Well-being self-efficacy and complier average causal effect estimation: A substantive-methodological synergy ✩

Nicholas D. Myers [a], *, Isaac Prilleltensky [b], Christopher R. Hill [a], Deborah L. Feltz [a]

[a] Department of Kinesiology, Michigan State University, United States
[b] School of Education and Human Development, University of Miami, United States

ABSTRACT

Objectives: The purpose of this manuscript was to provide a substantive (i.e., well-being self-efficacy) — methodological (i.e., complier average causal effect estimation) synergy of potential importance to future research in the psychology of sport and exercise with secondary data analyses from the Fun For Wellness intervention. Fun For Wellness is a new on-line intervention designed to promote growth in well-being. Well-being self-efficacy is a proposed mechanism by which the effect of Fun For Wellness on well-being may be transmitted. Complier average causal effect estimation is a methodology that estimates the effect of complying with an intervention.
Design: The study design was a prospective, double-blind, parallel group randomized controlled trial (RCT) detailed in Myers, Prilleltensky, et al. (2016). Data were collected at baseline, 30 days- and 60 days-post baseline. A total of 479 adult employees at a major university in the southeast of the United States of America were enrolled.
Method: A two-class linear regression model with complier average causal effect estimation was fitted to well-being self-efficacy scores at 30- and 60-days.
Results: The adjusted mean difference in well-being self-efficacy scores for participants who complied with the intervention, as compared to potential compliers in the Usual Care group, was equal to 0.21, $p = 0.061$, Cohen's $d = 0.36$ at 30-days and 0.28, $p = 0.050$, Cohen's $d = 0.49$ at 60-days.
Conclusion: Complier average causal effect estimation may be a useful approach for RCTs in sport and exercise psychology when at least some of the participants do not comply with the intervention.

© 2017 Published by Elsevier Ltd.

The purpose of this manuscript was to provide a substantive—methodological synergy of potential importance to future research in the psychology of sport and exercise with secondary data analyses from the Fun For Wellness intervention. The substantive focus of this manuscript was well-being self-efficacy, which we defined as the extent to which a person believes that he or she has the ability to achieve a positive state of affairs in important areas of his or her life. The methodological focus of this manuscript was complier average causal effect estimation, which is designed to estimate the effect of complying with — and not merely just the effect of being assigned to — an intervention. Fun For Wellness is a new on-line universal intervention designed to promote growth in subjective well-being by providing capability-enhancing learning opportunities to participants. The theoretical framework upon which the Fun For Wellness intervention was conceptualized was self-efficacy theory (Bandura, 1997). Compliance with the Fun For Wellness intervention was viewed as a source of well-being self-efficacy information that directly influenced well-being self-efficacy beliefs. A scale tailored to the Fun For Wellness intervention was developed to measure well-being self-efficacy. From this point forward the structure of this manuscript will follow that of a substantive-methodological synergy (e.g., Myers, 2013).

1. Intervention and original research

1.1. Fun For Wellness (FFW)

Fun For Wellness (FFW) is a new on-line universal intervention designed to promote growth in subjective well-being by providing

capability-enhancing learning opportunities to participants (Myers, Prilleltensky, et al., 2016). Within the FFW intervention, the conceptualization of subjective well-being was based on the seven-dimension factor structure proposed by Prilleltensky et al. (2015) in the development of the I COPPE Scale. Prilleltensky et al. labeled these dimensions of subjective well-being as follows: Interpersonal, Community, Occupational, Physical, Psychological, Economic and overall. There is evidence that each of the dimensions of subjective well-being (except for economic) purportedly measured by the I COPPE Scale is relevant within the study of exercise science (Myers, Park, et al., 2016). From this point forward, we generally omit the term "subjective" from the expression "subjective well-being" for textual parsimony.

Myers, Prilleltensky, et al. (2016) provided an initial evaluation of the efficacy of the FFW intervention to increase well-being in an adult population. Measures of well-being were constructed based on responses to the I COPPE Scale. Participants who complied with the FFW intervention had significantly higher well-being, as compared to compliers in the Usual Care group, in the following dimensions: interpersonal at 60-days, community at 30- and 60-days, psychological at 60-days and economic at 30- and 60-days post baseline. Although promoting growth in well-being was the main purpose of the FFW intervention (see Myers, Prilleltensky, et al. for a detailed explanation) the intervention may also have increased participants' self-efficacy beliefs because the capability-enhancing learning opportunities provided to participants in the FFW intervention were created based on Bandura's (1997) self-efficacy theory.[1]

Self-efficacy has been defined as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). The capability-enhancing learning opportunities that the FFW intervention provided to participants came in the form of 152 interactive and scenario-based challenges organized in the on-line environment by the acronym BET I CAN (Myers, Prilleltensky, et al., 2016). The behavior-focused challenges were intended to increase a participant's capabilities to set a goal and to create positive habits (e.g., Watson & Tharp, 2014). The emotion-focused challenges were intended to increase a participant's capabilities to cope with negative emotions and to collect positive emotions (e.g., Seligman, 2011). The thought-focused challenges were intended to increase a participant's capabilities to challenge negative assumptions and to create a new narrative for their life (e.g., Hays, 2014). The interaction-focused challenges were intended to increase a participant's capabilities to communicate and connect with others (e.g., Gander, Proyer, Ruch, & Wyss, 2013). The context-focused challenges were intended to increase a participant's capabilities to read cues and to change cues in the environment (e.g., Dolan, 2014). The awareness-focused challenges were intended to increase a participant's capabilities to know their self and to know the issue (e.g., Brown & Ryan, 2003). The next steps-focused challenges were intended to increase a participant's capabilities to make a plan and to stick with a plan (e.g., Norcross, 2012). In summary, the BET I CAN challenges in the FFW intervention provided opportunities for a participant to increase his or her capabilities to organize and execute actions required to increase his or her well-being.

The capability-enhancing learning opportunity within each of the BET I CAN challenges in the FFW intervention was designed in such a way as to provide a participant with exposure to one or more

of Bandura (1997) potential sources of self-efficacy information. More specifically, each of the BET I CAN challenges created by Myers, Prilleltensky, et al. (2016) required participants to do one of the following activities: (a) play an interactive game (conceptualized as an enactive mastery experience); (b) watch vignettes performed by professional actors (conceptualized as a vicarious experience); (c) listen and/or read mini-lectures narrated by a coach (conceptualized as verbal persuasion); and (d) engage in self-reflection exercises and chat rooms (conceptualized as an opportunity to assess relevant physiological and/or emotional states). There is a rich extant literature on sources of self-efficacy information in sport and exercise psychology (see Feltz, Short, & Sullivan, 2008; for a review). The efficacy of the FFW intervention to increase self-efficacy was not tested by Myers, Prilleltensky, et al. (2016). Myers, Prilleltensky, et al. did, however, speculate that participants in their FFW intervention may have experienced increased self-efficacy and that future research in this area was warranted.

1.2. Original research with some secondary data analyses

The self-efficacy data reported in subsequent sections of this manuscript were collected as a pilot project within the large-scale RCT detailed by Myers, Prilleltensky, et al. (2016). All procedures performed involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. An institutional review board provided necessary permission to conduct this study.

Consistent with the publication manual of the American Psychological Association ([APA], 2010), the results reported in subsequent sections of this manuscript were viewed by the authors of this manuscript as an empirical study of original research because such studies include "secondary analyses that test hypotheses by presenting novel analyses of data not considered or addressed in previous reports" (p. 10). The self-efficacy data reported in subsequent sections of this manuscript were not considered in the Myers, Prilleltensky, et al. manuscript. The covariates and compliance data briefly reported in subsequent sections of this manuscript were considered in the Myers, Prilleltensky, et al. manuscript. Because, however, all of the aforementioned data within this paragraph were collected within the same RCT we provide only a summary of the study design and methods because "it is not necessary to repeat the description of the design and methods of a longitudinal or large-scale project in its entirety" (APA, 2010, p. 14). Readers are referred to Myers, Prilleltensky, et al. for a full description of the design, methods and data collection details (e.g., participant flow from screening to randomization to retention).

1.2.1. Procedures and participants

The study design was a prospective, double-blind, parallel group randomized controlled trial (RCT). Recruitment, eligibility verification and data collection were conducted on-line. Data were collected at baseline (T1), 30 days-post baseline (T2) and 60 days-post baseline (T3). Upon completion of the battery, each participant received an Amazon electronic gift card worth $10 at T1, an additional $15 at T2 and an additional $25 at T3. Eligible employees at a major research university in the southeast of the United States of America were randomly assigned to the intervention (FFW) or usual care (UC) groups by computer software that was specified to achieve a 1:1 group (i.e., FFW:UC) assignment. Participants who were randomly assigned to the UC group were provided with 30 days (i.e., from T1 to T2) of 24 h access to a webpage that provided links to several well-established websites that focused on well-being. Participants who were randomly assigned to the FFW

---

[1] Other theoretical approaches (e.g., positive psychology etc.) influenced the creation of some of the specific content in some of the capability-enhancing learning opportunities but self-efficacy theory was the theoretical framework within which the FFW intervention was constructed.

group were provided with 30 days (i.e., from T1 to T2) of 24 h access to the BET I CAN challenges.

A total of 479 eligible participants were randomized to UC ($n = 242$) or FFW ($n = 237$).

A total of 462 participants ($n = 226$) or FFW ($n = 236$) provided at least partial data during the RCT. A total of 429 participants provided data at T1 ($n_{UC} = 217$, $n_{FFW} = 212$). A total of 303 participants provided data at T2 ($n_{UC} = 163$, $n_{FFW} = 140$). A total of 267 participants provided data at T3 ($n_{UC} = 161$, $n_{FFW} = 126$). There were no statistically significant differences in the observed demographic characteristics (i.e., gender, race, ethnicity, education level, marital status and salary) by randomization group. A majority of the participants ($M_{age} = 41.75$, $SD_{age} = 11.67$) were full-time employees (96.6%), female (76.0%), Hispanic or White, non-Hispanic (81.8%) and earned a salary of greater than or equal to \$50,000 (65.3%). Approximately one-half of the participants had a graduate degree (48.2%) and were married (47.8%).

2. Well-being self-efficacy

The substantive focus of this manuscript was well-being self-efficacy. We developed a scale to measure well-being self-efficacy based on guidelines for constructing tailored self-efficacy scales (e.g., Bandura, 2006). In Appendix A, which is available as an on-line supplement, we describe the development of the well-being self-efficacy (WBSE) Scale. In this section we describe only the methods and results used to provide some initial evidence for the reliability and validity of scores derived from responses to the WBSE Scale consistent with relevant recommendations in sport and exercise psychology (e.g., Feltz et al., 2008).

2.1. Method

Power for each of the focal parameters (i.e., the seven pattern coefficients) was estimated for the sample size observed at T1 ($N = 429$) using Monte Carlo methods as implemented in M*plus* 7.4 (Muthén & Muthén, 1998–2015) and as advocated for in exercise science (e.g., Myers, Ahn, & Jin, 2011; Schweizer & Furley, 2016). The population parameter value for each of the focal parameters was set equal to 0.50. The 0.50 population value of each focal parameter was considered a medium sized effect (e.g., MacCallum, Widaman, Zhang, & Hong, 1999). More broadly, the population model assumed that 25% of the variance in each item would be accounted for by the factor and that participants would, on average, tend to use the upper end of the rating scale. The number of replications was set to 10,000. Replications were drawn from a conditionally multivariate normal distribution. The power estimation value for each of the seven focal parameters was equal to 1.00. Thus, we assumed that there would be high power to detect a truly false null hypothesis for each of the focal parameters.

Responses to the WBSE Scale at T1 were analyzed. Missing data were addressed from this point forward with the relevant default approach (e.g., full information maximum likelihood using the observed information matrix) in M*plus* 7.4 under the assumption of missing at random (Schafer & Graham, 2002). Type I error rate was set to 0.05 from this point forward in this manuscript for each null hypothesis test.

The measurement theory described in Appendix A was fit to the observed data at T1 under maximum-likelihood estimation with a correction for non-normality so that standard errors for parameter estimates and the test statistic for exact fit could be appropriately adjusted. Indexes of model-data fit considered were: χ_R^2, RMSEA, SRMR, CFI, and TLI. A pattern coefficient had to meet two criteria to be determined to be meaningfully large: (a) statistically significant and (b) the absolute value of the standardized coefficient > 0.20.

The latter criterion is somewhat arbitrary but is consistent with Jennrich and Bentler (2012). Latent variable reliability was measured with coefficient Hancock and Mueller (2001).

2.2. Results

There was evidence that the measurement theory for responses to the WBSE Scale fit the observed data at T1: χ_R^2 (10) = 3.04, $p = 0.980$, RMSEA = 0.000 (CI$_{90\%}$ = 0.000-0.000), $p \approx 1.00$, SRMR = 0.010, CFI = 1.00, and TLI = 1.02. Table 1 provides parameter estimates (and the content for each item) from the model. Standardized pattern coefficient estimates ranged from 0.46 (*your main occupation*) to 0.82 (*your life as a whole*). Residual variance estimates ranged from 0.20 (*your life as a whole*) to 0.57 (*your main occupation*). The four residual covariance estimates ranged from −0.03 (*your physical health and wellness* with *relationships with important people in your life*) to 0.08 (*your main occupation* with *your economic situation*). The *p*-values for residual covariance estimates ranged from 0.016 to 0.339. Intercept estimates, which were equal to the observed means, ranged from 2.29 (*the community where you live*) to 2.89 (*relationships with important people in your life*). Variance accounted for estimates ranged from 0.21 (*your main occupation*) to 0.68 (*your life as a whole*). Coefficient *H* was equal to 0.86. Thus, we concluded that there was some initial validity evidence for responses to the WBSE Scale.

3. Complier average causal effect (CACE) estimation

The methodological focus of this manuscript was CACE estimation. CACE estimation is designed to estimate the effect of complying with − and not merely just the effect of being assigned to − an intervention. The effect of being assigned to an intervention is also known as the intent to treat (ITT) effect (e.g., Hollis & Campbell, 1999). Estimating the CACE, in addition to the ITT effect can be important because "Individuals not fully complying with their assigned treatments is a common problem encountered in randomized evaluations of behavioral interventions" (Stuart, Perry, Le, & Ialongo, 2008, p. 288). In general terms, and assuming less than full compliance in at least some participants, a large effect for compliers may manifest as a less than large overall effect under an ITT only approach (e.g., Bloom, 1984).

The testing of large-scale theory-based RCTs in both clinical and community settings is of interest to the *Psychology of Sport and Exercise* (Strauss & Ntoumanis, 2015). It seems plausible that at least some participants within some large-scale theory-based RCTs in sport and exercise psychology may not comply with an intervention. The authors of this manuscript, however, are unaware of a single application of CACE estimation in sport and exercise psychology. The possible underuse of CACE estimation in RCTs within sport and exercise psychology may be due to the fact that many of the seminal methodological papers describing CACE estimation were written in journals focused on statistics (e.g., Angrist, Imbens, & Rubin, 1996; Rubin, 1978). More recently, however, several more applied papers have been published to increase the use of CACE estimation in fields such as public health (e.g., Little & Rubin, 2000) prevention science (e.g., Stuart et al., 2008) and psychology (e.g., Huang et al., 2014).

Highlighting the potential and limitations of new analytical approaches to sport and exercise psychology is of interest to the *Psychology of Sport and Exercise* (Strauss & Ntoumanis, 2015). The remainder of this section provides an applied and brief introduction to CACE estimation (see Stuart et al., 2008 for a fuller introduction) for sport and exercise psychologists by using concepts from the RCT for the FFW intervention described previously in this manuscript. The core idea in the introduction that follows is that

Table 1
Parameter estimates from the model imposed on responses to the well-being self-efficacy (WBSE) scale at baseline.

WBSE scale item	Intercept(SE)	Pattern coefficient(SE)	Standardized pattern coefficient(SE)	Residual variance(SE)	Residual covariance(SE)	R^2
1. Relationships with important people in your life	2.89(0.04)***	0.52(0.04)***	0.66(0.04)***	0.34(0.03)***	with Item 4, -0.03(0.03)	0.44
2. The community where you live	2.29(0.04)***	0.40(0.04)***	0.48(0.04)***	0.53(0.04)***	—	0.23
3. Your main occupation	2.72(0.04)***	0.39(0.04)***	0.46(0.05)***	0.57(0.05)***	with Item 6, 0.08(0.03)*	0.21
4. Your physical health and wellness	2.56(0.04)***	0.56(0.05)***	0.61(0.04)***	0.54(0.04)***	with Item 5, 0.07(0.03)*	0.37
5. Your psychological and emotional well-being	2.65(0.04)***	0.69(0.04)***	0.79(0.03)***	0.28(0.04)***	with Item 1, 0.02(0.03)	0.63
6. Your economic situation	2.34(0.04)***	0.50(0.04)***	0.58(0.04)***	0.49(0.04)***	—	0.33
7. Your life as a whole	2.74(0.04)***	0.64(0.03)***	0.82(0.03)***	0.20(0.03)***	—	0.68

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

the authors of this manuscript would like to estimate an effect (i.e., the CACE) that compared participants who were assigned to the FFW group and who complied with the FFW intervention to participants who were assigned to the UC group and who would have complied with the FFW intervention had they been assigned to the FFW group. Limitations of CACE analyses will be reviewed in the Concluding Remarks section.

3.1. Compliance types in the FFW intervention

We begin our brief introduction to CACE estimation by categorizing participants into four potential compliance types based on a categorization scheme provided by Angrist et al. (1996). Suppose that in an RCT similar to that detailed by Myers, Prilleltensky, et al. (2016), that each participant, i, is from a set of N participants (i.e., $i = 1,2,3, …N$) and that each participant is randomly assigned to either the UC group ($FFW_i = 0$) or the FFW group ($FFW_i = 1$). Let the outcome variable be well-being self-efficacy ($WBSE$). The effect of being assigned to the FFW intervention, the ITT effect (τ), can be written as follows:

$$\tau = \frac{1}{N} \sum_{i=1}^{N} WBSE_i(FFW_i = 1) - WBSE_i(FFW_i = 0) \tag{1}$$

Let the FFW treatment actually received (i.e., dose received) be represented by DR, where DR is equal to 0 (i.e., no FFW treatment) or $= 1$ (i.e., full FFW treatment). Given this notation, then, $DR_i(FFW_i = 0)$ connotes the dose of FFW treatment that participant i receives if assigned to the UC group, while $DR_i(FFW_i = 1)$ connotes the dose of FFW treatment that participant i receives if assigned to the FFW group. For participants assigned to the UC group the observed $WBSE$ data can be denoted, $WBSE_i(FFW_i = 0, DR_i(FFW_i = 0))$. For participants assigned to the FFW group the observed $WBSE$ data can be denoted, $WBSE_i(FFW_i = 1, DR_i(FFW_i = 1))$.

3.1.1. Compliers (C)

Participants who fully participate in the FFW intervention when assigned to the FFW group (i.e., $DR_i(FFW_i = 1) = 1$)) and who do not participate in the FFW intervention when assigned to the UC group (i.e., $DR_i(FFW_i = 0) = 0$)). In words, and parallel in meaning to the notation used in the previous sentence, for compliance type C the random assignment to the FFW group or the UC group by computer software is equivalent to subsequent compliance behavior (i.e., a complier behaves as instructed by her or his specific group assignment). Constructing an operational definition for full compliance within the FFW RCT, based on both substantive and methodological considerations, is an important design-based consideration for the purpose of attempting to maximize the possibility for the presence of compliance type C.

3.1.2. Always-takers (AT)

Participants who fully participate in the FFW intervention when assigned to the FFW group (i.e., $DR_i(FFW_i = 1) = 1$)) or when assigned to the UC group (i.e., $DR_i(FFW_i = 0) = 1$)). In words, and parallel in meaning to the notation used in the previous sentence, for compliance type AT the random assignment to the FFW group or the UC group by computer software is equivalent to the subsequent compliance behavior only when a participant is assigned to the FFW group (i.e., an always-taker finds a way to access the treatment irrespective of his or her specific group assignment). Creating structural impediments to accessing the FFW intervention for participants assigned to the UC group (e.g., withholding unique and secure log-in information) is an important design-based consideration for the purpose of attempting to minimize the possibility for the presence of compliance type AT.

3.1.3. Never-takers (NT)

Participants who do not participate in the FFW intervention when assigned to the FFW group (i.e., $DR_i(FFW_i = 1) = 0$)) or when assigned to the UC group (i.e., $DR_i(FFW_i = 0) = 0$)). In words, and parallel in meaning to the notation used in the previous sentence, for compliance type NT the random assignment to the FFW group or the UC group by computer software is equivalent to the subsequent compliance behavior only when a participant is assigned to the UC group (i.e., a never-taker does not engage in the treatment irrespective of her or his specific group assignment). Creating structural enticers to encourage fully complying with the FFW intervention for participants assigned to the FFW group (e.g., BET I CAN challenges are perceived as engaging) is an important design-based consideration for the purpose of attempting to minimize the possibility for the presence of compliance type NT.

3.1.4. Defiers (D)

Participants who do not participate in the FFW intervention when assigned to the FFW group (i.e., $DR_i(FFW_i = 1) = 0$)) and who fully participate in the FFW intervention when assigned to the UC group (i.e., $DR_i(FFW_i = 0) = 1$)). In words, and parallel in meaning to the notation used in the previous sentence, for compliance type D the random assignment to the intervention (FFW) or usual care (UC) groups by computer software is never equivalent to the subsequent compliance behavior (i.e., a defier behaves in the opposite manner as instructed by his or her specific group assignment). Sampling from a population where a contrarian reaction to random assignment within the FFW RCT can reasonably be considered to be an unlikely observation is an important design-based consideration for the purpose of attempting to minimize the possibility for the presence of compliance type D.

3.2. τs by compliance types

Using the compliance type acronyms just defined (e.g.,

C = compliers, etc.) and letting p denote proportion, the τ defined in Equation (1) can be viewed as an overall average of multiple τs by compliance type:

$$\tau = p_C\tau_C + p_{AT}\tau_{AT} + p_{NT}\tau_{NT} + p_D\tau_D \qquad (2)$$

Note that τ_C is equivalent to what has previously been referred to as "the CACE." In the event that $p_C = 1.00$ (and by extension $p_{AT} = p_{NT} = p_D = 0$), that is, each participant's compliance behavior is equivalent to the random assignment to the FFW group or the UC group by computer software then:

$$\tau = \tau_C \qquad (3)$$

Or, equivalently, the ITT effect is the CACE. Clearly Equation (3) would be ideal as an inference from the equation could be that the FFW intervention is at least acceptable to all participants (i.e., $p_C = 1.00$). In the event that $p_C < 1.00$, that is, not every participant's compliance behavior is equivalent to the random assignment to the FFW group or the UC group by computer software, then $\tau \neq \tau_C$ and τ_C can be written as:

$$\tau_C = \frac{\tau}{p_C} \qquad (4)$$

If several key assumptions are made.

3.3. Assumptions of CACE analyses

There are at least two key challenges to estimating the CACE in the FFW RCT. First, compliance behavior is not randomly assigned in the FFW RCT by the investigators. For example, it is unclear if a participant who was assigned to the FFW group and subsequently fully complied with the FFW treatment is a compliance type C or is a compliance type AT. Second, each participant in the FFW RCT is assigned to either the UC group or the FFW group. Thus, we cannot directly compare participant i's WBSE score if assigned to the UC group versus if assigned to the FFW group. These two key challenges make many of the parameters in Equation (2) difficult to estimate unless some relevant assumptions are made.

3.3.1. Assumption 1
Rubin's (1978) stable unit treatment value assumption is made. This assumption implies that each participant's potential WBSE scores are not affected by the randomized assignment (i.e., to the UC group or to FFW group) of other participants. The apparent lack of clustering in the FFW RCT design (e.g., no cluster-level randomization) reported by Myers, Prilleltensky, et al. (2016) may be consistent with this assumption of independence of observations. Readers are referred to Jo, Asparouhov, Muthén, Ialongo, and Brown (2008) for more complex modeling options with non-independent data.

3.3.2. Assumption 2
Access to the FFW intervention is randomly assigned to participants Myers, Prilleltensky, et al. (2016) reported that participants were randomly assigned to either the UC group or the FFW group by computer software.

3.3.3. Assumption 3
Assignment to the FFW group yields at least some participants who comply with the FFW intervention (i.e., the estimate of $p_C > 0$) Myers, Prilleltensky, et al. (2016) electronically tracked access to the FFW intervention (e.g., BET I CAN challenges completed) by requiring each participant to use her/his unique and secure log-in information. Myers, Prilleltensky et al. reported p_C estimates from

0.156 to 0.549 across the I COPPE scores. An estimate of p_C is observed in the treatment group (i.e., FFW group in this manuscript) and unobserved (though assumed to be equal to the value observed in the treatment group due to random assignment) in the control group (i.e., UC group in this manuscript).

3.3.4. Assumption 4
There are no observations of compliance type D in the FFW intervention (i.e., $p_D = 0$). Imbens and Angrist (1994) refer to the critical assumption, $p_D = 0$, as the monotonicity assumption. Simply, the act of being assigned to the UC group cannot itself increase subsequent participation in the FFW intervention. Similarly, the act of being assigned to the FFW group cannot itself decrease subsequent participation in the FFW intervention. This assumption allows for a reduced form of Equation (2):

$$\tau = p_C\tau_C + p_{AT}\tau_{AT} + p_{NT}\tau_{NT} \qquad (5)$$

3.3.5. Assumption 5
The effect of random assignment in the FFW RCT for compliance type AT equals zero (i.e., $\tau_{AT} = 0$). Always taking participants invariably find a way to access the FFW intervention regardless of random assignment to either the UC group or the FFW group. Thus, WBSE scores are unrelated to random assignment in the FFW RCT for compliance type AT. This assumption allows for a reduced form of Equation (5):

$$\tau = p_C\tau_C + p_{NT}\tau_{NT} \qquad (6)$$

Readers are referred to Hirano, Imbens, Rubin, and Zhou (2000) for possible exceptions to this assumption.

3.3.6. Assumption 6
The effect of random assignment in the FFW RCT for compliance type NT equals zero (i.e., $\tau_{NT} = 0$). Never taking participants invariably do not access the FFW intervention regardless of random assignment to either the UC group or the FFW group. Thus, WBSE scores are unrelated to random assignment in the FFW RCT for compliance type NT. This assumption allows for a reduced form of Equation (6):

$$\tau = p_C\tau_C \qquad (7)$$

Assumption 5 and Assumption 6 are collectively referred to as exclusion restrictions. Along with the exclusion restrictions it is typically assumed that $p_{AT} = 0$, a result of which is that 1-p_C equals p_{NT} (recall that in Assumption 4, $p_D = 0$). Thus CACE estimation typically proceeds with a focus on two compliance type classifications (i.e., classes): C and NT (Jo, 2002). More broadly, this specific specification of CACE estimation can be viewed as a 2-class mixture model where class membership is observed for participants in the treatment group and unobserved (and therefore estimated) for participants in the control group. Readers are referred to Morin and Wang (2016) for a general introduction to mixture models for sport and exercise scientists.

4. Well-being self-efficacy and CACE estimation

The substantive-methodological synergy focus of this manuscript was to provide an initial evaluation of the efficacy of the FFW intervention to increase well-being self-efficacy with CACE estimation with some secondary data analyses from the Myers, Prilleltensky, et al. (2016) FFW RCT. The operational definition for compliance with the FFW intervention was based on the FFW

compliance scoring system created by Myers, Prilleltensky et al. Most broadly, the definition of full participation was based on both substantive (e.g., it would take approximately 2 h of interacting with the FFW intervention to earn sufficient participation points) and methodological (e.g., the presence of some compliers) considerations. More specifically, full participation was defined as (a) completing four introductory (e.g., orientation to the website; an introduction to the characters that appear in the vignettes; etc.) BET I CAN challenges and (b) earning at least 21 additional participation points by completing BET I CAN challenges related to interpersonal well-being.

Well-being self-efficacy was measured with the 7 items that define the WBSE Scale at T1, T2 and T3. For each participant an average observed score for well-being self-efficacy was created at each time point. The reliability of well-being self-efficacy composite scores was measured with coefficient omega (McDonald, 1970). Data on proposed pre-treatment demographic covariates of well-being (Rubenstein et al., 2016) were collected at baseline and included participant gender, age, race, education-level, marital status and salary.

4.1. Data analytic approach

Three models were fit in M*plus* 7.4 under maximum-likelihood (ML) estimation with robust standard errors. The primary purpose of each model was to estimate the effect of the FFW intervention to increase well-being self-efficacy over time and under some alternative model specifications (Jo, 2002). Model 0 estimated the effect of being assigned to the FFW intervention (i.e., the ITT effect or τ). Model 1 estimated the effect of being assigned to the FFW intervention for those who fully participated in the FFW intervention (i.e., the CACE or τ_C). Model 2 estimated τ_C and the effect of being assigned to the FFW intervention for those who did not fully participate in the FFW intervention (i.e., τ_{NT}). In each model, an effect size was calculated by dividing the mean difference by the square root of the variance pooled across the UC and FFW groups. This effect size was equal to Cohen's *d* (1988) in Model 0 and was viewed as an analog to Cohen's *d* in Model 1 and Model 2 in a mixture model framework. For the sake of continuity, we refer to the effect size as Cohen's *d* throughout the manuscript and we adopt commonly used heuristics to assist in the interpretation of an absolute value of Cohen's *d*: 0.20 (small), 0.50 (medium) and 0.80 (large). Missing data were handled with the default approach under the assumption that data were missing at random (Jo, Ginexi, & Ialongo, 2010). In Model 1 and Model 2 the demographic covariates were specified as predictors of compliance.

Model 0 imposed a regression model with well-being self-efficacy at T2 and T3 as the outcome variables. The demographic covariates, well-being self-efficacy at T1 and treatment assignment (UC = 0, FFW = 1) were specified as predictors of well-being self-efficacy at T2 and T3 and these regression coefficients were freely estimated. The intercepts for well-being self-efficacy at T2 and T3 were freely estimated. Residual (co-)variance between well-being self-efficacy at T2 and T3 was freely estimated. The direct effects from treatment assignment to well-being self-efficacy at T2 (τ_{T2}) and T3 (τ_{T3}) were the focal parameters and a positive value indicated that the FFW group had a higher adjusted mean for well-being self-efficacy as compared to the UC group. The expression "adjusted mean", is used to acknowledge the statistical adjustment made by including the covariates in the model.

Model 1 imposed a two-class regression model with CACE estimation with well-being self-efficacy at T2 and T3 as the outcome variables. Class 1 was conceptualized as the never-taking class (i.e., compliance type NT). Class 2 was conceptualized as the complier class (i.e., compliance type C). A binary latent class

indicator was created where compliers (i.e., at least 21 post-introductory participation points) in the FFW group had a value of 1, non-compliers (i.e., less than 21 post-introductory participation points) in the FFW group had a value of 0 and participants in the UC group had a missing value. The categorical latent variable representing compliance classification was regressed on the demographic covariates. The demographic covariates, well-being self-efficacy at T1 and treatment assignment were specified as predictors of well-being self-efficacy at T2 and T3 and these regression coefficients were freely estimated in each class. The two direct effects from treatment assignment to well-being self-efficacy at T2 and T3 were fixed to 0 in Class 1 (i.e., the exclusion restrictions (i.e. $\tau_{NT_{r2}} = \tau_{NT_{r3}} = 0$), and were freely estimated in Class 2 (i.e. $\tau_{C_{r2}}, \tau_{C_{r3}}$). The intercepts for well-being self-efficacy at T2 and T3 were freely estimated in each class. Residual (co-)variances for well-being self-efficacy at T2 and T3 were freely estimated in each class. The direct effects from treatment assignment to well-being self-efficacy at T2 and T3 in Class 2 (i.e., $\tau_{C_{r2}}, \tau_{C_{r3}}$) were the focal parameters and a positive value indicated that compliers in the FFW group had a higher adjusted mean for well-being self-efficacy as compared to compliers in the UC group. Readers are referred to example 7.24 of the Mplus User Guide (www.statmodel.com/HTML_UG/chapter7V8.htm) for a short version (in terms of the number of variables) of the input file required to impose this model. The input file used in the current study (which is considerably longer due to the number of variables) is available by request to the corresponding author.

Model 2 estimated all of the parameters specified in Model 1 while removing the exclusion restrictions for never-takers. Model 1 was nested within Model 2 and these models were empirically compared with the change in the likelihood ratio χ^2 (robust) test $\Delta\chi^2_R$. Considering the possibility that the exclusion restriction may not hold in the current social-behavioral FFW intervention was viewed as reasonable for two reasons. First, from a methodological perspective it has been shown that an estimate of τ_C can be biased if the true τ_{NT} effect is non-zero but is fixed to zero in a model and compliance is low (Jo, 2002). Second, from a conceptual perspective the research team expected that at least some of the participants assigned to the FFW group would earn more than 0 participation points but not enough participation points to be considered to be fully participating. The $\tau_{C_{r2}}$ and $\tau_{C_{r3}}$ effects and the $\tau_{NT_{r2}}$ and $\tau_{NT_{r3}}$ effects were the focal parameters. A negative τ_{NT} value indicated that never-takers in the FFW group had a lower adjusted mean for well-being self-efficacy as compared to never-takers in the UC group. Readers are referred to example 7.24 of the Mplus User Guide (www.statmodel.com/HTML_UG/chapter7V8.htm) for a short version (in terms of the number of variables) of the input file required (sans the command y on x2@0) to impose this model. The input file used in the current study (which is considerably longer due to the number of variables) is available by request to the corresponding author.

Power for both focal parameters in Model 1 (i.e., $\tau_{C_{r2}}, \tau_{C_{r3}}$) was estimated for a sample size of 462 using Monte Carlo methods as implemented in M*plus* 7.4. The population parameter value for both focal parameters was set equal to 0.33. The 0.33 population value for both focal parameters was considered a medium sized effect as the population value of Cohen's *d* was 0.50 at T2 and T3. More broadly, the population model assumed a 1:1 group (i.e., FFW:UC) randomized assignment, a compliance rate of 27% and the exclusion restrictions for never takers (i.e., $\tau_{NT_{r2}} = \tau_{NT_{r3}} = 0$). The number of replications was set to 10,000. Replications were initially drawn from a conditionally multivariate normal distribution. The power estimation value for both focal parameters was equal to 0.74. Thus, we assumed that there would be a modest level of power to detect a truly false null hypothesis for both of the focal parameters.

4.2. Results

Sixty-four participants who were randomized to the FFW group were classified as a complier (i.e., the estimate of p_C was equal to 0.27). Coefficient omega for well-being self-efficacy scores was equal to 0.82 at T1, 0.82 at T2 and 0.84 at T3. Mean well-being self-efficacy at T1 for the UC group, 2.58, was not statistically significantly different, $p = 0.582$, than mean well-being self-efficacy at T1 for the FFW group, 2.61.

4.2.1. Model 0

Table 2 provides unstandardized parameter estimates from Model 0. None of the demographic covariates were statistically significant predictors of well-being self-efficacy at T2 or at T3. The focal parameter at T2, τ_{T2}, was equal to 0.03, $SE = 0.05$, $p = 0.528$ and Cohen's $d = 0.05$. More descriptively, the adjusted well-being self-efficacy mean at T2 for the UC group equaled 2.48 while the adjusted well-being self-efficacy mean at T2 for the FFW group equaled 2.52. The focal parameter at T3, τ_{T3}, was equal to 0.04, $SE = 0.05$, $p = 0.411$ and Cohen's $d = 0.07$. More descriptively, the adjusted well-being self-efficacy mean at T3 for the UC group equaled 2.78 while the adjusted well-being self-efficacy mean at T3 for the FFW group equaled 2.82. In summary, there was evidence that the effect of simply being assigned to the FFW intervention, without considering actual participation in the FFW intervention, resulted in a null (with regard to hypothesis testing) and negligible (with regard to effect size) effect at 30-days and at 60-days since the onset of the intervention.

4.2.2. Model 1

Table 3 provides unstandardized parameter estimates for the complier class from Model 1 where well-being self-efficacy at T2 and T3 were the outcomes.[2] None of the demographic covariates were statistically significant predictors of well-being self-efficacy at T2 or at T3 for compliers. The focal parameter at T2, $\tau_{C_{r2}}$, was equal to 0.21, $SE = 0.11$, $p = 0.061$ and Cohen's $d = 0.36$. More descriptively, the adjusted mean at T2 for compliers assigned to the UC group equaled 2.30 while the adjusted mean at T2 for compliers

assigned to the FFW group equaled 2.51. The focal parameter at T3, $\tau_{C_{r3}}$, was equal to 0.28, $SE = 0.14$, $p = 0.050$ and Cohen's $d = 0.49$. More descriptively, the adjusted mean at T3 for compliers assigned to the UC group equaled 3.06 while the adjusted mean at T3 for compliers assigned to the FFW group equaled 3.34. In summary, there was evidence that participants who complied (i.e., compliers) with the FFW intervention in the FFW group had higher well-being self-efficacy (i.e., a small to moderate positive effect size), as compared to compliers in the UC group, at 30-days (although the difference was not statistically significant) and at 60-days since the onset of the intervention. Unstandardized parameter estimates for the never-taker class from Model 1 are available in Table 4.

4.2.3. Model 2

Model 1 was the accepted model because its fit to the data was not statistically significantly worse than Model 2, $\Delta\chi^2_R(2) = 2.51$, $p = 0.286$. Given the pedagogical intent of this manuscript, however, focal parameter estimates from Model 2 are briefly reported. The focal parameter at T2 for never-takers, $\tau_{NT_{r2}}$, was equal to 0.00, $SE = 0.07$, $p = 0.960$ and Cohen's $d = -0.01$. The focal parameter at T3 for never-takers, $\tau_{NT_{r3}}$, was equal to -0.08, $SE = 0.07$, $p = 0.196$ and Cohen's $d = -0.15$. The focal parameter at T2 for compliers, $\tau_{C_{r2}}$, was equal to 0.14, $SE = 0.11$, $p = 0.204$ and Cohen's $d = 0.25$. The focal parameter at T3 for compliers, $\tau_{C_{r3}}$, was equal to 0.37, $SE = 0.11$, $p = 0.001$ and Cohen's $d = 0.64$. In summary, there was evidence that participants who did not comply (i.e., never-takers) with the FFW intervention in the FFW group had similar well-being self-efficacy, as compared to never-takers in the UC group, at 30-days and at 60-days since the onset of the intervention. A negative effect of assignment to the treatment group for never-takers, while not observed in the current study, has been observed and explained in previous research: "… never-takers assigned to the treatment condition could be demoralized by failing to take the intervention opportunity. This negative psychological effect would not occur for never-takers assigned to the control condition, since the treatment is never offered." (Jo, 2002, p. 386). A full set of parameter estimates from Model 2 are available upon request to the lead author of this manuscript.

Table 2
Unstandardized Parameter Estimates from Model 0 where Well-Being Self-Efficacy at Time 2 (T2) and at Time 3 (T3) were the Outcomes.

Variable	Time 2			Time 3		
	Regression coefficients	Intercepts	Residual (Co)-variance	Regression coefficients	Intercepts	Residual variance
1. Fun For Wellness	0.03(0.05)[a]	–	–	0.04(0.05)[b]	–	–
2. Well-Being Self-Efficacy at Time 1	0.65(0.05)***	–	–	0.63(0.05)***	–	–
3. Female	−0.01(0.06)	–	–	−0.01(0.06)	–	–
4. Age	0.00(0.00)	–	–	0.00(0.00)	–	–
5. Hispanic	0.14(0.08)	–	–	0.05(0.09)	–	–
6. White	0.00(0.08)	–	–	−0.11(0.09)	–	–
7. Graduate	0.10(0.05)	–	–	0.04(0.05)	–	–
8. Married	−0.04(0.05)	–	–	−0.01(0.06)	–	–
9. Salary ≥ $50,000	−0.06(0.07)	–	–	−0.12(0.07)	–	–
10. Well-Being Self-Efficacy at T2	–	2.48(0.13)***	0.18(0.02)***	–	–	–
11. Well-Being Self-Efficacy at T3	–	–	with Variable 10, 0.09(0.01)***	–	2.78(0.14)***	0.17(0.02)***

Note. Values in parentheses are standard errors.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.
[a] The intent-to-treat effect at Time 2 (τ_{T2}).
[b] The intent-to-treat effect at Time 3 (τ_{T3}).

---

[2] Consistent with the results from Myers, Prilleltensky, et al. (2016), none of the demographic covariates were statistically significant predictors of compliance.

Table 3
Unstandardized Parameter Estimates for the Complier Class from Model 1 where Well-Being Self-efficacy atTime 2 (T2) and at Time 3 (T3) were the Outcomes.

Variable	Time 2			Time 3		
	Regression coefficients	intercepts	Residual (Co)-variance	Regression coefficients	Intercepts	Residual variance
1. Fun For Wellness	0.21(0.11)[a]	–	–	0.28(0.14)[*][b]	–	–
2. Well-Being Self-Efficacy at Time 1	0.84(0.11)***	–	–	0.51(0.10)***	–	–
3. Female	0.07(0.11)	–	–	−0.08(0.12)	–	–
4. Age	0.00(0.01)	–	–	−0.01(0.01)	–	–
5. Hispanic	−0.13(0.13)	–	–	−0.06(0.14)	–	–
6. White	−0.15(0.16)	–	–	−0.23(0.15)	–	–
7. Graduate	0.17(0.11)	–	–	0.20(0.16)	–	–
8. Married	0.11(0.11)	–	–	0.17(0.12)	–	–
9. Salary ≥ $50,000	−0.13(0.15)	–	–	−0.24(0.13)	–	–
10. Well-Being Self-Efficacy at T2	–	2.30(0.28)***	0.11(0.03)***	–	–	–
11. Well-Being Self-Efficacy at T3	–	–	with Variable 10, 0.06(0.03)*	–	3.06(0.28)***	0.15(0.03)***

Note. Values in parentheses are standard errors.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.
[a] The complier average causal effect at Time 2 (τ_{C_T2}).
[b] The complier average causal effect at Time 3 (τ_{C_T3}).

Table 4
Unstandardized Parameter Estimates for the Never-Taker Class from Model 1 where Well-Being Self-Efficacy at Time 2 (T2) and at Time 3(T3) were the Outcomes.

Variable	Time 2			Time 3		
	Regression coefficients	Intercepts	Residual (Co)-variance	Regression coefficients	Intercepts	Residual variance
1. Fun For Wellness	–	–	–	–	–	–
2. Well-Being Self-Efficacy at Time 1	0.61(0.07)***	–	–	0.70(0.06)***	–	–
3. Female	−0.01(0.08)	–	–	0.02(0.09)	–	–
4. Age	0.00(0.00)	–	–	0.00(0.00)	–	–
5. Hispanic	0.33(0.11)**	–	–	0.14(0.13)	–	–
6. White	0.17(0.12)	–	–	−0.02(0.13)	–	–
7. Graduate	0.06(0.08)	–	–	−0.02(0.08)	–	–
8. Married	−0.11(0.07)	–	–	−0.09(0.07)	–	–
9. Salary ≥ $50,000	−0.05(0.10)	–	–	−0.09(0.10)	–	–
10. Well-Being Self-Efficacy at T2	–	2.43(0.20)***	0.19(0.02)***	–	–	–
11. Well-Being Self-Efficacy at T3	–	–	with Variable 10, 0.09(0.02)***	–	2.58(0.21)***	0.16(0.02)***

Note. Values in parentheses are standard errors.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

4.3. Results under complete cases only

The approach to missing data used in the previous analyses, missing at random (MAR) conditional on the observed information (under maximum likelihood estimation), is a modern approach that can be implemented in randomized experiments with noncompliance (e.g., Jo et al., 2010). Unfortunately, a complete cases only approach to missing data (which can decrease power) may be commonly implemented in practice (e.g., Jo et al., 2010). For pedagogical reasons we refitted Model 0 and Model 1 under a complete cases only approach to compare these results to the results already reported in this manuscript which are provided below in parentheses (i.e., "versus") for convenience.

4.3.1. Model 0

The focal parameter at T2, τ_{T2}, was equal to −0.02 (versus 0.03), $SE = 0.06$ (versus 0.05), $p = 0.750$ (versus 0.528) and Cohen's $d = −0.03$ (versus 0.05). The focal parameter at T3, τ_{T3}, was equal to 0.02 (versus 0.04), $SE = 0.05$ (versus 0.05), $p = 0.659$ (versus 0.411) and Cohen's $d = 0.03$ (versus 0.07). In summary, at both T2 and T3 key results for the focal parameter were somewhat less encouraging under a complete cases only approach.

4.3.2. Model 1

The focal parameter at T2, $\tau_{C_{r2}}$, was equal to −0.05 (versus 0.21), $SE = 0.14$ (versus 0.11), $p = 0.715$ (versus 0.061) and Cohen's

$d = −0.07$ (versus 0.36). The focal parameter at T3, $\tau_{C_{r3}}$, was equal to 0.21 (versus 0.28), $SE = 0.12$ (versus 0.14), $p = 0.077$ (versus 0.050) and Cohen's $d = 0.28$ (versus 0.49). In summary, at both T2 and T3 key results for the focal parameter were somewhat less encouraging under a complete cases only approach.

5. Concluding remarks

The substantive focus of this manuscript was well-being self-efficacy, which we operationally defined to span a general domain within the context of a particular intervention. Our operational definition of well-being self-efficacy and the WBSE Scale that we developed both should be viewed as preliminary and perhaps context-specific efforts — particularly with regard to the well-being facet of well-being self-efficacy — in an area of research that may evolve over time. In the broader field of exercise science, Myers, Park, et al. (2016) noted that the "conceptualization of the subjective well-being construct is somewhat amorphous (e.g., operational definition, dimensionality, etc.) and the measurement of the construct often is inconsistent (e.g., instrument selected to measure the construct across studies)" (p. 230). Given the lack of a unified definition of well-being in exercise science and more generally (see Diener, 2012; for a summary), other useful operational definitions of well-being self-efficacy, particularly outside of the FFW intervention context, are likely.

The methodological focus of this manuscript was CACE

estimation. While we believe that CACE estimation may be a useful statistical approach for RCTs in sport and exercise psychology when at least some of the participants do not comply with the intervention, we also believe that there are at least three limitations to CACE estimation. First, we do not advocate a CACE analysis as a replacement for an ITT analysis. Both statistical approaches can provide important information about the efficacy of an intervention and we advocate that results from an ITT analysis should be reported in a manuscript that reports results from a CACE analysis. Second, there are several assumptions typically made in a CACE analysis and the context-specific perceived viability of these assumptions (e.g., definition of compliance; etc.) should be discussed in a manuscript that reports results from a CACE analysis. Third, the fact that a CACE analysis is even undertaken may be viewed as possible evidence that the intervention in a RCT may not have been viewed as acceptable to at least some participants in the treatment group. An idealized RCT likely would include that each participant fully complies with the treatment assigned to her or him, and hence, an ITT analysis would be sufficient (as it would be equivalent to a CACE analysis).

The substantive-methodological synergy focus of this manuscript was to provide an initial evaluation of the efficacy of the FFW intervention to increase well-being self-efficacy with CACE estimation. Results from this study provided some initial evidence for both the efficacy of, and possible revisions to, the FFW intervention. First, we provided some evidence that the FFW intervention may produce a small to moderate increase in the well-being self-efficacy of complying participants in the treatment group, as compared to complying participants in the control group, at 30-days and at 60-days since the onset of the intervention. It should be noted, however, that the small positive difference in well-being self-efficacy at 30-days was not statistically significant but we believe that this was likely attributable to a Type II error. Under the assumption that our belief about a Type II error is correct, then, this finding is potentially important because well-being self-efficacy is a proposed mechanism by which the effect of the FFW intervention on well-being may be transmitted. A future RCT that tests the indirect effect of the FFW intervention on well-being through well-being self-efficacy would build upon the initial efficacy results provided by Myers, Prilleltensky, et al. and the results in the current manuscript. Testing such a model within the CACE framework, however, may present additional complications such as the definition of compliance would have to accommodate both the intervening variable (e.g., well-being self-efficacy) and the main outcome (e.g., well-being). We suspect that related complications, in addition to complications related to mixture modeling more generally, may have factored in to the fact that substantive-CACE methodology synergies in other fields (e.g., Huang et al., 2014; Stuart et al., 2008) have also focused on direct effects. Second, we also provided evidence that compliance with the FFW intervention was relatively low as only about one-quarter of participants were defined as compliers. Revamping the FFW intervention to include more effective structural enticers to encourage fully complying with the FFW intervention for participants assigned to the treatment group is an important design-based consideration for a future RCT.

5.1. Broader implications

While the focus of the substantive and synergy sections of this paper is specific to a particular intervention, Fun For Wellness, and a particular outcome, well-being self-efficacy, the focus of the methodological section, CACE, may have much broader implications for sport and exercise psychology researchers. Although the testing of large-scale theory-based RCTs appears to be on the rise in sport and exercise psychology with regard to a variety of outcomes

such as mental health (e.g., Eather, Morgan, & Lubans, 2016) and physical activity (e.g., Prestwich et al., 2017) modern approaches to modeling the compliance behavior of participants within similarly-focused RCTs appears to be in its infancy. Given both the recent surge in RCTs, and the long history of theory-based interventions delivered within less than fully controlled research design environments, in sport and exercise psychology (e.g., Landers, 1983), it seems plausible that estimating the effect of complying with (i.e., CACE) — and not only just the effect of being assigned to (i.e., ITT) — an intervention could have major implications for both theory testing and policy implications (e.g., government sponsored incentives to promote compliance with efficacious behavioral interventions that promote individual and public health) in important areas of research within sport and exercise psychology. We hope that the content of this manuscript may promote an increased interest in the potential use of CACE estimation in sport and exercise psychology research when at least some participants do not comply with an intervention.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.psychsport.2017.02.010.

References

American Psychological Association. (2010). *Publication manual of the american psychological association* (6th ed.). Washington, DC: American Psychological Association.

Angrist, J., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*, 444—472. http://dx.doi.org/10.2307/2291629.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. C. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307—337). Information Age Publishing.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8*, 225—246. http://dx.doi.org/10.1177/0193841X8400800205.

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*, 822—848. http://dx.doi.org/10.1037/0022-3514.84.4.822.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Diener, E. (2012). *New findings and future directions for subjective well-being research* (Vol. 67, pp. 590—597). American Psychologist. http://dx.doi.org/10.1037/a0029541.

Dolan, P. (2014). *Happiness by design* (New York, NY: Penguin).

Eather, N., Morgan, P. J., & Lubans, D. R. (2016). Effects of exercise on mental health outcomes in adolescents: Findings from the CrossFit TM teens randomized controlled trial. *Psychology of Sport and Exercise, 26*, 14—23. http://dx.doi.org/10.1016/j.psychsport.2016.05.008.

Feltz, D. L., Short, S. E., & Sullivan, P. J. (2008). *Self-efficacy in sport.* Champaign, IL: Human Kinetics.

Gander, F., Proyer, R. T., Ruch, W., & Wyss, T. (2013). Strength-based positive interventions: Further evidence for their potential in enhancing well-being and alleviating depression. *Journal of Happiness Studies, 14*, 1241—1259. http://dx.doi.org/10.1007/s10902-012-9380-0.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. H. C. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Past and present. A festschrift in honor of Karl G. Jöreskog* (pp. 195—261). Chicago: Scientific Software International, Inc.

Hays, P. A. (2014). *Creating well-being: Four steps to a happier, healthier life.* Washington, DC: American Psychological Association.

Hirano, K., Imbens, G. W., Rubin, D. B., & Zhou, X. H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics, 1*, 69—88. http://dx.doi.org/10.1093/biostatistics/1.1.69.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ: British Medical Journal, 319*, 670—674. http://dx.doi.org/10.1136/bmj.319.7211.670.

Huang, S., Cordova, D., Estrada, Y., Brincks, A. M., Asfour, L. S., & Prado, G. (2014). An application of the complier average causal effect analysis to examine the effects of a family intervention in reducing illicit drug use among high-risk Hispanic adolescents. *Family Process, 53*, 336—347. http://dx.doi.org/10.1111/famp.12068.

Imbens, G. W., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62*, 467—476. http://dx.doi.org/10.2307/2951620.

Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika, 77*, 442—454. http://dx.doi.org/10.1007/s11336-012-9269-1.

Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics, 27*, 385—409. http://dx.doi.org/10.3102/10769986027004385.

Jo, B., Asparohouv, T., Muthén, B., Ialongo, N. S., & Brown, C. H. (2008). Cluster randomized trials with treatment noncompliance. *Psychological Methods, 13*, 1—18. http://dx.doi.org/10.1037/1082-989X.13.1.1.

Jo, B., Ginexi, E. M., & Ialongo, N. S. (2010). Handling missing data in randomized experiments with noncompliance. *Prevention Science, 11*, 384—396. http://dx.doi.org/10.1007/s11121-010-0175-4.

Landers, D. M. (1983). Whatever happened to theory testing in sport psychology. *Journal of Sport & Exercise Psychology, 5*, 134—151.

Little, R. J., & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health, 21*, 121—145. http://dx.doi.org/10.1146/annurev.publhealth.21.1.121.

MacCallum, R., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84—99. http://dx.doi.org/10.1037/1082-989X.4.1.84.

McDonald, R. P. (1970). The theoretical foundations of common factor analysis, principal factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 23*, 1—21. http://dx.doi.org/10.1111/j.2044-8317.1970.tb00432.x.

Morin, A. J. S., & Wang, J. C. K. (2016). A gentle introduction to mixture modeling using physical fitness performance data. In N. Ntoumanis, & N. D. Myers (Eds.), *An introduction to intermediate and advanced statistical analyses for sport and exercise scientists* (pp. 183—209). Chichester, UK: Wiley.

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Myers, N. D. (2013). Coaching competency and (exploratory) structural equation modeling: A substantive-methodological synergy. *Psychology of Sport and Exercise, 14*, 709—718. http://dx.doi.org/10.1016/j.psychsport.2013.04.008.

Myers, N. D., Ahn, S., & Jin, Y. (2011). Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: A Monte Carlo approach. *Research Quarterly for Exercise and Sport, 82*, 412—423. http://dx.doi.org/10.5641/027013611X13275191443621.

Myers, N. D., Park, S. E., Lefevor, G. T., Dietz, S., Prilleltensky, I., & Prado, G. J. (2016a).

Measuring multidimensional well-being with the I COPPE Scale in a Hispanic sample. *Measurement in Physical Education and Exercise Science, 20*, 230—243. http://dx.doi.org/10.1080/1091367X.2016.1226836.

Myers, N. D., Prilleltensky, I., Prilleltensky, O., McMahon, A., Dietz, S., & Rubenstein, C. L. (2016b). *Efficacy of the fun for wellness online intervention to promote multidimensional well-being: A randomized controlled trial* (Manuscript submitted for publication).

Norcross, J. C. (2012). *Changeology: 5 steps to realizing your goals and resolutions.* New York, NY: Simon & Schuster.

Prestwich, A., Conner, M., Morris, B., Finlayson, G., Sykes-Muskett, B., & Hurling, R. (2017). Do web-based competitions promote physical activity? Randomized controlled trial. *Psychology of Sport and Exercise, 29*, 1—9. http://dx.doi.org/10.1016/j.psychsport.2016.11.003.

Prilleltensky, I., Dietz, S., Prilleltensky, O., Myers, N. D., Rubenstein, C. L., Jin, Y., et al. (2015). Assessing multidimensional well-being: Development and validation of the I COPPE scale. *Journal of Community Psychology, 43*, 199—226. http://dx.doi.org/10.1002/jcop.21674.

Rubenstein, C. L., Duff, J., Prilleltensky, I., Jin, Y., Dietz, S., Myers, N. D., et al. (2016). Demographic group differences in domain-specific well-being. *Journal of Community Psychology, 44*, 499—515. http://dx.doi.org/10.1002/jcop.21784.

Rubin, D. B. (1978). Bayesian inference for causal effect: The role of randomization. *Annals of Statistics, 6*(1), 34—58. Retrieved from http://www.jstor.org/stable/2958688.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147—177. http://dx.doi.org/10.1037/1082-989X.7.2.147.

Schweizer, G., & Furley, F. (2016). Reproducible research in sport and exercise psychology: The role of sample sizes. *Psychology of Sport and Exercise, 23*, 114—122. http://dx.doi.org/10.1016/j.psychsport.2015.11.005.

Seligman, M. (2011). *Flourish: A visionary new understanding of happiness and well-being.* New York, NY: Simon and Schuster.

Strauss, B., & Ntoumanis, N. (2015). Our PSE journey: Looking back and forward. *Psychology of Sport and Exercise, 16*, 181—182. http://dx.doi.org/10.1016/j.psychsport.2014.11.002.

Stuart, E. A., Perry, D. F., Le, H.-N., & Ialongo. (2008). Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science, 9*, 288—298. http://dx.doi.org/10.1007/s11121-008-0104-y.

Watson, D. L., & Tharp, R. G. (2014). *Self-directed behavior: Self-modification for personal adjustment* (10th ed.). Belmont, CA: Cengage Learning.