Routledge
Taylor & Francis Group

The Journal of Positive Psychology

Dedicated to furthering research and promoting good practice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rpos20

Emphasizing scientific rigor in the development, testing, and implementation of positive psychological interventions

Samantha J. Heintzelman & Kostadin Kushlev

To cite this article: Samantha J. Heintzelman & Kostadin Kushlev (2020) Emphasizing scientific rigor in the development, testing, and implementation of positive psychological interventions, The Journal of Positive Psychology, 15:5, 685-690, DOI: 10.1080/17439760.2020.1789701

To link to this article: https://doi.org/10.1080/17439760.2020.1789701

THE JOURNAL OF POSITIVE PSYCHOLOGY
2020, VOL. 15, NO. 5, 685–690
https://doi.org/10.1080/17439760.2020.1789701

Routledge
Taylor & Francis Group

Emphasizing scientific rigor in the development, testing, and implementation of positive psychological interventions

Samantha J. Heintzelman[a] and Kostadin Kushlev[b]

[a]Department of Psychology, Rutgers University, Newark, NJ, USA; [b]Department of Psychology, Georgetown University, Washington, DC, USA

ABSTRACT

It is a critical period in the burgeoning field of positive psychological interventions (PPIs) to establish best research practices for the area. In this piece, we outline key features of intervention research that we believe have been underutilized in PPI science and offer recommendations to attain and evaluate key objectives for study design and analyses, measurement, sampling, and research personnel. We review work with one PPI – ENHANCE – to provide concrete examples of both successes and shortcomings in each of these aspirational areas. Our goal is to offer actionable recommendations for PPI researchers and guidance for practitioners evaluating this research for application.

ARTICLE HISTORY
Received 30 May 2019
Accepted 17 February 2020

KEYWORDS
Subjective well-being; intervention; measurement; happiness intervention; life satisfaction

A majority of people worldwide rate happiness as an 'extremely important' personal goal (Diener, 2000). As individuals (and organizations and societies) seek to become happier, the field of psychology has responded with the development and implementation of a number of different positive psychological interventions (PPIs). It is, therefore, critical to establish best practices in PPI development and research. We argue that a broader impact of PPIs in mental health promotion depends on the employment of best practices in established clinical sciences.

In this piece, we outline a selection of key features of intervention research that we believe have been underutilized in PPI science and offer recommendations to attain and evaluate these objectives as the field develops. We focus on four elements of PPI research: study design and analyses, measurement, sampling, and research personnel. For each of these elements, we supplement our recommendations with implementation examples from our own work with one PPI – ENHANCE (Heintzelman et al., 2020; Kushlev et al., 2020, 2017) – to provide concrete representations of both successes and shortcomings in these aspirational areas. Our goal is to offer actionable recommendations for PPI researchers and guidance for practitioners evaluating this research for application.

Study design & analyses

We believe it is imperative that PPI research adheres to established best practices in intervention science. Within the clinical and medical intervention sciences, randomized controlled trials (RCTs) are placed at the highest levels of evidence strength classification systems (e.g., Howick et al., 2009; US Preventive Services Task Force, 1989). Yet, the gold-standard RCT design is rarely used in PPI research (Quoidbach et al., 2015). One reason for this underutilization of the best clinical research practices may be that PPI research often focuses on nonclinical populations. However, the standards in RCT design and methodology are established to support inferences about the efficacy of any intervention, including those with nonclinical samples. We believe that adherence to best RCT practices is crucial for developing robust PPIs.

Intention-to-treat analyses

One key RCT feature, often overlooked in PPI research, is the use of intention-to-treat analyses (Fisher et al., 1990), which examine intervention effects on outcomes regardless of participants' treatment compliance (Heritier et al., 2003). Thus, a participant's data are analyzed regardless of whether they completed 5%, 50%, or 100% of the treatment sessions (or 'doses'). This intention-to-treat approach tests the efficacy of an intervention in a manner sensitive to external validity, situating the intervention in a real world context in which participants may face barriers, including the time and motivation, to complete an intervention.

Intention-to-treat analyses are also beneficial from a basic scientific perspective. Decisions to remove participants who did not complete a certain dosage

CONTACT Samantha J. Heintzelman samantha.heintzelman@rutgers.edu

threshold are often made by researchers post-hoc (perhaps only after full sample analyses did not yield positive results). These decisions about analyzing and reporting data, including the selective reporting of analyses from a sample subset without pre-specified exclusion criteria or the discriminatory presentation of only those outcome variables producing positive effects, are called 'researcher degrees of freedom.' Researcher degrees of freedom can lead to an overestimation of effect sizes and an increase in false-positive results, which can accumulate to form an unreliable and nonreplicable body of science (John et al., 2012; Simmons et al., 2011). Intention-to-treat analyses limit a researcher's latitude in analysis decisions and protect against problematic biases that can (even inadvertently) creep into research.

An example of intention-to-treat analyses in PPI research can be found in our ENHANCE trial report (Heintzelman et al., 2020). Using intention-to-treat analyses made us more confident about the external validity of ENHANCE as a treatment program that can be efficaciously administered to people with busy lives and various levels of motivation.

CONSORT diagrams

The use of Consolidated Standards of Reporting Trials (CONSORT) diagrams is another RCT practice that can boost the quality and replicability of PPI research (Moher et al., 2012). CONSORT diagrams require researchers to broadly consider, document, and report on the number of participants at all stages of the trial, from recruitment to completion. CONSORT diagrams include recruitment details not typically reported in traditional research, such as the number of people who expressed interest in the study, and the number who were ineligible due to explicit exclusion criteria. CONSORT diagrams also include information about enrolled participants, including condition assignments and detailed information regarding trial attrition. In line with the increasing focus on replicability in psychological science (Pashler & Wagenmakers, 2012), CONSORT diagrams are a standardized method promoting careful reflections about the populations represented by particular treatment samples, importantly informing the conclusions that can – and cannot – be drawn from a specific trial. Constructing CONSORT diagrams for the ENHANCE trial made us carefully consider the population represented by our sample and allowed us to concisely communicate a large amount of information about our sample to readers.

Steps to building a cumulative PPI Science

Another key issue in building a robust and respectable literature on the efficacy of PPIs is ensuring a cumulative approach to science. In our view, a cumulative approach includes two key characteristics: (1) reporting results from unsuccessful interventions and, (2) carefully selecting control comparison treatments.

First, we encourage PPI researchers to preregister their studies on www.clinicaltrials.gov. While laudable alternatives, such as the Open Science Framework, can be used in tandem, the preregistration on clinicaltrials.gov is standardized to the best practices specific to intervention science. For example, researchers are required to clearly identify primary outcomes to evaluate the general intervention efficacy, further controlling researcher degrees of freedom by mandating specificity before beginning a trial. Researchers must also report, following the completion of a trial, whether the intervention tested was, indeed, efficacious. These trial reports become part of a public repository that can be used by researchers to survey interventions across their area of interest. Beyond compiling information about efficacious PPIs, wide use of clinicaltrials.gov would also provide valuable insights regarding those intervention strategies that *aren't* typically efficacious, information that might not otherwise be readily accessible beyond researchers' file drawers, though essential for an efficient cumulative PPI science.

Preregistering the ENHANCE trial on clinicaltrials.org required us to distinguish between a carefully selected primary outcome – subjective well-being (Diener, 2000) – a broader range of secondary outcomes, and an even wider range of exploratory measures. These practices provided us with a system of accountability for transparently drawing appropriate conclusions from our data given the explicit contours of our sample and hypotheses.

A second key to cumulative PPI science is the selection of appropriate comparison groups. As the PPI literature continues to develop, researchers should compare any new PPIs to existing PPIs whose efficacy has already been established in a push towards ever-improving methods for increasing happiness in the most innovative, efficient, and enduring manner possible. While we carefully examined the PPI literature for an appropriate comparison group for ENHANCE, we concluded that at the time, there were no PPIs established as a standard for comparison. We, therefore, decided to utilize a wait-list control group. Our hope is that by following best practices in clinical science, ENHANCE can be used in the future as a comparison treatment group to build towards ever-improved PPI methods.

Measurement

Measuring subjective well-being

PPI researchers must also make a series of decisions regarding measurement for program evaluation and broader scientific purposes, including which variables to measure and how and when to measure them. Foremost, we urge researchers to continue the widely-used practice of measuring subjective well-being in PPI research with well-validated measures such as the Satisfaction with Life Scale (Diener et al., 1985), the Scale of Positive and Negative Experience (Diener et al., 2010), Cantril's self-anchoring ladder (Cantril, 1965) or the Scales of Psychological Well-Being (Ryff, 1989), depending on the goals of the particular program being tested.

In addition, we recommend that researchers supplement self-report measures of subjective well-being with indirect measures, such as peer reports. While there are many existing psychometrically strong subjective well-being scales with high levels of reliability and established validity (Lucas, 2018), PPIs can exacerbate reporting biases, such as demand characteristics. Thus, PPI participants may inflate their reports of subjective well-being following treatment to please the researchers or to justify their own time and efforts. A multi-method strategy for measuring subjective well-being in PPI research can allay reporting biases concerns.

In our research with ENHANCE, we supplemented the self-report measures with peer reports of participants' well-being (Pavot et al., 1991), which we collected from several of each participant's close peers (e.g., their spouses, family, close friends, as well as by using a memory bias task. In the Positive and Negative Memory Task (Seidlitz & Diener, 1993), participants are asked to list as many positive and negative life events as they can recall in three minutes each. The memory accessibility ratio of positive to negative events serves as an indirect measure of subjective well-being, and has been shown to be stable across time (Sandvik et al., 1993). These non-self-report measures supported the efficacy of the ENHANCE intervention, suggesting that our PPI produced reliable gains in subjective well-being.

Beyond non-self-report measures, more fine-grained data collection strategies, such as the Day Reconstruction Method (Kahneman et al., 2004) or the Experience Sampling Method (Larson & Csikszentmihalyi, 1983) could offer a more extensive examination of the processes underlying changes in subjective well-being over the course of an PPI and beyond (e.g., Fredrickson et al., 2008).

Measuring downstream outcomes of increasing happiness

PPI research can also benefit from increased inclusion of outcomes that may be influenced by subjective well-being. In their influential examination of the subjective well-being literature, Lyubomirsky, King et al. (2005) reviewed the substantial body of evidence for the relationships between happiness and success in life domains such as physical health, relationships, income, and work performance. While drawing on a multitude of studies using correlational and longitudinal designs and laboratory manipulations of short term mood, they also note an important gap in the literature: ' … if these same behaviors are also increased by long-term interventions to enhance global happiness, the case for happiness being causally related to success will be strengthened even more' (p. 841). Measuring possible downstream consequences of increasing happiness in PPI research could fill this crucial gap in basic research by boosting our understanding of the benefits of subjective well-being.

Researchers have already begun to include measures of physical health in PPI trials, primarily in trials targeting clinical samples with existing health issues (e.g., Moskowitz et al., 2017). We believe it is important to include such outcome measures within samples of healthy adults as well. In the ENHANCE trial, for example, we included cognitive and physical health measures. These have allowed us to provide experimental evidence, for example, that long-term changes in well-being drove improvements on some physical health outcomes (Kushlev et al., 2020). We encourage PPI researchers to include a host of downstream outcomes linked to happiness to build a body of work clarifying its causal role in important life outcomes.

Focusing on processes underlying efficacy

While specifying primary outcomes to define evidence for program efficacy is paramount, we argue that it is also important to understand the psychological processes that make a PPI efficacious. We recommend that researchers aim to gain a deeper understanding of *why* and *how* a PPI drove changes in subjective well-being. In the existing PPI literature, only rarely have studies addressed the mechanisms of action of a specific intervention in increasing subjective well-being. Yet, understanding the processes behind change is essential for the development of the PPI field. Researchers should, therefore, consider including not only simple treatment

adherence measures, but also measures of the psychological processes hypothesized to drive subjective well-being gains.

In the ENHANCE trial, we examined the process of well-being development in two central ways. First, we measured participants' skills in the areas targeted by each of the intervention activities prior to and following program participation. This allowed us to test whether participants were, in fact, improving on those skills targeted by the program and, furthermore, whether their improvement on those skills mediated improvements in subjective well-being as hypothesized (Heintzelman et al., 2020). Second, we collected data regarding subjective well-being throughout the course of the treatment through weekly surveys, which provided another avenue to examine the time course of well-being development throughout the program participation period (Kushlev et al., 2020). These process-level measures are essential for understanding the mechanisms of subjective well-being change through PPI.

Long-term follow-up assessments

To properly examine PPI effects given the well documented processes of hedonic adaptation (e.g., Sheldon & Lyubomirsky, 2012), it is essential to understand the durability of PPI intervention effects. After the completion of treatment, therefore, we advocate for follow-up assessments to be administered for as long as possible.

In the ENHANCE trial, we conducted extensive assessments at baseline, at immediate post-treatment, and at a follow-up point three months later (six months from baseline). We also collected data from a subset of our sample at a later follow-up six months after treatment (nine months from baseline). Future research should extend assessments even further post-treatment, including one-year and two-year follow-ups. We suggest gradually reducing the assessment frequency over time, but increasing the incentives for participants to complete these long-term assessments to improve completion rates.

In summary, in evaluating the efficacy of PPIs, we recommend the use of well-validated measures of subjective well-being coupled with indirect measures of well-being. Furthermore, we urge PPI researchers to examine both the processes and the durability of well-being gains.

Sampling

Another aspect of study design that is important for rigorous PPI research is the matter of sample selection. In general, researchers must employ careful recruitment strategies to minimize participant selection biases and demand characteristics that can plague PPI research. Furthermore, PPI research has problematically maintained an overreliance on a convenience sample of participants often representing socioeconomically advantaged, highly educated, young, white, women. The general issues with relying on such WEIRD (Western, Educated, Industrialized, Rich, and Democratic cultures) samples are well documented elsewhere (Henrich et al., 2010).

An issue of WEIRD samples specifically related to PPI research is that increasing one's happiness may simply be a bourgeois desire. Researchers have attempted to address this critique by pointing out that the beneficial consequences linked to happiness suggesting that happiness is central to human adaptation rather than solely 'a symptom of Western comfort and self-centeredness' (Lyubomirsky, Sheldon et al., 2005, p. 111). The role of happiness in adaptive functioning, however, has little bearing on the key issue of whether PPIs can increase happiness in more disadvantaged populations or whether increasing happiness in these populations leads to similar beneficial outcomes across life domains. To make findings from the PPI more generalizable, it is essential that research seek to recruit from populations with diverse characteristics including age, gender, ethnicity, background, life circumstances, socioeconomic status, mental health status, and so forth.

We see substantial shortcomings in this regard in our own work and throughout the PPI field. Diversifying PPI samples will require deep commitment to approach the challenges inherent in broad recruitment efforts. Achieving this goal will likely require partnerships with community groups, community outreach to establish trusting relationships with target populations, and generous financial support from funding agencies. Furthermore, additional considerations are required to eliminate procedural barriers to participation facing particular populations including, but not limited to, reimbursements for time, allocations of any needed equipment, relevant access to technology, education-level adaptable materials and assessments, transportation, and childcare provision.

Personnel

Finally, drawing on our own experience, we believe that the successful execution of rigorous PPI research is greatly facilitated by collaborations between team members with diverse scholarly expertise. For example, the ENHANCE team consisted not only of content

experts – well-being scholars – but also of clinical interventionists. These collaborations were essential in building a project at the intersection of basic and applied science. Furthermore, diverse research personnel selection is imperative in carrying out the aims we outlined above. For instance, health psychologists and applied psychologists will certainly be assets to teams developing projects including health and work related outcome variables. Additionally, population experts with knowledge of the cultural norms of a targeted community will facilitate successful PPIs sensitive to particular needs of non-WEIRD samples.

Moving forward

We believe that each of the aspirations we have outlined above – for strong research designs, thoughtful measurement plans, inclusive sampling strategies, and collaborative research teams – can contribute to the development of a rigorous, replicable, and cumulative PPI science with broad impact. We are enthusiastic about the growth of PPI research and foresee that continued advancement depends upon meeting the highest standards for intervention research.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Kostadin Kushlev http://orcid.org/0000-0003-1025-3258

References

Cantril, H. (1965). *The pattern of human concerns*. Rutgers University Press.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, *55*(1), 34–43. https://doi.org/10.1037/0003-066X.55.1.34

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97*(2), 143–156. https://doi.org/10.1007/s11205-009-9493-y

Fisher, L. D., Dixon, D. O., Herson, J., Frankowski, R. K., Hearron, M. S., & Peace, K. E. (1990). Intention to treat in clinical trials. *Statistical Issues in Drug Research and Development*, 331–350. CRC Press. https://doi.org/10.1201/9780203738610

Fredrickson, B. L., Cohn, M. A., Coffey, K. A., Pek, J., & Finkel, S. M. (2008). Open hearts build lives: Positive emotions, induced through loving-kindness meditation, build consequential personal resources. *Journal of Personality and Social Psychology*, *95*(5), 1045–1062. https://doi.org/10.1037/a0013262

Heintzelman, S. J., Kushlev, K., Lutes, L. D., Wirtz, D., Kanippayoor, J. M., Leitner, D., Oishi, S., & Diener, E. (2020). ENHANCE: Evidence for the efficacy of a comprehensive intervention program to promote subjective well-being. Journal of Experimental Psychology: Applied, 26(2), 360-383.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, *33*(2–3), 111–135. https://doi.org/10.1017/S0140525X10000725

Heritier, S. R., Gebski, V. J., & Keech, A. C. (2003). Inclusion of patients in clinical trial analysis: The intention-to-treat principle. *Medical Journal of Australia*, *179*(8), 438–440. https://doi.org/10.5694/j.1326-5377.2003.tb05627.x

Howick, J., Phillips, B., Ball, C., Sackett, D., & Badenoch, D. (2009). *Oxford centre for evidence-based medicine levels of evidence*. University of Oxford, Centre for Evidence-Based Medicine.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, *306*(5702), 1776–1780. https://doi.org/10.1126/science.1103572

Kushlev, K., Heintzelman, S. J., & Diener, E. (2020). Does happiness improve health? Evidence from a randomized controlled trial. Psychological Science, 0956797620919673.

Kushlev, K., Heintzelman, S. J., Lutes, L. D., Wirtz, D., Oishi, S., & Diener, E. (2017). ENHANCE: Design and rationale of a randomized controlled trial for promoting enduring happiness & well-being. *Contemporary Clinical Trials*, *52*, 62–74. https://doi.org/10.1016/j.cct.2016.11.003

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In H. T. Reis (Ed.), New Directions for Methodology of Social and Behavioral Sciences (vol. 15, pp. 41–56). San Francisco: Jossey-Bass.

Lucas, R. E. (2018). Reevaluating the strengths and weaknesses of self-report measures of subjective well-being. In E. Diener, S. Oishi, & L. Tay (Eds..), *Handbook of well-being*. DEF Publishers. DOI:nobascholar.com. https://nobascholar.com/chapters/2/download.pdf

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin*, *131*(6), 803–855. https://doi.org/10.1037/0033-2909.131.6.803

Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology*, *9*(2), 111–131. https://doi.org/10.1037/1089-2680.9.2.111

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2012). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *International Journal of Surgery*, *10*(1), 28–55. https://doi.org/10.1016/j.ijsu.2011.10.001

Moskowitz, J. T., Carrico, A. W., Duncan, L. G., Cohn, M. A., Cheung, E. O., Batchelder, A., Martinez, L., Segawa, E., Acree, M., & Folkman, S. (2017). Randomized controlled trial of a positive affect intervention for people newly diagnosed with HIV. *Journal of Consulting and Clinical Psychology*, *85*(5), 409–423. https://doi.org/10.1037/ccp0000188

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253

Pavot, W., Diener, E., Colvin, C. R., & Sandvik, E. (1991). Further validation of the Satisfaction with Life Scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment*, *57*(1), 149–161. https://doi.org/10.1207/s15327752jpa5701_17

Quoidbach, J., Mikolajczak, M., & Gross, J. J. (2015). Positive interventions: An emotion regulation perspective. *Psychological Bulletin*, *141*(3), 655–693. https://doi.org/10.1037/a0038648

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*(6), 1069–1081. https://doi.org/10.1037/0022–3514.57.6.1069

Sandvik, E., Diener, E., & Seidlitz, L. (1993). Subjective well-being: The convergence and stability of self-report and non-self-report measures. *Journal of Personality*, *61*(3), 317–342. https://doi.org/10.1111/j.1467-6494.1993.tb00283.x

Seidlitz, L., & Diener, E. (1993). Memory for positive versus negative life events: Theories for the differences between happy and unhappy persons. *Journal of Personality and Social Psychology*, *64*(4), 654–664. https://doi.org/10.1037/0022-3514.64.4.654

Sheldon, K. M., & Lyubomirsky, S. (2012). The challenge of staying happier: Testing the hedonic adaptation prevention model. *Personality & Social Psychology Bulletin*, *38*(5), 670–680. https://doi.org/10.1177/0146167212436400

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

US Preventive Services Task Force. (1989). *Guide to clinical preventive services: Report of the US preventive services task force*. DIANE publishing.