interpersonal and group contingencies. *Perspectives on Behavioral Science*, 43, 115–135. <https://doi.org/10.1007/s40614-020-00245-z>
- Dallery, J., Cassidy, R. N., & Raiff, B. R. (2013). Single-case experimental designs to evaluate novel technology-based health interventions. *Journal of Medical Internet Research*, 15(2), e22. <https://doi.org/10.2196/jmir.2227>
- Dallery, J., Kurti, A., & Erb, P. (2015). A new frontier: Integrating behavioral and digital technology to promote health behavior. *The Behavior Analyst*, 38(1), 19–49. <https://doi.org/10.1007/s40614-014-0017-y>
- Garcia, S. M., & Tor, A. (2009). The N-effect: More competitors, less competition. *Psychological Science*, 20(7), 871–877. <https://doi.org/10.1111/j.1467-9280.2009.02385.x>
- Garcia, S. M., Tor, A., & Schiff, T. M. (2013). The psychology of competition: A social comparison perspective. *Perspectives on Psychological Science*, 8(6), 634–650. <https://doi.org/10.1177/1745691613504114>
- Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin*, 144(2), 177–197. <https://doi.org/10.1037/bul0000127>
- Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology*, 76(1), 129–142. <https://doi.org/10.1007/s11205-012-0227-1>
- Gordon-Larsen, P., Nelson, M. C., Page, P., & Popkin, B. M. (2006). Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics*, 117(2), 417–424. <https://doi.org/10.1542/peds.2005-0058>

- Locke, K. D. (2007). Personalized and generalized comparisons: Causes and consequences of variations in the focus of social comparisons. *Personality & Social Psychology Bulletin*, 33(2), 213–225. <https://doi.org/10.1177/0146167206293492>
- Miller, R., & Brown, W. (2004). Steps and sitting in a working population. *International Journal of Behavioral Medicine*, 11(4), 219–224. https://doi.org/10.1207/s15327558ijbm1104_5
- Muller, D., & Fayant, M. P. (2010). On being exposed to superior others: Consequences of self-threatening upward social comparisons. *Social and Personality Psychology Compass*, 4(8), 621–634. <https://doi.org/10.1111/j.1751-9004.2010.00279.x>
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, 42(2), 284–299. <https://doi.org/10.1016/j.beth.2010.08.006>
- Parker, R. I., & Vannest, K. J. (2012). Bottom-up analysis of single-case research designs. *Journal of Behavioral Education*, 21(3), 254–265. <https://doi.org/10.1007/s10864-012-9153-1>
- Salmon, J., Owen, N., Crawford, D., Bauman, A., & Sallis, J. F. (2003). Physical activity and sedentary behavior: A population-based study of barriers, enjoyment, and preference. *Health Psychology*, 22(2), 178–188. <https://doi.org/10.1037/0278-6133.22.2.178>
- Schmitt, D. R. (1976). Some conditions affecting the choice to cooperate or compete. *Journal of the Experimental Analysis of Behavior*, 25(165), 178. <https://doi.org/10.1901/jeab.1976.25-165>
- Schmitt, D. R. (1981). Performance under cooperation or competition. *American Behavioral Scientist*, 24(5), 649–679. <https://doi.org/10.1177/000276428102400505>
- Schmitt, D. R. (2000). Effects of competitive reward distribution on auditing and competitive responding. *Journal of the Experimental Analysis of Behavior*, 74(1), 115–125. <https://doi.org/10.1901/jeab.2000.74-115>
- Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect? *Current Directions in Psychological Science*, 11(5), 159–163. <https://doi.org/10.1111/1467-8721.00191>
- U.S. Department of Health and Human Services. (2018). *Physical Activity Guidelines for Americans, 2nd edition*. Washington, DC: U.S. <https://www.hhs.gov/fitness/be-active/physical-activity-guidelines-for-americans/index.html>
- Valbuena, D., Miller, B. G., Samaha, A. L., & Miltenberger, R. G. (2017). Data presentation options to manage variability in physical activity research. *Journal of Applied Behavior Analysis*, 50, 622–640. <https://doi.org/10.1002/jaba.397>
- Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, 90(2), 245–271. <https://doi.org/10.1037/0033-2909.90.2.245>
- Wood, J. V., Taylor, S. E., & Lichtman, R. R. (1985). Social comparison in adjustment to breast cancer. *Journal of Personality and Social Psychology*, 49(5), 1169–1183. <https://doi.org/10.1037//0022-3514.49.5.1169>
- Zhang, J., Brackbill, D., Yang, S., Becker, J., Herbert, N., & Centola, D. (2016). Support or competition? How online social networks increase physical activity: A randomized controlled trial. *Preventive Medicine Reports*, 4, 453–458. <https://doi.org/10.1016/j.pmedr.2016.08.008>
- Zerger, H. M., Miller, B. G., Valbuena, D., & Miltenberger, R. G. (2017). Effects of student pairing and public review on physical activity during school recess. *Journal of Applied Behavior Analysis*, 50(3), 529–537. <https://doi.org/10.1002/jaba.389>
- Zuckerman, O., & Gal-Oz, A. (2014). Deconstructing gamification: Evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. *Personal and Ubiquitous Computing*, 18(7), 1705–1719. <https://doi.org/10.1007/s00779-014-0783-2>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.